COMPUTERS, AUTOMATIC CONTROL, SIGNAL PROCESSING and SYSTEMS SCIENCE

Proceedings of the 2014 International Conference on Systems, Control, Signal Processing and Informatics II (SCSI '14)

Prague, Czech Republic
April 2-4, 2014

Recent Advances in Electrical Engineering Series - 33

ISSN: 1790-5117
ISBN: 978-1-61804-233-0

COMPUTERS, AUTOMATIC CONTROL, SIGNAL PROCESSING and SYSTEMS SCIENCE

Proceedings of the 2014 International Conference on Systems, Control, Signal Processing and Informatics II (SCSI '14)

Prague, Czech Republic
April 2-4, 2014

All papers of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

ISSN: 1790-5117
ISBN: 978-1-61804-233-0

COMPUTERS, AUTOMATIC CONTROL, SIGNAL PROCESSING and SYSTEMS SCIENCE

Proceedings of the 2014 International Conference on Systems, Control, Signal Processing and Informatics II (SCSI '14)

Prague, Czech Republic
April 2-4, 2014

Organizing Committee

General Chairs (EDITORS)

- Professor Kleanthis Psarris,
  The City University of New York,
  USA
- Professor Pierre Borne
  IEEE France Section Chair, IEEE Fellow,
  IEEE/SMC Past President
  Ecole Centrale de Lille, BP 48, 59651
  Villeneuve d'Ascq, France
- Professor Imre Rudas,
  Obuda University, Budapest,
  Hungary
- Professor Yuriy S. Shmaliy
  IEEE Fellow
  The University of Guanajuato, Mexico

Associate Editors

- Prof. Fareed H. Felemban,
  Vice-President,
  Taif University,
  Kingdom of Saudi Arabia
- Prof. Bassant M. Elbagoury,
  College of Computers & Information Technology,
  Taif University,
  Kingdom of Saudi Arabia

Senior Program Chair

- Professor Leon Chua,
  IEEE Fellow,
  University of Berkeley, USA

Program Chairs

- Prof. Wasfy B Mikhael,
  IEEE Fellow,
  University of Central Florida Orlando,
  USA
- Professor Zoran Bojkovic,
  Univ. of Belgrade, Serbia
- Professor Panos Pardalos,
  IEEE Fellow, University of Florida,
  USA

Tutorials Chair

- Professor Kamisetty Rao
  IEEE Fellow
  Univ. of Texas at Arlington
  USA

Special Session Chair

- Prof. H. M. Srivastava
  University of Victoria
  Victoria, British Columbia V8W 3R4
  Canada

Workshops Chair

- Prof. Ryszard S. Choras
  Institute of Telecommunications
  University of Technology & Life Sciences
  Bydgoszcz, Poland

Local Organizing Chair

- Assistant Prof. Klimis Ntalianis,
  Tech. Educ. Inst. of Athens (TEI),
  Athens, Greece

Publication Chair

- Professor Maurice Margenstern,
  Université de Lorraine,
  France

Publicity Committee

- Professor James Lam, IEEE Fellow,
  IET Fellow, IMA Fellow,
  The University of Hong Kong, Hong Kong
- Professor Erol Gelenbe, IEEE Fellow,
  ACM Fellow, IET Fellow (former IEE),
  Imperial College,
  University of London, London, UK

International Liaisons

- Professor Theodore B. Trafalis,
  University of Oklahoma, USA
- Professor Olga Martin
  Applied Sciences Faculty
  Politehnica University of Bucharest
  Romania
- Professor Vincenzo Niola
  Departement of Mechanical Engineering for Energetics
  University of Naples "Federico II"
  Naples, Italy
- Professor M. Affan Badar,
  Indiana State University,
  Terre Haute, Indiana, USA

Steering Committee

- Prof. Erol Gelenbe, IEEE Fellow, ACM Fellow, IET Fellow (former IEE), Imperial College, University of London, London, UK
- Prof. James Lam, IEEE Fellow, IET Fellow, IMA Fellow, The University of Hong Kong, Hong Kong
- Prof. M. Affan Badar, Indiana State University, Terre Haute, Indiana, USA
- Prof. Chih-Cheng Hung, Southern Polytechnic State University, Marietta, GA, USA
- Prof. Martin Bohner, Missouri University of Science and Technology, Rolla, Missouri, USA
- Prof. Ravi P. Agarwal, Texas A&M University-Kingsville, Kingsville, Texas, USA

Program Committee

Prof. Lotfi Zadeh (IEEE Fellow,University of Berkeley, USA)
Prof. Leon Chua (IEEE Fellow,University of Berkeley, USA)
Prof. Michio Sugeno (RIKEN Brain Science Institute (RIKEN BSI), Japan)
Prof. Dimitri Bertsekas (IEEE Fellow, MIT, USA)
Prof. Demetri Terzopoulos (IEEE Fellow, ACM Fellow, UCLA, USA)
Prof. Georgios B. Giannakis (IEEE Fellow, University of Minnesota, USA)
Prof. George Vachtsevanos (Georgia Institute of Technology, USA)
Prof. Abraham Bers (IEEE Fellow, MIT, USA)
Prof. Brian Barsky (IEEE Fellow, University of Berkeley, USA)
Prof. Aggelos Katsaggelos (IEEE Fellow, Northwestern University, USA)
Prof. Josef Sifakis (Turing Award 2007, CNRS/Verimag, France)
Prof. Hisashi Kobayashi (Princeton University, USA)
Prof. Kinshuk (Fellow IEEE, Massey Univ. New Zeland),
Prof. Leonid Kazovsky (Stanford University, USA)
Prof. Narsingh Deo (IEEE Fellow, ACM Fellow, University of Central Florida, USA)
Prof. Kamisetty Rao (Fellow IEEE, Univ. of Texas at Arlington,USA)
Prof. Anastassios Venetsanopoulos (Fellow IEEE, University of Toronto, Canada)
Prof. Steven Collicott (Purdue University, West Lafayette, IN, USA)
Prof. Nikolaos Paragios (Ecole Centrale Paris, France)
Prof. Nikolaos G. Bourbakis (IEEE Fellow, Wright State University, USA)
Prof. Stamatios Kartalopoulos (IEEE Fellow, University of Oklahoma, USA)
Prof. Irwin Sandberg (IEEE Fellow, University of Texas at Austin, USA),
Prof. Michael Sebek (IEEE Fellow, Czech Technical University in Prague, Czech Republic)
Prof. Hashem Akbari (University of California, Berkeley, USA)
Prof. Yuriy S. Shmaliy, (IEEE Fellow, The University of Guanajuato, Mexico)
Prof. Lei Xu (IEEE Fellow, Chinese University of Hong Kong, Hong Kong)
Prof. Paul E. Dimotakis (California Institute of Technology Pasadena, USA)
Prof. Martin Pelikan (UMSL, USA)
Prof. Patrick Wang (MIT, USA)
Prof. Wasfy B Mikhael (IEEE Fellow, University of Central Florida Orlando,USA)
Prof. Sunil Das (IEEE Fellow, University of Ottawa, Canada)
Prof. Panos Pardalos (University of Florida, USA)
Prof. Nikolaos D. Katopodes (University of Michigan, USA)
Prof. Bimal K. Bose (Life Fellow of IEEE, University of Tennessee, Knoxville, USA)
Prof. Janusz Kacprzyk (IEEE Fellow, Polish Academy of Sciences, Poland)
Prof. Sidney Burrus (IEEE Fellow, Rice University, USA)
Prof. Biswa N. Datta (IEEE Fellow, Northern Illinois University, USA)
Prof. Mihai Putinar (University of California at Santa Barbara, USA)
Prof. Wlodzislaw Duch (Nicolaus Copernicus University, Poland)
Prof. Tadeusz Kaczorek (IEEE Fellow, Warsaw University of Tehcnology, Poland)
Prof. Michael N. Katehakis (Rutgers, The State University of New Jersey, USA)
Prof. Pan Agathoklis (Univ. of Victoria, Canada)
Dr. Subhas C. Misra (Harvard University, USA)
Prof. Martin van den Toorn (Delft University of Technology, The Netherlands)
Prof. Malcolm J. Crocker (Distinguished University Prof., Auburn University,USA)
Prof. Urszula Ledzewicz, Southern Illinois University , USA.
Prof. Dimitri Kazakos, Dean, (Texas Southern University, USA)
Prof. Ronald Yager (Iona College, USA)
Prof. Athanassios Manikas (Imperial College, London, UK)
Prof. Keith L. Clark (Imperial College, London, UK)
Prof. Argyris Varonides (Univ. of Scranton, USA)
Prof. S. Furfari (Direction Generale Energie et Transports, Brussels, EU)

Prof. Constantin Udriste, University Politehnica of Bucharest , ROMANIA
Dr. Michelle Luke (Univ. Berkeley, USA)
Prof. Patrice Brault (Univ. Paris-sud, France)
Prof. Jim Cunningham (Imperial College London, UK)
Prof. Philippe Ben-Abdallah (Ecole Polytechnique de l'Universite de Nantes, France)
Prof. Photios Anninos (Medical School of Thrace, Greece)
Prof. Ichiro Hagiwara, (Tokyo Institute of Technology, Japan)
Prof. Andris Buikis (Latvian Academy of Science. Latvia)
Prof. Akshai Aggarwal (University of Windsor, Canada)
Prof. George Vachtsevanos (Georgia Institute of Technology, USA)
Prof. Ulrich Albrecht (Auburn University, USA)
Prof. Imre J. Rudas (Obuda University, Hungary)
Prof. Alexey L Sadovski (IEEE Fellow, Texas A&M University, USA)
Prof. Amedeo Andreotti (University of Naples, Italy)
Prof. Ryszard S. Choras (University of Technology and Life Sciences Bydgoszcz, Poland)
Prof. Remi Leandre (Universite de Bourgogne, Dijon, France)
Prof. Moustapha Diaby (University of Connecticut, USA)
Prof. Brian McCartin (New York University, USA)
Prof. Elias C. Aifantis (Aristotle Univ. of Thessaloniki, Greece)
Prof. Anastasios Lyrintzis (Purdue University, USA)
Prof. Charles Long (Prof. Emeritus University of Wisconsin, USA)
Prof. Marvin Goldstein (NASA Glenn Research Center, USA)
Prof. Costin Cepisca (University POLITEHNICA of Bucharest, Romania)
Prof. Kleanthis Psarris (University of Texas at San Antonio, USA)
Prof. Ron Goldman (Rice University, USA)
Prof. Ioannis A. Kakadiaris (University of Houston, USA)
Prof. Richard Tapia (Rice University, USA)
Prof. Milivoje M. Kostic (Northern Illinois University, USA)
Prof. Helmut Jaberg (University of Technology Graz, Austria)
Prof. Ardeshir Anjomani (The University of Texas at Arlington, USA)
Prof. Heinz Ulbrich (Technical University Munich, Germany)
Prof. Reinhard Leithner (Technical University Braunschweig, Germany)
Prof. Elbrous M. Jafarov (Istanbul Technical University, Turkey)
Prof. M. Ehsani (Texas A&M University, USA)
Prof. Sesh Commuri (University of Oklahoma, USA)
Prof. Nicolas Galanis (Universite de Sherbrooke, Canada)
Prof. S. H. Sohrab (Northwestern University, USA)
Prof. Rui J. P. de Figueiredo (University of California, USA)
Prof. Valeri Mladenov (Technical University of Sofia, Bulgaria)
Prof. Hiroshi Sakaki (Meisei University, Tokyo, Japan)
Prof. Zoran S. Bojkovic (Technical University of Belgrade, Serbia)
Prof. K. D. Klaes, (Head of the EPS Support Science Team in the MET Division at EUMETSAT, France)
Prof. Emira Maljevic (Technical University of Belgrade, Serbia)
Prof. Kazuhiko Tsuda (University of Tsukuba, Tokyo, Japan)
Prof. Milan Stork (University of West Bohemia , Czech Republic)
Prof. C. G. Helmis (University of Athens, Greece)
Prof. Lajos Barna (Budapest University of Technology and Economics, Hungary)
Prof. Nobuoki Mano (Meisei University, Tokyo, Japan)
Prof. Nobuo Nakajima (The University of Electro-Communications, Tokyo, Japan)
Prof. Victor-Emil Neagoe (Polytechnic University of Bucharest, Romania)
Prof. P. Vanderstraeten (Brussels Institute for Environmental Management, Belgium)
Prof. Annaliese Bischoff (University of Massachusetts, Amherst, USA)
Prof. Virgil Tiponut (Politehnica University of Timisoara, Romania)
Prof. Andrei Kolyshkin (Riga Technical University, Latvia)

Prof. Fumiaki Imado (Shinshu University, Japan)
Prof. Sotirios G. Ziavras (New Jersey Institute of Technology, USA)
Prof. Constantin Volosencu (Politehnica University of Timisoara, Romania)
Prof. Marc A. Rosen (University of Ontario Institute of Technology, Canada)
Prof. Thomas M. Gatton (National University, San Diego, USA)
Prof. Leonardo Pagnotta (University of Calabria, Italy)
Prof. Yan Wu (Georgia Southern University, USA)
Prof. Daniel N. Riahi (University of Texas-Pan American, USA)
Prof. Alexander Grebennikov (Autonomous University of Puebla, Mexico)
Prof. Bennie F. L. Ward (Baylor University, TX, USA)
Prof. Guennadi A. Kouzaev (Norwegian University of Science and Technology, Norway)
Prof. Eugene Kindler (University of Ostrava, Czech Republic)
Prof. Geoff Skinner (The University of Newcastle, Australia)
Prof. Hamido Fujita (Iwate Prefectural University(IPU), Japan)
Prof. Francesco Muzi (University of L'Aquila, Italy)
Prof. Claudio Rossi (University of Siena, Italy)
Prof. Sergey B. Leonov (Joint Institute for High Temperature Russian Academy of Science, Russia)
Prof. Arpad A. Fay (University of Miskolc, Hungary)
Prof. Lili He (San Jose State University, USA)
Prof. M. Nasseh Tabrizi (East Carolina University, USA)
Prof. Alaa Eldin Fahmy (University Of Calgary, Canada)
Prof. Gh. Pascovici (University of Koeln, Germany)
Prof. Pier Paolo Delsanto (Politecnico of Torino, Italy)
Prof. Radu Munteanu (Rector of the Technical University of Cluj-Napoca, Romania)
Prof. Ioan Dumitrache (Politehnica University of Bucharest, Romania)
Prof. Corneliu Lazar (Technical University Gh.Asachi Iasi, Romania)
Prof. Nicola Pitrone (Universita degli Studi Catania, Italia)
Prof. Miquel Salgot (University of Barcelona, Spain)
Prof. Amaury A. Caballero (Florida International University, USA)
Prof. Maria I. Garcia-Planas (Universitat Politecnica de Catalunya, Spain)
Prof. Petar Popivanov (Bulgarian Academy of Sciences, Bulgaria)
Prof. Alexander Gegov (University of Portsmouth, UK)
Prof. Lin Feng (Nanyang Technological University, Singapore)
Prof. Colin Fyfe (University of the West of Scotland, UK)
Prof. Zhaohui Luo (Univ of London, UK)
Prof. Wolfgang Wenzel (Institute for Nanotechnology, Germany)
Prof. Weilian Su (Naval Postgraduate School, USA)
Prof. Phillip G. Bradford (The University of Alabama, USA)
Prof. Ray Hefferlin (Southern Adventist University, TN, USA)
Prof. Gabriella Bognar (University of Miskolc, Hungary)
Prof. Hamid Abachi (Monash University, Australia)
Prof. Karlheinz Spindler (Fachhochschule Wiesbaden, Germany)
Prof. Josef Boercsoek (Universitat Kassel, Germany)
Prof. Eyad H. Abed (University of Maryland, Maryland, USA)
Prof. F. Castanie (TeSA, Toulouse, France)
Prof. Robert K. L. Gay (Nanyang Technological University, Singapore)
Prof. Andrzej Ordys (Kingston University, UK)
Prof. Harris Catrakis (Univ of California Irvine, USA)
Prof. T Bott (The University of Birmingham, UK)
Prof. T.-W. Lee (Arizona State University, AZ, USA)
Prof. Le Yi Wang (Wayne State University, Detroit, USA)
Prof. Oleksander Markovskyy (National Technical University of Ukraine, Ukraine)
Prof. Suresh P. Sethi (University of Texas at Dallas, USA)
Prof. Hartmut Hillmer(University of Kassel, Germany)

Prof. Bram Van Putten (Wageningen University, The Netherlands)
Prof. Alexander Iomin (Technion - Israel Institute of Technology, Israel)
Prof. Roberto San Jose (Technical University of Madrid, Spain)
Prof. Minvydas Ragulskis (Kaunas University of Technology, Lithuania)
Prof. Arun Kulkarni (The University of Texas at Tyler, USA)
Prof. Joydeep Mitra (New Mexico State University, USA)
Prof. Vincenzo Niola (University of Naples Federico II, Italy)
Prof. Ion Chryssoverghi (National Technical University of Athens, Greece)
Prof. Dr. Aydin Akan (Istanbul University, Turkey)
Prof. Sarka Necasova (Academy of Sciences, Prague, Czech Republic)
Prof. C. D. Memos (National Technical University of Athens, Greece)
Prof. S. Y. Chen, (Zhejiang University of Technology, China and University of Hamburg, Germany)
Prof. Duc Nguyen (Old Dominion University, Norfolk, USA)
Prof. Tuan Pham (James Cook University, Townsville, Australia)
Prof. Jiri Klima (Technical Faculty of CZU in Prague, Czech Republic)
Prof. Rossella Cancelliere (University of Torino, Italy)
Prof. Dr-Eng. Christian Bouquegneau (Faculty Polytechnique de Mons, Belgium)
Prof. Wladyslaw Mielczarski (Technical University of Lodz, Poland)
Prof. Ibrahim Hassan (Concordia University, Montreal, Quebec, Canada)
Prof. Stavros J.Baloyannis (Medical School, Aristotle University of Thessaloniki, Greece)
Prof. James F. Frenzel (University of Idaho, USA)
Prof. Vilem Srovnal,(Technical University of Ostrava, Czech Republic)
Prof. J. M. Giron-Sierra (Universidad Complutense de Madrid, Spain)
Prof. Walter Dosch (University of Luebeck, Germany)
Prof. Rudolf Freund (Vienna University of Technology, Austria)
Prof. Erich Schmidt (Vienna University of Technology, Austria)
Prof. Alessandro Genco (University of Palermo, Italy)
Prof. Martin Lopez Morales (Technical University of Monterey, Mexico)
Prof. Ralph W. Oberste-Vorth (Marshall University, USA)
Prof. Vladimir Damgov (Bulgarian Academy of Sciences, Bulgaria)
Prof. P.Borne (Ecole Central de Lille, France)

Additional Reviewers

Lesley Farmer	California State University Long Beach, CA, USA
Kei Eguchi	Fukuoka Institute of Technology, Japan
James Vance	The University of Virginia's College at Wise, VA, USA
Eleazar Jimenez Serrano	Kyushu University, Japan
Zhong-Jie Han	Tianjin University, China
Minhui Yan	Shanghai Maritime University, China
George Barreto	Pontificia Universidad Javeriana, Colombia
Tetsuya Shimamura	Saitama University, Japan
Shinji Osada	Gifu University School of Medicine, Japan
Genqi Xu	Tianjin University, China
Jose Flores	The University of South Dakota, SD, USA
Philippe Dondon	Institut polytechnique de Bordeaux, France
Imre Rudas	Obuda University, Budapest, Hungary
Abelha Antonio	Universidade do Minho, Portugal
Tetsuya Yoshida	Hokkaido University, Japan
Sorinel Oprisan	College of Charleston, CA, USA
Xiang Bai	Huazhong University of Science and Technology, China
Francesco Rotondo	Polytechnic of Bari University, Italy
Valeri Mladenov	Technical University of Sofia, Bulgaria
Stavros Ponis	National Technical University of Athens, Greece
Matthias Buyle	Artesis Hogeschool Antwerpen, Belgium
José Carlos Metrôlho	Instituto Politecnico de Castelo Branco, Portugal
Kazuhiko Natori	Toho University, Japan
Ole Christian Boe	Norwegian Military Academy, Norway
Alejandro Fuentes-Penna	Universidad Autónoma del Estado de Hidalgo, Mexico
João Bastos	Instituto Superior de Engenharia do Porto, Portugal
Masaji Tanaka	Okayama University of Science, Japan
Yamagishi Hiromitsu	Ehime University, Japan
Manoj K. Jha	Morgan State University in Baltimore, USA
Frederic Kuznik	National Institute of Applied Sciences, Lyon, France
Dmitrijs Serdjuks	Riga Technical University, Latvia
Andrey Dmitriev	Russian Academy of Sciences, Russia
Francesco Zirilli	Sapienza Universita di Roma, Italy
Hessam Ghasemnejad	Kingston University London, UK
Bazil Taha Ahmed	Universidad Autonoma de Madrid, Spain
Jon Burley	Michigan State University, MI, USA
Takuya Yamano	Kanagawa University, Japan
Miguel Carriegos	Universidad de Leon, Spain
Deolinda Rasteiro	Coimbra Institute of Engineering, Portugal
Santoso Wibowo	CQ University, Australia
M. Javed Khan	Tuskegee University, AL, USA
Konstantin Volkov	Kingston University London, UK
Moran Wang	Tsinghua University, China
Angel F. Tenorio	Universidad Pablo de Olavide, Spain

Table of Contents

ISBN: 978-1-61804-233-0

ISBN: 978-1-61804-233-0

Keynote Lecture

Interpolation and Projective Representation in Computer Graphics, Visualization and Games

Vaclav Skala
University of West Bohemia
Plzen, Czech Republic
E-mail: skala@kiv.zcu.cz

Rongjiang Pan
Shandong University
Jinan, China
E-mail: panrj@sdu.edu.cn

Abstract: Today's engineering problem solutions are based mostly on computational packages. However the computational power doubles in 18 months. In 15 years perspective the computational power will be of 2^10 = 1024 of today's computational power. Engineering problems solved will be more complicated, complex and will lead to a numerically ill conditioned problems especially in the perspective of today available floating point representation and formulation in the Euclidean space.
Homogeneous coordinates and projective geometry are mostly connected with geometric transformations only. However the projective extension of the Euclidean system allows reformulation of geometrical problems which can be easily solved. In many cases quite complicated formulae are becoming simple from the geometrical and computational point of view. In addition they lead to simple parallelization and to matrix-vector operations which are convenient for matrix-vector hardware architecture like GPU.
In this short tutorial we will introduce "practical theory" of the projective space and homogeneous coordinates. We will show that a solution of linear system of equations is equivalent to generalized cross product and how this influences basic geometrical algorithms. The projective formulation is also convenient for computation of barycentric coordinates, as it is actually one cross-product implemented as one clock instruction on GPU. Selected examples of engineering disasters caused by non-robust computations will be presented as well.

Brief Biography of the Speaker: Prof.Vaclav Skala is a Full professor of Computer Science at the University of West Bohemia, Plzen, Czech Republic. He received his Ing. (equivalent of MSc.) degree in 1975 from the Institute of Technology in Plzen and CSc. (equivalent of Ph.D.) degree from the Czech Technical University in Prague in 1981. In 1996 he became a full professor in Computer Science. He is the Head of the Center of Computer Graphics and Visualization at the University of West Bohemia in Plzen (http://Graphics.zcu.cz) since 1996.
Prof.Vaclav Skala is a member of editorial board of The Visual Computer (Springer), Computers and Graphics (Elsevier), Machine Graphics and Vision (Polish Academy of Sciences), The International Journal of Virtual Reality (IPI Press, USA) and the Editor in Chief of the Journal of WSCG. He has been a member of several international program committees of prestigious conferences and workshops. He is a member of ACM SIGGRAPH, IEEE and Eurographics Association. He became a Fellow of the Eurographics Association in 2010.

ISBN: 978-1-61804-233-0

Prof.Vaclav Skala has published over 200 research papers in scientific journal and at international research conferences. His current research interests are computer graphics, visualization and mathematics, especially geometrical algebra, algorithms and data structures. Details can be found at http://www.VaclavSkala.eu

Prof. Rongjiang Pan is a professor in the School of Computer Science and Technology, Shandong University, China. He received a BSc in computer science, a Msc in computer science, a PhD in computer science from Shandong University, China in 1996, 2001 and 2005, respectively. During 2006 and 2007, he was a visiting scholar at the University of West Bohemia in Plzen under a program supported by the international exchange scholarship between China and Czech governments. He is now a visiting professor at the School of Engineering, Brown University from 2014 to 2105 under the support of China Scholarship Council.

He is a Member of the ACM. His research interests include 3D shape modeling and analysis, computer graphics and vision, image processing. He has published over 20 research papers in journal and at conferences

ISBN: 978-1-61804-233-0

ISBN: 978-1-61804-233-0

Notes on Bounded Harmonic Mapping Related to Starlike Functions

Durdane Varol, Melike Aydoğan and Yaşar Polatoğlu

***Abstract*—**.Let $f = h(z) + \overline{g(z)}$ be a sense-preserving harmonic mapping in the open unit disc $\mathcal{D} = \{z : |z| < 1\}$. If $f(z)$ satisfies the condition $\left| \frac{\frac{1}{b_1} g'(z)}{h'(z)} - M \right| < M$, $M > \frac{1}{2}$, then $f(z)$ is called bounded harmonic mapping. In this work, we will give some main properties of the class of bounded harmonic mappings.

***Keywords*—**Coefficient inequality, Distortion theorem, Growth theorem, Harmonic mappings, Subordination principle.

I. Introduction

Let Ω be the family of functions $\phi(z)$ regular in D and satisfying the conditions $\phi(0) = 0, |\phi(z)| < 1$ for every $z \in \mathcal{D}$.

Next, denote by P the family of functions $p(z) = 1 + p_1 z + p_2 z^2 + \dots$ regular in D and such that $p(z)$ is in P if and only if

$$p(z) = \frac{1 + \phi(z)}{1 + \phi(z)} \tag{1}$$

for some $\phi(z) \in \Omega$ and every $z \in D$.

Moreover, let S^* denote the family of functions $h(z) = z + c_2 z^2 + \dots$ regular in D and such that $h(z)$ is in S^* if and only if

$$z \frac{h'(z)}{h(z)} = p(z)$$

for some $p(z) \in P$ and $z \in \mathcal{D}$. Let $s_1(z) = z + d_2 z^2 + \dots$ and $s_2(z) = z + e_2 z^2 + \dots$ be analytic functions in the open unit disc D. If there exists a function $\phi(z) \in \Omega$ such that $s_1(z) = s_2(\phi(z))$ for all $z \in \mathcal{D}$, then we say that $s_1(z)$ is subordinate to $s_2(z)$ and we write $s_1(z) \prec s_2(z)$. Specially if $s_2(z)$ is univalent in D, then $s_1(z) \prec s_2(z)$ if and only if $s_1(\mathcal{D}) \subset s_2(\mathcal{D})$ implies $s_1(\mathcal{D}_r) \subset s_2(\mathcal{D}_r)$, where $\mathcal{D}_r = \{z : |z| < r, 0 < r < 1\}$. ([2]-[4])

Finally, a planar harmonic mapping in the open unit disc D is a complex-valued harmonic function $f(z)$ which maps D onto the some planar domain $f(\mathrm{D})$. Since D is a simply connected domain, the mapping f has a canonical decomposition $f(z) = h(z) + \overline{g(z)}$, where $h(z)$ and $g(z)$ are analytic in D and have the following power series expansion,

$$h(z) = \sum_{n=0}^{\infty} a_n z^n,$$

$$g(z) = \sum_{n=0}^{\infty} b_n z^n$$

where $a_n, b_n \in \mathbb{C}$, $n = 0, 1, 2, \dots$ as usual we call $h(z)$ the analytic part of $f(z)$ and $g(z)$ is co-analytic part of $f(z)$. An elegant and complete treatment theory of the harmonic mapping is given Duren's monograph [3].

Lewy [7] proved in 1936 that the harmonic mapping of $f(z)$ is locally univalent in D if and only if its Jacobien $J_f = |h'(z)|^2 - |g'(z)|^2$ is different from zero in D. In the view of this result, locally univalent harmonic mappings in the open unit disc D are either sense-preserving if $|h'(z)| > |g'(z)|$ in D or sense-reversing if

D. Varol is with the Department of Mathematics, Işık University, Meşrutiyet Köyü Şile, Istanbul, Turkey(e-mail: durdane.varol@isik.edu.tr).

M. Aydoğan is with the Department of Mathematics, Işık University, Meşrutiyet Köyü Şile, Istanbul, Turkey(e-mail: melike.aydogan@isikun.edu.tr).

Y.Polatoğlu is with the Department of Mathematics and Computer Science, Istanbul Kültür University, Istanbul, Turkey(e-mail: y.polatoglu@iku.edu.tr).

ISBN: 978-1-61804-233-0

$|g'(z)| > |h'(z)|$ in D. Throughout this paper, we will restrict ourselves to the study of sense-preserving harmonic mappings. We also note that $f(z) = h(z) + \overline{g(z)}$ is sense-preserving in D if and only if $h'(z)$ does not vanish in D and the second dilatation $\omega(z) = \frac{g'(z)}{h'(z)}$ has the property $|\omega(z)| < 1$ for all $z \in D$. Therefore, the class of all sense-preserving harmonic mappings in the open unit disc D with $a_0 = b_0 = 0$ and $a_1 = 1$ will be denoted by S_H. Thus, S_H contains standard class S of univalent functions. The family of all mappings $f \in S_H$ with the additional property $g'(0) = 0$, i.e., $b_1 = 0$ is denoted by S_H^0. Hence, it is clear that $S \subset S_H^0 \subset S_H$.

The main purpose of this paper is to investigate the class of harmonic mappings

$$S_H^*(M) = \left\{ \begin{array}{l} f = h(z) + \overline{g(z)} \in S_H \mid \\ \left| \dfrac{\frac{1}{b_1} g'(z)}{h'(z)} - M \right| < M, \\ M > \dfrac{1}{2}, h(z) \in S^*(M) \end{array} \right\}.$$

For this investigation we will need the following theorem and lemma.

Theorem 1 ([2]-[4])

Let $h(z)$ be an element of S^*, then

$$\frac{r}{(1+r)^2} \le |h(z)| \le \frac{r}{(1-r)^2}$$

and

$$\frac{1-r}{(1+r)^3} \le |h'(z)| \le \frac{1+r}{(1-r)^3}.$$

Theorem 2 ([2])

If $F(z)$ and $G(z)$ are regular in D, $F(0) = G(0)$, $G(z)$ maps D onto a many-sheeted region which is starlike with respect to the origin and $\frac{F'(z)}{G'(z)} \in P$ and $\frac{F(z)}{G(z)} \in P$.

Lemma 1 ([5])

Let $\phi(z)$ be regular in the unit disk D with $\phi(0) = 0$. Then if $|\phi(z)|$ attains its maximum value on the circle $|z| = r$ at the point z_1, one has $z_1 \phi'(z_1) = k\phi(z_1)$ for $k \ge 1$.

II. Main Results

Theorem 3

Let $f(z) = h(z) + \overline{g(z)}$ be an element of $S_H^*(M)$, then

$$\frac{g'(z)}{h'(z)} \prec b_1 \frac{1+z}{1+\alpha z} \tag{1}$$

where $\alpha = 1 - \frac{1}{M}$.

Proof. Since $f(z) = h(z) + \overline{g(z)}$ be an element of $S_H^*(M)$, then

$$h(z) = z + a_2 z^2 + \ldots \Rightarrow h'(z) = 1 + 2a_2 z + \ldots$$

$$g(z) = b_1 z + b_2 z^2 + \ldots \Rightarrow \frac{1}{b_1} g'(z) = 1 + 2\frac{b_2}{b_1} z + \ldots.$$

Thus

$$\frac{\frac{1}{b_1} g'(z)}{h'(z)} = \frac{G'(z)}{h'(z)}$$

$$\Rightarrow \left| \frac{G'(z)}{h'(z)} - M \right| < M$$

$$\Rightarrow \left| \frac{1}{M} \cdot \frac{G'(z)}{h'(z)} - 1 \right| < 1$$

$\varphi(z) = \frac{1}{M} \cdot \frac{G'(z)}{h'(z)} - 1$, $\varphi(z)$ is analytic and $\varphi(0) = \frac{1}{M} - 1$, therefore we consider the function

$$\phi(z) = \frac{\varphi(z) - \varphi(0)}{1 - \overline{\varphi(0)}\varphi(z)} = \frac{\frac{1}{M}\left(\frac{G'(z)}{h'(z)} - 1 \right)}{1 - \left(1 - \frac{1}{M}\right)\left(\frac{1}{M} \frac{G'(z)}{h'(z)} - 1 \right)}.$$

This function satisfies the conditions of Schwarz lemma. Then we can write

$$\frac{G'(z)}{h'(z)} = \frac{1 + \phi(z)}{1 + \alpha\phi(z)}, \quad \alpha = \frac{1}{M} - 1. \tag{2}$$

The equality (2) shows that

$$\frac{g'(z)}{h'(z)} \prec b_1 \frac{1+z}{1+\alpha z}.$$

Corollary 1

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S_H^*(M)$, then

$$F\left(|b_1|,\alpha,-r\right) \le |g'(z)| \le F\left(|b_1|,\alpha,r\right) \qquad (3)$$

where

$$F\left(|b_1|,\alpha,r\right)=\frac{1+r}{(1-r)^3}\cdot\frac{|b_1|(1+r)}{1+\alpha r}.$$

Proof. Since the transformation $\omega(z)=\frac{1+z}{1+\alpha z}$ maps $|z|=r$ onto the disc with the centre $C(r)=\left(\frac{1-\alpha r^2}{1-\alpha^2 r^2},0\right)$ and the radius $\rho(r)=\frac{(1-\alpha)r}{1-\alpha^2 r^2}$, then using Subordination or Lindelöf principle we can write

$$\left|\frac{G'(z)}{h'(z)}-\frac{1-\alpha r^2}{1-\alpha^2 r^2}\right| \le \frac{(1-\alpha)r}{1-\alpha^2 r^2}. \qquad (4)$$

After the simple calculations from (4) and using Theorem 1 we obtain (3).

Theorem 4

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S_H^*(M)$, then

$$\frac{g(z)}{h(z)} \prec b_1 \frac{1+z}{1+\alpha z}.$$

Proof. Using Corollary 2.2 we have

$$\left|\frac{1}{b_1}\frac{g'(z)}{h'(z)}-\frac{1-\alpha r^2}{1-\alpha^2 r^2}\right| \le \frac{(1-\alpha)r}{1-\alpha^2 r^2}$$

$$\Rightarrow \left|\frac{g'(z)}{h'(z)}-\frac{b_1(1-\alpha r^2)}{1-\alpha^2 r^2}\right| \le \frac{|b_1|(1-\alpha)r}{1-\alpha^2 r^2}.$$

Therefore we can write

$$\omega(\mathcal{D}_r)=\left\{z \,|\, \left|\frac{g'(z)}{h'(z)}-\frac{b_1(1-\alpha r^2)}{1-\alpha^2 r^2}\right| \le \frac{|b_1|(1-\alpha)r}{1-\alpha^2 r^2}\right\} \qquad (5)$$

Now we define the function $\phi(z)$ by

$$\frac{g(z)}{h(z)}=b_1\frac{1+\phi(z)}{1+\alpha\phi(z)}, \qquad (6)$$

then we have

$$\frac{g(0)}{h(0)}=b_1=b_1\frac{1+\phi(0)}{1+\alpha\phi(0)} \Rightarrow 1=\frac{1+\phi(0)}{1+\alpha\phi(0)} \Rightarrow \phi(0)=0.$$

$\phi(z)$ is analytic.

Now we show that $|\phi(z)|<1$ for all $z\in\mathcal{D}$. Indeed assume the contrary; there exists a point on $|z|=r$ such that $|\phi(z_1)|=1$. Taking the derivative from (6) and using Jack lemma (Lemma 1) we obtain

$$\frac{g'(z)}{h'(z)}=b_1\frac{1+\phi(z)}{1+\alpha\phi(z)}+b_1\frac{(1-\alpha)z\phi'(z)}{(1+\alpha\phi(z))^2}\cdot\frac{h(z)}{zh'(z)}$$

$$\Rightarrow \omega(z_1)=\frac{g'(z_1)}{h'(z_1)}$$

$$=b_1\left(\frac{1+\phi(z_1)}{1+\alpha\phi(z_1)}+\frac{(1-\alpha)z\phi'(z_1)}{(1+\alpha\phi(z_1))^2}\cdot\frac{1-r^2}{\left(1+r^2\right)+2re^{i\theta}}\right)\notin\omega(\mathcal{D}_r).$$

because $|\phi(z_1)|=1$ and $k\ge 1$. But this contradicts the condition (5) and so our assumption is wrong, i.e., $|\phi(z)|<1$ for all $z\in\mathcal{D}$.

Corollary 2

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S_H^*(M)$, then

$$\frac{r}{(1+r)^2}F_1\left(|b_1|,\alpha,-r\right) \le |g(z)| \le \frac{r}{(1-r)^2}F_1\left(|b_1|,\alpha,r\right) \qquad (7)$$

where

$$F_1\left(|b_1|,\alpha,r\right)=\frac{|b_1|(1+r)}{1+\alpha r}.$$

Proof. Since $\frac{g(z)}{h(z)} \prec b_1\frac{1+z}{1+\alpha z}$, then we have

$$\left|\frac{g(z)}{h(z)}-\frac{b_1(1-\alpha r^2)}{1-\alpha^2 r^2}\right| \le \frac{|b_1|(1-\alpha)r}{1-\alpha^2 r^2}. \qquad (8)$$

Using Theorem 1 and the inequality (8) we obtain (7).

Corollary 4

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S_H^*(M)$, then

$$\int \frac{|b_1|(1-r)}{1-\alpha r}\cdot\frac{1-r}{(1+r)^3}dr \le |f| \le \int \frac{|b_1|(1+r)}{1+\alpha r}\cdot\frac{1+r}{(1-r)^3}dr.$$

Proof. Using (5) we obtain

$$\frac{|b_1|(1-r)}{1-\alpha r} \le |\omega(z)| \le \frac{|b_1|(1+r)}{1+\alpha r}.$$

Therefore, we have

$$\begin{aligned} &\frac{(1+\alpha r)-|b_1|(1+r)}{1+\alpha r} \le 1-|\omega(z)| \\ &\le \frac{(1-\alpha r)-|b_1|(1-r)}{1-\alpha r}, \end{aligned} \tag{10}$$

$$\begin{aligned} &\frac{(1-\alpha r)+|b_1|(1-r)}{1-\alpha r} \le 1+|\omega(z)| \\ &\le \frac{(1+\alpha r)+|b_1|(1+r)}{1+\alpha r}. \end{aligned} \tag{11}$$

On the other hand, we have

$$\left(1-|\omega(z)|\right)|h'(z)||dz| \le |df| \le \left(1+|\omega(z)|\right)|h'(z)||dz|. \tag{12}$$

Considering (10), (11), (12) and Theorem 1, we get the desired result.

Theorem 5

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S_H^*(M)$, then

$$\begin{aligned} &\sum_{k=1}^{n}(k+1)^2\left|\frac{b_{k+1}}{b_1}-a_k\right|^2 \le (2-\frac{1}{M})^2 \\ &+\sum_{k=1}^{n-1}(k+1)^2\left|a_k+\left(1-\frac{1}{M}\right)\frac{b_{k+1}}{b_1}\right|. \end{aligned} \tag{13}$$

Proof. Since

$$\begin{aligned} &\frac{g'(z)}{h'(z)} \prec b_1\frac{1+z}{1+\alpha z} \\ &\Rightarrow \frac{g'(z)}{h'(z)} \prec b_1\frac{1+z}{1+\left(\frac{1}{M}-1\right)z} \\ &\Rightarrow \frac{g'(z)}{h'(z)} = b_1\frac{1+\phi(z)}{1-\left(1-\frac{1}{M}\right)\phi(z)}. \end{aligned}$$

$$\frac{\frac{1}{b_1}g'(z)}{h'(z)} = \frac{1+\phi(z)}{1-\left(1-\frac{1}{M}\right)\phi(z)}. \tag{14}$$

The equality (14) can be written in the following form

$$\frac{\frac{1}{b_1}g'(z)}{h'(z)} = \frac{G(z)}{H(z)} = \frac{1+\phi(z)}{1-\left(1-\frac{1}{M}\right)\phi(z)} \Rightarrow \frac{G(z)}{H(z)} = \frac{1+\phi(z)}{1-\alpha\phi(z)},$$

where $\alpha=\left(1-\frac{1}{M}\right)$.

Therefore, we have

$$\begin{aligned} &G(z)-H(z) = \left(H(z)+aG(z)\right)\phi(z) \\ &= \sum_{k=1}^{n}(d_k-e_k)z^k + \sum_{k=n+1}^{\infty}(d_k-e_k)z^k \\ &-\left(\sum_{k=n}^{\infty}(e_k+ad_k)z^k\right)\left(\sum_{k=1}^{\infty}c_kz^k\right) \\ &= \left[(a+1)+\sum_{k=1}^{n-1}(e_k+ad_k)z^k\right] \end{aligned} \tag{15}$$

where $d_k=\frac{(k+1)b_{k+1}}{b_1}$, $e_k=(k+1)a_{k+1}$, $k=1,2,\ldots$.

Equality (15) can be written in the following manner

$$\sum_{k=1}^{n}(d_k-e_k)z^k + \sum_{k=n+1}^{\infty}s_kz^k = \left[(a+1)+\sum_{k=1}^{n-1}(e_k+ad_k)z^k\right]\phi(z) \tag{16}$$

where the coefficients s_k have been chosen suitably and the equality (16) can be written in the form

$$F_1(z)=F_2(z)\phi(z), \quad |\phi(z)|<1$$

then we have

$$\begin{aligned} &|F_1(z)|^2 = |F_2(z)\phi(z)|^2 = |F_2(z)|^2|\phi(z)|^2 < |F_2(z)|^2 \\ &\Rightarrow \frac{1}{2\pi}\int_0^{2\pi}|F_1(re^{i\theta})|^2\,d\theta \le \frac{1}{2\pi}\int_0^{2\pi}|F_2(re^{i\theta})|^2\,d\theta \\ &\Rightarrow \sum_{k=1}^{n}|d_k-e_k|^2r^{2k} + \sum_{k=n+1}^{\infty}|s_k|^2r^{2k} \\ &\le \left[(a+1)^2r^{2k}+\sum_{k=1}^{n-1}\left(|e_k+ad_k|\right)^2r^{2k}\right] \end{aligned}$$

passing to the limit as $r\to 1$ we obtain (13). The method of this proof has been based on the Clunie method[1]

REFERENCES

[1] Duren, P. ,*Univalent Functions*, Springer Verlag, (1983).

[2] Duren, P. ,*Harmonic Mappings in the Plane*, Vol. 156 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge UK, (2004

[3] Goodman A. W. , *Univalent Functions*, Volume I, Mariner publishing Company INC, Tampa Florida, (1983).

[4] Jack, I.S. ,*Functions starlike and convex of order alpha*, J. London Math. Soc. 3, (1971), 369-374.

[5] Janowski, W. ,*Some extremal problems for certain families of analytic functions I,* Annales Policini Mathematici 28, (1973), 297-326.
[6] Lewys, H. ,*On the non-vanishing of the Jacobian in certain one-to-one mappings,* Bull. Amer. Math. Soc. 42, (1936), 689-692.

Turning a Serial Forward Code into a Parallel Inverse Code: A Case Study from Geothermal Engineering

H. Martin Bücker, M. Ali Rostami, and Ralf Seidler

***Abstract*—**Significant human effort is spent in developing numerical simulation codes that represent a forward problem in computational science and engineering. Typically, the only focus of these developments is on issues of the forward model, exclusively addressing how the output of interest is efficiently and accurately simulated for given input parameters. Though these issues may be numerous and intricate we argue that, most likely, the forward model will later also be used in a framework addressing inverse problems. Therefore, any design of a forward model should consider issues related to inverse modeling as well. We show by example of the geothermal engineering code SHEMAT that this requires to rethink design decisions that were originally taken for the forward model.

***Keywords*-**Inverse problems, optimal experimental design, automatic differentiation, software engineering, message passing, OpenMP, SHEMAT.

I. Introduction

ADVANCED mathematical models to describe and analyze complicated phenomena of interest have led to a tremendous increase of the understanding in various disciplines, including physical, engineering, financial, and life sciences. There is also a long tradition in transforming these mathematical models into computer models to be efficiently executed on state-of-the-art computer architectures. These substantial advances were only possible by bringing together profound knowledge and sophisticated techniques from applied mathematics, computer science, and the corresponding application disciplines.

Often, the design and implementation of these forward models require substantial human effort that is justified by the scientific findings resulting from using these models to better understand the given application problem at hand. It is not uncommon to use these forward models over and over again while varying certain input parameters of interest. However, these computer models are increasingly used as the starting point for further, more advanced investigations that go far beyond executing the forward model multiple times with different input parameters. The design of these forward models should therefore consider the requirements of an inverse framework as well. Unfortunately, today, most forward models are designed without taking into account that they will be used to solve inverse problems.

Here, we focus on the interplay between parallel computing and automatic differentiation when designing a forward model.

H. M. Bücker, M. A. Rostami, and R. Seidler are with the Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Jena, Germany.

Manuscript received January 31, 2014; revised March 31, 2014.

To the best of our knowledge, there is no previous publication dealing with this topic. Somehow related are approaches that try to reduce the computational complexity of a forward model that is used in an optimization framework. An example in the context of a geoscientific application is the approach in [1] that introduces a simplification of forward models in magnetotellurics to enable evolutionary optimization algorithms in three space dimensions. However, the reduction of the computational complexity is not the focus of the present article. The new contribution of the present article is to demonstrate the mutual dependencies between parallel computing used in various different ways for a forward model and automatic differentiation.

The outline of this paper is as follows. We first give a short overview on a forward model developed for the solution of geothermal engineering problems in Sect. II. We then show how the forward model is used in the context of inverse problems in Sect. III. Finally, in Sect. IV we describe the enabling technologies that need to fit together for the solution of problems arising from geothermal engineering.

II. Forward Modeling

The Institute for Applied Geophysics and Geothermal Energy, E.ON Energy Research Center, at RWTH Aachen University is currently developing a geothermal simulation package. This software is called Simulator for HEat and MAss Transport (SHEMAT) [2], [3]. It solves the coupled transient equations for groundwater flow, heat transport, and the transport of reactive solutes in porous media at high temperatures in three space dimensions. Here, we consider a mathematical model for fluid flow [2] and heat transport. The equation of fluid flow for the hydraulic potential (head) h_0 at location (x, z) and time t is given by

$$\rho_f g(\alpha+\psi\beta)\frac{\partial h_0}{\partial t} - \nabla\cdot\left(\frac{\rho_f g\kappa}{\mu_f}(\nabla h_0 + \rho_r \nabla z)\right) = W. \quad (1)$$

Here, ρ_f and ρ_r denote the fluid and the rock density, α and β represent the compressibilities of rock and fluid phase, respectively, and ψ is the porosity. The hydraulic permeability tensor is denoted by κ, while the fluid dynamic viscosity is represented by μ_f. The symbol g is used for the gravitational acceleration and W corresponds to a mass source term due to externally inflowing water.

ISBN: 978-1-61804-233-0

The temperature T is governed by the conductive-advective heat transport equation

$$(\rho c)_e \frac{\partial T}{\partial t} - \nabla \cdot (\lambda_e \nabla T) - (\rho c)_f \, \mathbf{a} \cdot \nabla T = H, \qquad (2)$$

where $(\rho c)_e$ denotes the effective heat capacity of the saturated porous medium and the fluid, λ_e is the effective thermal conductivity, $(\rho c)_f$ represents the volumetric heat capacity of the fluid, and H corresponds to a possible heat source term. Equations (1) and (2) are coupled via the Darcy velocity $\mathbf{a}$ satisfying Darcy's law

$$\mathbf{a} = -\frac{\kappa}{\mu_f}(\nabla P + \rho_f \, g \nabla z), \qquad (3)$$

where the pressure P depends on the head h_0. Given suitable initial and boundary conditions, a numerical scheme for the solution of (1)–(3) computes approximations of head h_0 and temperature T.

The numerical scheme implemented in SHEMAT consists of a finite-difference approach. After the discretization of the domain, the following iteration for the solution to the head equation is presented in [2]

$$-(1-\omega)K \cdot h_{\text{old}} - R \cdot h_{\text{old}} - W = \omega K \cdot h_{\text{new}} - R \cdot h_{\text{new}}, \quad (4)$$

in which ω is the time weighting parameter controlling the implicity or explicity of the solution approach, the symbol R is given by

$$R = \frac{\rho_f g(\alpha + \psi\beta)}{\Delta t},$$

and K is the conductivity matrix. This matrix is computed via

$$K = Ah_{k-1} + Bh_{j-1} + Ch_{i-1} + Dh + Eh_{i+1} + Fh_{j+1} + Gh_{k+1},$$

where A, B, C, D, E, F and G are the coefficients achieved by discretization and $h_{i,j,k}$ refers to the head in the three-dimensional cell (i, j, k). In this formula, the the indices of the head h are simplified in an obvious fashion, for example h_{i+1} means $h_{i+1,j,k}$ and h is used to denote $h_{i,j,k}$.

Equation (4) solves the steady-state flow equation. For a solution of the more general case, another time step iteration is needed. The following algorithm is a fixed-point method for this nonlinear problem implemented in SHEMAT.

1: **for** $t = 0, \Delta t, t_{\text{end}}$ **do**
2: Initial guess for h at the current time step, $h^t = h_0^{t+\Delta t}$, usually using the head value from the previous step
3: Compute time step control Δt
4: **for** $k = 1, 2, ..., \max$ (Fixed-point iteration) **do**
5: Setup the coefficient matrix $M(h_{k-1}^{t+\Delta t})$
6: The right-hand side: $B = -(1-\omega)Mh^t - Rh^t - W$
7: Solve $Mh_k^{t+\Delta t} = B$ for $h_k^{t+\Delta t}$
8: Compute the residuals: $RE_k = |h_k^{t+\Delta t} - h_{k-1}^{t+\Delta t}|$
9: Adaptive Relaxation: $h_k^{t+\Delta t} = (1 - \delta_k)h_{k-1}^{t+\Delta t} + \delta_k h_k^{t+\Delta t}$
10: **if** $\max |RE_k| \geq \epsilon$ **then**
11: $h^t = h_k^{t+\Delta t}$ and go to the step 3
12: **end if**
13: **end for**
14: **end for**

III. Going beyond Forward Modeling

Once a sophisticated forward model is available, a typical situation is to address—in some form or the other—an inverse problem related to that forward model. A prominent example is a parameter estimation problem. That is, the input parameters to the simulation are estimated so that they represent the real world as close as possible. In particular, these problems are ubiquitous in the geosciences. In that application area, a large number of properties of the subsurface are to be estimated from only a few data points obtained from some measurements.

More formally, a forward model computes approximations of, say, the temperature T at spatial position x for some unknown input parameters p. Throughout this article, we assume that the unknown parameter p is the hydraulic permeability κ. Conceptually, this forward model, which is described by (1)–(3) together with appropriate boundary and initial conditions, is nothing but a function

$$f(x, p). \qquad (5)$$

So, a parameter estimation problem is represented by minimizing a suitably-defined distance function between some measurements $m(x)$ and the solution of the forward problem $f(x, p)$. That is, the unknown parameters p are obtained from the solution of

$$\min_p \|m(x) - f(x, p)\|, \qquad (6)$$

where $\|\cdot\|$ is some norm.

In the context of SHEMAT, various algorithms for the solution of different inverse problems are implemented. These algorithms that make use of the forward model fall into the classes of stochastic techniques and deterministic techniques.

Stochastic or Monte-Carlo-like methods offer the advantage that large systems can be sampled in a number of random configurations, and that the data obtained from multiple forward models can be used to describe the system as a whole. In the context of SHEMAT, several stochastic methods are implemented, including sequential Gaussian simulation and ensemble Kalman filtering techniques [4]. These techniques make use of the forward model only in the form of evaluations of the forward model. A disadvantage of these techniques is that they may require a large number of these evaluations. The textbook [5] covers stochastic methods from a geoscientific point of view.

Deterministic methods for the solution of inverse problems tend to involve a smaller number of evaluations of the forward model. However, in addition to the evaluation of the forward model, they also need the derivatives of the forward model. For instance, a deterministic algorithm for the solution of (6) requires the evaluation of

$$\frac{\partial f(x, p)}{\partial p}. \qquad (7)$$

See the textbooks [6], [7], [8] for more details on these methods.

Optimal Experimental Design (OED) refers to a set of techniques to minimize the uncertainty in a parameter estimation

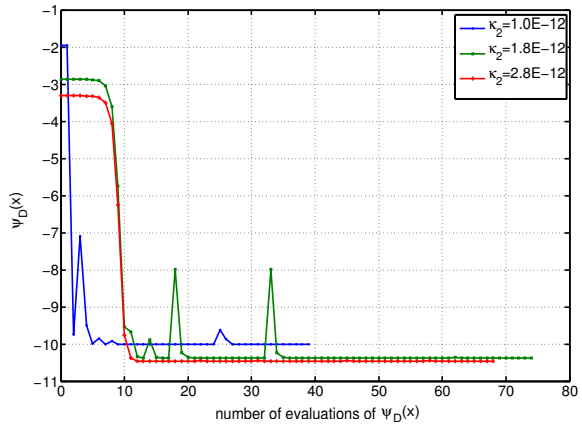


Fig. 1. Values of the objectives Ψ_D plotted versus the number of evaluations of Ψ_D carried out during the optimization process. Each curve corresponds to a different permeability value κ. See [9] for more details.

problem with respect to some experimental conditions. In the context of geothermal engineering, one can ask for questions

- Where to drill a borehole?
- When to take the measurements?

such that the uncertainty when estimating the hydraulic permeability in a certain rock layer from temperature measurements in a borehole is minimized.

More formally, OED solves an optimization problem of the form

$$\min_x \Psi(x) \tag{8}$$

where the objective $\Psi(x)$ quantifies the uncertainty of an experimental condition x. In the above example, the experimental condition is the position of a borehole. Typical objectives are

$$\Psi_A(x) = \log\bigl(\mathrm{trace}(F^{-1})\bigr)$$
$$\Psi_D(x) = -\log\bigl(\det(F)\bigr),$$

where

$$F = \left(\frac{\partial f(x,p)}{\partial p}\right)^T \left(\frac{\partial f(x,p)}{\partial p}\right) \tag{9}$$

is the so called Fisher Matrix. It is important to observe that gradient-based algorithms for the solution of (8) need the forward model not only in the form of evaluations of $f(x,p)$, and first-order derivatives like (7); they also require evaluations of mixed second-order derivatives

$$\frac{\partial^2 f(x,p)}{\partial p\,\partial x}. \tag{10}$$

As an example we consider the OED problem addressed in [9]. The problem there is to find the location to drill a borehole such that the uncertainty to estimate the hydraulic permeability from temperature measurements taken in the borehole is minimized. A geothermal reservoir in western Australia is considered in two space dimensions. Figure 1 shows one result of that particular discussion.

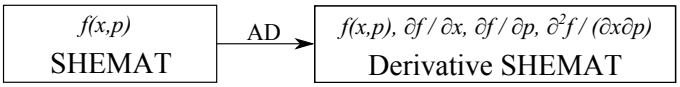


Fig. 2. The overall principle of automatic differentiation (AD) where a new program for the derivatives of the forward model is generated mechanically from the program of the forward model.

IV. Enabling Technologies

To successfully solve the above mentioned optimization problems, various building blocks are necessary. First of all, a forward model $f(x,p)$ needs to be developed. However, the development of forward problems—though complicated, challenging, and often needing considerable human effort—is a "standard procedure." Therefore, we do not address this building block in this article.

Another building block for the algorithms sketched in the previous discussion is the availability of derivatives of the forward model. More precisely, we need first-order derivatives of the form (7) and mixed second-order derivatives (10). One option is to derive these derivatives in an analytic way and implement another computer model for these derivatives by hand. Since this requires too much human effort for most real-world problems, we do not consider this to be feasible. Numerical differentiation by divided differencing is another option, but this inherently involves truncation error so that the derivatives are always inaccurate. Therefore, we rely on derivatives computed by automatic differentiation (AD) [10], [11]. Given a program to evaluate the forward model $f(x,p)$, the set of AD techniques mechanically generate another program for the evaluation of its derivatives without truncation error. The overall AD transformation process is schematically depicted in Fig. 2.

As soon as the problem sizes increases, for instance by increasing the resolution of the grid or the number of parameters to be estimated, the time to compute the overall solution process becomes prohibitively large. Often, the storage requirement exceeds the amount of available memory on a conventional computer architecture too. So, it is quite common that parallelism has to be introduced to solve these challenging problems. SHEMAT involves various different forms of parallelism. From the programming model point of view, it is parallelized with MPI [12] for distributed-memory computers as well as with OpenMP [13] for shared-memory computers.

In particular, the Monte Carlo methods are parallelized using OpenMP. The reason is that these methods rely on (mostly) independent evaluations of the forward model at different input values. This inherently-parallel structure is easily parallelized with OpenMP which is the de facto standard for shared-memory parallel programming. For the Monte Carlo methods, OpenMP is used outside of the forward model. We stress that there are parts inside the forward model that are also parallelized with OpenMP.

Shared-memory computers do not scale to a very large number of processes. Distributed-memory computers allow to increase the number of processors further. Therefore, the forward model of SHEMAT is currently redesigned to also

involve parallelization based on MPI. This is tricky since the usual hybrid parallel programming paradigm is based on using OpenMP from within an MPI process. However, in the Monte Carlo methods of the current SHEMAT version, there would be need to use a hybrid MPI/OpenMP program within a parallel OpenMP region which is currently considered to be unreasonable.

Furthermore, there is currently no AD software that is capable of reliably transforming programs using MPI and/or OpenMP. Our strategy is therefore to implement a library that hides most of these parallel programming constructs form the AD transformation process. There is also another problem arising from the various levels of parallelism currently implemented in SHEMAT. Given a serial version of the forward model, there is a way to automatically parallelize the derivatives in the AD-generated code. This parallelization strategy is orthogonal to the parallelization strategy used to parallelize multiple instances of forward solvers in the Monte Carlo methods [14]. However, both these parallelization strategies need to work together. Moreover, they also need to be compliant to the new MPI parallelization of the forward model.

V. Conclusion

If a forward model is used in the context of an inverse problem, there are many intricate issues that have to be considered. In particular, we focus on a case study from geothermal engineering where a serial code for the forward model is turned into a parallel code used in an inverse framework. In fact, some parts of the forward model are already parallelized. We demonstrate the need for a clear design of the forward model and show the various implications that make difficult the redesign of this forward code in the context of its integration into a parallel inverse framework. The major difficulties arise from the various levels of parallelism. Some of these parallel levels are orthogonal to each other while others are nested and, thus, involve dependencies. Another challenge is to cope with parallelism in combination with automatic differentiation.

Acknowledgment

The authors thank our colleagues from the E.ON Energy Research Center and the Institute for Mathematics (Continuous Optimization), both at RWTH Aachen University, for a fruitful collaboration. This work is partially supported by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) within the project MeProRisk II, contract number 0325389 (F) as well as by the German Federal Ministry of Education and Research (BMBF) within the project HYDRODAM, contract number 01DS13018.

References

[1] B. Alexander, S. Thiel, and J. Peacock, "Application of evolutionary methods to 3D geoscience modelling," in *Proceedings of the Fourteenth International Conference on Genetic and Evolutionary Computation Conference*, ser. GECCO '12. New York, NY, USA: ACM, 2012, pp. 1039–1046. [Online]. Available: http://doi.acm.org/10.1145/2330163.2330308

[2] J. Bartels, M. Kühn, and C. Clauser, "Numerical simulation of reactive flow using SHEMAT," in *Numerical Simulation of Reactive Flow in Hot Aquifers*, C. Clauser, Ed. Springer Berlin Heidelberg, 2003, pp. 5–74.

[3] V. Rath, A. Wolf, and H. M. Bücker, "Joint three-dimensional inversion of coupled groundwater flow and heat transfer based on automatic differentiation: Sensitivity calculation, verification, and synthetic examples," *Geophysical Journal International*, vol. 167, no. 1, pp. 453–466, 2006.

[4] C. Vogt, G. Marquart, C. Kosack, A. Wolf, and C. Clauser, "Estimating the permeability distribution and its uncertainty at the EGS demonstration reservoir Soultz-sous-Forêts using the ensemble Kalman filter," *Water Resources Research*, vol. 48, no. 8, p. W08517, 2012. [Online]. Available: http://dx.doi.org/10.1029/2011WR011673

[5] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004.

[6] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[7] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: John Wiley & Sons, 1987.

[8] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs: Prentice-Hall, 1983.

[9] R. Seidler, H. M. Bücker, K. Padalkina, M. Herty, J. Niederau, G. Marquart, and A. Rasch, "Redesigning the EFCOSS framework towards finding optimally located boreholes in geothermal engineering," accepted for publication at TMCE 2014, Budapest.

[10] L. B. Rall, *Automatic Differentiation: Techniques and Applications*, ser. Lecture Notes in Computer Science. Berlin: Springer, 1981, vol. 120.

[11] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd ed., ser. Other Titles in Applied Mathematics. Philadelphia, PA: SIAM, 2008, no. 105.

[12] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI: Portable Parallel Programming with the Message-passing Interface*, ser. Scientific and engineering computation. MIT Press, 1999, no. 1.

[13] B. Chapman, G. Jost, and R. van der Pas, *Using OpenMP: Portable Shared Memory Parallel Programming*, ser. Scientific Computation Series. MIT Press, 2008, no. 10.

[14] A. Wolf, "Ein Softwarekonzept zur hierarchischen Parallelisierung von stochastischen und deterministischen Inversionsproblemen auf modernen ccNUMA-Plattformen unter Nutzung automatischer Programmtransformation," Dissertation, Department of Computer Science, RWTH Aachen University, 2011.

Hybrid Algorithm for Clustering of Microarray Data

Emir Buza[1], Zikrija Avdagic[1], Samir Omanovic[1] and Aida Hajdarpasic[2]
[1]Faculty of Electrical Engineering, Department for Computer Science and Informatics
[2]Faculty of Medicine, Center for Genetics
University of Sarajevo
{emir.buza, zikrija.avdagic, samir.omanovic}@etf.unsa.ba, aida.hajdarpasic@gmail.com

***Abstract*—Clustering is a crucial step in the analysis of gene expression data. Its goal is to identify the natural clusters and provide a reliable estimate of the number of distinct clusters in a given data set. In this paper we propose new hybrid algorithm for clustering of microarray data based on spectral clustering and k-means. Our algorithm consist of four steps, including preprocessing or filtering step, and finding optimal number of clusters by using two different clustering methods based on hierarchical and partition-based approaches. Then, we cluster data based on similarity/dissimilarity metrics with spectral clustering. In the final step, we select centroid genes based on k-means results. The proposed method was tested on six data sets from GEMS microarray database. When compared with existing single or combination of clustering methods, our results indicate about 10% improvement in selection of representative genes.**

Keywords Bioinformatics, Clustering Analysis, Gene expression, Data mining, Spectral clustering, Microarray data

I. Introduction

Microarrays have been used in various biomedical application such as gene discovery [1], [2], disease diagnosis [3], pharmacogenomics (drug discovery) [4], and toxicology [5]. DNA microarray technology is currently the most popular technology widely used for gene expression profiles. It is common to process thousands of genes from a large set of information in a microarray dataset in one experiment [6], [7]. Microarray data is usually represented as a matrix ($N \times M$), where each row represents a gene and each column an experiment. There are two approaches for detecting gene groups: *biclustering* and *clustering* [8]. Row-column clustering is performed simultaneously in biclustering approach, while in clustering approach either row (gene) or column (experiment) clustering is performed. In this paper we focus on clustering approach.

It is very difficult to make the right classification of data by using only pure statistical methods, especially in cases where the number of dimensions (number of samples) reaches up to several tens of thousands. The first problem is *prepocessing* or *filtering* less important information (information that do not contribute to good classification of genes/samples) from those that are crucial for further analysis. The goal is to transform data set into a improved, appropriate and reliable form for clustering. Thus, the quality of the clustering is increased. The second problem is *selection of representative genes* which have highly relevant information about disease from a large number of genes in a data set.

The general problem for multidimensional data sets whose deep structure is unknown is that the number and distribution of clusters is unknown in advance. Thus, the well-known clustering problem is how can one decide how many representative weight vectors should be used? When a number and distribution of the input clusters is known in advance, the selection of a representative can be found in a few simple computations.

In general, there are three main methods for microarray classification: *supervised*, *unsupervised* and *statistical* methods. The first category includes methods [9], [10], [11] for classification which require training and testing steps. However, in practice, microarray data sets usually contain a large number of genes and relatively small number of samples, so it is very difficult to make a good selection microarray data on subsets for training and testing. On the other hand, these methods usually suffer from a problem of overfitting, thus the characteristics of training set are memorized instead of capturing the desired pattern. Thus, additional knowledge about structure of data set and its classes is usually needed.

The second category includes clustering methods [12], [8], [13], [14] which do not require special knowledge for classification of data into a predefined number of groups. However, these methods usually suffer from a problem of initial number of clusters and starting centroid points, especially partitioning based clustering methods. Statistical methods [22], [24] are very popular and usually used as the first step of preprocessing [25] for all above methods. These kind of methods require a lot of time for execution, especially if microarray has a lot of dimensions. The second problem is that they need the help of other methods for completing final classification task. Thus, we need a more effective method and software tools for analyzing gene expression profiles of the microarray experimental samples [26].

In this paper we propose new hybrid algorithm for clustering of microarray data based on spectral clustering and k-means. Our hybrid algorithm is composed of four steps. In the first step, we pre-process data into a reduced number of data dimensions (genes) by using only statistical methods such as average value and standard deviation. In the second step, we determine optimal number of clusters by using two different clustering methods based on hierarchical and partition-based approaches. We use two different instances with two distance metrics. In the third step, we cluster data by similarity/dissimilarity metrics with spectral clustering, based on previously determined number of clusters. In the final step, we select centroid genes based on k-means results from the second step and results from spectral clustering.

The paper is organized as follows: background information on clustering is briefly reviewed in section II. In section III, we propose new hybrid algorithm for clustering of microarray

ISBN: 978-1-61804-233-0

data. Implementation and results are presented in section IV, and in section V we conclude the paper with some remarks.

II. Background

Data clustering is a method of grouping objects into meaningful categories. Clustering is used for discovery of similarity degree among forms, data exploration to discover underlying structures, compression for organizing data and other application for any scientific field that collects data. Supervised clustering methods use labels for easier cluster identification. However, labeling a large set of sample patterns can be costly, thus *unsupervised* methods are used. Some advantages of unsupervised methods are detection of the gradual change of pattern over time, identification of features useful for categorization, gaining insight into the nature or structure of the data during the early stages of an investigation.

In the context of expression profile for microarray data clustering has four main steps: 1) feature selection or extraction, 2) design of clustering algorithm, 3) validation of clusters and 4) interpretation of results. In recent years, there are many published studies in this field [8], [13], [14], [15], [16], [17]. In summary [6], these approaches can be grouped in five major groups based on: a) *similarity and distance measures*, b) *hierarchical clustering*, c) *partition clustering*, d) *model-based* and e) *feature-based methods*. The goal of these approaches is to group objects into disjoint clusters in such a way that objects with high similarity to each other belong to one cluster.

Methods based on similarity and distance measures vary in the way the distance between data is measured. The examples of distances often used are: Euclidean distance, correlation distance based on Pearson correlation coefficient, Mahalanobis distance, cosine angle and squared Pearson correlation. Hierarchical clustering is divided into agglomerative and divisive. Agglomerative algorithms begin with individual gene expression pattern cluster and successively merge them into smaller groups (bottom-up approach), while divisive approach start with whole set and divide it into smaller groups (top-down approach).

Partition clustering starts with pre-defined number of clusters with initial cluster centroids, allocate each data point to the cluster which has the nearest centroid, and compute the new centroids of the clusters. Then, alternate between second and third step until no data points change clusters [18]. Two classic representative algorithms are k-center and k-means. Model-based approaches are based on the fundamental relation between expression profile and a function of model parameters such as mixture-model and hidden Markov model. Feature-based methods focus on overall shapes of characteristic features, such as local shape-based similarity measure and scale-space signals. An exhaustive reviews of exiting approaches for the analysis of gene expression data can be found in [8], [14].

III. Hybrid algorithm

Our hybrid algorithm is composed of four steps as illustrated in Fig. 1: A. Data Set Prepocessing, B. Selection of the number of clusters, C. Spectral clustering and D. Selection of representative genes. In this section, we describe each step in detail.

A. Data Set Prepocessing

The first and preliminary step for classification and clustering methods is data cleaning and reduction of data dimensionality. Experimental studies have shown that microarray data sets contain bigger number of genes than the number of samples. Most of these genes are without appropriate expression values, thus they do not carry any important information for experiments.

In general, methods for dimension reduction can be categorized using different criteria like: normalization, discretization to binary values, simple statistical techniques (average value, standard variation, standard deviation, or combination of these), selection of the best original features and selection of a representative gene.

For our case we used statistical methods such as average value and standard deviation for cleaning (reduction) of mircoarray data as described in Eq. 1, Eq. 2 and Algorithm 1.

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^{N}(x_{ij}), \forall j = 1, 2, ..., M \tag{1}$$

$$s_j = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2}, \forall j = 1, 2, ..., M \tag{2}$$

Algorithm 1 Cleaning (reduction) microarray Data set

Require:
$data$ - microarray dataset;
$\bar{x}_j$ - average value by j column;
s_j - standard deviation by columns;

$k \leftarrow 0$ ▷ Set k to 0
for $j \leq$ size of number columns of $data$; $j \leftarrow j + 1$ **do**
  if $(s_j > \delta \cdot \bar{x}_j)$ **then**
    $k \leftarrow k + 1$ ▷ increment k when if statement is true
    for $i \leq$ size of number rows of $data$; $i \leftarrow i + 1$ **do**
      $g_{ik} \leftarrow x_{ij}$
    end for
  end if
end for
return (G) ▷ Return reduced Data set.

In Algoritam 1, $1 \leq \delta \leq 2$. For our case, δ is set to 1.75.

B. Selection of the number of clusters

For determining the optimal number of clusters, we used four clustering algorithms: two k-means and two Hierarchical Agglomerative Clustering (HAC), as illustrated in Fig. 1.B. In other words, we use two clustering methods with different similarity (disimilarity) measures. In our case, for the distance measure we selected Euclidean distance and the correlation distance based on Pearson correlation coefficient. After that the process of clustering of reduced microarray data set is repeated C_{max} times for each of four selected clustering algorithms. Upper boundary C_{max} in the worst case could be specified as

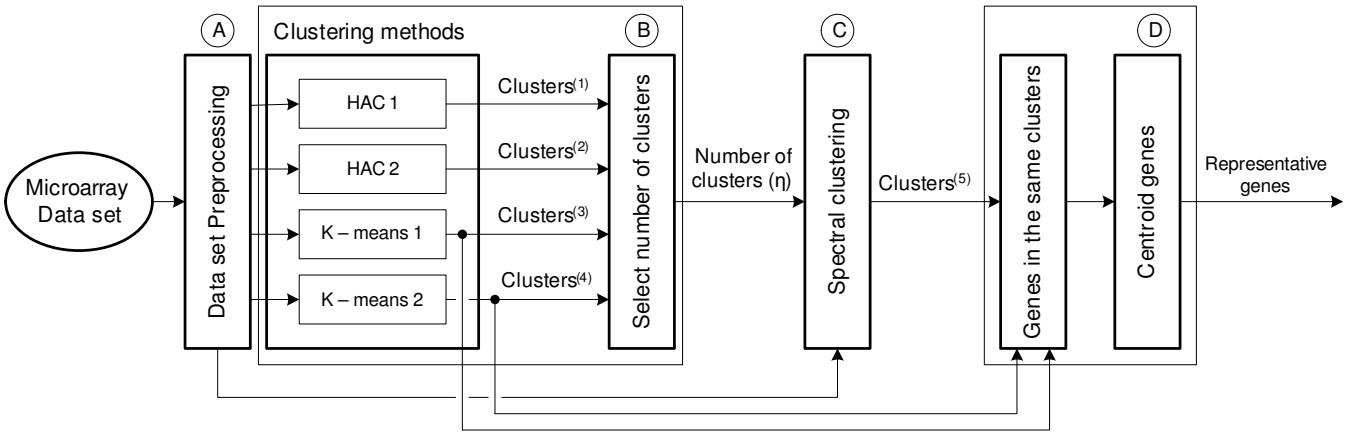


Fig. 1. Basic steps of Hybrid Algorithm for Clustering of Microarray data and selection of representative genes.

$N/2$, where N is number of experiments in the data set. In better case C_{max} should be selected so that it is sufficiently large than the optimal number of clusters (this upper boundary could be specified based on our knowledge of the data set).

During the process of execution of methods clustering, for each clustering algorithm l $(l = 1, 2, ..., 4)$ we calculated the sum of the variances of points within predefined number of clusters $k = 1, 2, ..., C_{max}$ based on Eq. 3.

$$Q_{lk} = \sum_{k=1}^{C_{max}} ((\sum_{x_i \in C_k} 1) - 1) \times \frac{1}{\sum_{x_i \in C_k} 1} \times \sum_{j=1, x_i \in C_k}^{M} (x_{ij} - \bar{x}_j)^2) \quad (3)$$

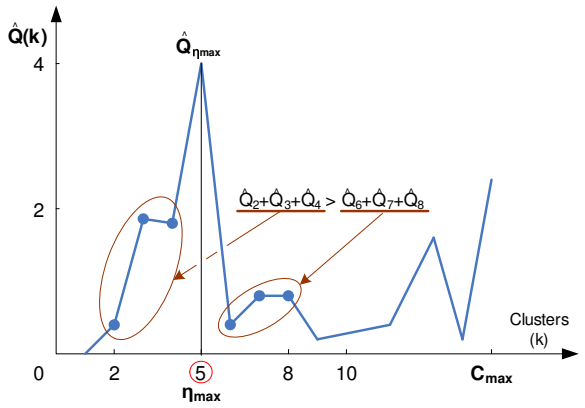
Suppose Q_{1k}, Q_{2k}, Q_{3k} and Q_{4k} are sums of variances for each of four clustering methods executed based on Eq. 3, where is $k = 1, 2, ..., C_{max}$, respectively. Finally, the optimal number of clusters η is calculated based on Eq. 4, Eq. 5, and Eq. 6.

$$\hat{Q}_k = \frac{1}{4} \sum_{l=1}^{4} (Q_{lk} - (\frac{1}{4} \sum_{j=1}^{4} Q_{jk}))^2, \forall k = 1, 2, ..., C_{max} \quad (4)$$

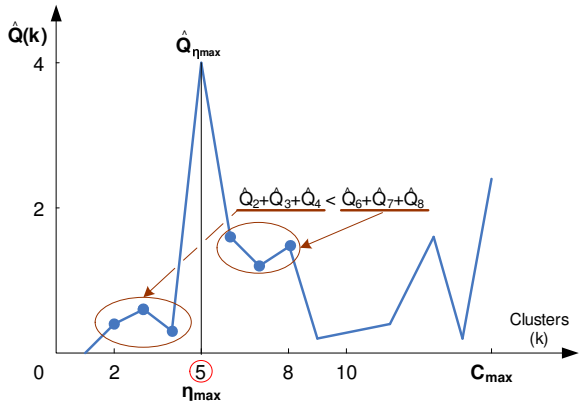
$$\eta_{max} = \arg \max_{2 \le k \le C_{max}} \{\hat{Q}_k\}, \quad (5)$$

$$\eta = \begin{cases} \eta_{max} - 1, & \text{if } \sum_{i=2}^{\eta_{max}-1} \hat{Q}_i > \sum_{\eta_{max}+1}^{2(\eta_{max}-1)} \hat{Q}_i \\ \eta_{max}, & \text{if } \sum_{i=2}^{\eta_{max}-1} \hat{Q}_i \le \sum_{\eta_{max}+1}^{2(\eta_{max}-1)} \hat{Q}_i \end{cases} \quad (6)$$

From Eq. 6, the maximal number of clusters C_{max} in the best case is selected as $C_{max} = 2 \times \eta_{max}$. A graphical representation of the process of finding the optimal number of clusters is shown in Fig. 2. Point η_{max} represents the value of k $(k = 1, 2, ..., C_{max})$, where the variance $\hat{Q}_k$ has the maximal value in point $\hat{Q}_{\eta_{max}}$.



(a) Left side is greater than the right side, i.e. the optimal number of clusters is calculated as $\eta = \eta_{max} - 1$.



(b) Right side is greater than the left side, i.e. the optimal number of clusters is calculated as $\eta = \eta_{max}$.

Fig. 2. Graphical representation of the process of finding the maximal and optimal number of clusters.

C. Spectral clustering

Recently, spectral clustering [19] became one of the most popular clustering algorithms. Traditional clustering algorithms, such as the k-means algorithm, use only simple metrics such as Euclidean distance to calculate distances between points in a data set and then make clusters by distance values (max or min). Unfortunately, they are not good enough for

clustering data into a predefined number of clusters.

In this paper we used the most commonly used normalized spectral clustering algorithm according to Ng et.al [20] and [21]. The basic technique of the spectral clustering is to perform dimensionality reduction before clustering in fewer dimensions. As input in the spectral clustering we used affinity matrix $A = a(_{ij}), i, j = 1, 2, .., n$ [19], [20], [23], which contains relative similarity for each pair of n points in the data set. This affinity matrix [23] is typically defined as $e^{\frac{-d^2}{\sigma^2}}$, in a way similar to the Gaussian kernel based on inter-point euclidean distance, where d is Euclidean distance between points and σ is a scale factor.

The basic steps of this algorithm are presented below:

- input: number k of clusters which need to construct
- for a given dataset of n points $X = x_1, ..., x_n \in R^l$ form the affinity matrix $A \in R^{n \times n}$ which is defined by $a_{ij} = e^{\frac{-d^2(x_i, x_j)}{\sigma^2}}$, $i, j = 1, 2, ..., n$, where $d(x_i, x_j)$ is some distance function (e.g. Euclidean) between points x_i and x_j
- compute degree matrix $D = diag(d_i)$ where $d_i = \sum_{j=1}^{n} a_{ij}$
- compute the normalized Laplacian matrix $L = D^{\frac{-1}{2}} \times L \times D^{\frac{1}{2}}$, where L is Laplacian matrix defined as $L = D - S$
- perform the eigen value decomposition by equation $L \times \upsilon = \lambda \times \upsilon$, where $\upsilon \in R^{n \times n}$ matrix of eigen vectors and $\lambda \in R^{n \times n}$ is matrix of eigen values
- form matrix $U \in R^{n \times k}$ from matrix υ, $u_{ij} = \upsilon_{im}$ where $i = 1, 2, ..., n; j = 1, 2, ..., k$ and m is k largest eigen vectors which are selected as the last k columns from matrix υ
- construct the normalized matrix Y from the obtained matrix U $y_{ij} = \frac{u_{ij}}{(\sum_{l=1}^{k} u_{il}^2)^{\frac{1}{2}}}$, where $i = 1, 2, ..., n; j, l = 1, 2, ..., k$
- clustering n points $y_i \in R^k, i = 1, 2, ..., n$ with K-means algorithm into $C_1, C_2, ..., C_k$ clusters

For our case, we found that the affinity matrix A (Eq. 7) gives the good results for microarray data set.

$$A = e^{-\sin(\frac{\arccos(R)}{2})^2} \tag{7}$$

where R is Pearson's correlation coefficient given by Eq. 8, x and y are points in d-dimensional space, $\bar{x}, \bar{y}$ are average values for x and y, and n is number of variables.

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{8}$$

D. Selection of representative genes

Selection of representative genes is executed by Algorithm 2. Selection is performed based on results of Spectral clustering and selective k-means results. Genes which exists in the same clusters (from results of Spectral and two k-means clustering) are selected as potential representative genes. In the second step, only genes with minimal Euclidean distance between potential selected genes ($g_k \in G$) and center of Spectral clusters are selected as the best representative genes ($\hat{X}$).

Algorithm 2 Selection of representative genes

Require:
X - microarray data set;
$C^{(e)}$ - k-means (Euclidean) clusters;
$C^{(c)}$ - k-means (Correlation) clusters;
$C^{(s)}$ - Spectral clustering clusters for number of clusters;

Relabel $C_k^1 \leftarrow C_k^{(e)}$ according to centers of clusters $C^{(s)}$ and $C^{(e)}$, where $C_k^{(e)} = \{C_1^{(e)}, C_2^{(e)}, ..., C_\eta^{(e)}\}$ for k = 1, 2, ... η
Relabel $C_k^2 \leftarrow C_k^{(c)}$ according to centers of clusters $C^{(s)}$ and $C^{(c)}$, where $C_k^{(c)} = \{C_1^{(c)}, C_2^{(c)}, ..., C_\eta^{(c)}\}$ for k = 1, 2, ... η
for i=1 to 2; $i \leftarrow i + 1$ **do**
  for k=1 to η; $k = k + 1$ **do**
    $G_k^i \leftarrow X_{C_k^i} \bigcap X_{C_k^{(s)}}$
  end for
end for
for k=1 to η; $k = k + 1$ **do**
  $G_k \leftarrow G_k^1 \bigcup G_k^2$
  $\hat{x}_k \leftarrow \min_{i, g_{ik} \in G_k} d^2(m_k^{(s)}, g_{ik})$, where $\{\hat{x}_k\} \in \hat{X}$
  ▷ d^2 is Euclidean square distance between $m_k^{(s)}$ (center of cluster $C_k^{(s)}$) and g_{ik}.
end for
return $(\hat{X})$ ▷ Return representative genes

IV. Implementation and Results

A. Implementation

Our method has been implemented in MATLAB version 7.11.0 (R2010b) with Statistics Toolbox. The clustering was performed on Intel Core2 1.80 GHz CPU with 4GB of RAM.

The effectiveness of our method has been verified on data sets selected from [27]. This approach allows convenient verification and comparison with similar algorithms or methods.

B. Results

In order to evaluate our hybrid algorithm we used several data sets. Data sets are used for evaluation our hybrid algorithm as described in Table I. The first column is data set identification, the second is number of clusters, the third is number of experiments, and the forth is number of genes in a data set.

Some results are presented in Fig. 3, 4, 5, where (a) depicts the sum of the variances of points within predefined number of clusters $k = 1, 2, ..., C_{max}$ (Eq. 3), and (b) depicts the optimal number of clusters η (Eq. 4).

We compare our results manually with other existing clustering models including the spectral clustering, k-means and hierarchical agglomerative clustering with different selected measures. We emphasize the fact that our main goal it to

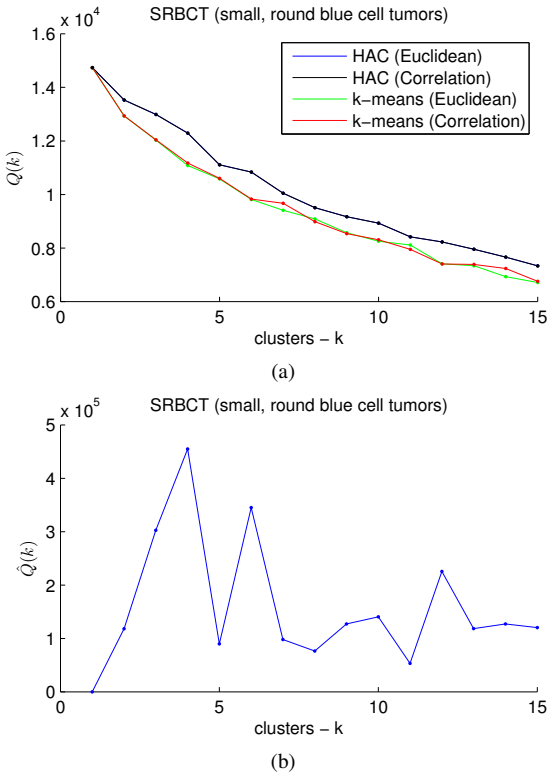


Fig. 3. Representation of Small, Round Blue Cell Tumors (SRBCT) Data set, (a) the sum of the variances of points within predefined number of clusters $k = 1, 2, ..., C_{max}$ (Eq. 3), (b) the optimal number of clusters $\eta = 4$ (Eq. 4).

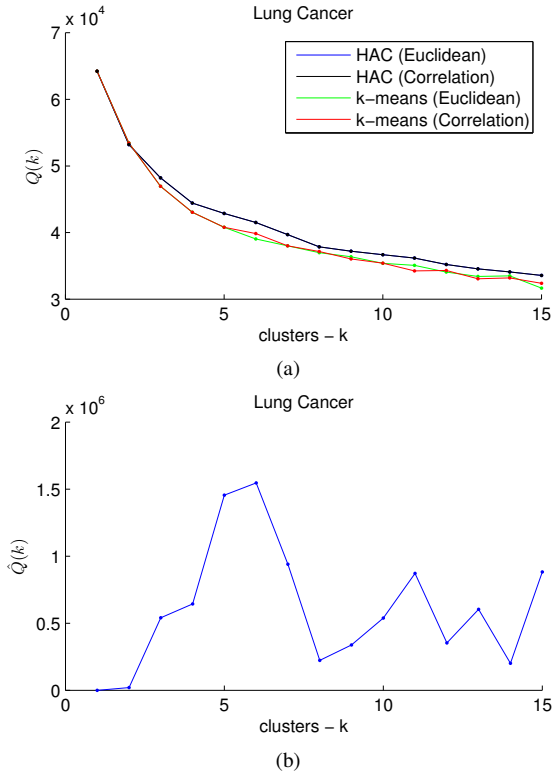


Fig. 4. Representation of Lung Cancer Data set, (a) the sum of the variances of points within predefined number of clusters $k = 1, 2, ..., C_{max}$ (Eq. 3), (b) the optimal number of clusters $\eta = 5$ (Eq. 4).

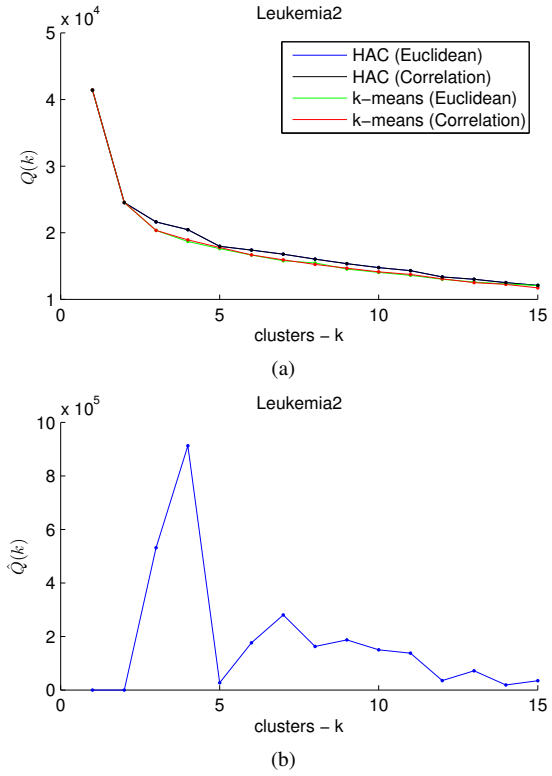


Fig. 5. Representation of Leukemia2 Data set, (a) the sum of the variances of points within predefined number of clusters $k = 1, 2, ..., C_{max}$ (Eq. 3), (b) the optimal number of clusters $\eta = 3$ (Eq. 4).

TABLE I. DATA SETS USED FOR EXPERIMENTATION

Data set name	Clusters	Experiments	Genes
Prostate Tumor	2	102	10509
SRBCT	4	83	2308
Lung Cancer	5	203	12600
DLBCL	2	77	5469
Brain Tumor2	4	50	10367
Leukemia2	3	72	11225

show the effectiveness of our hybrid algorithm in selection of representative genes [28], rather than application of single clustering method or their combination when using large microarray data sets [27].

The estimation accuracy when compared with existing single or combination of clustering methods is about 10% improved. We think that this information can be used for better prediction of unknown microarray genes.

V. CONCLUSION

In this paper we propose new hybrid algorithm for clustering of microarray data based on spectral clustering and k-means. Spectral clustering is a simple method for finding structure in data, which uses spectral properties of an associated pairwise similarity matrix.

The proposed method was tested on six Data sets from GEMS microarray data sets [27]. On a given set of data, our method improved selection of representative genes for about 10% when compared with existing single or combination of clustering methods. Thus, it is suitable for classification of unknown microarray genes.

ISBN: 978-1-61804-233-0

REFERENCES

[1] Nan, Xiaofei and Wang, Nan and Gong, Ping and Zhang, Chaoyang and Chen, Yixin and Wilkins, Dawn, *Biomarker Discovery Using 1-norm Regularization for Multiclass Earthworm Microarray Gene Expression Data*, Neurocomputing, Vol. 92, pp. 36–43, 2012

[2] Chu, Wei and Ghahramani, Zoubin and Falciani, Francesco and Wild, David L., *Biomarker Discovery in Microarray Gene Expression Data with Gaussian Processes*, Bioinformatics, Vol. 21(16), pp. 3385–3393, 2005

[3] Xin Zhao and Leo Wang-Kit Cheung, *Multiclass Kernel-Imbedded Gaussian Processes for Microarray Data Analysis*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, IEEE/ACM Transactions, Vol. 8(4), pp. 1545–5963, 2011

[4] Paul F. Predki, *Functional Protein Microarrays in Drug Discovery*, CRC Press, 2007

[5] Teresa Lettieri, *Recent Applications of DNA Microarray Technology to Toxicology and Ecotoxicology*, Environmental Health Perspectives, Vol. 114(1), pp. 4–9, 2006

[6] Amphun Chaiboonchoe and Sandhya Samarasinghe and Don Kulasiri, *Machine Learning for Childhood Acute Lymphoblastic Leukaemia Gene Expression Data Analysis: A Review*, Current Bioinformatics, Vol. 5(2), pp. 118–133, 2010

[7] Michael J. Korenberg, *Microarray data analysis: methods and applications*, Humana Press, 2007

[8] Jiang, Daxin and Tang, Chun and Zhang, Aidong, *Cluster analysis for gene expression data: A survey*, Knowledge and Data Engineering, IEEE Transactions, Vol. 16(11), pp. 1370–1386, 2004

[9] Perez M., Marwala T., *Microarray data feature selection using hybrid genetic algorithm simulated annealing*, Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention, Humana Press, pp. 1–5, 2012

[10] Greer, Braden, and Javed Khan, *Online analysis of microarray data using artificial neural networks*, Microarray Data Analysis, Humana Press, pp. 61–73, 2007

[11] Wutao Chen, Huijuan Lu, Mingyi Wang, Cheng Fang, *Gene Expression Data Classification Using Artificial Neural Network Ensembles Based on Samples Filtering*, Artificial Intelligence and Computational Intelligence, 2009. AICI '09. International Conference, Vol. 1, pp. 626–628, 2009

[12] Riccardo De Bin and Davide Risso, *A novel approach to the clustering of microarray data via nonparametric density estimation*, BMC Bioinformatics, pp. 1–8, 2011

[13] JXu, Rui and Wunsch, Donald, *Survey of clustering algorithms*, Neural Networks, IEEE Transactions, Vol. 16(3), pp. 645–678, 2005

[14] Madeira, Sara C and Oliveira, Arlindo L, *Biclustering algorithms for biological data analysis: a survey*, Computational Biology and Bioinformatics, IEEE/ACM Transactions, Vol. 1(1), pp. 24–45, 2004

[15] Shamir, Ron and Sharan, Roded, *Algorithmic approaches to clustering gene expression data*, Current Topics in Computational Biology, 2001

[16] Tibshirani, Robert and Hastie, Trevor and Eisen, Mike and Ross, Doug and Botstein, David and Brown, Pat, *Clustering methods for the analysis of DNA microarray data*, Dept. Statist., Stanford Univ., Stanford, CA, Tech. Rep, 1999

[17] Yin, Longde and Huang, Chun-Hsi *Clustering of Gene Expression Data: Performance and Similarity Analysis*, Computer and Computational Sciences, 2006. IMSCCS '06. First International Multi-Symposiums , Vol. 1, pp. 20–24, 2006

[18] Roger K. Blashfield and Mark S. Aldenderfer *Computer Programs for Performing Iterative Partitioning Cluster Analysis*, APPLIED PSYCHOLOGICAL MEASUREMENT, Vol. 2, No. 4, pp. 533-541, 1978

[19] U. Luxburg, *A Tutorial on Spectral Clustering*, Statistics and Computing 17(4), pp. 395 - 416, 2007

[20] A. Y. Ng., M. I. Jordan, Y. Weiss, *On spectral clustering: analysis and an algorithm*, Advances in Neural Information Processing Systems 14, pp. 849-856, 2001

[21] Implementation of four key algorithms of Spectral Graph Clustering using eigen vectors, *http://www.mathworks.com/matlabcentral/fileexchange/ 26354-spectral-clustering-algorithms/content/Spectral%20Clustering/ Jordan_Weiss.m*

[22] Matlab Statistics Toolbox, *http://www.mathworks.com/help/stats/index.html*

[23] D. Xu; P. Zhao; W. Gui; Ch. Yang; Y. Xie, *Research on spectral clustering algorithms based on building different affinity matrix*, Control and Decision Conference, pp. 3160-3165, 2013

[24] Hautaniemi S., Lehmussola A., Yli-Harja O., *DNA microarray data preprocessing*, First International Symposium on Control, Communications and Signal Processing, pp. 751–754, 2004

[25] Stiglic G., Kocbek S., Kokol P. *Unsupervised variance based preprocessing of microarray data*, Computer-Based Medical Systems, CBMS 2009. 22nd IEEE International Symposium, pp. 1–4, 2009

[26] Emanuel Weitschek, Giovanni Felici and Paola Bertolazzi, *MALA: A Microarray Clustering and Classification Software*, Database and Expert Systems Applications (DEXA), 23rd International Workshop, pp. 201–205, 2012

[27] GEMS (Gene Expression Model Selector) data sets, *http://www.gems-system.org/*

[28] Hanczar, Blaise and Courtine, Mélanie and Benis, Arriel and Hennegar, Corneliu and Clément, Karine and Zucker, Jean-Daniel., *Improving Classification of Microarray Data Using Prototype-based Feature Selection*, SIGKDD Explor. Newsl., Vol. 5(2), pp. 23–30, 2003

[29] Hong Chang; Dit-Yan Yeung, *Robust path-based spectral clustering with application to image segmentation*,Tenth IEEE International Conference on Computer Vision, vol. 1, pp.278-285, 2005

Multiobjective optimization in problems of quarry design and planning: models, methods and practical experience

Andrey M. Valuev

***Abstract*—**The paper analyzes existing practice of quarry design and long-term planning problems setup and solution and proposes new approach on the basis of continuous-valued models. Sectoral and contour models aimed at solution of these problems are presented, comparison of them and their preferences in adequacy and accuracy with respect to discrete models are given. Characteristic features of optimization problems based on presented models are shown resulting in specific features of direct methods application. Practical significance of multiobjective optimization in these problems is substantiated. The solution of that type of problems is exemplified by multiobjective optimization of mining works for deep Neriungrinskii" coal quarry, main features of intelligent optimization software developed by author and used in these computations being presented as well.

***Keywords*—** Decomposition, feasible direction method, gradient-restoration method, mining work development, multiobjective optimization

I. Introduction

IT seems quite astonishing that the vast and very important area of solid minerals production very rare attracts attention of professional mathematicians — specialists in control theory, optimization, numerical analysis. The most widely known and practically used method, namely method of optimal ultimate pit limits establishment by Lerchs and Grossman [1], [2] was proposed almost fifty years ago. It regards both mined-out space of a quarry and a deposit as sets of blocks of regular shape. Based on correct use of dynamic programming, the method belongs to a very scarce set of optimization methods for mining industry problems that have strict mathematical substantiation. Since that time there were many perfections and similar methods [3]–[5], but all of them do not change the original setup of the problem. However, it is known for many years that this problem setup reflects the property of a real quarry in a very limited way, so the solution may give only preliminary recommendation as to quarry final contour and depth. It is practically useless for determination of stage quarry contours and mining work regime, what are the main problems of feasibility study of opencast mining design.

On the contrary, there are many methods of industrial mathematics with much more narrow field of application of each. They are not more than heuristics. Worst of all, usually we have not only the absence of any proof of their convergence, but even the exact formulation of the problem being solved. It is sometimes stated that practical results of their application turned to be satisfactory but no analysis confirming such assertions can be found.

So we have a strange situation. There are a lot of strict methods for a problem with a clear formulation that, however, cannot give more than a crude assessment of a quarry design as a whole and hardly can be spread to assessment of stages of mining works. On the other hand, here are a lot of methods without any substantiation for various design and planning problems. And, after all, several widely used software packages aimed at design and planning problems solution with very little (if any) application of optimization methods, with which most decisions are made by users-engineers with a trial-and-error method in a tedious low-level man-machine dialogue. Another question is whether general methods of linear, nonlinear or integer programming may be applied to problems in question without any adaptation to its specific features. My experience convinces me that more usual applications of linear programming require oversimplified models. On the contrary, more adequate models of nonlinear and mixed integer programming belong to the area where methods are less universal and the success of their application depends usually on taking into account specific features of a problem. In the regarded domain it is not likely to encounter problems with favourable features, e.g., smooth convex programming problems.

For this reason, choices of more adequate model and more efficient method in the regarded area are interrelated and often must be performed simultaneously. My observations for the period of more than 20 years, discussion on several most important international scientific conferences — 3rd, 5th and 20th International Symposium on Mine Planning and Equipment Selection (1994, 1996, 2011}, 1st and 2nd regional APCOM Symposium on Application of Computers and Operations Research in the Mineral Industries (1994, 1997) —

This work was supported by Ministry of Education and Science of Russian Federation as a part of mandatory section of state Moscow State Mining University.

Andrey M. Valuev is with Department of Management in Mining Industry, Moscow State Mining University CO 80305 USA (corresponding author phone: +7962-922-8580; fax: +7499-237-3163; e-mail: valuev.online@gmail.com).

show that there are more difficulties than successes on this way [6].

After all, necessity to choose design solutions on the basis of several criteria makes the problem of optimization method application to these choices much more complicated. In Russia, the principal problem exist in balancing interests of the state — the owner of mineral resources — and manufacturers that running their business of mining them [7]. Interest of both cannot be reduced to the sole criterion as well.

The present paper presents author's approach to setup and solution of the quarry stage contour and some examples of its practical application as well author's understanding of ways of its further development.

II. Surface Mining Modelling

During the process of surface mining the rock massif is permanently excavated and pit mined-out space grows, part of which can be used as overburden dumps. If a quarry is represented in more detail, we must distinguish collapses of a blasted rock that quickly change their shapes and sizes during excavation and slopes of rock massif sufficiently changing almost instantaneously by blasts separated in time with periods of several days or even weeks. In design and long-time planning problems it is sufficient to reduce consideration to shapes of these slopes as well as horizontal grounds (*berms*) used to allocate extraction machines and provide freight traffic by trucks or trains. The whole surface of mined-out space is named *quarry contour*. Its stepwise shape reflects separation of mining work places into a succession of benches, each corresponding to a certain layer.

Supposing that slope inclination does not change from the bottom to the top of one bench we fully represent an open pit contour with lower and higher edges of bench slopes. Moreover, for known slope angles depending on properties of local rock it is sufficient to establish only the lower edge of each bench, the higher edge being wholly determined with the lower one. To guarantee the possibility of road allocation on grounds we assume that these edges are smooth lines on planes which curvature does not exceeds $R_{\min}^{-1}$ where $R_{\min}$ is the maximum radius of road turns. This representation of a quarry contour in plan with bench edges is shown on Fig.1.

So a pit contour may be approximately treated as an irregular step pyramid. Its representation as a combination of the same rectangular blocks or hexagonal prisms is universal but crude. Horizontal grounds used for excavators and dumpers allocation and horizontal segments of truck or train roads must be composed of block upper bases. So to obtain high accuracy very small blocks must be used. There is no need for high accuracy in the design of ultimate pit contour. Moreover, broad bases of large blocks constitute grounds enabling to locate roads with relatively small curvature that may be practically used. However, in design of stage pit contours so low accuracy is unsatisfactory.

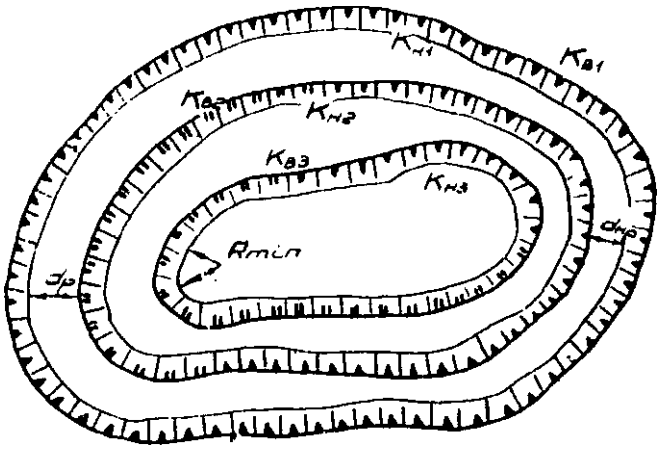


Fig. 1. The general view of a quarry in horizontal projection (as in design drawings)

However, it was understood many years ago that the model in use expresses the only principal property of a quarry shape: it becomes wider from its bottom to top and inclination of its slopes corresponds to rock properties. Blocks models do not include any conditions for transport access that is the next practical requirement for a quarry design and its implementation. So these method may give satisfactory shape for upper benches, but for lower ones often quite non-implementable edges are obtained (see Fig. 2).

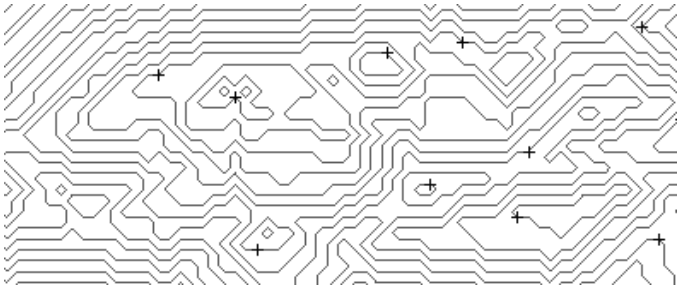


Fig.2. An example of ultimate pit contour generated by Korobov's method with MINEFRAME software package [8]

In turn, we must emphasize, after all, that methods of this type admit the only criterion, namely the total profit of mining the deposit. Moreover, this profit is calculated in a conditional way, without possibility of taking into account discounting of the income and expenses.

Bench edges must have no self-intersections and narrow loops (the minimum width is g). We consider edges in their projection on the same horizontal plane. Then between the *l*-th and (*l*+1)-th edge must be a ground with minimum width $d_l = b + (H_{l+1} - H_l) / \tan \alpha_{l+1}$ (where b is the minimum width of each bench ground, H_l is the absolute height of the *l*-th bench bottom and α_l is the inclination angle for the *l*-th bench slope). Here we consider a simplified model with uniform rock on the entire stretch of each bench. Using representation of each curve with dependencies of its points coordinates from natural parameter *s* (i.e., the length of the curve between point and conditional initial point) we formulate the model of a

ISBN: 978-1-61804-233-0

quarry shape in such a way

$$dx_l/ds=\cos\varphi_l,\ \ dy_l/ds=\sin\varphi_l,\ \ d\varphi_l/ds=k_l, \tag{1}$$

$$x_l(S_l)=x_l(0),\ \ y_l(S_l)=y_l(0),\ \ \varphi_l(S_l)=\varphi_l(0), \tag{2}$$

$$|k_l|\le R_{\min}^{-1}, \tag{3}$$

$$\rho(x_l(s),y_l(s);x_l(s'),y_l(s'))\ge g,\quad s,s'\in[0,S_l],|s-s'|\ge\pi R_{\min}, \tag{4}$$

$$\rho(x_l(s),y_l(s);x_{l+1}(s'),y_{l+1}(s'))\ge d_l,\quad s\in[0,S_l],\ s'\in[0,S_{l+1}],\ l=1,\dots,L-1. \tag{5}$$

Relationships (1)–(5) describe a typical quarry having the only bottom and located on the territory with horizontal surface and may be modified in a natural way for other cases: for quarry on a mountain slope or between hills, for a quarry which lower area consists of several separate pits etc. In these cases either some line may be open or extra lines representing contours of separate pits must be introduced. We name a model of this type represented with a *set of curves* a *SC model*.

One can see that the model (1)–(5) and a generic SC model resembles a dynamic system described with ODE. So optimization problems for SC models must be a kind of optimum control problems. For example, a problem of optimization the entire value of minerals within the quarry bound corresponds to the criterion

$$\sum_{l=1}^{L}H_l\iint_{Q_l}C_l(x,y)dxdy= \\ =\sum_{l=1}^{L}H_l\int_0^{S_l}E(x_l(s),y_l(s))\cos\varphi_l(s)ds\to\max, \tag{6}$$

where $C_l(x,y)$ characterizes value of minerals in an elementary column of the *l*-th layer with base coordinates x, y (including cost of extraction, transportation and overburden dumping) and

$$E(x,y)=\int_0^y C(x,Y)dY.$$

However, restrictions (4)–(5) are very untypical and look like restrictions in semi-infinite programming problems but are imposed on continuum sets of curves points. There is no hope to solve the problem of this type analytically. Recent work [9] in the area of numerical solution of optimum control problem show that reliable strictly substantiated methods enable to solve more simple problems only. On the contrary, methods of optimum control problems reduction to mathematical programming show their efficiency three decades ago still [10]. A similar approach by author based on discrete-time approximation of optimum control problems with constraints on state coordinates showed its efficiency too [11].

So it was proposed to approximate a quasi-dynamic SC model with a finite-dimensional *SPL model* in which curves are approximated with *polylines* (i.e., broken lines). Variables in a SPL model are coordinates of points — straight segments of polylines. They are denoted as x_{il}, y_{il}, and segment lengths as

$s_{il}=\rho(x_{il}, y_{il}, x_{i+1l}, y_{i+1l})$.

To guarantee that the maximum approximation error has the order $\varepsilon^2 R_{\min}$ for a given value ε a condition is introduced

$$s_{i+1l}\le\varepsilon R_{\min} \tag{7}$$

Then (3) is approximated with

$$2\mathrm{tg}(\psi_l/2)/s_{il}\le(R_{\min})^{-1},\quad 2\mathrm{tg}(\psi_l/2)/s_{i+1l}\le(R_{\min})^{-1}, \tag{8}$$

(4) and (5) — respectively with

$$\rho_1(x_{il}, y_{il}, x_{jl}, y_j, x_{j+1l}, y_{j+1l})\ge g,\ \rho_1(x_{i+1l}, y_{i+1l}, x_j, y_{j+1l}, x_{j+1l}, y_{j+1l})\ge g,$$
$$\rho_1(x_{jl}, y_{jl}, x_{il}, y_{il}, x_{i+1l}, y_{i+1l})\ge g,\ \rho_1(x_{j+1l}, y_{j+1l}, x_{il}, y_{il}, x_{i+1l}, y_{i+1l})\ge g \tag{9}$$

for

$$i<j,\ \ s_{i+1l}+\dots+s_{j-1l}\le\pi R_{\min},\quad s_{j+1l}+\dots+s_{n_l l}+s_{1l}+\dots+s_{i-1l}\le\pi R_{\min}. \tag{10}$$

and

$$\rho_1(x_{il}, y_{il}, x_{jl+1}, y_{jl+1}, x_{j+1l+1}, y_{j+1l+1})\ge d_{\min l},$$
$$\rho_1(x_{i+1l}, y_{i+1l}, x_{jl+1}, y_{j+1l}, x_{j+1l+1}, y_{j+1l+1})\ge d_{\min l},$$
$$\rho_1(x_{jl+1}, y_{jl+1}, x_{il}, y_{il}, x_{i+1l}, y_{i+1l})\ge d_{\min l},$$
$$\rho_1(x_{j+1l+1}, y_{j+1l+1}, x_{il}, y_{il}, x_{i+1l}, y_{i+1l})\ge d_{\min l}, \tag{11}$$

The listed restrictions are posed: (7), (8) —on all vertices of all polylines; (9) on all pairs (a vertex, a segment) of each polyline satisfying (10), (11) — on all pairs (a vertex, a segment of polyline of an adjacent level).

Here $\rho_1(x_0, y_0, x_1, y_1, x_2, y_2)$ denotes the distance between a point with coordinates x_0, y_0 and a straight segment which ends have coordinates x_1, y_1 and x_2, y_2. Thus relationships (7)–(11) constitute a SPL model, i.e., a finite-dimensional contour model of a quarry. We shall the following notation: u_{il} for a pair (x_{il}, y_{il}), u_l for a vector comprising u_{il} for all vertices of the *l*-th polyline; u for an *N*-dimensional vector comprising components of all u_l.

III. Optimization Problems of Quarry Design on a Finite-Dimensional Contour Model

If an optimization problem of quarry design is formulated on the basis of a SPL model, then constraints (7)–(10) constitute most of relationships in it. Beside them, a problem may contain a relative small number of additional restrictions depending either on all variables or on variables representing a certain polyline or a pair of polylines; these restrictions may be equalities as well as inequalities. This fact is the most important in the choice of the proper optimization method. In particular, we must emphasize that:

1) The number of all constraints is great, much more than *N*.
2) Nevertheless, on each u satisfying (7)–(11) the number of active and ε-active constraints has the same order that *N* (an inequality constraint $F(u)\le 0$ is named *active* on vector v if $F(v)=0$ and *ε-active* if $-\varepsilon\le F(v)\le 0$).
3) For constraints (7)–(11) its residual $F(u)$ depends on a limited of components of u which number is not greater than 6.
4) Therefore, matrix $M_\varepsilon(u)$ which rows are gradients of $F(u)$ for ε-active constraints is a band-cellwise matrix. Besides, each cell is a sparse and usually triangular matrix resembling analogous matrices for discrete-time optimum control problems.
5) An u satisfying (7)–(11) may be easily found in many

ISBN: 978-1-61804-233-0

ways. It suffices to determine a succession of concentric circles with proper radii and determine each polyline as an arc polygon for respective circle. Then there are many ways of deformation of such polygons retaining (7)–(11) validity. As to additional constraints, it is much more difficult to satisfy them (it may be impossible, if the problem setup is unreasonable).

For this reason, non-direct methods like modified Lagrange multipliers method [12] in its standard form seem to be inefficient since modified Lagrange function must be a sum of a too large number of terms. More crude methods like penalty functions methods may useful to find an admissible u satisfying all additional constraints [13]. Recent years showed new achievements in convex programming, including nonsmooth optimization [14]. It must be noted, however, that some of these constraints must reflect relations between quarry and deposit shapes and so are very unlike to be expressed with convex functions. So a penalty or modified Lagrange function is very likely to have many local minima that are useless for obtaining the optimum (and even feasible) solution.

Feasible directions method (FDM) and direct methods combining features of FDM and gradient-restoration method (for problems with equality constraints) have some preferences. Starting from a u^0 satisfying (7)–(11) as an initial point, these methods enable to retain their validity. It is necessary to take into account on each iteration only ε-active constraints. Search of descent direction may be performed efficiently taking into account sparsity of matrix $M_\varepsilon(u)$ determining the respective LP problem. More efficiency of the above mentioned direct methods is gained with the use of decomposition techniques.

The general form of optimization problems in question (without equality constraints) is the following

$$F_0(u_1,...,u_n) \to \min, \tag{12}$$

$$F_j(u_l) \le 0, \quad j \in I_{1l}, l = 1,..., n, \tag{13}$$

$$F_j(u_l, u_{l-1}) \le 0, \quad j \in I_{2l}, l = 2,..., n. \tag{14}$$

$$F_j(u_1,...,u_n) \le 0, \quad j \in I_0. \tag{15}$$

With a formal equation of dynamics

$$x_l = u_{l-1}, l = 2,..., n, \tag{16}$$

it is transformed into a problem of discrete-time optimum control for which a FDM with decomposition determination of descent direction was proposed in [11] and evolved later in [15]. Principally it is based on representation of control vector variation in the form

$$w=H_1y_1+H_2y_2+\ldots+H_{n+1}y_{n+1} \tag{17}$$

where the set of ε-active constraints is divided into $2n+1$ subsets

$$I_\varepsilon(u) = J_{11} \cup J_{21} \cup J_{21} \cup ... \cup J_{2n} \cup J_{n+1} \tag{18}$$

where

$$J_{1j} \subseteq I_{1l\varepsilon}(u), \quad J_{2j} \subseteq I_{2l\varepsilon}(u_l, u_{l-1}), \tag{19}$$

and matrices $H_1,\ldots, H_{n+1}$ are determined from the condition: for any w

$$(F_{ju}(u_i), w) = (F_{ju}(u_i), H_i y_i), \quad j \in J_{1l}, \tag{20}$$

$$(F_{ju}(u_i, u_{i-1}), w) = (F_{ju}(u_i), H_i y_i), \quad j \in J_{2l}, \tag{21}$$

$$(F_{ju}(u), w) = (F_{ju}(u), H_{n+1} y_{n+1}), \quad j \in J_{n+1}. \tag{22}$$

Then the problems of descent directions determination for given $\delta \le \varepsilon$ are

$$s_l \to \min; \quad (F_{0u}(u), H_i y_i) \le -s_l, \tag{23}$$

$$(F_{ju}(u_l), H_l y_l) \le -K_{lj} s_l, \quad j \in I_{1l\delta}(u_i) \cap J_l, \tag{24}$$

$$(F_{ju}(u_l, u_{l-1}), H_l y_l) \le -K_{lj} s_l, \quad j \in I_{2l\delta}(u_i, u_{l-1}) \cap J_l, \tag{25}$$

$$-1 \le y_{il} \le 1, \ i=1,\ldots, M_l. \tag{26}$$

and analogous problem for $l=n+1$ determined by (23), (26) and

$$(F_{ju}(u), H_l y_l) \le -K_{lj} s_l, \quad j \in I_\delta(u_i) \cap J_l. \tag{27}$$

It is proved that for regular ways of determination of $H_1(u)$, $H_2(u)$, …,$H_{n+1}(u)$ necessary condition of optimality is equivalent to the condition that for solutions of all the problems (23)–(26) and (23), (26), (27) with $\delta = 0$ takes place $s_l = 0$. So the FDM with decomposition is equivalent to the original FDM by requires several times less computations.

This approach was applied to optimization of stage contours of Ekibastuz quarry for variant of its mining with a sole quarry that was earlier put forward and developed with engineering methods by a group of scientists from Moscow State Mining University. With the proposed method, starting from their design contours it was achieved a significant improvement in main parameters. E.g., for the 7-th stage (see Fig. 3) stripping ratio was reduced from 2.77 to 2.72 m^3/ton.

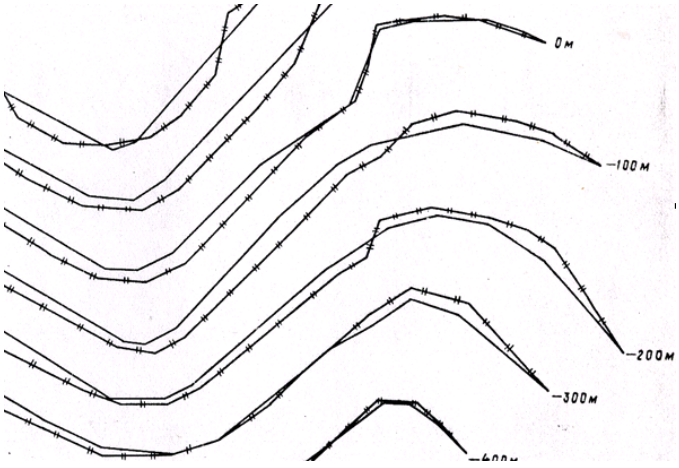


Fig. 3. A fragment of general view of design contour of Ekibastuz quarry for depth 480 m before and after optimization

These results seem promising, but many serious difficulties arise. Most of them result from calculation of quantitative and qualitative indices of mineral production for design contours obtained during optimization process. Deposit is usually consists of a lot of separate seams (orebodies) with a complicated shape of bounds. Besides, these parts of a deposit are not uniform. Note, that in optimization process it is necessary calculate not only required indices but as well their derivatives with respect to x_{il}, y_{il}. Representing orebody

bounds for a certain horizontal section with polylines and performing preliminary triangulation of its internal domain we reach moderate computational complexity of required calculations. Nevertheless, in many cases non-differentiable dependences take place; and worst feature of non-smoothness is that it is most likely in the vicinity of optimum contours. The following illustration (Fig. 4) show cases in which the square of intersection of polygonal domains is continuously differentiable with respect to their vertices coordinates or not. One can see that the square of a polygon domain which bound includes polyline ABC is continuously differentiable with respect to x_B, x_C and y_C and not continuously differentiable with respect to x_A, y_A and y_B. The same properties displays formula (6). So regularization techniques are required for optimization reliability.

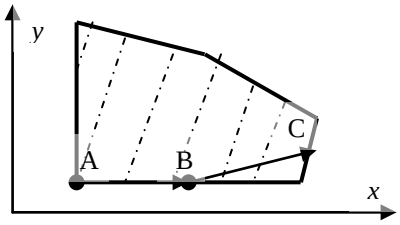


Fig. 4. Mutual disposition of a bench edge and a seam

IV. A SECTORAL REPRESENTATION OF A QUARRY CONTOUR IN OPTIMIZATION PROBLEMS

Sectoral representation was proposed by I. B. Tabakman [16] for the sake of annual planning for quarries. The main idea consists in representation of each edge with points of its intersection with a system of lines. This lines may be either parallel or forming an acute angle between neighboring lines. Domain between two lines is named *sector*. Dependencies of qualitative and quantitative characteristics of mineral along the sector axis are determined before solution of any optimization problems and used in further computations. For rounded columnar deposits it was sufficient to use the same system of sectors for all benches [16], in general it is impossible (see Fig. 5).

It must be emphasized that sectoral representation is much less universal than (7)–(11) and enable to represent a subset of possible variants, their main property is that each edge intersects each sector axis one time. It would be better if intersection angles are close to the right angle. Nevertheless, on deep quarries with regular technology directions of mining works change slowly, and for periods of several years sectoral representation is satisfactory.

In works by I. B. Tabakman and his followers [17] constraints expressing technological requirements on shape of a quarry were represented with linear inequality constraints. This representation turns to be crude. Another form of a sectoral model is based on the use of relationships (7)–(11) supposing that each polyline vertex lie on border line of a corresponding sector. Additional relationship is the constraint on angles between sector border lines and edge polylines in vertices of the latter.

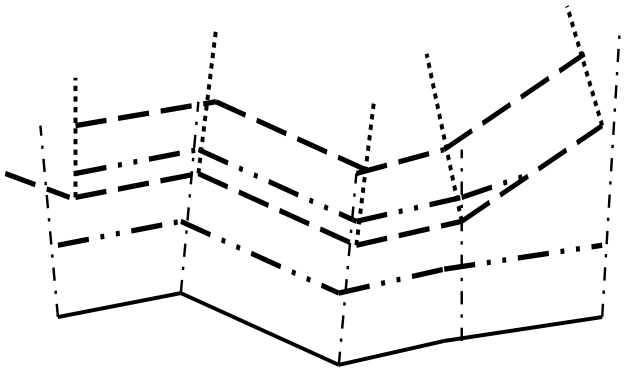


Fig. 5. Sectoral representation of a quarry contour

It was developed, rather occasionally, a research sofware package aimed at solution of quarry contour optimization problems on the basis of sectoral models [18]. The initial objective was to work out computerized means for investigation of variants of mining work development on coal quarry “Neriungrinskii” in connection with equipment selection. Practical requirements resulted in multiobjective optimization implementation in the software package. In practice, it meant that a simple language of problem determination was proposed and each problem, including format and order of its inputted and resulting data is expressed not with program modules, but with special text files treated with program modules, as it is described in [19]. See below the fragment of language representation of a planning problem:

```
1pit
 pit_name   c15:0:input
 coal_dens  r:0:input
 overburden_dens r:0:input
 yr_advance r:0:input
 CC_mass    r:2:sumobj(zone.CC_mass)
 K9_mass    r:2:sumobj(zone.K9_mass)
 ash_mass   r:2:sumobj(zone.ash_mass)
 coal_mass  r:2:sumobj(zone.coal_mass)
overburden_vol  r:2:sumobj(zone.overburden_vol)
 ash        r:2:100*pit.ash_mass/pit.coal_mass
 sr         r:2:pit.overburden_vol/pit.coal_mass
2 zone
 CC_mass    r:2:sumobj(block.CC_mass)
 K9_mass    r:2:sumobj(block.K9_mass)
 ash_mass   r:2:sumobj(block.ash_mass)
3 block
 CC_mass    r:0:input
 K9_mass   r:0:input
 ash_mass   r:0:input
 coal_mass   r:2: CC_mass+K9_mass
 ash_mass   r:2:ash_content*
. ash_content    r:0:input
```

In this fragment we see description of both data structure and problem relationships expressed with arithmetic operators

ISBN: 978-1-61804-233-0

and a set of functions (here it is *sumobj* for the sum of an array components).

V. Experience for Coal Quarry "Neriungrinskii"

In complex study of prospects of "Neriungrinskii" coal quarry production the use of our software package played a very important role. Main aim of the study consisted in the choice of transportation fleet for the next four-year period, but this choice demanded elaboration and comparison of mining work development variants. These variants were generated with the package as successions of optimum year plans with respect to given criteria and restrictions sets.

Characteristic feature of this quarry is the presence of resources of two type of coal, coking coal K9 and coal for power generation CC. Production of both types plays important role and is run simultaneously. Quarry is very deep; the number of benches for the period in question was 15. To represent mining work development a non-linear model was elaborated and prepared for processing by our software package.

The first cycle of optimization was devoted to establishment of maximum production of coal and its types when the maximum advance of mining operation was limited to 90 m per year. It was established that maximum production of K9 coal results in total stripping ratio 8,7 m^3 per ton. Decreasing K9 coal production by 2,75 million ton, or 13%, it is possible to diminish total stripping ratio to 4,74 m^3 per ton.

Mining work modeling was effected in the following way: first, four-year plan was calculated by optimization with respect to indices for the entire four-year period *P*, then step-wise optimization was run for succession of years within a period with posing restrictions that guarantee achievement of the optimum values of indices for the entire period.

In the second cycle of research modeling was effected under condition that stripping ratio for each year is equal to the fixed value (its variants were 5,6 and 5,8 m^3 per ton), that gave the chance of maximum possible achievement of demands for quarry production of both types. Possible proportion between K9 and CC production were established by maximization of production of any of these types. It was shown that stable production of CC coal in the range from 4,0 to 4,2 million ton from year to year and minimization of deflection of stripping ratio for a year from its average value yields stabilization of total coal production as well. This variant demands, however, purchase of all new trucks in the beginning of the period. In comparison, another variant was calculated with step-wise minimization of overburden extraction and given range of coal production. Some general results of the above mentioned computations are presented in tables I–II, the first presenting results of optimization for period *P* and the second for subsequent years.

Table I. Results of solution of some optimization problems in complex investigation of "Neriungrinskii" coal quarry

Problem objective	Total coal, 10^6 ton	Including output of coal sorts		Over-burden, 1000 m^3	Ash content, %	Strip-ping ratio, m^3/ton
		CC	K9			
Maximization of coking coal output for period *P*	41.8	21.2	20.6	36.4	17,38	8,71
Minimization of stripping ratio for period *P*	35.8	18.2	17.5	17.0	17.2	4,74
Minimization of over-burden volume for fixed coal output	38.7	19.6	19.1	23.7	17,29	6,13

Table II. Minimization of year stripping ratio deviation from its mean value in the contours of the four-year period *P*

Total coal, 1000 ton	Including		Overburden, 1000 m^3	Ash content, %	Mining operation advance, m
	CC	K9			
7730	3511	4218	43288	17,16	97
7708	3421	4286	43167	17,51	87
7817	3507	4309	43779	17,60	98
7496	3257	4239	41983	17,50	98

VI. Conclusion

In the paper the author tried to show practical significance of the problem in question, some particular achievements and still existing problems. In particular, author's approach based on domain-oriented modifications of direct optimization methods yielded ome practical results. Nevertheless, use of optimization methods, outside a very narrow domain of ultimate pit limits determination, stays occasional. One of the reason is that they are not integrated in practically used software packages. On the other hand, to perform this integration it is necessary to persuade software developers in reliability of these methods. It must be emphasized that reliable methods are required for solution of complicated problems, nonconvex and as a rule nonsmooth.

As to models, all of them were developed for regular technologies on deep quarries with relatively great number of extraction units thus closely interlinked in their operation, and with the use of freight transport. There exist a lot of smaller quarries with less and less interrelated machines and mixed transportation systems. Models for these conditions must combine features of discrete-valued and continuous-valued model.

As for multiobjective optimization, we showed that it is

ISBN: 978-1-61804-233-0

needed and really takes place in practical sense: solutions of problems with various criteria and additional constraints are obtained and compared. Nevertheless, now it is possible to calculate and compare a limited set of variants with parameters determined via engineer's intuition and practical experience. The obstacle in use of more regular approaches to multiobjective optimization is computational complexity as well as absence of special methods (e.g., for Pareto optimization) taking into account specific features of classes of problems in question.

And, after all, design and planning problems must be as a rule dynamic, not static. The theory of opencast mining many decades ago put forward the idea of mining work regime [20] that is the general dynamic representation of the mining operation for the whole period of quarry work. In fact, dynamic optimization in opencast mining design is still restricted to conditonally-dynamic approach, as formulated in [16]. To solve dynamic problems of relatively great dimension it is necessary to pay more attention to computational efficiency. Hierarchical decomposition (probably, based on [11] and similar decomposition schemes) seems to be one of promising ways to diminish amount of computations.

REFERENCES

[1] H. Lerchs and I. F. Grossman, "Optimum design of open-pit mines", *The Canad. mining and metallurg. bull.*, vol. 58, no. 633, pp. 47–54, 1965.

[2] C. G. Alford and J. Whittle, "Application of Lerchs-Grossmann pit optimization to the design of open pit mines", in *Proceedings of the AusIMM/IE Aust. Newman Combined Group, Large Open Pit Mining Conference*, Oct. 1986.

[3] T. B. Johnson, "A comparative study of methods for determining ultimate open pit mining limits", in *Proceedings of the 11-th APCOM Symposium, Tucson, Arizona, April 15-20, 1973*.

[4] M. David, P. A Dowd., and S. D. Korobov, "Forecasting departure from planning in open pit design and grade control", in *Proceedings of the XII-th APCOM Symposium. Denver, Colorado, USA, April 1974*

[5] S. D. Korobov, "Analysis of computer-aided methods of open pit limits design", *Gornyi zhurnal (Mining Journal)*, no. 4, pp. 59–62, 1981 (In Russian).

[6] S. Rakhimbekov, "Problem of optimization in mining", *Proceedings of the Twentieth International Symposium on Mine Planning and Equipment Selection MPES 2011*, A. Zharmenov, R. Singhal, and S. Yefremova, Ed. Almaty, October 12–14, 2011. Almaty, 2011, pp. 537–546.

[7] A. A. Ashikhmin and Yu. V. Sytnik, "Economical evaluation and choice of implementation variants of solid mineral deposits mining designs", *Ratsional'noe osvoenie nedr (Rational use of mineral resources)*, no. 2, pp. 15–19, 2013. (In Russian)

[8] O. V. Nagovitsyn and S. V. Lukichev, "Computerized instruments of mining work engineering support in MINEFRAME software system", *Mining informational and analytical bulletin*, no. 7, pp. 184–192, July 2013. (In Russian).

[9] V. A. Srochko and V. G. Antonik, *Methods of multi-extremal optimum control problems. Manual*. Irkusk, Irkusk State University Publishers, 2012, pp. 1–104. (In Russian)

[10] Yu. G. Evtushenko, *Methods of extremal problems solution and their application in optimization software packages*, Moscow, Nauka, 1982, pp. 284–371. (In Russian)

[11] A. M. Valuev, "A numerical method for multistage optimization problems with stepsize computation of the descent direction", *U.S.S.R. Comput. Math. Math. Phys.*, vol. 27, no. 5, pp. 128–137, 1987; [translation from Zh. Vychisl. Mat. Mat. Fiz., vol. 27, no. 10, pp. 1474–1488, Oct. 1987].

[12] Yu. G. Evtushenko, A. M. Rubinov, and V. G. Zhadan. "General Lagrange-type functions in constrained global optimization. Part I: Auxiliary functions and optimality conditions", *Optimization Methods and Software*, vol. 16, no. 1-4, pp. 193-230, 2001.

[13] J. F. Bonnans, Ch. J. Gilbert, C. Lemaréchal, and C. Sagastizábal., *Numerical optimization. Theoretical and practical aspects*, Berlin, Springer-Verlag, 2003, pp. 141–196.

[14] Yu. Nesterov, *Introductory lectures on convex optimization. A basic course*, Boston, Kluwer, 2004, pp. 1–236.

[15] A. M. Valuev, "A new model of resource planning for optimal project scheduling", *Mathematical Modelling and Analysis*, vol. 12, no. 2, pp. 255–266, 2007.

[16] I. B. Tabakman, *Principles of CAM systmems design for quarries*, Tashkent, Fan, 1977, pp. 64–72. (In Russian)

[17] A. Melamud and D.S. Young, Optimizing "Interdependence of operating cost", in *Proc. of the 24-th APCOM Symposium*, Montreal (Canada), vol. 2, 1993, pp.75–82.

[18] A. M. Valuev,"Method and a computer program for coal quarry operation zone optimization", *Supplement papers of the Mining informational and analytical bulletin*, no 8, 2003, pp. 1–22. (In Russian).

[19] A. M. Valuev, "Intelligent programming and informational means for representation and solution of adaptive organizational planning problems for open pits", in *the Second Regional APCOM'97 Symposium*, Moscow, 1997, pp. 217–221.

[20] V. V. Rzevsky, *Opencast mining. Technology and integrated mechanisation*, Moscow, Mir Publishers, pp. 412–437, 1987.

Andrey M. Valuev graduated from the Moscow Institute of Physics and Technology (MIPT) in 1978 and received a degree of Candidate of Physical and Mathematical Sciences in Computational Mathematics from MIPT in 1984. Doctor of Sciences in Mathematical Modeling, Numerical Methods and Software Packages (2008).

Professor of Departments of Management in Mining Industry of MSMU and Computational Modeling of Technological Processes of MIPT. For 30 years participated in many research projects related with mathematical modeling and application of control theory and computational methods, mainly in mining industry.

Member of Organizational Bureau of monthly seminar "Scientific and practical problems of automobile-road complex of Russia" under leadership of vice president of the Russian Academy of Sciences V. V. Kozlov.

Advanced method of noncontact measurement of shrouded blade vibration in steam turbine: evaluation of bladed disc mode shape

Jaromir Strnad, Jindrich Liska

***Abstract*—** This paper deals with an advanced method for the evaluation of a bladed disc behavior in terms of the wheel vibration and blade service time consumption. This method is developed as a part of the noncontact vibration monitoring system of the steam turbine shrouded blades. The proposed method utilizes the time-frequency processing (cross spectra) to analyse the data from the optical and magnetoresistive sensors, which are mounted in the stator radially above the rotor blades. Fundamentally, the blade vibrations are detected during the blade passages under the sensors and the following signal processing, which covers also the proposed method, leads to the estimation of the blade residual service life. The prototype system implementing above mentioned techniques was installed into the last stage of the new steam turbine (LP part). The method for bladed disc mode shape evaluation was successfully verified on the signals, which were obtained during the commission operation of the turbine.

***Keywords*—** Blade tip-timing, cross spectrum, nodal diameter, travelling wave.

I. Introduction

Any mechanical stress of rotating blades reduces their service life. A damage of the blades causes changes in their frequency characteristics, frequency and intensity of the excitation forces, which are measured using contact or contactless method. The contact measurement by strain gauges gives the direct information about the blades stress during a whole revolution [1]. This method is unsuitable for long-term measurements, because a sensor service life is short in a corrosive environment. Furthermore, it is very difficult and expensive to monitor all blades of the bladed disc. The contactless method BTT (blade tip-timing) is based on the measurements with the sensors that are mounted radially on the circumference of the stator above the rotor blades [2]-[4]. The times of blade tip passages under the sensors are analyzed. Changes of the frequency characteristics, which are associated with their normal wear and tear, are reflected very slowly. We can not measure these changes using conventional methods. In contrast, the immediate blade damages are manifested by the rapid changes, which enforce the immediate shutdown and repair of the turbine.

It is important to monitor the blades frequency characteristics, during the turbine runup, rundown and power changes. The measurement of vibrations gives us the information about the blade deflection and blade residual service life. Based on these informations, we monitor the state of the blades and other parts of the turbine. The long-term measurement of the blade deflection is important for the planning shutdowns and optimization of the cost of maintenance and operation of the turbine.

The blade deflections are determined by the difference between the real and expected passage times of the blades under the sensors. The deflection of the specific blade is sampled once per revolution. It can not be used the antialiasing filter in the measurement chain, because the turbine rotation frequency is in fact the blade vibration sampling frequency and furtermore, the rotation speed changes. For these reasons, the vibrations of the specific blade are measured in the limited frequency bands with lot images of higher frequencies and the low frequency resolution. An increasing of the sensor amount, and thus a broadening of the frequency bands, has little impact for the expended resources.

Instead of one blade behavior analysis, the whole bladed disc could be analyzed. The deflections are sampled by the passages of all the blades. Then, the width of the frequency band is sufficient and the aliasing manifests itself not so much. The allblade spectrum, which is spectrum involving all blades of one disc, is calculated for each sensor from deflections measured for all the blades passed under the sensor using the short-time Fourier transform. Consequently, cross spectra are calculated from the couples of the allblade spectra from different sensors.

II. Blade vibrations manifestation

The blades interact with each other and their behavior affects the behavior of the whole bladed disc. A mode shape of a bladed disc describes the disc behavior, because the blade vibrations are not random. There is the specific relative

This work was supported by the Technology Agency of the Czech Republic (Grant TA02020728 - Research and Development of Methods and Aplliences for contactless identification of turbine blade state) and also by the European Regional Development Fund (ERDF), project "NTIS – New Technologies for the Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090.

J. Strnad is with the University of West Bohemia, NTIS - European Centre of Excellence, Pilsen, Univerzitni 22, 30614 Czech Republic (phone: +042-377632551; fax: +420-377632502; e-mail: jastrnad@ ntis.zcu.cz).

J. Liska is with the University of West Bohemia, NTIS - European Centre of Excellence, Pilsen, 30614 Czech Republic (e-mail: jinliska@ntis.zcu.cz).

position of the blades, which is periodically repeated. In such cases, there are points on the disc, which are not in a motion in one time. These points lie on the straight lines, which pass through the disc centre, or on the concentric circles around the disc center. These points are known as the nodal diameter and nodal circle. The nodal diameters and circles characterize the vibrations shapes. In practice, the rotating and standing shapes (travelling or standing waves) are monitored in term of the disc vibration. The following Fig. 1 illustrates the selected nodal diameters including the special case - zeroth nodal diameter.

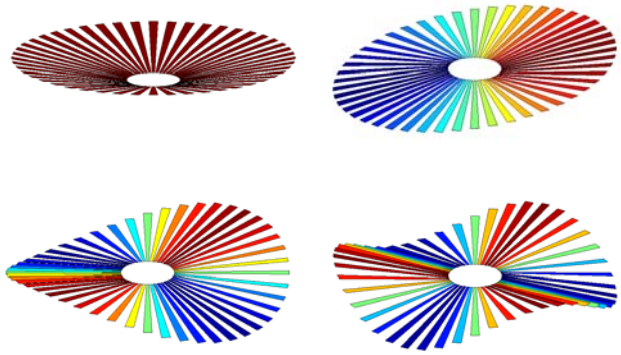


Fig. 1 An example of several nodal diameters.

III. Standing Wave

The bladed disc vibrates with *ND*th nodal diameter with the frequency f_{osc}. This modulation is monitored using the timing of the blades with the stationary sensor

$$f_{mod} = ND \cdot f_{rot} \pm f_{osc} \tag{1}$$

There are two frequencies due to the modulation.

$$f_k = f_{osc} \tag{2}$$
$$f_d = ND \cdot f_{rot} \tag{3}$$

Using two sensors s_1 and s_2, we obtain following signals

$$s_1 = cos(2\pi f_k t_1 + \varphi_k) \cdot cos(2\pi f_d t_1 + \varphi_{d1}) \tag{4}$$
$$s_2 = cos(2\pi f_k t_1 + \varphi_k) \cdot cos(2\pi f_d t_1 + \varphi_{d2}) \tag{5}$$

The wave with the frequency f_{osc} has the same phase shift for both signals because the sensors measure synchronously. We analyze different blades at the same time.The wave with the frequency f_d corresponds to the bladed disc shape. Different blades have different deflections at the same time. If we consider only the bladed disc shape, then the individual blade have always the same deflection. However, different blades pass under different sensors at the same time. Therefore, the vibrations on the frequency f_d have a different phase shift at the same time, see (6)

$$\varphi_{d2} - \varphi_{d1} = \rho \cdot ND, \tag{6}$$

where ρ is the angle formed by sensors. According to the goniometric equality

$$\cos\alpha \cdot \cos\beta = \frac{1}{2}\left[\cos(\alpha - \beta) + \cos(\alpha + \beta)\right], \tag{7}$$

we monitor the signals according to expressions (4) and (5) at each sensor.

$$s_1 = \frac{1}{2}\{\cos[2\pi(f_k - f_d)t_1 + \varphi_k - \varphi_{d1}] + \cos[2\pi(f_k + f_d)t_1 + \varphi_k + \varphi_{d1}]\} \tag{8}$$
$$s_2 = \frac{1}{2}\{\cos[2\pi(f_k - f_d)t_1 + \varphi_k - \varphi_{d2}] + \cos[2\pi(f_k + f_d)t_1 + \varphi_k + \varphi_{d2}]\} \tag{9}$$

After the discrete Fourier transform calculation, the complex frequency signals are obtained from the measured time signals. From these complex signals, we can calculate the amplitude and angle on the given frequency. Further, we focus on the determination of phase. From the expressions (8) and (9), we determine the phases (φ) of both frequency components. For the first sensor, we apply the expression (8)

$$\Phi(S_1(f_k - f_d)) = \varphi_k - \varphi_{d1} \tag{10}$$
$$\Phi(S_1(f_k + f_d)) = \varphi_k + \varphi_{d1} \tag{11}$$

and for the second sensor the expression (9)

$$\Phi(S_2(f_k - f_d)) = \varphi_k - \varphi_{d2} \tag{12}$$
$$\Phi(S_2(f_k + f_d)) = \varphi_k + \varphi_{d2} \tag{13}$$

It holds true for the multiply of two complex numbers c_1 and c_2, that the angle determined by the product is equal to the sum of the angles determined from the individual factors.
If we will consider complex conjugate number c_1, we will get the difference of the angles.

$$\Phi(c_1^* \cdot c_2) = \Phi(c_2) - \Phi(c_1) \tag{14}$$

We obtain the cross-correlation spectrum by the multiplying of the complex conjugate Fourier image of one measured signal with the Fourier image of the other measured signal. The phase of the frequency components corresponds to the following formulas

$$\Phi(S_1(f_k - f_d)^* \cdot S_2(f_k - f_d)) = -(\varphi_k - \varphi_{d1}) + (\varphi_k - \varphi_{d2}) = -ND \cdot \rho \tag{15}$$
$$\Phi(S_1(f_k + f_d)^* \cdot S_2(f_k + f_d)) = -(\varphi_k + \varphi_{d1}) + (\varphi_k + \varphi_{d2}) = ND \cdot \rho \tag{16}$$

If we find two integer multiples of the angle between the sensors (one positive, the other negative) in the cross-spectrum, we can identify the founded nodal diameter.

ISBN: 978-1-61804-233-0

IV. Travelling Wave

The travelling wave is constituted by the deflections of the circumferentially blades. All blades vibrate with the same amplitude and the adjacent blades vibrate with different phase. If the total circumferentially phase delay achieves 2π, we can see a rotating diameter, on which the blades have zero deflection at the moment. If phase delay achieves 4π, we can see two diameters, etc. The deflection of one blade is described as follows

$$\sin(2\pi f_{osc} t + \varphi) \tag{17}$$

The deflection of each blade for any time is described by the (18), where b is the blade number, N the blade quantity and D the diameters number.

$$\sin(2\pi f_{osc} t + 2\pi \frac{b}{N} \cdot D + \varphi_{begin}) \tag{18}$$

If we bring the disc rotation to the (18), we will detect the position of only one blade at one time. Assuming the uniform blades distribution around the disc circumference and constant speed, we can write

$$\sin(2\pi f_{osc}(t + b\Delta t) + 2\pi \frac{b}{N} \cdot D + \varphi_{begin}) \tag{19}$$

Because the initial time is constant, it is expressed as the part of the initial phase delay without loss of a generality. Correspondingly, we take the initial time as zero.

$$\sin(2\pi f_{osc}(b\Delta t) + 2\pi \frac{b}{N} \cdot D + \varphi_{begin}) \tag{20}$$

Due to the mentioned assumptions, the time differences are replaced

$$\Delta t = \frac{1}{N \cdot f_{rot}} \tag{21}$$

$$\sin(2\pi f_{osc}(\frac{b}{N \cdot f_{rot}}) + 2\pi \frac{b}{N} \cdot D + \varphi_{begin}) \tag{22}$$

After the modifications, we achieve

$$\sin(2\pi f_{osc} \frac{b}{N}(\frac{f_{osc}}{f_{rot}} + D) + \varphi_{begin}) \tag{23}$$

The sequentially coming blades, are replaced with time T

$$\frac{p}{N f_{rot}} = T \tag{24}$$

$$\sin(2\pi T(f_{osc} + D \cdot f_{rot}) + \varphi_{begin}) \tag{25}$$

In the spectrograms (whether allblade or cross), we monitor the travelling wave on the frequency

$$f_{obs} = f_{osc} + D \cdot f_{rot} \tag{26}$$

The monitored wave has the form

$$\sin(2\pi T f_{obs} + \varphi_{begin}) \tag{27}$$

From the cross spectrogram, we find out with which diameter (D) and on what frequency (f_{obs}) the travelling wave seemingly vibrates. The real frequency of the blades movement is

$$f_{osc} = f_{obs} - D \cdot f_{rot} \tag{28}$$

So far, we consider the travelling wave in one yet unspecified direction. The travelling waves are forward or backward, that depends on the direction, in which the wave propagates according to the disc rotation direction.

Let us take into account the rotation direction such that the blades sequentially pass under the sensor with increasing number (p). Then, the forward wave propagates against the direction of the blades numbering. The maximum deflection passes from the blades with a higher number to the blades with a lower number p. This means, that the phase φ is the higher the more blade numbers increase.

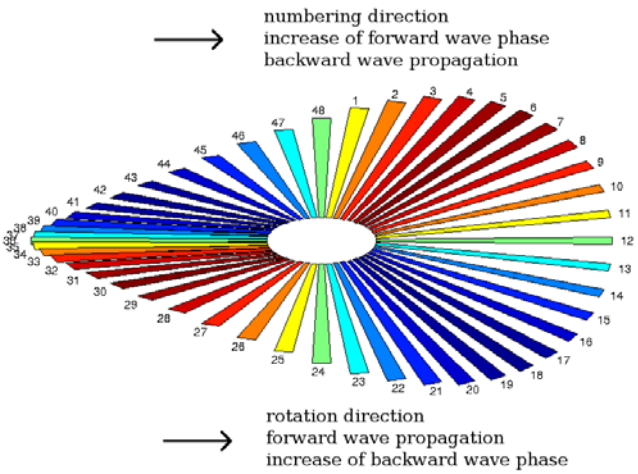


Fig. 2 Definition of the directions on bladed disc.

The backward wave propagates in the direction of the blades numbering. It follows, that the phase is the more downward the more blade numbers increase. The opposite wave propagation is achieved by the negative phase φ. The directions are shown in Fig. 2. Let's assume that φ is positive, we can rewrite (17) into folowing forms:

$$\sin(2\pi f_{osc} t + \varphi) \tag{29}$$

$$\sin(2\pi f_{osc} - \varphi) \tag{30}$$

ISBN: 978-1-61804-233-0

Equation (29) or (30) corresponds with forward or backward wave. In (18) the quantity of the rotating diameters D was defined. Further, to the sign of φ from the previous expression, we extend the value D to the negative numbers, which maintain the validity of the (18) for the forward and backward wave.

V. EVALUATION OF NODAL DIAMETERS

Based on the angle ρ between the sensors, the phases are defined for each of the searched nodal diameter.

$$\pm\rho \cdot ND \tag{31}$$

The phases on the individual frequency components of the cross spectrum, which was discussed in the section III., are compared with the searched phases. In the conformity with the phase tolerance, the calculated nodal diameter is assigned to the corresponding coefficient. When we process the signals from more than two sensors, we compare the coefficients for the individual frequency components of all cross spectra, which combine the signals from each sensor. The coefficient for each frequency component is assigned to the corresponding color that will be used for plotting.

The colors of the frequency components with low amplitudes are suppressed, because of the differentiation of the nodal diameters from the amplitude insignificant frequency components and noise. From these cross spectra, we obtain the cross spectrogram in Fig. 3, which gives us the information about the substractional and additional components (same color) of individual nodal diameters.

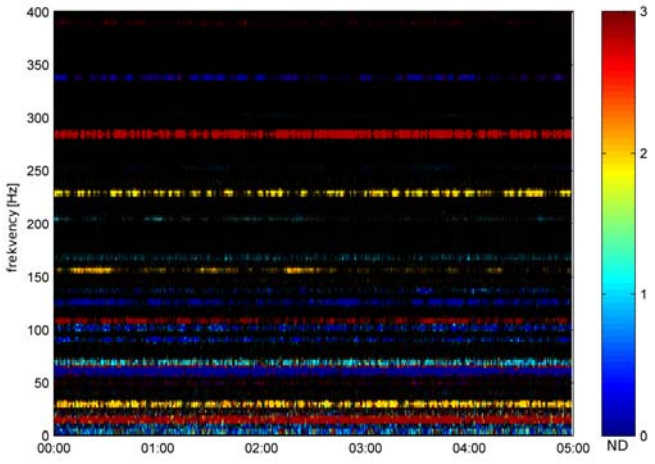


Fig. 3 An example of cross spectrogram.

The Fig. 3 shows strong second nodal diameter 130 Hz in yellow (additional and substractional component of 230 Hz and 30 Hz) or third nodal diameter 135 Hz in red (additional and substractional component 285 Hz and 15 Hz). Less apparent is the first nodal diameter 120Hz in light blue (additional and substractional component 170 Hz and 70 Hz). The zeroth nodal diameter in dark blue (not observed with modulation, so without additional and substractional components) is visible in the spectrogram at several frequencies.

VI. AMPLITUDE EVALUATION

At the first, let's assume standing wave with amplitude A vibrating on the disc. Using basic goniometric relation we can write following

$$A \cdot \cos\alpha \cdot \cos\beta = A \cdot \frac{1}{2}[\cos(\alpha-\beta)+\cos(\alpha+\beta)] \tag{32}$$

The standing wave can be expressed as linear combination of two traveling waves with opposite direction of propagation. Let's define amplitudes of both traveling waves

$$A_{\oplus} = A_{\ominus} = \frac{A}{2}. \tag{33}$$

These amplitudes can be calculated directly from allblade spectrum. For standing wave the following equation holds.

$$A\cdot\cos\alpha\cdot\cos\beta = \left[A_{\ominus}\cdot\cos(\alpha-\beta)+A_{\oplus}\cdot\cos(\alpha+\beta)\right] \tag{34}$$

The amplitude of standing wave is monitored on the blades in the antinodes. This is the maximum amplitude across all blades of the disc. The minimal amplitude is monitored on the blades in the nodes of the standing wave. In the ideal case, minimal amplitude is zero.

In the case of traveling wave, the amplitudes $A_{\oplus}$ and $A_{\ominus}$ are different, $A_{\oplus} \neq A_{\ominus}$. This is interpreted as the superposition of the standing and travelling wave. This behavior of the disc can not be described by the equation (34). It is necessary to expand the equation with the description of the traveling wave. Superposition of waves travelling in opposite direction can be expressed as superposition of standing wave and single travelling wave. The travelling wave propagates in the direction with the higher amplitude of the frequency component. In following example holds $A_{\ominus} < A_{\oplus}$.

$$\begin{aligned}&\left[A_{\ominus}\cdot\cos(\alpha-\beta)+A_{\oplus}\cdot\cos(\alpha+\beta)\right]\\ &\quad=\underbrace{2\cdot A_{\ominus}\cdot\cos\alpha\cdot\cos\beta}_{standing}+\underbrace{(A_{\oplus}-A_{\ominus})\cdot\cos(\alpha+\beta)}_{travelling}\end{aligned} \tag{35}$$

In the case of $A_{\ominus} < A_{\oplus}$, the amplitude of the traveling wave is equal to $A_{\oplus} - A_{\ominus}$. In the same case, the amplitude of the standing wave is equal to

$$A = 2 \cdot A_{\ominus} \tag{36}$$

In the comparison to the standing wave, the blade amplitudes

are increased by the amplitude of the travelling wave. The blade amplitudes in the antinodes of the standing waves (maximum amplitude) are equal to $A_{\oplus} + A_{\ominus}$ and the blade amplitudes in the nodes of the standing wave (minimum amplitude) are equal to $|A_{\oplus} - A_{\ominus}|$.

VII. Sensors placement

The sensors should be placed to each other at the angle, which is the integer multiple of the angle formed by the blades. In this arrangement, the blades with the corresponding angle pass under the sensor approximately at the same time. Therefore, there is no increase of the phase delay.

It is also good to choose the angle depending on the expected nodal diameter. Higher nodal diameter represents larger monitored phase. It is necessary to choose the angle between the sensors such that *ND* ρ lies in the interval 0 to π rad. This prevents an overlap of the detected angles of different nodal diameters, for example the first and third nodal diameter (sensors form angle $\pi/2$ rad). It is recommended to choose the angle between the sensors on the value of prime multiples of the angle, which encloses the blades.

VIII. Conclusions

Any mechanical stress of blades reduces their service time. The blade vibration measurement provides the information about the long-lasting and rapid changes in the blades parameters. This allows early detection of the blade faults. With the blade tip timing method using the magnetoresistive sensors, the oscillations of the longest blades of the low-pressure part of the steam turbine 270MW were measured. The nodal diameters were identified and the amplitudes of their oscillation were calculated. The frequencies of the nodal diameters were verified by the model, which were provided by the manufacturer of the turbine. The calculated data will be used for the determination of the blade residual service times.

Acknowledgment

The authors would like to thank to the company Doosan Skoda Power for the possibility to do the research in the area of industrial interest and to have an opportunity to solve the actual problems on the field of steam turbine troubleshooting.

References

[1] M. Zielinski and G. Ziller, *Noncontact vibration measurements on compressor rotor blades*, Electronics and Measurement Technology Division. MTU München, A DaimlerChrysler Company, Dachauer Strasse 665, D-80995 München, Germany

[2] M. Zielinski and G. Ziller, *Noncontact Blade Vibration Measurement System for Aero Engine Application*. MTU Aero Engines GmbH Engine Control and Testing Division 80995 München, Germany

[3] P. Beauseroy, R. Lengellé, *Nonintrusive turbomachine blade vibration measurement system.* System Modelling and Dependability Group - FRE CNRS 2848, Charles Delaunay Institute, Universite´ de Technologie de Troyes BP2060 10010 TROYES Cedex, France

[4] Y. Kadoy, M. Mase, Y. Kaneko, S. Umemura, T. Oda, M. C. Johnson, "Noncontact Vibrational Measurement Technology of Steam Turbine Blade", *JSME International Journal*,Vol. 38, No. 3, 1995

[5] G. Dimitriadis, I. B. Carrington, J. R. Wright, J. E. Cooper, Blade-tip *Timing Measurement of Synchronous Vibrations of Rotating Bladed Assemblies*, Mechanical Systems and Signal Processing (2002),599-622

[6] De-chao Ye, Fa-jie Duan, Hao-tian Guo, Yangzong Li, Kai Wang, Turbine blade tip clearance measurement using a skeweddual-beam fiber optic sensor, Optical Engineering 51(8), 081514 (August 2012)

Deterministic phase retrieval for determination of a simple defect in L/S type mask inspection

Wooshik Kim and Younghun You

***Abstract*—**In this paper, we consider the phase retrieval problem in extreme ultraviolet lithography. Using diffraction pattern of a defected L/S type mask and the defect-free pattern, we derive equations for determination of the location and value of a one-point wide defect occurred in the LS type pattern mask. We have made a model for this problem and derived closed form equations to determine the location and value using techniques developed in solving phase retrieval problems. We find that the location can be determined to within a symmetry ambiguity and the value uniquely. We have presented an example that shows the validity of the equations.

***Keywords*—**Phase retrieval, actinic mask inspection, coherent diffraction imaging, extreme ultraviolet lithography (EUVL).

I. Introduction

In Extreme ultraviolet lithography (EUVL) detecting and characterizing defects on the multilayer mask blank is very important [1]. There have been several approaches to detect the defects. One approach is by using coherent diffraction imaging method [2, 3]. In this approach, since we can measure its diffraction intensity, this approach is basically the same as phase retrieval and thus we can exploit the algorithms that are developed in this area such as iterative algorithms in [4] and non-iterative, deterministic algorithm derived in [5]. For example, [6] presented some results after they applied an iterative algorithm of [4] to this problem. However, apparently it looks like that no attempt has made to apply some deterministic approach to this problem.

In [5], some of deterministic algorithms are developed for the phase retrieval problem where the two sheets of Fourier transform magnitude or Fourier intensity are assumed to be known, i.e., the FTM of a desired unknown signal and the signal obtained by the addition of the unknown desired signal and a reference signal. Also the reference signal is assumed to be known.

In this paper, we consider the application of a deterministic algorithm to this problem. First of all, we assume that the mask has a LS pattern and some of the information of the original defect-free pattern such as its diffraction pattern is available. Secondly we assume that the shape of the defect is a pixel-wide line type. The goal is to determine the location of stripe type defect and its amount.

This paper is organized as follows. In section II, we build a mathematical model for this problem. In section III, we derive some of equations for determining the values that we are interested in. IN section IV, we show one example that shows how the algorithm works.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2010664).

W. Kim and Y. You are with the Dept of Information and Communication Engineering, Sejong University, Seoul Korea (phone: (+822) 3408-3199; e-mail: wskim@sejong.ac.kr).

II. Problem Description

In this paper, we consider a simple stripe type defect occurred in an L/S type pattern mask [1]. Figure 1 is a simple diagram of a model that we are considering.

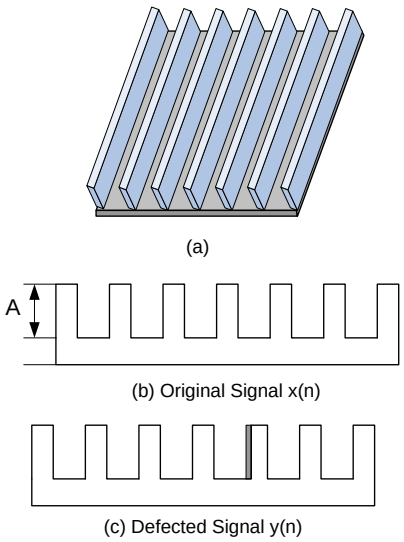


Fig. 1 Schematic model of the problem. (a) Schematic diagram of the model. (b) The cross section of the defect-free pattern, and (c) cross section of a pattern with defect.

Let the mathematical notation of the defect-free normal model be $x(n)$. Then the pattern with a defect can be represented as

$$y(n) = x(n) - A\delta(n - n_0) \tag{1}$$

where A is the amount of the defection and its value is real and

ISBN: 978-1-61804-233-0

positive and n_0 is the location of the defection that we want to know. We assume that the defect is a pixel thick. We assume that we have the information of the defect-free pattern, i.e., the diffraction pattern of $x(n)$. To derive these values we assume that the ROS (Region of Support) of x(n) is R(N)=[0, N-1] and the ROS of the signal is at least twice the ROS of the original signal. Also, we assume that the diffraction pattern of these patterns can be measured at least twice the region of support of those of the mask patterns, i.e., R(-N,+N).

III. Derivation

In this section, we derive equations for the values of A and n_0. To do this, we first derive some of equations.

. Since we assume that we have the information about the diffraction patter of the defection-free pattern. So, if we take the inverse Fourier transform of the intensity pattern, i.e., its diffraction pattern, we can get the autocorrelation function of the signal. In mathematical terms,

$$\begin{aligned} r_x(n) &= F^{-1}\left\{\left|X(\omega)\right|^2\right\} \\ &= x(n) * x(-n) \end{aligned} \tag{2}$$

Likewise, if we get the inverse Fourier transform of the square of the diffraction pattern of the defected signal y(n), then we get

$$\begin{aligned} r_y(n) &= F^{-1}\left\{\left|Y(\omega)\right|^2\right\} \\ &= y(n) * y(-n) \end{aligned} \tag{3}$$

Using the relationship (1), the autocorrelation function $r_y(n)$ is related with $r_x(n)$, A, and $x(n)$ as follows.

$$\begin{aligned} r_y(n) &= y(n) * y(-n) \\ &= [x(n) - A\delta(n-n_0)] * [x(-n) - A\delta(-n-n_0)] \\ &= r_x(n) - Ax(-n+n_0) - Ax(n+n_0) + A^2\delta(n) \end{aligned} \tag{4}$$

A. Determination of n_0

The value of n_0 may be determined from the given information to within an ambiguity. We consider the function

$$\begin{aligned} r_{xy}(n) &= r_x(n) - r_y(n) \\ &= Ax(n+n_0) + Ax(-n+n_0) - A^2\delta(n) \end{aligned} \tag{5}$$

If we find the region of support of each term, we get

$$\begin{aligned} x(-n+n_0): &\quad [-N+1+n_0, n_0] \\ x(n+n_0): &\quad [-n_0, N-1-n_0] \end{aligned} \tag{6}$$

Thus, the region of support of $r_{xy}(n)$ is given as $[\min\{-N+1+n_0, -n_0\}, \max\{n_0, N-1-n_0\}]$. If we write this condition more specifically, then we get

$$\min\{-N+1+n_0, -n_0\} = \begin{cases} -N+1+n_0 & n_0 < \dfrac{N-1}{2} \\ -\dfrac{N-1}{2} & n_0 = \dfrac{N-1}{2} \\ -n_0 & n_0 > \dfrac{N-1}{2} \end{cases} \tag{7}$$

Likewise,

$$\max\{n_0, N-1-n_0\} = \begin{cases} N-1-n_0 & n_0 < \dfrac{N-1}{2} \\ \dfrac{N-1}{2} & n_0 = \dfrac{N-1}{2} \\ n_0 & n_0 > \dfrac{N-1}{2} \end{cases} \tag{8}$$

So, we can determine n_0 to within an ambiguity, i.e., the upper limit of the R.O.S. of $r_{xy}(n)$ is either n_0 if $n_0 > \frac{N-1}{2}$ or $N-1-n_0$ if $n_0 < \frac{N-1}{2}$.

B. Determining A

From the assumption, we can determine the value of A. From Eq (5), if we let n=0, then we can build a quadratic equation and we can determine A by solving a quadratic equation if we know $x(n_0)$, which may be calculated from $r_x(n)$. In other words,

$$A^2 - 2Ax(n_0) + r_x(0) - r_y(0) = 0 \tag{9}$$

Since $x(n_0)$ is the signal value of the defection-free signal $x(n)$, this value is the same as $x(0)$ as well as $x(N-1)$, i.e.,

$$\begin{aligned} x(0)x(N-1) &= x(n_0)^2 = r_x(N-1) \\ x(n_0) &= \sqrt{r_x(N-1)} \end{aligned} \tag{10}$$

IV. Example

In this section, we consider the validity of the equations by sowing an example.

A. Pre calculation

We assume that the defect-free original signal and a defected signal is given as follows.

x= [10 10 10 10 3 3 3 10 10 10 10 3 3 3 10 10 10 10 3 3 3 10
10 10 10 3 3 3 10 10 10 10];
y= [10 10 10 10 3 3 3 5 10 10 10 3 3 3 10 10 10 10 3 3 3 10
10 10 10 3 3 3 10 10 10 10];

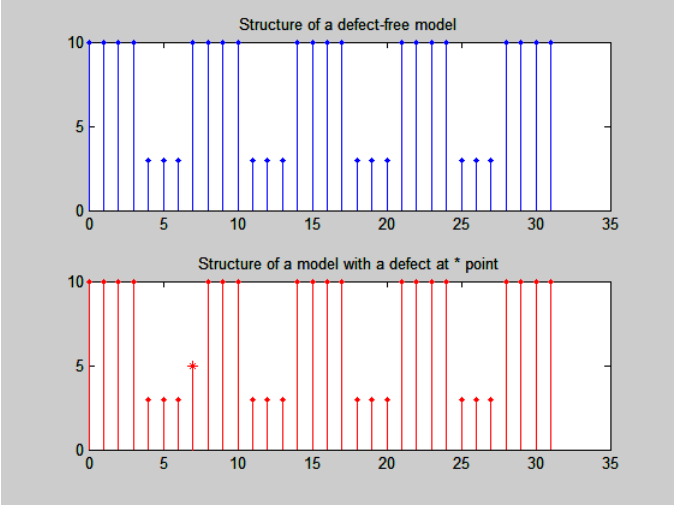


Fig. 2 Graphs of the defect-free original signal and a defected signal. The defect is marked as *.

As we can see in this figure, the location of the defect is at $n_0 = 7$ and $A = 5$. So, if we represent these signals mathematically, then we get

$$y(n) = x(n) - A\delta(n - n_0) = x(n) - 5\delta(n - 7) \quad (11)$$

Now we are going to find out whether we can determine the location and value of the defect. The following table is the autocorrelations of x(n), y(n) and the difference between them, i.e., $r_{xy}(n) = r_x(n) - r_y(n)$.

Table 1. Values of autocorrelation functions and its difference

n	-31	-30	-29	-28	-27	-26	-25	-24	-23	-22	-21
r_x	100	200	300	400	360	320	280	380	529	678	827
r_y	100	200	300	400	360	320	280	330	479	628	777
$r_{xy}(n)$	0	0	0	0	0	0	0	50	50	50	50
n	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10
r_x	738	649	560	660	858	1056	1254	1116	978	840	940
r_y	723	634	545	610	808	1006	1204	1101	963	825	890
$r_{xy}(n)$	15	15	15	50	50	50	50	15	15	15	50
n	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1
r_x	1187	1434	1681	1494	1307	1120	1220	1516	1812	2108	1812
r_y	1137	1384	1581	1429	1242	1055	1155	1451	1747	2033	1747
$r_{xy}(n)$	50	50	100	65	65	65	65	65	65	75	65
n	2	3	4	5	6	7	8	9	10	11	12
r_x	1516	1220	1120	1307	1494	1681	1434	1187	940	840	978
r_y	1451	1155	1055	1242	1429	1581	1384	1137	890	825	963
$r_{xy}(n)$	65	65	65	65	65	100	50	50	50	15	15
n	13	14	15	16	17	18	19	20	21	22	23
r_x	1116	1254	1056	858	660	560	649	738	827	678	529
r_y	1101	1204	1006	808	610	545	634	723	777	628	479
$r_{xy}(n)$	15	50	50	50	50	15	15	15	50	50	50
n	24	25	26	27	28	29	30	31			
r_x	380	280	320	360	400	300	200	100			
r_y	330	280	320	360	400	300	200	100			
$r_{xy}(n)$	50	0	0	0	0	0	0	0			

B. *Determination of* n_0

Since the support of $r_{xy}(n)$ is given as [-24, 24], n_0 will be either 24 or $N-1-n_0$ = 24, i.e., n_0 =7.

C. *Determination of* A

The value of A can be determined from Eq. (9) by solving a quadratic equation. To do this, we need to know the values of $x(n_0)$ and $r_x(0) - r_y(0)$. $x(n_0)$ can be determined from Eq. (10). In this example, $r_x(N-1) = r_x(31) = 100$ and $x(n_0)^2 = 100$, $x(n_0) = 10$. Also, $r_x(0) - r_y(0)$ is given from the table as $r_x(0) - r_y(0) = 2108 - 2033 = 75$. From this we can build the equation , i.e.,

$$A^2 - 2Ax(n_0) + r_x(0) - r_y(0) = A^2 - 20A + 75 = 0 \quad (12)$$

If we solve this equation, we get either A = 15 or A = 5. Since A=15 exceeds $x(n_0)$, which is the maximum value, A=15 is not realistic and so we drop this value. So A should be 5. So the defect can be represented as either $5\delta(n-5)$ or $5\delta(n-24)$. Right now, these two are not distinguishable. However, if we have little more information such as non-symmetric property of the desired pattern, we may distinguish between them.

So, we determined the location of the defect to within a symmetry ambiguity and its value uniquely.

ISBN: 978-1-61804-233-0

V. Conclusion

In this paper, we considered the determination of the location and value of a one-point wide defect. We have made a model for this problem and derived closed form equations to determine the location and value. We have presented an example that shows how the algorithm works. Future work may be the generalization of this algorithm to mode generalized defects including the defects with wider than 1 point and possibly the problems with more than one defects as well as non-stripe type defects.

References

[1] M. S. Yi, T. Haga, C. Walton, and J. Bokor, "High sensitivity actinic detection of native defects on extreme ultraviolet lithography mask blanks," Journal of Vacuum Science & Technology B, vol. 19, pp. 2401-2405, Nov-Dec 2001.

[2] Anna Tchikoulaeva, et al., "EUV actinic blank inspection: from prototype to production", Proc. SPIE 8679, Extreme Ultraviolet (EUV) Lithography IV, 86790I (April 1, 2013); doi:10.1117/12.2011776

[3] Tetsuo Harada, et al., « The Coherent EUV Scatterometry Microscope for Actinic Mask Inspection and Metrology", Proc. of SPIE Vol. 8081, 2011

[4] J. R. Fienup, "Phase retrieval algorithms: a comparison", Applied Optics, Vol. 21, Issue 15, pp. 2758-2769 (1982) http://dx.doi.org/10.1364/AO.21.002758

[5] Kim, Wooshik and Monson H. Hayes. "Phase retrieval using two Fourier transform intensities." Journal of the Optical Society of America A:7(3) 441-449. March, 1990.

[6] Ki-Hyuk Kim, Jung-Guen Jo, Min-Chul Park, et al., "Coherent scattering stereoscopic microscopy for mask inspection of extreme ultra-violet lithography", Proc. SPIE 8738, Three-Dimensional Imaging, Visualization, and Display 2013, 87380Z (29 May 2013); doi: 10.1117/12.2018588

A new Mammography segmentation technique based on Watershed, Wavelet and Curvelet transform

Mohamed Ali HAMDI , Karim Saheb Ettabaa and Mohamed Lamine HARABI
Laboratoires Des Matériaux, Molécules, Université de Carthage
Institut National des Sciences Appliquées et de Technologies, Tunisia
[2]Laboratoire de Recherche en Informatique Arabisée
et Documentique Intégrée (R.I.AD.I)
Ecole Nationale des Sciences de l'Informatique.

Campus Universitaire de Manouba, 3020 Manouba, Tunis, Tunisie

[3] Télécom Bretagne, laboratoire ITI
Technopôle Brest Iroise CS 81828,
39318 Brest Cedex France
Time University, Tunisia

***Abstract*—**With the repaid advancements of computing technology, any use of the computer-based technologies. Increase in different scientific fields. The segmentation's image is an important problem in different fields of computer vision and image processing. This paper presents a new approach for computer aided detection of microcalcifications clusters in digital mammograms. The proposed microcalcification detection method is done With MATLAB. New result of watershed segmentation entirely relay on the image contrast. Image contrast may be degraded during image acquisition. So watershed algorithm can generate under or over segmentation on badly contrast images. In order to reduce these deficiencies of watershed algorithm a preprocessing step using Curvelet or Wavelet transform is performed on input images. The proposed approach is applied to a database of some dense mammographic images, originating from the MIAS database. Results shows that the proposed approach gives a satisfactory detection performance

***Keywords*—**Image processing, Curvelet transform, segmentation, watershed.

I. INTRODUCTION

Breast cancer is one of the frequent and leading causes of mortality among woman, in developed countries. The age is one of the risk factor of breast cancer and women within the age between 40-69 have more risk of the breast cancer. In west about 53%-92% of the population has this disease. In a study a mammogram was done to 151,198 women. Though early detection of breast cancer can increase the survival rate. The current diagnostic method for early detection of breast cancer is the mammography, so mammographie is low dose X-ray projections of the breast, and it's the best method for detecting cancer at the early stage. Microcalcifications are tiny bits of the calcium, and may be show in patterns or in clusters and they are associated with extra cell activities in breast .Usually an extra cell growth isn't cancerous, yet sometimes tight clusters of microcalcifications can indicate early breast cancer. Macrocalcifications are usually a sign of benign breast cancer. 82% of the microcalcification is benign. microcalcifications in the breast show up as white speckles on breast X-rays image. The calcifications are usually small; varying from 100 micrometer to 300 micrometer, Therefore it's very difficult to detect the microcalcifications as such, when more than 10 microcalcifications are clustered together, it becomes possible to diagnose malignant disease. So the survival depends on how early the cancer is detected. But any calcification formation should be detected at the benign stage. Hence, the Computer Aided Diagnosis system is used to detect calcification clusters. Many algorithms have been proposed for automatic detection of breast cancer in mammographies. Features extracted from mammograms can be used for detection of cancers.

The method for preprocessing on the images is Curvelet or Wavelet transform is a multiscale directional transform which allows an almost optimal non adaptive sparse representation of objects with edges. It has generated increasing interest in the community of applied mathematics and signal processing over the past years. It is an approach used to enhance the image contrast when image is degraded.

II. THE CURVELET TRANSFORM

The Curvelet transform is a multiscale directional transform and a higher dimensional of the Wavelet transform which allows an optimal non adaptive representation of edges

designed to represent images at different scales and different angles. [1][2]Curvelet has two mathematical properties [3] [4]: Please submit your manuscript electronically for review as e-mail attachments.

a- The curved singularities should be approximated with some coefficients and in a non-adaptive representation named «curvelets."

b- The curvelets remain coherent waveforms under the action of the wave equation in a smooth medium A curvelet frame [15] {φ_γ} is a wave packet frame on L2 R2 based on second dyadic decomposition:

$$f(x) = \sum_\gamma C_\gamma \varphi_\gamma(x) \qquad C_\gamma = \int f(x)\varphi_\gamma(x)\,dx$$

Then the Frequency shells is $2k<|\xi|<2\ (k+1)$; and the Angular Sectors is $\propto(\omega, \xi)\leq 2\ (-k/2)$ finally the approximation rate is optimal:

a- We choose n largest coefficients)

b- And no frame can do better for jumps along C2curves. C_γ *in* $f = \sum_\gamma C_\gamma \varphi_\gamma$ and $\|f - f_n\|_{L^2}^2 \leq n^{-2}\log(n)^3$

c- The wavelet expansion : $\|f - f_n\|_{L^2}^2 \leq n^{-1}$

III. THE WAVELET TRANSFORM

The wavelet transform is a mathematical tool that can be used to describe images in multiple resolutions. The wavelet decomposition is a complete representation, since it allows a perfect reconstruction of the original image . Also, since a low-pass filter is involved, noise suppression is inherent to this transform. According to Mallat's pyramid algorithm the input image is convolved with low-pass and high-pass filters associated with a mother wavelet, and downsampled afterwards. Four images (each one with half the size of the original image) are produced, corresponding to high frequencies in the horizontal direction and low frequencies in the vertical direction (HL), low frequencies in the horizontal direction and high frequencies in the vertical direction (LH), high frequencies in both directions (HH) and low frequencies in both directions (LL). This last image is a low-pass version of the original image, and will be called the approximation image. This procedure is repeated for the approximation image at each resolution 2j (please note that dyadic scales are used). The four images HL, LH, HH and LL are denoted, respectively, by , and A2j. If the wavelet transform is applied up to the scale 2J, the original image can be reconstructed using images A2J and . In this paper, the Haar wavelet was chosen because of its orthogonality and, more important, its small support. Also, it requires small computational complexity (linear with respect to the size of the input image) to compute the wavelet decomposition with the Haar wavelet. The expressions for the low-pass h[n] and high-pass g[n] filters for the Haar wavelets are provided in this equation

$$h[n] = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right], \qquad g[n] = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]$$

IV. THE WATERSHED TRANSFORM

The watershed transform [6] is a morphological based tool for image segmentation. In grey scale the mathematical morphology watershed transform for segmentation is originally proposed by Digabel and Lantuejoul in1977 and later improved by Li et Al in 2003. The watershed transform can be classified as a region-based segmentation approach.

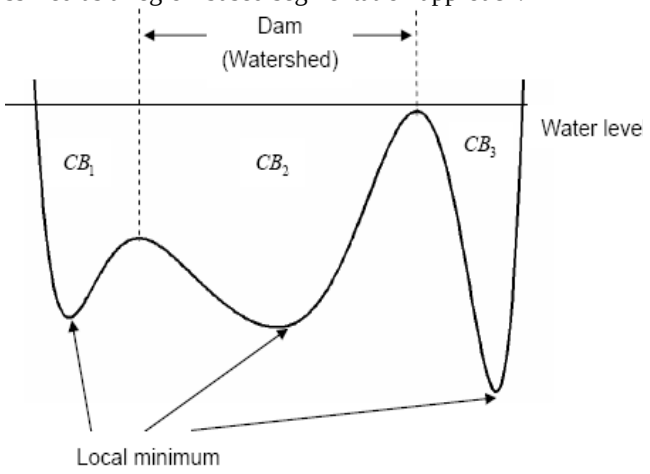


Fig. 1 Illustration of immersion process of watershed transforms. (CB: Catchment basins)

The idea [5] of watershed can be view as a landscape immersed in a lake; catchment basins will be filled up with water starting at each local minimum. Dams must be built where the water coming from different catchment basins may be meeting in order to avoid the merging of catchment basins.

The water shed lines are defined by the catchment basins divided by the dam at the highest level where the water can reach in the landscape [7]. As a result, watershed lines can separate individual catchment basins in the landscape. The idea is described in Figure 1 which describes the flooding or rain falling process of watershed algorithm. The process of rain falling is described in Figure 2.

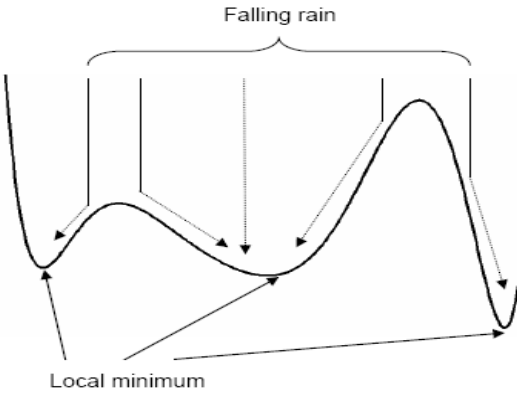


Fig. 2 Illustrations of flooding (process of watershed transform)

V. PROPOSED TECHNIQUE

The proposed Mammography segmentation algorithm here is able to segment the mammograms with minimum drawbacks of under segmentation and over segmentation. The steps of the proposed system are shown in Figure 3.

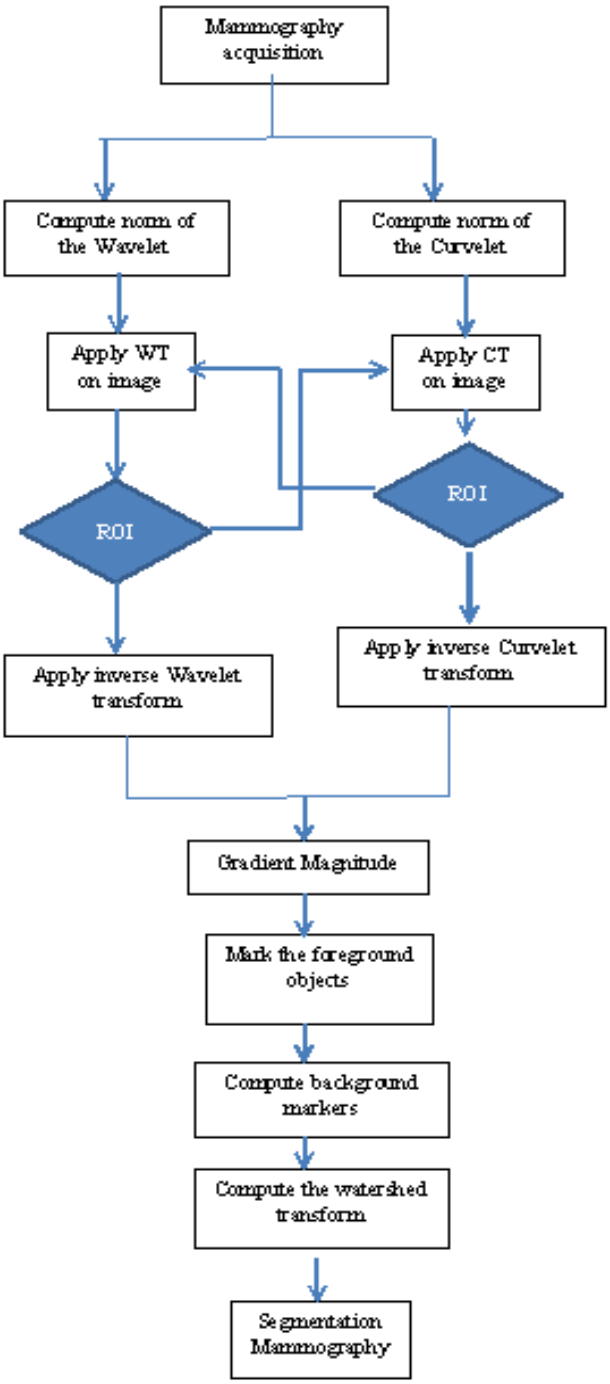
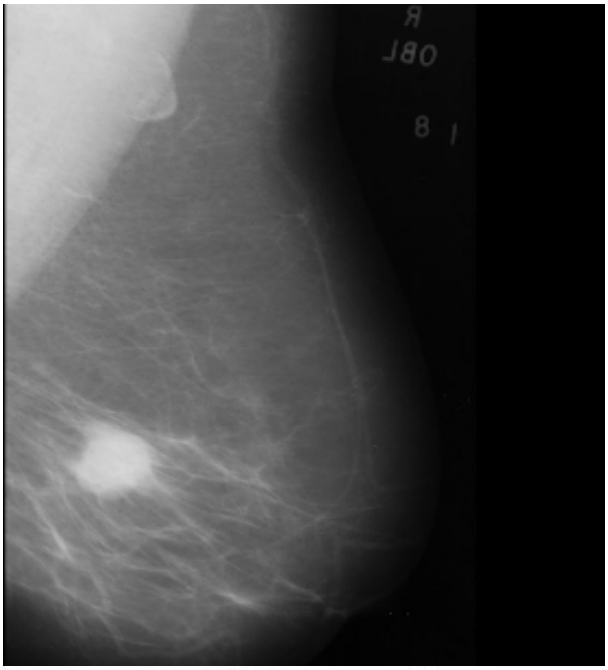
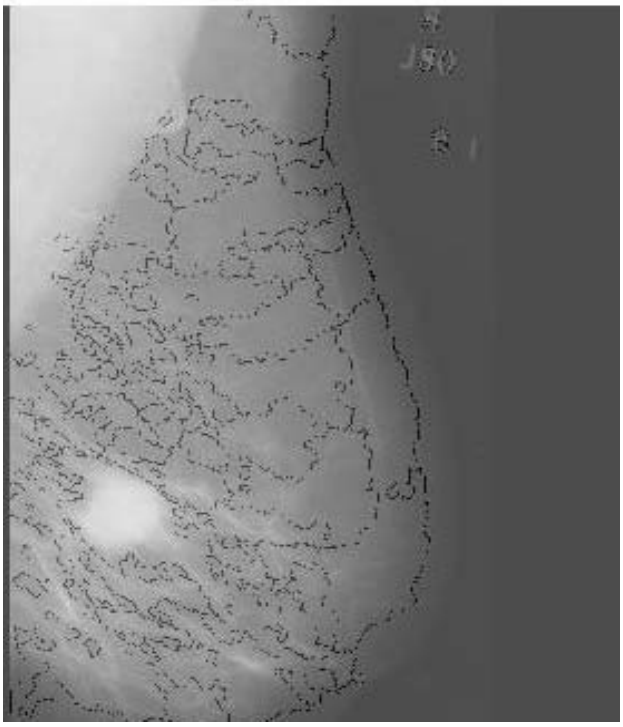


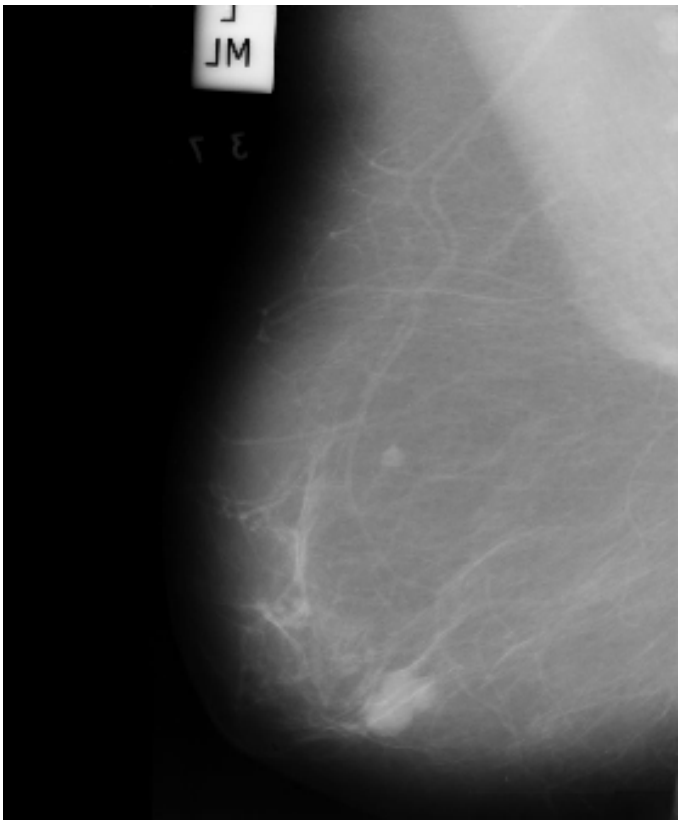
Fig. 3 Proposed process for image segmentation

VI. Results

The results of the proposed techniques are described in this section, we used MIAS database for mammograms:







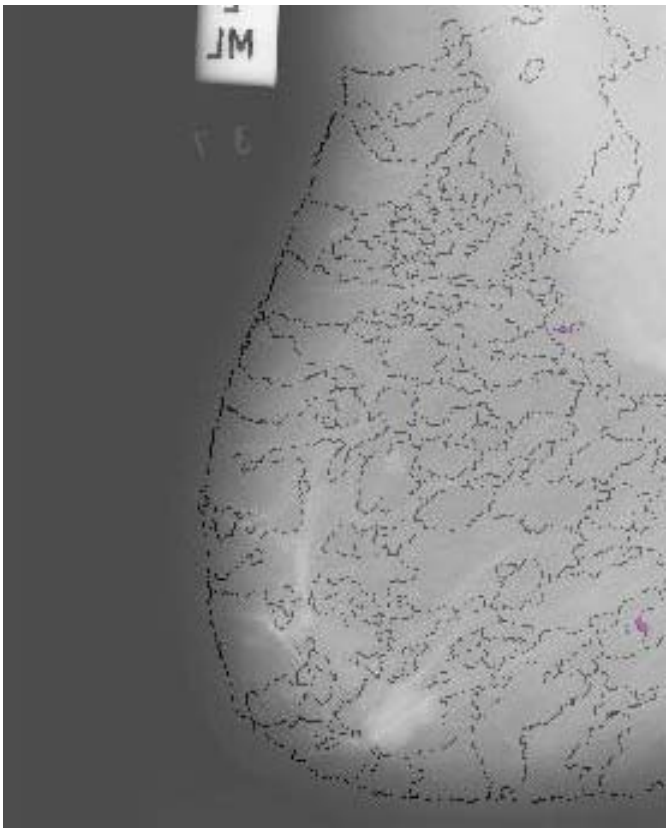


Fig. 4 Original Mammograhies and segmentation results

VII. Conclusion

The goal of Mammography segmentation process is to identify the segments of the image according to the image characteristic e.g. Cancer shape. The simplified working of the image segmentation system is stated here. The most important step is the mammography acquisition. Any lack during mammography acquisition can cause many problems in the result. The image used in this process is taken from MIAS database. The input images are of low contrast. This segmentation process deals with the problem caused by these low contrast images by applying a preprocessing step using Curvelet or Wavelet transform. This step enhances the contrast of the input image so that the gradient of the image is strong enough to properly segment the image by using the watershed. After preprocessing step the gradient of the image is finding by converting the input image to grey scale. And this gradient of image is used as the input the image. The results show the improvement in the segmentation results using Curvelet or Wavelet transform. The system will inspected only one mammography at a time. This system can be very helpful for the segmentation of the mammography which are used in different computer's diagnosis. And the mammography analysis process can be enabled by this system. The research content of this system was segmentation and image enhancement.

References

[1] Donoho D. L. and Duncan M. R. 1999. Digital Curvelet Transform Strategy, Implementation and Experiments. http://www.curvelet.org/papers/DCvT99.pdf

[2] Starck J.L., Candes E. and Donoho D.L. 2002. The Curvelet Transform for Image Denoising. IEEE Transactions on Image Processing. Vol. 11(6): 670-684.

[3] Jean-Luc Starck, Fionn Murtagh, Jalal M. Fadili, "SPARSE IMAGE AND SIGNAL PROCESSING Wavelets, Curvelets, Morphological Diversity" cambridge university press Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, S˜ao Paulo, Delhi, Dubai, Tokyo Cambridge University Press 32 Avenue of the Americas, New York, NY 10013-2473, USA www.cambridge.org Information on this title: www.cambridge.org/9780521119139

[4] Starck J.L., Candes E., Murtagh F. and Donoho D.L. 2003. Gray and Color Image Contrast Enhancement by the Curvelet Transform. IEEE Transaction on Image Processing. Vol. 12(6): 706-717.

[5] Yuqian Zhao; Jianxin Liu; Huifen Li; Guiyuan Li; " Improved watershed algorithm for dowels image segmentation" Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on Digital Object Identifier: 10.1109/WCICA.2008.4594115 Publication Year: 2008 , Page(s): 7644 - 7648 IEEE CONFERENCES

[6] Shuang Wang; Xiuli Ma; Xiangrong Zhang; Licheng Jiao " Watershed-based textural image segmentation" Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on Digital Object Identifier: 10.1109/ISPACS.2007.4445886 Publication Year: 2007 , Page(s): 312 – 315 IEEE CONFERENCES

[7] M.A. HAMDI and M.A. MRABTI "Combination of Morpho-logical Snakes and Curvelet Transform for IVUS Contours Detection", International Journal of Tomography and simulation, 2012, volume 21, issue3, 48-57

[8] M.A. HAMDI, "Combining Contourlet Transform and Water-shed for Robust Multi-Scale Image Segmentation", International Journal of Imaging and robotics, Volume 8, Issue 2, 2012, 49-54.

Multiplayer Games, Competitive Models, and Descriptive Computing

Cyrus F Nourani *and Oliver Schulte**

Abstract— Agent game plan competitive model computing. agent game trees and agent intelligent languages are presented with description computing logic models. A basis to model discovery and prediction planning is stated. The computing model is based on a novel competitive learning with agent multiplayer game tree planning. The techniques are developed on a descriptive game logic where model compatibility is characterized on von Neumann, Morgenstern, Kuhn VMK game descriptions model embedding and game goal satisfiability. The techniques apply to both zero-sum and arbitrary games. The new encoding, with a *VMK game function situation,* where agent sequence actions, are have embedded measures. The import is that game tree modeling on VMK encodings are based on computable partition functions on generic game model diagrams. The encoding allows us to reach newer computability areas on game trees based on agent game trees on VMK game models. Novel payoff criteria on game trees and game topologies are obtained. Epistemic accessibility is addressed on game model diagrams payoff computations. Furthermore criteria are presented to reach a sound and complete game logic based on VMK and hints on applications to compute Nash equilibrium criteria and a precise mathematical basis is stated.

Keywords— Game trees, Game Sequent Description
VMK Games, Game Tree Models, Competitive Model
Computing, Game Description Logic
Nash Games, VMK Game Trees

* SFU, Computation Logic Labs, Burnaby, BC, Canada Acdmkrd.tripoid.com: cyrusfn@alum.mit,edu and OSchulte@cs.sfu.ca

I. INTRODUCTION

Modeling, objectives, and planning issues are examined with agent planning and competitive models. Model discovery and prediction is applied to compare models and get specific confidence intervals to supply to goal formulas. Competitive model learning is presented based on the new agent computing theories the first author had defined since 1994. The foundations are applied to present precise decision strategies on multiplayer games with only perfect information between agent pairs. The game tree model is applied to train models. The computing model is based on a novel competitive learning with agent multiplayer game tree planning. Specific agents transform the models to reach goal plans where goals are satisfied based on competitive game tree learning. Intelligent and/or trees and means-end analysis were applied with agents as the hidden –step computations.

A novel multiplayer game model is presented where "intelligent" agent enriched languages can be applied to address game questions on models in the mathematical logic sense. The game tree model is applied to encode von Neumann, Morgenstern, Kuhn games that were rendering on a situation calculus on (Schulte and Delgrande 2004). The sections outline is as follows. Section two briefs on knowledge representation on game tress and games as plans.

Minimal predictions and predictive diagrams are examined on applications to game trees as goal satisfaction. Section 3 briefs on competitive models as a basis for realizing agent game trees as goals. Agent and/or trees are presented to carry on game tree computations. Infinitary multiplayer games are examined and modeling questions with generic diagrams are studied. Basic two person agent games and the group characterizations based on the first authors preceding publications are briefed to lead onto newer area the second author and company had developed on von Neumann, Morgenstern, Kuhn games. Games, situations, and compatibility are stated with a new encoding on the above game situations that are based on game tuple ordinals computed on generic diagrams. Game descriptions based on first author's description logic on generic diagrams are applied to KVM to reach new model- theoretic game tree bases. The new KVM encoding, *VMK game function situation, and game tree model diagrams,* embeds aggregate measures as lifted preorders based on the game tree nodes information partition functions allows us embed measures on game trees where agent sequence actions, be that random sequences, are modeled. The import is that game tree modeling, KVM encodings based on partition functions on generic game model diagrams, allows us to reach newer computability areas on game trees based on agent game trees on KVM game models. Novel payoff criteria on game trees and game topologies are obtained. Applying that to a game tree diagram models and generic diagram embedding's, payoff function characterizations are presented based on game tree topologies. Epistemic accessibility is addressed on game model diagrams payoff computations. Section 5 presents criteria on a sound and complete game logic based on KVM and hints on applications to compute Nash equilibrium criteria. Furthermore, games and tractability areas are addressed based on descriptive computing, cardinality on concept descriptions, and descriptive computability. Sections 6 briefs on computing games on proof trees based on plan goal satisfaction, predictions and game tree models, following the first author's publications past decade.

II. KR PLAN DISCOVERY

Modeling with agent planning is applied where uncertainty, including effector and sensor uncertainty, are relegated to agents, where competitive learning on game trees determines a confidence interval. The incomplete knowledge modeling is treated with KR on predictive model diagrams. Model discovery at KB's are with specific techniques defined for trees. Model diagrams allow us to model- theoretically characterize incomplete KR. To key into the incomplete knowledge base we apply generalized predictive diagrams whereby specified diagram functions a search engine can select onto localized data fields. The predictive model diagrams (Nourani 1995) could be minimally represented by the set of functions {f1,...,fn} Data discovery from KR on diagrams might be viewed as satisfying a goal by getting at relevant data which instantiates a goal. The goal formula states what relevant data is sought. We propose methods that

ISBN: 978-1-61804-233-0

can be applied to planning (Nourani 1991) with diagrams to implement discovery planning. In planning with G-diagrams that part of the plan that involves free Skolemized trees is carried along with the proof tree for a plan goal. Computing with diagram functions allows us to key to active visual databases with agents. Diagrams are well-known concepts in mathematical logic and model theory. The diagram of a structure is the set of atomic and negated atomic sentences that are true in that structure. Models uphold to a deductive closure of the axioms modeled and some rules of inference. A generic model diagram (G-diagram) (Nourani 1991,94a)
is a diagram in which the elements of the structure are all represented by a specified minimal set of function symbols and constants, such that it is sufficient to define the truth of formulas only for the terms generated by the minimal set of functions and constant symbols. Such assignment implicitly defines the diagram. It allows us to define a canonical model of a theory in terms of a minimal family of function symbols. The minimal set of functions that define a G- diagram are those with which a standard model could be defined.

A. Predictions and Discovery

Minimal prediction is an artificial intelligence technique defined since the author's model-theoretic planning project. It is a cumulative nonmontonic approximation (Nourani 1999c) attained with completing model diagrams on what might be true in a model or knowledge base. A *predictive diagram* for a theory T is a diagram D (M), where M is a model for T, and for any formula q in M, either the function f: q → {0,1} is defined, or there exists a formula p in D (M), such that T∪{p} proves q; or that T proves q by minimal prediction. A *generalized predictive diagram* is a predictive diagram with D (M) defined from a minimal set of functions. The predictive diagram could be minimally represented by a set of functions {f1,...,fn} that inductively define the model. The free trees we had defined by the notion of provability implied by the definition, could consist of some extra Skolem functions {g1,...,gm} that appear at free trees. The f terms and g terms, tree congruences, and predictive diagrams then characterize partial deduction with free trees. The predictive diagrams are applied to discover models to the intelligent game trees. Prediction is applied to plan goal satisfiability and can be combined with plausibility (Nourani 1991) probabilities, and fuzzy logic to obtain, for example, confidence intervals.

B. Business Modeling Examples

The following examples from (Nourani 2010) can be motivating for business applications. Example: A business manager has 6 multitalented players, designed with personality codes indicated with codes on the following balls. The plan is to accomplish 5 tasks with persons with matching personality codes to the task, constituting a team. Team competitive models can be generated with comparing teams on specific assignments based on the task area strength. The optimality principles outlined on the first author's publications might be to accomplish the goal with as few a team grouping as possible, thereby minimizing costs.

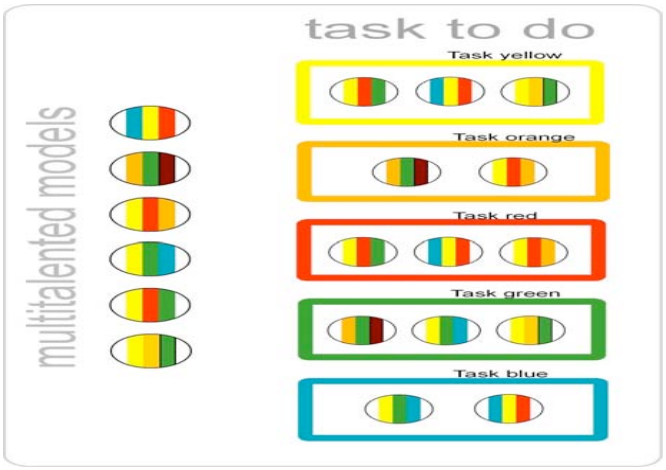


Figure 1: Multitalented models

The following section presents new agent game trees the author had put forth [26]. Applying game theory to business is tantamount to an interactive decision theory. Decisions are based on the world as given. However, the best decision depends on what others do, and what other do may depend on what they think you do. Hence games and decisions are intertwined. A second stage business plan needs to specify how to assign resources with respect to the decisions, ERP plans, and apply that to elect supply chain policies, which can in part specify how the business is to operate. A tactical planning model that plans critical resources up to sales and delivery is a business planner's dream. Planning and tasking require definition of their respective policies and processes; and the analyses of supply chain parameters. The above are the key elements of a game, anticipating behavior and acquiring advantage. The players on the business planned must know their options, the incentives, and how do the competitors think.
Example premises: Strategic Interactions
Strategies :{ Advertise, Do Not Advertise}
Payoffs:Companies' Profits Advertising costs $ million
The branches compute a Boolean function via agents. The Boolean function is what might satisfy a goal formula on the tree. An intelligent AND/OR tree is solved if the corresponding Boolean functions solve the AND/OR trees named by intelligent functions on the trees. Thus node m might be f (a1,a2,a3) & g(b1,b2), where f and g are Boolean functions of three and two variables, respectively, and ai's and bi's are Boolean valued agents satisfying goal formulas for f and g. The agent trees are applied to satisfy goals to attain competitive models for business plans and ERP.

The And vs. Or principle is carried out on the above trees with the System to design ERP systems and manage as Cause principle decisions. The agent business modeling techniques the author had introduced [25,28] apply the exact 'system as cause' and 'system as symptom' based on models (assumptions, values, etc.) and the 'system vs. symptom' principle via tracking systems behavior with cooperating computational agent trees. The design might apply agents splitting trees, where splitting trees is a well-known decision tree technique. Surrogate agents are applied to splitting trees. The technique is based on the first author's intelligent tree

project ECAI 1994 and European AI Communication journal are based on agent splitting tree decisions like what is designed later on the CART system: The ordinary splitting tree decisions are regression based, developed at Berkeley and Stanford (Brieman, Friedman, et.al. 1984), (Breiman 1996). The agent splitting agent decision trees have been developed independently by the author since 1994 at the Intelligent Tree Computing project. For new directions in forecasting and business planning (Nourani 2002). Team coding example diagram from reach plan optimal games where a project is managed with a competitive optimality principle is to achieve the goals minimizing costs with the specific player code rule first author and company 2005.

III. COMPETITIVE MODELS GAMES

Planning is based on goal satisfaction at models. Multiagent planning, for example as (Muller-Pischel 1994, Bazier et.al. 1997), in the paper is modeled as a competitive learning problem where the agents compete on game trees as candidates to satisfy goals hence realizing specific models where the plan goals are satisfied**.** When a specific agent group "wins" to satisfy a goal the group has presented a model to the specific goal, presumably consistent with an intended world model. For example, if there is a goal to put a spacecraft at a specific planet's orbit, there might be competing agents with alternate micro-plans to accomplish the goal. While the galaxy model is the same, the specific virtual worlds where a plan is carried out to accomplish a real goal at the galaxy via agents are not. Therefore, Plan goal selections and objectives are facilitated with competitive agent learning. The intelligent languages (Nourani 1993d) are ways to encode plans with agents and compare models on goal satisfaction to examine and predict via model diagrams why one plan is better than another or how it could fail. Virtual model planning is treated in the author's publications where plan comparison can be carried out at VR planning (Nourani 1999b).

Games play an important role as a basis to economic theories. Here the import is brought forth onto decision tree planning with agents. Intelligent tree computing theories we have defined since 1994 can be applied to present precise strategies and prove theorems on multiplayer games. Game tree degree with respect to models is defined and applied to prove soundness and completeness. The game is viewed as a multiplayer game with only perfect information between agent pairs. Upper bounds on determined games are presented. The author had presented a chess-playing basis in 1997 to a computing conference.

For each chess piece a designating agent is defined. The player P makes its moves based on the board B it views. <P,B> might view chess as if the pieces on the board had come alive and were autonomous agents carrying out two-person games as in Alice in Wonderland. Game moves are individual tree operations.

A. Intelligent AND/OR Trees and Search

AND/OR trees Nilsson (1969) are game trees defined to solve a game from a player's standpoint.

Figure2:And/OrTrees

Formally a node problem is considered solved if one of the following conditions hold. 1. The node is the set of terminal nodes (primitive problem- the node has no successor). 2. The node has AND nodes as successors and the successors are solved. 3. The node has OR nodes as successors and any one of the successors is solved. A solution to the original problem is given by the subgraph of AND/OR graph sufficient to show that the node is solved. A program which can play a theoretically perfect game would have task like searching and AND/OR tree for a solution to a one-person problem to a two-person game. An intelligent AND/OR tree is and AND/OR tree where the tree branches are intelligent trees (Nourani 1991). The branches compute a Boolean function via agents. The Boolean function is what might satisfy a goal formula on the tree. An intelligent AND/OR tree is solved iff the corresponding Boolean functions solve the AND/OR trees named by intelligent functions on the trees. Thus node m might be f(a1,a2,a3) & g(b1,b2), where f and g are Boolean functions of three and two variables, respectively, and ai's and bi's are Boolean valued agents satisfying goal formulas for f and g.

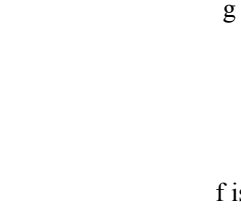


Figure 3: Agent And/Or Trees

The chess game trees can be defined by agent augmenting AND/OR trees (Nourani 1990's). For the intelligent game trees and the problem solving techniques defined, the same model can be applied to the game trees in the sense of two person games and to the state space from the single agent view. The two person game tree is obtained from the intelligent tree model, as is the state space tree for agents. To obtain the two-person game tree the cross-board-coboard agent computation is depicted on a tree. Whereas the sate-space trees for each agent is determined by the computation sequence on its side of the board-coboard. A **game tree degree** is the game state a tree is at with respect to a model truth assignment, e.g. to the parameters to the Boolean functions above. Let generic diagram or G-diagrams be diagrams definable by specific functions. Intelligent signatures (Nourani 1996a) are signatures with designated multiplayer game tree function symbols. A soundness and completeness theorem is proved on the intelligent signature language Nourani (1999a). The techniques allowed us to present a novel model- theoretic basis to game trees, and generally to the new intelligent game trees. The following

specifics are from (Nourani 1999). Let N be the set of all functions from w to w. Let A be a subset of N. (Gale-Stewart,1953) associated with A a 2-person game of prefect information G<A>. Player I begins by choosing n0 in w; player two chooses n1 in w; then I chooses n2 in w; so on. Let a(i) = ni. I wins G<A> if and only if a in A. We say that G<A> is determined if one of the players has a winning strategy.

Proposition 1 If G<A> is determined, the complexity upper bound on the number of moves to win is A's cardinality.

Theorem 2 For every pair p of opposing agents there is a set A<p> (N. The worse case bound for the number of moves for a determined game based on the intelligent game tree model is the sum ({|A<p>|:p agent pairs}.

Proof Sum over the proposition (Nourani 1999).

At the intelligent game trees the wining agents determine the specific model where the plan goals are satisfied.

On the present paper we are considering game models, zero-sum, or not, however not carrying on with intent agent analytics (Nourani AAAI 2002). At the agent intelligent game trees the wining agents determine the specific model where the plan goals are satisfied. On what follows we first examine game situations and models that can be applied to zero-sum or nondeterministic based on game goals are satisfied on game tree models.

B. Games, Situations, and Compatibility

We begin with a basic mathematical definition on what situations are followed with what model
diagrams are and we can characterize situations with generic model diagrams.

Definition 3.1 A situation consists of a nonempty set D, the domain of the situation, and two mappings: g,h. g is a mapping of function letters into functions over the domain as in standard model theory. h maps each predicate letter, pn, to a function from Dn to a subset of {t,f}, to determine the truth value of atomic formulas as defined below. The logic has four truth values: the set of subsets of {t,f}.{{t},{f},{t,f},0}. The latter two is corresponding to inconsistency, and lack of knowledge of whether it is true or false.

Due to the above truth values, the number of situations exceeds the number of possible worlds. The possible worlds being those situations with no missing information and no contradictions. From the above definitions the mapping of terms and predicate models extend as in standard model theory. Next, a **compatible set of situations** is a set of situations with the same domain and the same mapping of function letters to functions. In other worlds, the situations in a compatible set of situations differ only on the truth conditions they assign to predicate letters. Models uphold to a deductive closure of the axioms and some rules of inference as specified. By the definition of a diagram they are a set of atomic and negated atomic sentences. Thus the diagram might be considered as a basis for a model, provided we can by algebraic extension, define the truth value of arbitrary formulas instantiated with arbitrary terms. Thus all compound sentences build out of atomic sentences then could be assigned a truth value, handing over a model. It will be made clearer in the following subsections. The following examples would run throughout the paper. Consider the primitive first order language (FOL)

From the first authors a decades publications encoding with generic world models sepifics are as follows.

L = {c},{f(X)},{p(X),q(X)}

Let us apply Prolog notation convention for constants and variables) and the simple theory {for all X: p(X) → q(X),p(c)}, and indicate what is meant by the various notions.

[model] = {p(c),q(c),q(f(c)),q(f(f(c))),...},{p(c) ∧q(c), p(c) ∧ p(X), p(c) &p(f(X)), ...}, {p(c) v p(X), p(c) v p(f(X)), p(c) → ¬p(c)...}

[diagram] = {p(c),q(c),p(c),q(f(c)),q(f(f(c))),...},...,q(X)} i..e. diagram = the set of atomic formulas of a model.

Thus the diagram is [diagram]= {p(c),q(c),q(f(c)),q(f(f(c))), ...,q(X)}

Based on the above, we can define generalized diagrams. The term generalized is applied to indicate that such diagrams are defined by algebraic extension from basic terms and constants of a language. The fully defined diagrams make use of only a minimal set of functions. Generalized diagram is

[generalized diagram]= {p(c),q(c),p(f(t)),q(f(t))} for t defined by induction, as {t0=c , and tn= {f(t(n- 1))} for n>0.

It is thus not necessary to redefine all f(X)'s since they are instantiated. Nondeterministic diagrams are those in which some formulas are assigned an indeterminate symbol, neither true nor false, that can be arbitrarily assigned in due time.

[nondeterministic diagram] = {p(c),q(c),p(f(t)),q(f(c)),q(f(f(c))),I_q(f(s)))}, t is as defined by induction before and I_q(f(s)) = I_q for some indeterminate symbol I_q, for {s=t n, n>=2. These G-diagrams are applicable for KR in planning with incomplete knowledge with backtracking **Free Skolemized diagrams** are those in which instead of an indeterminate symbols there are free Skolem functions, that could be universally quantified.

[free Skolemized diagram] = {p(c),q(c),p(f(t)),q(f(c)),q(f(f(c))),q_F.s(s)}, where t and s are as defined in the sections before. These G-diagrams are applicable to KR for planning with incomplete knowledge [11] and free proof trees.

A **generalized free diagram** (GF-diagram) is a diagram that is defined by algebraic extension form a minimal set of function symbols.

[generalized free diagram] = {p(c),q(c),p(f(t)),q(f(t)), q_F.s(s)} for t and s as before. A generalized free plausible diagram, is a GFD that has plausibility degrees assigned to it. These G-diagrams are applied for KR to planning with free proof trees [11,18]

C. Game Tree Projections on Model Diagrams

Game tree diagrams are encoded with model diagram functions. Repeating the example from the preceding sections. A game tree node m might be f(a1,a2,a3) & g(b1,b2), where f and g are Boolean functions of three and two variables, respectively, and ai's and bi's are Boolean valued agents satisfying goal. On a the above game model diagram the tree is defined with two terms f(a1,a2,a3) & g(b1,b2) on a two stage collapse to a model diagram that can be assigned a Boolean value. Typical assignments might be in general: f(t1,...,tn)àà{T,F, H(t1,...,tn)}, when t1,...,tn are free

ISBN: 978-1-61804-233-0

Skolemized agent trees, and h is a tupled morphism <h1,...,hn> on the undetermined trees on a Boolean algebra, T,F, are True,False assignments,

Definition 3.2 Let M be a structure for a language L, call a subset X of M a generating set for M if no proper substructure of M contains X, i.e. if M is the closure of X U {c[M]: c is a constant symbol of L}. An assignment of constants to M is a pair <A,G>, where A is an infinite set of constant symbols in L and G: A → M, such that {G[a]: a in A} is a set of generators for M. Interpreting a by g[a], every element of M is denoted by at least one closed term of L[A]. For a fixed assignment <A,G> of constants to M, the diagram of M, D<A,G>[M] is the set of basic [atomic and negated atomic] sentences of L[A] true in M. [Note that L[A] is L enriched with set A of constant symbols.]

Definition 3.3 A Generic diagram for a structure M is a diagram D<A,G>, such that the G in definition 3.2 has a proper definition by specific function symbols.

Remark: The functions above are those by which a standard model could be defined by inductive definitions. The dynamics of epistemic states as formulated by generalized diagrams (Nourani 1998, 91) is exactly what addresses the compatibility of situations. What it takes to have an algebra and model theory of epistemic states, as defined by generalized diagram of possible worlds is what (Nourani 1998, 1991) carrys on with. To decide compatibility of two situations we compare their generalized diagrams. Thus we have the following Theorem.

Theorem 3 Two situations are compatible iff their corresponding generalized diagrams are compatible with respect to the Boolean structure of the set to which formulas are mapped (by the function h above, defining situations).

Proof The generic diagrams, definition 4.3, encode possible worlds and since we can define a one- one correspondence between possible worlds and truth sets for situations (Nourani 1998, 1991, 1984), computability is definable by the G-diagrams.

The computational reducibility areas were briefed at (Nourani 1998, and 2009). An implication is that with infinitary logic we can form an infinite conjunction of beliefs with respect to possible worlds. The worlds are represented with the compatible generalized diagrams that can characterize reaching for an object form the model-theoretic point of view, for example in a Kantian sense, or reaching goals on game trees.

IV. Descriptions and Games

Von Neumann and Morgenstern VM-Short designed game theory as a very general tool for modeling agent interactions. VM game theory in a logical formalism, so as to enable computational agents to employ game-theoretical models and reasoning for the interactions that they engage in. Let us consider VM games in which the agents can take at most countably many actions and in which the agents' payoff functions are continuous. (Schulte et. al. 2004) show that every such VM game G has an axiomatization Axioms(G) in the situation calculus that represents the game in the following strong sense: The game G itself is a model of Axioms(G), and all models of Axioms(G) are isomorphic. It follows that the axiomatization is correct, in the sense that Axioms (G) entails only true assertions about the game G, and complete in the sense that Axioms(G) entails all true assertions about the game G.

Let us call axiomatizations A(G) where all models of A(G) are isomorphic a categorical axiomatization.

Definition 4.4 (von Neumann, Morgenstern, Kuhn). A sequential game G is a tuple

⟨N,H,player,fc,{Ii},{ui}⟩ whose components are as follows.

1. A finite set N (the set of *players).*

2. A set H of sequences satisfying the following three properties. (a) The empty sequence ∅ is a member of H .

(b) If σ is in H, then every initial segment of σ is in H. (c) If h is an infinite sequence such that every finite initial segment of h is in H , then h is in H .

Each member of H is a *history;* each element of a history is an *action* taken by a player. A history σ is *terminal* if there is no history s ∈ H such that σ ⊂ s. (Thus all infinite histories are terminal.) The set of terminal histories is denoted by Z. The set of actions available at a finite history s is denoted by A(s) = {a: s∗a ∈ H}.

A function *player* that assigns to each nonterminal history a member of N ∪ {c}. The function *player* determines which player takes an action after the history s. If player(s) = c, then it is "nature's" turn to make a chance move.

3. A function fc that assigns to every history s for which player(s) = c a probability measure

fc(·|s) on A(s). Each probability measure is independent of every other such measure. (Thus fc(a|s) is the probability that "nature" chooses action a after the history s.)

4. For each player i ∈ N an *information partition* Ii defined on {s ∈ H : player(s) = i}. An element Ii of Ii is called an *information set* of Player i. We require that if s, s′ are members of the same information set Ii, then A(s) = A(s′).

5. For each player i ∈ N a *payoff function* ui : Z → R that assigns a real number to each terminal history.

An element *Ii* of *Ii* is called an *information set* of Player *i.* We require that if *s, s* are members of the same information set *Ii,* then *A(s) = A(s).* 6. For each player *i* ∈ N a *payoff function ui* :Z → R that assigns a real number to each terminal history.

To examine game models we present an algebraic characterizations that is not specific on the game so far as the probability measure assignment on game sequences are concerned. The information set on the game nodes are encoded on the game diagram. Whatever the game degree is at the node is how the following sequence is carried on. On that basis when can examine games based on satisfying a goal, where the compatibilist are characterized on specific VMK model embedding that are stated on a following section.

Let us state a categorical VMK situation. For the time being consider modeling zero-sum games and not adversarial games not to confuse the modeling questions. Modeling agent games based on game tree satisfiability does not require explicit probability measures. The information partition can be encoded such that whatever the unknown probabilities are, that in any case are very difficult to ascertain, the game tree

node moves can be determined as the game develops. In section 3.7 we present a treatment on measures and payoffs.

Definition 4.5 A ***VMK game function situation*** consists of a domain *N x H, where N is the natural numbers, H a set of game history sequences, and a mapping pair g,f. f maps function letters to (agent) functions and g maps pairs from <N,H> to {t.f}*.

Notes: On condition 5 at VMK definition 4.4 the information partition for a player I is determined at each game node by the game diagram game tree degree (section 4.1) at that node that is known to that player, c.f. (Nourani 1994, 1996).

The basic intricacy is what can you say about conditions on H on the encoding on a basic definition: H for example: that agent player plays to increases the pay on ordinals on every move. That is, there is an ordering on the functions on <N,H>, that is from a mathematical absolute. That is essentially a pre-ordered mapping to {t,f} where the ordering is completed on stages. However, the ordering is not known before a game goal is satisfied.

A. Models Preliminaries

To have precise statements on game models we begin with basic model theoretic preliminaries. Agent models and morphisms on agent models had been treated in the first author's publications since the past decade. We develop that area further to reach specific game models. We apply basic model \theory techniques to check upward compatibility on game tree goal satisfaction. The modeling techniques are basic game tree node stratification. Since VMK has a "categorical acclimatization" we can further explore game model embedding techniques with morphisms that the first author has developed on categorical models during the past decade and applied to agent computing (Nourani1996, 2005).

Let us start from certain model-theoretic premises with propositions known form basic model theory. A **structure** consists of a set along with a collection of finitary functions and relations, which are defined on it. Universal algebra studies structures that generalize the algebraic structures such as groups, rings, fields, vector spaces and lattices. Model theory has a different scope than universal algebra, encompassing more arbitrary theories, including foundational structures such as models of set theory. From the model-theoretic point of view, structures are the objects used to define the semantics. of first-order logic. A **structure** can be defined as a triple consisting of a **domain** *A,* a signature Σ, and an **interpretation function** *I* that indicates how the signature is to be interpreted on the domain. To indicate that a structure has a particular signature σ one can refer to it as a σ-structure. The domain of a structure is an arbitrary set (non-empty); it is also called the **underlying set** of the structure, its **carrier** (especially in universal algebra), or its **universe** (especially in model theory). Sometimes the notation or is used for the domain of , but often no notational distinction is made between a structure and its domain, i.e. same symbol can refer to both. The signature of a structure consists of a set of **function symbols** and **relation symbols** along with a function that ascribes to each symbol *s* a natural number, called the **arity** of *s*.

For example, let *G* be a graph consisting of two vertices connected by an edge, and let *H* be the graph consisting of the same vertices but no edges. *H* is a subgraph of *G,* but not an induced substructure. Given two structures and of the same signature Σ`, a **Σ-homomorphism** is a map that can preserve the functions and relations. A Σ -homomorphism h is called a Σ-embedding if it is one-to-one and for every *n-ary* relation symbol *R* of Σ and any elements ai...,an, the following equivalence holds:

R(a1,...,an) iff R(h(a1),...h(an)). Thus an embedding is the same thing as a strong homomorphism, which is one-to-one. A structure defined for all formulas in the language consisting of the language of A together with a constant symbol for each element of *M,* which is interpreted as that element. A structure is said to be a **model** of a theory *T* if the language of M is the same as the language of *T* and every sentence in *T* is satisfied by M. Thus, for example, a "ring" is a structure for the language of rings that satisfies each of the ring axioms, and a model of ZFC set theory is a structure in the language of set theory that satisfies each of the ZFC axioms. Two structures *M* and *N* of the same signature σ are **elementarily equivalent** if every first-order sentence (formula without free variables) over σ is true in *M* if and only if it is true in *N,* i.e. if *M* and *N* have the same complete first-order theory. If *M* and *N* are elementarily equivalent, written $M \equiv N$.

A first-order theory is complete if and only if any two of its models are elementarily equivalent. N is an **elementary substructure** of *M* if *N* and *M* are structures of the same signature ∑ such that for all first-order ∑ -formulas φ(x1,...,xn) with free variables *x1,...,xn,* and all elements *a1,...,an* of *N,* φ(a1,...,an) holds in *N* if and only if it holds in *M: N* |= φ(a1,...,an) iff *M* |= φ(a1,...,an). Let *M* be a structure of signature R(a1,...,an), there are b1,...,bn such that R(h(a1),...,h(an)) and *N* a substructure of *M. N* is an elementary substructure of *M* if and only if for every first-order formula φ(x,y1,...,yn) over σ and all elements *b1,...,bn* from *N,* if *M* $x\varphi(x,b1,...,bn)$, then there is an element *a* in *N* such that *M* φ(a,b1,...,bn). An **elementary embedding** of a structure *N* into a structure *M* of the same signature is a map $h:N \rightarrow M$ such that for every first-order σ-formula φ(x1,...,xn) and all elements *a1,...,an* of *N, N* |= φ(a1,...,an) implies *M* |= φ(h(a1),...,h(an)). Every elementary embedding is a strong homomorphism, and its image is an elementary substructure.

Proposition 2 Let A and B be models for a language L. Then A is isomorphically embedded in B iff B can be expanded to a model of the diagram of A

B. VMK and Axiom Games

From the preceding sections we reminds ourselves that on VMK games when a game *G* is a model of *Axioms (G),* all models of *Axioms (G)* are isomorphic. *Axioms (G)* entails *only* true assertions about the game *G,* and *complete* in the sense that *Axioms (G)* entails *all* true assertions about the game *G.* Situation calculus is applied to model the structure of a multi-agent interaction, we consider agents *reasoning* about optimal actions in the multi-agent system. We show how to define the game- theoretic notion of a *strategy,* or policy, and introduce predicates describing which strategies are optimal in a given environment and which strategy combinations form Nash equilibriums.

The present paper we apply the generalized diagram encodings to situation calculus (Nourani 1994- 2010) on a

descriptive computing to VMK, presenting a computable model theoretic characterization. Descriptive definability on model diagrams were presented at (Nourani 2000) and modal diagram considerations at (Nourani 1998) on the Uppsala Logic Colloquium. To start with we are checking on VMK models where agent games can have variable function games assigned on arbitrary basis at game tree nodes. The game becomes nondeterministic in that sense but we are not assigning probability measures at all. We set-up a premise that that a player at a node takes game tree route that increases game tree degree, an ordinal, support. The VMK diagram models can carry that.

C. Compatibility Ordering and VMK Game Trees

Agent game tree generic diagram satisfiability based on the above compatibility criteria on game tree nodes implies the following.

Definition 4.6 Two generic diagrams D1 and D2 on the same logical language £ are compatible iff for ever assignment to free variables to models on £, every model M for £, M ⊨ D1 iff M ⊨ D2.

Proposition 3 A game tree node generic diagram on a player at an agent game tree node is satisfiable iff the subtree agent game trees situations have a compatible generic diagram to the node.

Proof Follows from game tress node satisfiability, definitions 4.3 4.4, and 4.6.

Proposition 4 Agent game tree generic diagrams encode compatibility on satisfying the VMK game
tree criteria for the node player.

Proof Follows from VMK definitions 4.5, 4.6 and proposition 4.3.

Proposition 5 The information partition for each player I at a node is is Ṿ D<fI > where D<fI > is the diagram definable to the player I's agent functions at that node, where Ṿ is the infinite conjunction.

Proof Since VMK has countable agent assignments at each game tree node there is a diagram 13
definable on the partition where countable conjunction can determine the next move. □

For example, viewing the diagrams on a Tableaux's (Nourani 2000), a conjunction indicates what the information partition state is for an agent.

Theorem 4 A game tree is solved at a node iff there is an elementary embedding on the generic diagrams, sequentially upward to the node, corresponding to VMK played at the node, that solves the root node.

Proof Elementary embedding has upward compatibility on tree nodes as a premise. Proposition 3.4 grants us that a VMK node check is encoded on diagram compatibility. Applying proposition 3.1 we can base diagram check the ordinal assignments in isomorphic embedding to compute a countable conjunction, hence an ordinal. On proposition 3.5 we have that compatibility is definable on a countable conjunction over the agent partition diagrams. Therefore at the root node, a game goal formula φ is solved, iff at each game tree node to the terminal node the D<fI> on the nodes, for every model Mj, and a game sequence <S>, Msj |= D<fI> iff Msj-1 |= D<fI> on the ordered sequence s ∈ <H x I>, j an ordinal < |S|.□

First author characterizes agent morphisms, in for example (Nourani 2005), to compute agent state- space model computations. So we can, for example, state the following.

Corollary 4.1 The agent game trees nodes at VMK are solved iff at the tree game sequence there are strong holomorphic game tree model diagram embeddings towards a terminal nodes where a goal diagram is satisfied. □

From proposition 3.5 we can address on a newer publications how model epistemic can be treated. The relationship to the diagram is an information set, an agent's knowledge is characterized by the diagram. What we have not stated is whether we can use modal logic or not. For starters, the diagram is no longer finite unless you make further assumptions, because you have to model what agents know and believe about each other. The above proposition allows us to treat infinite conditions. Model diagrams were treated on the first author's publications, for example, (Nourani 2000).

V. Game Tree Degrees, Model Diagrams, and Payoff's

From the VMK game definition we have the following criteria

*(a)For each player i ∈ N an *information partition* Ii defined on {s ∈ H : player(s) = i}. An element Ii of Ii is called an *information set* of Player i. We require that if s, s′ are members of the same information set Ii, then A(s) = A(s′).

* (b) For each player i ∈ N a *payoff function* $u_i : Z \rightarrow R$ that assigns a real number to each terminal history.

Recall that on condition 5 at VMK definition 3.4 the information partition for a player I is determined at each game node by the game diagram game tree degree (section 3.1) at that node that is known to that playerBased on definition 3.5 and consequences, e.g., on (a,b) above, at VMK definition 3.4 the information partition for a player I is determined at each game node by the game diagram game tree degree (section 4.1) at the node that is known to that player. Applying proposition 4 and 5, with considerations that generic model diagrams with VMK game function situation, definition 4.5, we have a second order lift on payoff functions. That is the measures on the nodes are game trees determined based on how the agent model sequent functions are satisfying the model diagrams at game tree node on theorem 4. The measures are not lost, whatever the measures are, but only embedded on the game tree lifting preorders. The measures can be created on random sequences for all we know, for example.

A. Topological Spaces and Games

Let A, B be two topological spaces. A mapping f : A → B is *continuous* if for every open set Y ⊆ B, its preimage $f^{-1}(Y)$ is an open subset of A. Thus to define the set of continuous payoff functions, we need to introduce a system of open sets on Z, the set of infinite histories, and R, the set of real

numbers.
We shall employ the standard topology on R, denoted by R. The *Baire topology* is a standard topology for a space that consists of infinite sequences, such as the set of infinite game histories. It is important in analysis, game theory. Let A be a set of actions. Let [s] = {h: s ⊂ h} be the set of infinite action sequences that extend s. The basic open sets are the sets of the form [s] for each finite sequence (situation) s. An open set is a union of basic open sets, including again the empty set ∅. We denote the resulting topological space by B(A).
For each rational q and natural number n > 0, define an open interval O(q,n) centered at q by O(q, n) = {x: |x − q| < 1/n}. Let Eu(q, n) = u−1(O(q, n)). Since u is continuous, each set Eu(q, n) is an open set in the Baire space and hence a union of situations. So the function u induces a characteristic relation *intervalu(s,* q, n) that holds just in case [s] ⊆ Eu(q, n). In other words, *intervalu(s,* q, n) holds iff for all histories h extending s the utility u(h) is within distance 1/n of the rational q. In the example with the discontinuous payoff function above, *intervalu(s,1,n)* does *not* hold for any situation s for n > 1.

From Schulte et.al,2005 on a given any situation s at which the disaster has not yet occurred, there is a history h ⊃ s with the disaster occurring, such that u(h) = 0 < |1 − 1/n|. Intuitively, with a continuous payoff function u an initial situation s determines the value u(h) for any history h extending s up to a certain "confidence interval" that bounds the possible values of histories extending s. As we move further into the history h, with longer initial segments s of h, the "confidence intervals" associated with s become centered around the value u(h).

B. Games and Topologies

Beginning from (Schulte- Delgrande 2004), let us examine the following: A successor operation + : *action* → *action* that takes an action **a** to the "next" action **a+.** We write **a(n)** as a shorthand for **a0**+⋯+ where the successor operation is applied n times to a distinguished constant **a0** (thus **a(0)** = **a0).** The constants **a(n)** may serve as names for actions. As in the case of finitely many actions, introduce unique names axioms of the form **a(i)** ≠ **a(j)** where i≠ j. The following induction axiom is examined.

Axiom: ∀P.[P(a0) ∧ ∀a.P(a) → **P(a+)]** → ∀a.P(a).

Lemma 1 (OSD). Let M = ⟨actions, +, ⌈⌉⟩ be a model of axiom and the unique names axioms. Then for every a ∈ *actions,* there is one and only one constant **a(n)** such that ⌈a(n)⌉ = a.

Consider ⌈⌉ that is a 1–1 and total assignment of actions in the game form A(F) to action constants.

Theorem (OSD) Let A be a set of actions, and let u:B(A)→R be a continuous function. Let *payoff* be the axiom ∀h, n.∃s h. interval u(s, u(h), n). Then

1. u satisfies *payoff,* and 2. if u′ : B(A) → R satisfies *payoff,* then u′ = u.

From section 3.4 given two structures and of the same signature Σ, a Σ-homomorphism is a map that can preserve the functions and relations. A Σ -homomorphism h is called a Σ-embedding if it is one-to-one and for every *n-ary* relation symbol *R* of Σ and any elements ai,…,an, the following equivalence holds: R(a1,...,an) iff R(h(a1),...h(an)). Thus an embedding is the same thing as a strong homomorphism which is one-to-one.

For example, applying corollary 3.1 to the above we have

Theorem 5 Let A be a set of actions on M, and let u:B(A)→R be a continuous function. Then u satisfies a *payoff* axiom iff u can be extended to a strong holomorphic embedding on the M model diagram towards a terminal node satisfying a goal on M.

Applying the VMK theorem 4 and noting that every elementary embedding is a strong homomorphism we carry on to the following.

Theorem 6. Let A be a set of actions on M, and let u:B(A)→R be a continuous function. The u satisfies a *payoff* axiom at M iff there is an elementary embedding on the generic diagrams for M, sequentially upward to the node, corresponding to VMK played at the node that solves the root node. state is for an agent. □

Considering epistemic accessibility based on the correspondence between possible worlds and situations encoded on generic diagrams (Nourani 1994). We can state the following proposition.

Proposition 6 Let A be a set of actions on M, and let u:B(A)→R be a continuous function. The u satisfies a *payoff* axiom iff there is a generic diagram on M that is epistemically accessible by A on M that satisfies the payoff axiom at least on one game tree node.

C. Completeness on Game Logic and Nash Criteria

We can state preliminary theorems on VMK agent games. Basic agent logic soundness and completeness areas were examined by the first author on (Nourani 2002). Making preliminary assumptions on VMK game situations let us examine soundness and completeness questions. Let us consider stratification as the process whereby generic diagrams are characterized with recursive computations on agent functions on game trees, for example compare to a notion on (Geneserth 2010). We can further consider backward chaining on game trees based on standard AI and (Schulte 2003) as the premises.

Theorem 7 There is a sound and complete agent logic on VMK game situations provided

ISBN: 978-1-61804-233-0

(i) VMK agent language is a countable fragment.

(ii) The agent information partitions are definable on a countable generic diagram,

(iii) Game tree node ordinal is definable with a countable conjunction on the generic diagrams.

Proof outline: Follows from theorem 4, the propositions, and by considering carrying soundness on forward model diagram stratifications and the embedding propositions above. Completeness is ascertained on upward model diagram compatibility, and compactness. To have a feel for that let us consider stratification as the process whereby generic diagrams are characterized with recursive computations on agent functions on game trees, for example compare to a notion on (Genesereth 2010). We can further consider backward chaining on games tree based on standard AI and (Schulte 2003).

From the above we can state theorems on Nash criteria based on the generic diagram characterizations.

D. Descriptions, Games, and Tractability

From the second author's briefs we might ponder how we have computability on game descriptions based on model diagrams on game node trees, since having a sound and complete logic can ascertain when and how the criteria are satisfied. Given that every action in M has a unique action constant naming it, it follows from the axioms that every situation in M has a unique situation constant naming it (as before, we may write *nameM(s)),* which implies that there is a 1–1 onto mapping between the situations in M and A<ω. The criteria accomplished are important on model characterizations. Looking back to theorem 2 and the correspondence of generic model diagrams encoding for possible worlds (Nourani 1994,1998, 2009), we can contemplate the following.

Proposition 7 A<ω has a unique encoding on generic model diagrams where game model compatibility can be characterized and the computability questions addressed on game models.

From the first author's Descriptive computing publications we have the following tractability consequences on game tree model diagram computations. We define descriptive computation to be computing with G- diagrams for the model and techniques for defining models with G-diagrams from the syntax of a logical language. G-diagrams are diagrams definable with a known function set. Thus the computing model is definable by generic model-diagrams with a function set. The analogous terminology in set theory refers to sets or topological structure definable in a simple manner. Thus by descriptive computation we can address artificial intelligence planning and theorem proving, for example.

The latter computational issues are pursued by the first author in (Nourani 1984,1992,2010). The logical representation for reaching goals in general might be infinitary only. The first author in Nourani (1994a,b, 1996) shows that the artificial intelligence problem is to acquire a decidable descriptive computation for the problem domain. The infinite tree premise is a condition in (Schulte-Delgrande 2004) on continuous payoff charachterizations as well. Hence logic on infinitary games were applied in our publications. The first author (Nourani 1996) proves two specific theorems for descriptive computing on diagrams. A compatibility theorem applies descriptive computing to characterize situation compatibility. Further, a computational epistemic reducibility theorem is proved by the descriptive computing techniques on infinitary languages by the first author. A deterministic epistemic is defined and it is proved not reducible to known epistemics. We further define apply infinitary logic and cardinality with transitive closure properties on sets and languages to define descriptive computable and admissible sets. We have defined a set to be Descriptive Computable iff it is definable by a G-diagram with computable functions and proved. What is accomplished above is that (A) For descriptive computable sets the interesting transitive closures on sentences are is definable from a generic-diagram by recursion. (B) For A an admissible computable set, A is descriptive computable. Cardinality restrictions on concepts are important areas explored by AI. The concept description logics systems allow users to express local cardinality on particular role filers. Global restrictions on the instances of a concept are difficult and not possible. Cardinality restrictions on concepts can be applied as an application domain description logic (Baader, et.al. 2003). The concept definitions with generic diagrams for localized KR and its relations to descriptive computable sets can be applied to concept cardinality restriction. By applying localized functions to define generic diagrams models for languages can be generated with cardinality restrictions.

VI. Proof Trees on Games

A.Game Tree Computational Heuristics and Stratification

Let us now view the deductive methods, for example the proof-theoretic example :SLDNF-resolution, a well-known deductive heuristic. A SLDNF-proof can be considered as the unfolding of an AND/OR-tree, which is rooted in the formula to be proven, whose branches are determined by formulas of the theory, and whose leaves are determined by atomic formulas which are true in a world. Partial deduction from our view point (Nourani-Hoppe 1995) usually computes from a formula and a theory an existential quantified diagram. In these papers and (Nourani 1995,2005) we also instantiate proof tree leaves with free Skolemized trees, where free trees are substituted for the leaves. By a free Skolemized tree we intend a term built with constant symbols and Skolem functions terms. Dropping the assumption that proof-tree leaves get instantiated with atomic formulas only yields an abstract and general notion of proof trees. The mathematical formalization that allows us to apply the method of free proof

trees is based on the first author's 1995-2005. In the present approach, as we shall further define, leaves could be free Skolemized trees. By a free Skolemized tree we intend a term made of constant symbols and Skolem function terms. Like models and diagrams, which where generalized above in different ways, we can generalize the notion of a proof. First author had developed free proof tree techniques since projects at TU Berlin, 1994. Free proof trees allow us to carry on Skolemized tress on game tree computing models, for example, that can have unassigned variables. The techniques allow us to carry on predictive model diagrams. Reverse Skolemization (Nourani 1986) that can be carried on with generic model diagrams corresponding to what since Genesereth (2011) is applying on game tree "stratified" recursion to check game tree computations. The free trees defined by the notion of provability implied by the definition, could consist of some extra Skolem functions {g1,...,gm},that appear at free trees. The f terms and g terms, tree congruences, and predictive diagrams then characterize partial deduction with free trees. To compare recursive stratification on game trees on what Geneserth calls recursive stratification we carry on models that are recursive on generic diagram functions where goal satisfaction is realized on plans with free proof trees (Nourani 1994-2007). Thus essentially the basic heurstics here is satisfying nodes on agent AND./OR game trees. The general heuristics to accomplish that are a game tree deductive technique based on computing game tree unfoldings projected onto predictive model diagrams. The soundness and completeness of these techniques, e,g, heuristics as a computing logic is published since 1994 at several events (Berlin logic colloquium, Potsdam Universitat Mathematik 1994, AISB 1995, and Systems and Cybernectics 2005). The heuristic nomenclature indicates that a heuristic function is called an *admissible- heuristic* if it never overestimates the cost of reaching the goal, i.e. the cost it estimates to reach the goal is not higher than the lowest possible cost from the current point in the path. An admissible heuristic is also known as an *optimistic heuristic* (Russell and Norvig 2002). We shall use the term optimistic heuristic from now on to save ambiguity with admissible sets from mathematics (Shoenfield 1967, or ,first author's publications on descriptive computing, for the time being. What is the cost estimate on satisfying a goal on an unfolding projection to model diagrams, for example with SLNDF, to satisfy a goal? Our heuristics are based on satisfying nondeterministic Skolemized trees. The heuristics aims to decrease the unknown assignments on the trees. Since at least one path on the tree must have all assignments defined to T, or F, and at most one such assignment closes the search, the "cost estimate," is no more than the lowest. Let us call that nomenclature as one compound word: "admissible heuristics," not to confuse that with the notion of admissible sets at mathematical logic in the first author's publications. We do not know if we can have a relationship to that mathematical area yet. To become more specific how game tree node degrees can be ranked, we state one example linear measure proposition.

Proposition 8 Supposing the degree rank is defined based on a linear counting measure-the intersection at a goal state, where there are true literals, counting the goal literals that become true, assuming that each action can make at most one goal literal true at a time, then the ranking is admissible heuristic ranking function.

Not being restricted to the above proposition on specific ranks, only that there is viable node degree rank, we have more general areas to address as follows, independent of the specific ranking notions.

Lemma 2 And/Or agent game tree satisfaction with unfolding projections onto predictive model diagrams is an optimistic heuristic.

We can prove that direct deploying more on models, or carry on with two corollaries. The process to reach an admissible heuristic based on SLDNF is done with defining a measure on what literals are selected first on assignments. The process entails that the minimal route to resolving literals are chosen so as to accomplish an admissible heuristic with SLDNF. The measures are defined in a precise mathematical manner (Ntienjem 1997).

Corollary 6.1 SLNDF deductive heuristics based on admissible selection functions is an optimistic heuristic.

Proof Follows from a report on SLNDF computing logic at the Universitat Augsburg (Ntienjem, 1997).

Corollary 6.2 Predictive diagram unfolding encodes SLDNF techniques on partial deductions.

Proof Nourani-Hoppe 1995, TU Berlin, Potsdam Universitat Logic colloquium, AISB 1995, and Nourani 2005 Systems and Cybernetics..

B. Game Node Degree Heuristics Example

A Nash Equilibrium (NE) is a joint strategy such that no agent may unilaterally change its strategy without lowering its expected payoff in the one shot play of the game. Nash (1951) showed that every n-player matrix game has at least one. Each node has a heuristic value which says how close a given state is to the goal state. With these values you can determine which node is the best to move to next. In a turn-based games like chess you can always calculate the exact number of possible states. The current state will not change until you do something. In a game that is not turn-based your opponent can change the current state at any given time. Since the moment is random the number of possible states can be considered infinite. In a non-turn based game tree you also can't have heuristic values based on closeness to goal. In a non-turn based game tree there is no point in looking far ahead because the current state can change radically at any moment. The heuristic algorithm must not only incorporate the closeness to goal but also estimate how the situation could change due to actions of the opponent. Let us consider examples towards specific game tree computations. Suppose a goal is described by a DNF formula G1 OR G2 OR … OR Gn.

An example linear degree assignment is that the node degree is the max over the disjuncts of the intersections. E.g. if goal formula Gi shares k literals with the literals that are known to be true at node u, then the degree of u is at least k. Let us now consider an example that was run at (Schulte-Delgrande 2005). Consider servers on the internet. Each server is connected to several sources of information and several users, as well as other servers. There is a cost to receiving and transmitting messages for the servers, which they recover by charging their users. We have two servers Srv1 and Srv2, and two users – journalists – U1 and U2. Both servers are connected to each other and to user U1; server Srv2 also serves user U2. There are two types of news items that interest the users: politics and showbiz. The various costs and charges for transmissions add up to payoffs for the servers, depending on what message gets sent where. For example, it costs Srv1 4 cents to send a message to Srv2, and it costs Srv2 2 cents to send a message to U2. If U2 receives a showbiz message from Srv2 via Srv1, he pays Srv1 and Srv2 each 6 cents. So in that case the overall payoff to Srv1 is −4 + 6 = 2, and the overall payoff to Srv2 is −2 + 6 = 4. (Schulte-Delgrande 2005) describe the specifics in detail; a summary is figure 1.

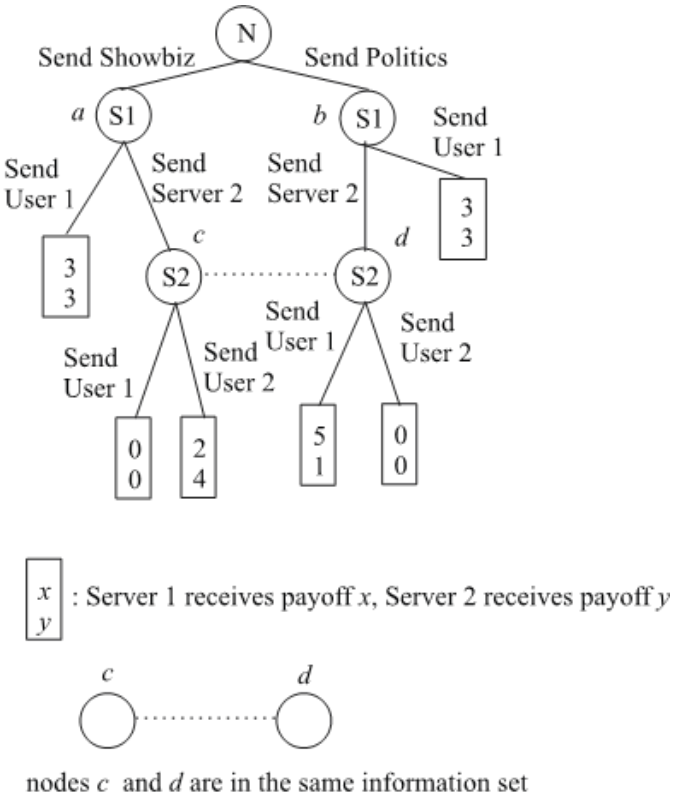


Figure 4. A game-theoretic model of the Node Diagram

a Action(N) = Showbiz
b Action(N) = Politics
c Action(S1) = Send Server 2,
Action(N) = Showbiz OR Action(N) = Send Politics
d Action(S1) = Send Server 2,
Action(N) = Showbiz OR Action(N) = Send Politics

Table 1. The model diagrams for each choice node in the game tree.

C. Matrix Agent Games and Nash Equilibriums

Examining Multiagent Reinforcement Learning for iterated play of games, more specifically matrix games, let us introduce some well-known concepts from game theory. In general, let S denote the set of states in the game and let Ai denote the set of actions that agent/player i may select in each state s ϵ S. Let a = (a1, a2,..., an), where ai ϵ Ai be a joint action for n agents, and let A = A1 x...x An be the set of possible joint actions. Zero-sum games are games where the rewards of the agents for each joint action sum to zero.

General sum games allow for any sum of values for the reward of a joint action. A strategy for agent i is a probability distribution π (.) over its actions set Ai. Let π (S) denote a strategy over all states s ϵ S and let π(s) (or πi) denote a strategy in a single state s.One way to look at equilibrium on diagrams is that we consider.

An n-player game on a matrix, where each row corresponds to a player diagram.

We focus on the more restricted matrix game, denoted by a set of matrices

R = R1,...Rn. Let R(π) = (R1(π); : : : ; Rn(π)) be a vector of expected

payoffs when the joint strategy π is played. Also, let Ri (π i, π-i) be the expected

payoff to agent i when it plays strategy πi and the other agents play π-i.

A stage game is a single iteration of a matrix game, and a repeated game is the indefinite repetition of the stage game between the same agents. While matrix games do not have state, agents can encode the previous w joint actions taken by the agents as state information. Each individual matrix game has certain classic game theoretic values. The minimax value for player i is the least reward that can be achieved if the game is known and the game is only played once. There is a natural strategy based on a single player's view to n-player game that does not have to have a paring characterization. Nash equilibrium naturally occurs based on how well each player heuristic on SLNDF for example unfolds. Based on the above we can have a proposition.

Proposition 9 A matrix game reaches a Nash equilibrium when all the player corresponding model diagram rows are defined and at least one player accomplishes the goal set(s).

D. More on Interactive Decisions

Games theory is theory of rational behavior for interactive decision problems. In a game, several agents strive to maximize their (expected) utility index by choosing particular courses of action, and each agent's final utility payoffs depend on the profile of courses of action chosen by all agents. The interactive situation, specified by the set of participants, the possible courses of action of each agent, and the set of all possible utility payoffs, is called a game. The agents 'playing' a game are called the players. In degenerate games, the

players' payoffs only depend on their own actions. For example, in competitive markets, it is enough that each player optimizes regardless of the behavior of other traders. As soon as a small number of agents are involved in an economic transaction, however, the payoffs to each of them depend on the other agents' actions. From Schulte-Delgrande 2005 notation, for example. We can cast the interactive Nash based on the above sections. Let us define the payoff to each player that results when two strategies are paired in situation s, which is just their utility from the resulting action sequence:
Ui(π1, π2, s) = ui
Play (π1, π2, s)
For the notion of Nash equilibrium we need the concept of a best reply.
bestReply 1 (π1, π2, s)≡$\forall\pi$1.valid(π1)) → U1(π1, π2, s) U1(π1, π2, s),
bestReply 2 (π1, π2, s)≡$\forall\pi$2 .valid(π2) → U2(π1, π2, s) U2(π1, π2, s).
A pair of strategies (π1, π2) forms a Nash equilibrium in a game overall just in
case each strategy is a best response to the other. The payoff from two strategies in the
overall game is just the result of starting in the initial situation S0. Thus we have nash
(π1, π2)≡valid(π1) ∧ valid(π2) ∧ bestReply1(π1, π2, S0) ∧ bestReply2(π1, π2, S0).

F. Predictive Plan Proof Trees

The paper accomplishes modeling a multiplayer cooperative game to reach a goal or modeling group action game where group A is playing group B where B's moves are not apriori known. Neither is both sides information set. Hence the information set that A group player carries on to base moves on, is changed on a diagram. The models and encoding apply to zero-sum and/or arbitrary games. We have viewed the game as zero- sum on a game from a group action stage that plays to reach a goal where certain propositions are true. New VMK encodings allow us to reach new mathematical goals on game model computability towards specific Nash equilibrium models. A novel basis to decision-theoretic planning is presented classical and non-classical planning techniques, see for example, (Hedeler et.al. 1990, Wilkins 1984) from artificial intelligence with games and decision trees providing a agent expressive planning model following the first authors competitive game model planning publication can be developed based on the new techniques here. We use a broad definition of decision- theoretic planning that includes planning techniques that deal with all types of uncertainty and plan evaluation. Planning with predictive model diagrams represented with keyed KR to knowledge bases is presented. Techniques for representing uncertainty, plan generation, plan evaluation, plan improvement, and are accommodate with agents, predictive diagrams, and competitive model learning. Modeling with effector and sensor uncertainty, incomplete knowledge of the current state, and how the world operates can be treated with agents and competitive models. Bounds on game trees can be stated based on the first author's preceding publications agent games and the second author's generalizations on VMK to agents. The heuristics area can be further developed based on a comprehensive treatment on game unfolding and computability criteria on model diagrams. Computability and reachability on Nash models are further areas to explore.

E. Proof Trees on Games

First author had developed free proof tree techniques since projects at TU Berlin , 1994. Free proof trees allow us to carry on Skolemized tress on game tree computing models, for example, that can have unassigned variables. The techniques allow us to carry on predictive model diagrams. Reverse Skolemization (Nourani 1986) that can be carried on with generic model diagrams might correspond to what since Generserth (2010?) is applying on game tree "stratified" recursion to check game tree computations. Existentially quantified goals on diagrams carry a main deficit. The Skolemized formulas are not characterized. A **predictive diagram** for a theory T is a diagram D[M], where M is a model for T, and for any formula q in M, either the function f: q → {0,1} is defined, or there exists a formula p in D[M], such that T U {p} proves q; or that T proves q by minimal prediction. A **generalized predictive diagram,** is an predictive diagram with D[M] defined from a minimal set of functions. The predictive diagram could be minimally represented by a set of functions {f1,...,fn} that inductively define the model. The free trees [11] defined by the notion of provability implied by the definition, could consist of some extra Skolem functions {g1,...,gl}that appear at free trees. The f terms and g terms, tree congruences, and predictive diagrams then characterize partial deduction with free trees. To compare recursive stratification on game trees on what Geneserth calls recursive stratification we carry on models that are recursive on generic diagram functions where goal satisfaction is realized on plans with free proof trees (Nourani 1994-2007).

Theorem 8 For the free proof trees defined for a goal formula from a generic diagram there is a canonical model satisfying the goal formulas. It is the canonical initial model created with the generic diagram.
Proof In planning with generic diagrams plan trees involving free Skolemized trees is carried along with the proof tree for a plan goal. The idea is that if the free proof tree is constructed then the plan has a model in which the goals are satisfied. There is analogy to SLD proofs. We can view on the one hand, SLD resolution type proofs on ground terms, where we go from p(0) to p(f(c)); or form p(f(c)) to p(f(g(c)). Whereas, while doing proofs with free Skolemized trees we are facing proofs of the form p(g(....)) proves p(f(g(....)) and generalizations to p(f(x)) proves For all x , p(f(x)). Since the proof trees are either proving plan goals for formulas defined on the G-diagram, or are computing with Skolem functions defining the G-diagram, by generic diagram definition, and theorems on (Nourani 1994) the model defined by the generic diagram applies and it is canonical,e.g. Initial, for the proofs.

VII. Conclusion and Areas to Explore

Modeling a multiplayer cooperative game to reach a goal or modeling group action game where group A is playing group B where B's moves are not apriori known. Neither is both

sides information set. Hence the information set is changed on a diagram, that A group player carries on to base moves on. The models and encoding apply to zero-sum and/or arbitrary games. We have viewed the game as zero- sum on a game from a group action stage that plays to reach a goal where certain propositions are true. A novel basis to decision-theoretic planning is presented classical and non-classical planning techniques, see for example, (Hedeler et.al. 1990, Wilkins 1984) from artificial intelligence with games and decision trees providing a agent expressive planning model. We use a broad definition of decision- theoretic planning that includes planning techniques that deal with all types of uncertainty and plan evaluation. Planning with predictive model diagrams represented with keyed KR to knowledge bases is presented. Techniques for representing uncertainty, plan generation, plan evaluation, plan improvement, and are accommodate with agents, predictive diagrams, and competitive model learning. Modeling with effector and sensor uncertainty, incomplete knowledge of the current state, and how the world operates is treated with agents and competitive models. Bounds on game trees can be stated based on the first author's preceding publications agent games and the second author's generalizations on VMK to agents.

REFERENCES

[1] Brazier, F.M.T. Dunin-Keplicz, B., Jennings, N.R. and Treur, J. (1997) DESIRE: modeling mulch- agent systems in a compositional formal framework, International Journal of Cooperative Information Systems, M. Huhns, M. Singh, (Eds.), special issue on Formal Methods in Cooperative Information Systems, vol. 1.152, Electronic Edition.

[2] Nourani,C.F.1996a, Slalom tree computing – a tree computing theory for artificial intelligence, AI Communications, European AI Journal, IOS Press.
Volume 9, Number 4/ 1996. 10.3233/AIC-1996-9402

[3] ADJ - Goguen, J.A., J.W. Thatcher, E.G. Wagner and J.B. Wright,"A Junction Between Computer Science and Category Theory," (parts I and II),IBM T.J. Watson Research Center, Yorktown Heights, N.Y. Research Report, RC4526,1973.

[4] Nilsson, N.J.1971, Problem Solving Methods in Artificial Intelligence," NewYork,McGraw-Hill,1971.

[5] Nourani,C.F. 2000, Descriptive Definability: A Model Computing Perspective on the *Tableaux*. CSIT 2000 Ufa, Russia. *de/~cp/p/zombie/*

[6] Chung, C.C. and Kiesler, H. Jerome 1973, *Model Theory,* Elsevier, ISBN 978-0-7204-0692-4.

[7] Nourani, C.F. 1998, "Intelligent Languages, A preliminary Syntactic Theory,"In: Proceedings of the Mathematical Foundations CS'98 Satellite workshop on Grammar systems. Ed. A.Kelemenová, Silesian University, Faculty of Philosophy and Sciences, Institute of Computer Science, Opava, 1998 pp. 281- 287.

[8] Kleene, S.A. "Introduction to metamathematics", North-Holland(1951).

[9] Nourani, C.F. 1995, "Free Proof Trees and Model-theoretic Planning," February 23, 1995, Automated Reasoning AISB, England, April 1995.

[10] Schulte,O (2003). Iterated Backward Inference: An Algorithm for Proper Rationalizability. Proceedings of TARK IX (Theoretical Aspects of Reasoning About Knowledge), Bloomington, Indiana, pp. 15--28. ACM, New York. Expanded version with full proofs.

[10] Nourani, C.F.1999a, "Agent Computing, KB For Intelligent Forecasting, and Model Discovery for Knowledge Management," June 1998. *AAAI-Workshop on Agent Based Systems in the Business Context,* Orlando, Florida, July 18-July 22,1999.

[11] Nourani, C.F. 1991, "Planning and Plausible Reasoning in AI," *Proceedings Scandinavian Conference in AI, May,* Denmark, 150-157, IOS Press

[12] Nourani, C.F. 1997 "Multiagent Chess Games," AAAI Chess Track, Providence, RI, July 1997.

[13] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Chapman & Hall (Wadsworth, Inc.): New York, 1984.

[14] Nourani, C.F.,"Equational Intensity, Initial Models, and AI Reasoning," Technical Report, 1983, : A: Conceptual Overview, in Proceedings Sixth European Conference in Artificial Intelligence, Pisa, Italy, September 1984, North-Holland.

[15] Nilsson, N.J. 1969,"Searching, problem solving, and game-playing trees for minimal cost solutions." In A.J. Morell (Ed.) IFIP 1968 Vol.2, Amsterdam, North-Holland, 1556-1562, 1969.

[16] Nourani, C.F. 1999a "Infinitary Multiplayer Games," Utrecht, August 1999.
The Bulletin of Symbolic Logic, Volume 6, Number 1, March 2000. EUROPEAN SUMMER MEETING OF THE ASSOCIATION FOR SYMBOLIC LOGIC.

[17] Genesereth, M.R. and N.J. Nilsson 1987, Logical Foundations of Artificial Intelligence,Morgan Kaufmann, 1987.

[18] Vulkan. N. 2012 ,Strategic Design of Mobile Agents, AI Magazine, Vol 23, No 3.AAAI.

[19] Nourani,C.F. 1994, "A Theory For Programming With Intelligent Syntax Trees And Intelligent Decisive Agents-Preliminary Report" 11th European Conference A.I., ECAI Workshop on DAI Applications to Decision Theory, Amsterdam, August 1994.

[20] Breiman, L. (1996). Some properties of splitting criteria, Machine Learning 24: 41–47.

[21] Gale, D. and F.M. Stewart,1953 "Infinite Games with Perfect Information," in Contributions to the Theory of Games, Annals of Mathematical Studies, vol. 28, Princeton.

[22] Wilkins, D., "Domain Independent Planning: Representation and Plan Generation," AI, 22(3): 269- 301, 1984.

[23] Nourani, C.F. and R.M Moudi.,Open Loop Control and Business Planning- A Preliminary Brief,International Conference Autonomous Systems, Tahiti, French Polynesia. October 23-28, 2005. www.iaria.org/conferences/ICAS/ICAS2005/ General Information/GeneralInformation.html, page 8-ISBN:0-7695-2450-8.

[24] Parikh,R. 1983, Propositional game logic, in: IEEE Symposium on Foundations of Computer Science (1983) pp. 195–200. D. Poole, Decision theory, the situation calculus, and conditional plans, Linköping Electronic Articles in Computer and Information Science 3(8) (1998).

[25] Koller and A. Pfeffer, "Representations and solutions for game-theoretic problems," Artificial Intelligence 94(1) (1997) 167–215.
[26] Nash, J.F. 1951, Non-cooperative games, Annals of Mathematics 54 (1951) 286–295.
[27] Nourani, C.F, 1997, "Descriptive Computing-The Preliminary Definition," Summer Logic Colloquium, July 1996, San Sebstian Spain. See AMS April 1997, Memphis.
[28] Schulte,O and J. Delgrande (2004), Representing von Neumann-Morgenstern Games in the Situation Calculus. Annals of Mathematics and Artificial Intelligence 42 (1-3): 73-101. (Special Issue on Multiagent Systems and Computational Logic). A shorter version of this paper appeared in the 2002 AAAI Workshop on Decision and Game Theory
[29] Nourani ,CF 1997 "Syntax Trees, Intensional Models, and Modal Diagrams, For Natural Language Models, Revised July 1997. Uppsala Logic Colloquium, August1998, Uppsala University,Sweden.
[30] Nourani,C.F. "Higher Stratified Consistency and Completeness Proofs", Summer Logic Colloquium, Helsinki, August 14-20 http://www.logic.univie.ac.at/cgi-bin/abstract/show.pl?new=e049a2efe0c1a4b7a3ddaa11a75d8 152, *Mathematicians,* Vol. II (Cambridge, 1913).
[31] Peter F. Patel-Schneider, A Decidable First-Order Logic for Knowledge Representation. Journal of Automated Reasoning 6, 1990, pages 361–388. A preliminary version published as AI Technical Report Number 45, Schlumberger Palo Alto Research, September, 1985.
[32] The Description Logic Handbook 2003, Theory, Implementation and Applications. Edited by: Franz Baader, et.al. Aachen University of Technology, ISBN:9780521781763 adder et.al. 96,
[33] Badder,F, M. Buchheit, B. Hollunder, 1996, Cardinality Restrictions on Concepts, AI, 1996.
[34] The Knowledge Engineering Review, 2012, - Special Issue 02 (Agent-Based Computational Economics) Volume 27 / Special Issue 02 / June 2012 . Cambridge Online Journals.
[35] Nourani,C.F. 2009, A Descriptive Computing, Information Forum, Liepzig, Germany, March 2009. SIWN2009 Program, 2009. The Foresight Academy of Technology Press International Transactions on Systems Science and Applications, Vol. 5, No. 1, June 2009, pp. 60-69.
[36]D.G. Kendall (1974) Foundations of a theory of random sets. In Stochastic Geometry (E.F. Harding and D.G. Kendall, editors) : 322-376. J.Wiley, New York.
[37] Nourani, C.F. 2012,. Competitive Models, Game Tree Degrees, and Projective Geometry on Random Sets - A Preliminary, ASL Annual Meeting, Wisconsin, March 2012.
[38] Genesereth, M, 2010 Stanford University Computer Science Lecture Notes on Games.
[39] LOFT 2008, Logic and the Foundations of Game and Decision Theory - 8th International Conference, Amsterdam, The Netherlands, July 3-5, 2008, Series: Lecture Notes in Computer Science, Vol. 6006 Subseries: Lecture Notes in Artificial Intelligence Bonanno, Giacomo; Löwe, Benedikt; van der Hoek, Wiebe (Eds.)1st Edition., 2010, XI, 207 p ISBN: 978-3-642-15163-7.
[40] von Neumann and O.Morgenstern, 1994, *Theory of Games and Economic Behavior* (Princeton UniversityPress, Princeton, NJ, 1944).
[41] Holland, J.H. and Miller, J.H., Artificial Adaptive Agents in Economic Theory, American Economic Association,Journal American Economic Review.Volume (Year): 81 (1991) Issue.
[42] Nash. J Equilibrium points in n-person games, Proc. National Academy of Sciences of the USA, 36:48,49,1950.
[43] Nourani, C.F. and Grace S.L. Loo ,2000, KR and Model discovery from active DB with Predictive logic, Data Mining 2000 Applications to Business and Finance Cambridge, UK, August 2000.
[44] Nourani,C.F. & R.M., Management Process Models and Game Trees Applications to Economic games, September 1999.Invited Paper SSGRR, L'Auquila, Rome, Italy, September 2001.
[45] Nourani, C.F. 2002, Game Trees, Competitive Models, and ERP, New Business Models and Enabling Technologies, Management School, St Petersburg, Russia, Keynote Address, June 2002. Fraunhofer Institute for Open Communication Systems, Germany 11:00-12:00. www.math.spbu.ru/user/krivulin/Work2002/Workshop.htm. Proceedings Editor Nikolai Krivulin,
[46] Nourani, C.F. 1997, Intelligent Tree Computing, Decision Trees and Soft OOP Tree Computing, September 2, 1997, Frontiers in Soft Computing and Decision Systems, Papers from the 1997 Fall Symposium, Boston, Technical Report FS-97-04 AAAI, ISBN: 1-57735-079-0, www.aaai.org/Press/Reports/Symposia/Fall/fs-97-04.html.
[47] Nourani, C.F., and Hoppe,.Th, GF-Diagrams for Models and Free Proof Trees, Technical Universitat Berlin, Informatiks, March 26, Berlin Logic Colloquium, May 1994, Universitat Potsdamm, Potsdamm, Germany. Humboldt Universtiat Mathematik sponsor.
[48] Nourani, C.F. 2003, KR, Predictive Model Discovery, and Schema Completion, th World Multiconference on Systemics,Cybernetics and Informatics (SCI 2002) Orlando, USA, July 14-18, 2003, http://www.iiis.org/sci2002.
[49] Mathematical Logic by J. R. Shoenfield (Paperback - Feb 9 2001)
ASL Publications. Hard cover 1967 Addison-Wesley Educational Publishers Inc (December 1967) ISBN-10: 0201070286 ISBN-13: 978-0201070286.
[50] Russell, S.J.; Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*. Prentice Hall. ISBN 0-13-790395-2.
[51] Ntienjem, E. 1997, "Completeness and Termination of SLNDF- Resolution and Determination of a Selection function using Mode." Universitat Ausburg, Report 1997-06

Using Simulation System AGNES for Modeling Execution of Parallel Algorithms on Supercomputers

Igor Chernykh, Boris Glinskiy, Igor Kulikov, Mikhail Marchenko, Alexey Rodionov, Dmitriy Podkorytov, and Dmitry Karavaev

***Abstract*—The aim scope of the research is a possibility of representation of parallel algorithms on different architectures of exaflops supercomputers based on simulation. AGent NEtwork Simulator (AGNES) proposed by authors for complex modeling of scalability on different possible architectures of exaflops supercomputers. Some examples of scalability (50K+ cores) modeling of algorithms are presented. Algorithms for modeling are taken from different sciences: gas dynamics, computational geophysics, chemical engineering.**

***Index Terms*—agent oriented simulation, scalable parallel algorithms, exaflops supercomputers**

I. Introduction

THE problem of studying the properties of scalability of parallel algorithms and programs when transferring their implementation from current computer to more powerful up to future exaflops class supercomputers reached the level of technological challenges and requires thorough researchs. Computational algorithms are generally more conservative compared with the development of computer technology. Evaluating future behavior of an algorithm developed by a given computational scheme can be done by simulating its execution. Such model can represent usage of thousands or millions of CPU (GPU) cores by a parallel program. By simulation it is possible to find bottlenecks in algorithms for understanding how to modify them for achieving efficient scaling onto a large number of cores. The problem of modeling scalable algorithms is not a new one. Many research groups around the world works on it starting from the begining of "parallel era" (see review [1], for example). Good example of early projects that aims in simulating execution of parallel programs in reconfigurable environment is *PARSIT* [2]. Some researches deal with simulating single algorithm or special class of algorithms [3], [4] that is, obviously, simpler than designing general tool for simulating high-performance computations. Most known project in this direction is BigSim (http://charm.cs.uiuc.edu/research/bigsim).

The BigSim project aims to creation of a simulation environment that enables the development, testing and proposing configuration of future generations supercomputers [5]. Another interesting project ParJava developed by Institute of system programming RAS in Moscow [6], [7]. They create a model of parallel program that effectively simulate its execution on non-existent computer with hypothetical architecture by using the instrumental computer. This simulation allows estimation of execution time and memory utilization. Unfortunately, this project is limited to simulation of programs developed by Java and MPI only.

In many cases we need estimating scalability rate of parallel programs only, not exact loading of computer's kernels and memory and/or computation time. For this purpose we can use simpler models than that used in BigSim and thus using simpler simulation tools and providing faster result. In this research we use simulation system of our design named AGNES (AGent NEtwork Simulator) [8], [9] that proves to be the efficient tool for comparative investigation of different variants of mapping parallel program and data onto high-performance computational system up to future exaflops supercomputers.

II. AGNES (AGent NEtwork Simulator)

Been primely designed for simulation of digital networks, AGNES proved to be suitable for simulating execution of parallel programs also [10]. JADE was chosen as a basis because it is a powerful tool for creating multi-agent systems in Java. AGNES consists of four parts: the runtime agents, the base class libraries needed for development of agent-based systems, a set of utilities for monitoring and administration of the MAS (multi-agent system) and GUI subsystem. For simulating high performance computing it is important that JADE is a FIPA-compliant distributed agent platform which can use one or more computers (nodes). Each computer (node) should use only one Java virtual machine. AGNES platform consists of a set of containers, distributed in the network. Typically each host holds one container (but if necessary there can be more than one). Agents exist inside containers.

All communications between agents occurs through the JADE application messaging, according to FIPA standard. The key property of a JADE agent is a set of its behavior. Agent's life cycle ends when the agent has no active behavior any more. AGNES uses all of these opportunities and expands MAS to system simulation. AGNES simulation environment (see Fig. 1) consists of agents which can be divided into two groups: control agents (CA), which create and support simulation environment, and functional agents (FA), which form a model that is executed inside this environment.

This work was supported in part by the RFBR grant no. 12-01-00034, 12-01-00727; 13-07-00589; grant of the President of Russia MK-4183.2013.9; and by Grants Council of the President of Russian Federation "Scientific school 5176.2010.9"

All authors are with the Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Lavrentjeva ave. 6, Novosibirsk, 630090, Russia e-mail: (see http://www.sscc.ru).

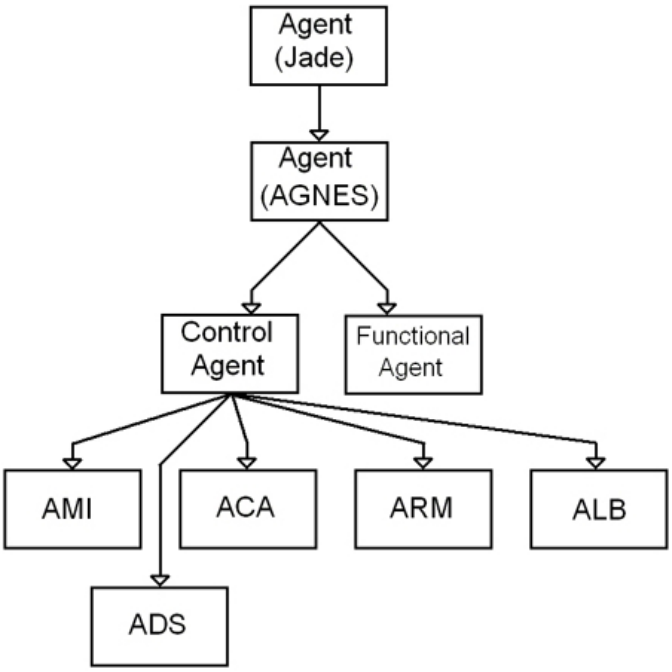


Figure 1. Architecture of AGNES. AMI is model initializer, ARM — resource manager, ALB — load balancer, ADS — data storage, ACA — main control agent (manager)

FAs designed for different models can be organized in the problem-oriented libraries.

A. Control agents of AGNES

The main tasks of CA: 1) Initialize and run the model; 2) Collecting and storing information about model's behaviour; 3) Synchronization of simulation time; 4) Load balancing between nodes involved in simulation; 5) Interaction with an user (output of reports and supporting his ability to interfere into simulation run); 6) Ensuring fault tolerance, possible recovery of a model.

When a model is executed, all FAs are divided into virtual clusters, and to each such cluster a supervising agent (SA) is assigned, which is the special kind of CA. The main functions of SA: 1) Detecting and identification of agent's failure in the simulation environment; 2) Forwarding control commands to FAs; 3) Storing information for agents' recovery; 4) Model's restoration at failures.

B. Data Collection and Storing

Messages of two types circulate inside an AGNES model: control commands and information messages. Special agents (loggers) are designed for collecting and storing information messages thus providing data collection and possible restoration of a model in case of failure. These loggers subscribe for certain kinds of messages and receive their copies. In different runs of a model different kinds of information may be of interest, so the opportunity is for filtering only such information that is needed. This opportunity is realized basing on structure of FIPA messages and comfortable mechanism of their classification.

Information about state of FAs is vital during the simulation. For obtaining this information CAs use "yellow pages" JADE's service that provides finding agents and querying and storing needed information from them.

C. Load Balancing

JADE application is an distributed one that runs on a network of processors. JADE allows changing execution environment of multi-agent system dynamically, that is attach it to some existing program or remove some working containers. For providing better performance of simulation environment, AGNES supervises these processes. When changes are detected in an executing environment, AGNES tries redistribute agents for uniform usage of available resources. This redistribution is realized as migration of agents between containers supported by JADE's means. For better performance and fault-tolerance, AGNES tries redistribute agents of all types uniformly.

D. User Interface

Besides available GUI tools of JADE environment, AGNES provides the following additional means for user interface:

- Workbench for constructing a model from library FAs and its tuning (see Fig. 2);
- Output of a model's graph whose nodes corresponds to FAs and edges to communication channels between them;
- Output of tables of logs and data collected by loggers;
- On-line tools for changing a model, adding or removing FAs;
- etc.

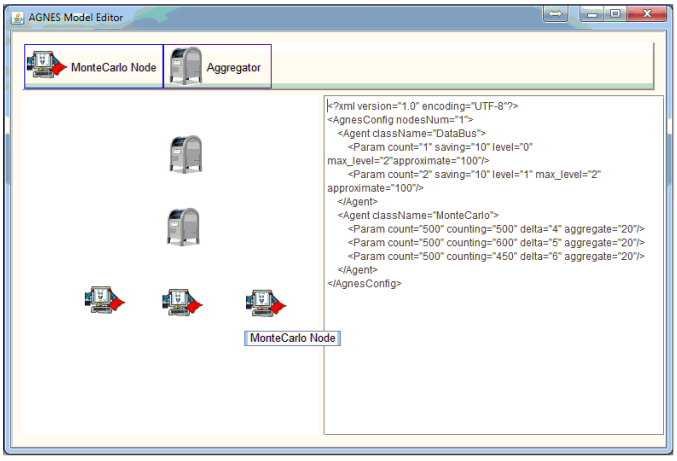


Figure 2. AGNES's workbench

E. AGNES's Agents Functions

Joint behavior of FAs simulate behavior of a simulated system. Thus AGNES most suits to simulation of systems that can be naturally decomposed onto simple components that interact with each other. System of public transportation and sensor network are good examples. FAs must be designed specially for each model or class of models. In last case special library

ISBN: 978-1-61804-233-0

of FAs can be designed. At the same time every FA must perform some actions that are common for all FAs and are needed for providing the simulation process itself (periodical backups, synchronization, routing of messages between agents etc.). Functions that are model-depended are realized by the following procedures:

- **`Sleep()`** – model suspending;
- **`Wakeup()`** – continue simulation;
- **`Start(array InitParameters)`** – model initialization and start;
- **`RestoreFrom(time Moment)`** – backtracking agent condition to special moment;
- **`CreateBackup(time Moment, agent Receiver)`** – creation of reserve copy of data needed for an agent recovery, parameter is the address of agent that is assigned for storing of this data;
- **`RestoreBackup(array BackupParameters)`** – reconstruction of agent from its backup data;
- **`GetStatus()`** – get data about the state of simulation.

The following system functions of AGNES's agent are model-independent:

- **`SaveBackup(agent Sender, time BackupMoment, array BackupData)`** – order to store backup data for "Sender" at time BackupMoment;
- **`SendBackup(agent Receiver, time BackupMoment, agent BackupAgent)`** – transferring backup data;
- **`SetTime(time ModelTime)`** – setting simulation time for the current agent;
- **`GetAgentTime()`** – returns agent's simulation time;
- **`SendLog()`** – broadcasts message with agent's current state;
- **`Ping()`** – command for checking if an agent is in working state.

The benefits of AGNES are: redundancy of simulating code, load balancing, availability of domain-specific libraries of agents, possibility of changing a model during its execution.

At present there is no definitive representation of architecture of future exaflops supercomputer. One of the possible architectures is based on the heterogeneous cluster architecture and consists from CPU and GPU units. We use this architecture for simulation modeling of parallel algorithms. Program model for simulation is a weighted graph of transitions between parallel blocks of the program. Time delays in the blocks are defined on the basis of measurements that was made in the real test runs of a programs on NKS-30T cluster based on 120 Tesla2090M GPUs and more than 1000 Intel Xenon CPU cores.

III. Simulation examples

We have used AGNES-based simulation models for evaluating scaling and/or tuning a number of very different numerical algorithms, among them:

1) Monte-Carlo method for simulation of a gas flow with chemical reactions;
2) grid methods for solving different problems of chemical kinetics, astrophysics and gas dynamics;
3) computing random graph's reliability;
4) genetic algorithm for unreliable network optimization;
5) etc.

Prior simulation of these algorithms allows choosing their parameters for improvement performance of their real execution and predicting performance for future runs on more powerful supercomputers. Let us discuss one of simulated algorithms, parallel Monte-Carlo Method, in details.

The test problem which has applications to rare gas dynamics simulation was considered. Main reason to use it is that we know the exact analytic solution to the problem and therefore can compare it with numerical results. More detailed description of the problem, algorithm and its model is presented in [10].

A simulation of one realization of the test particles ensemble consists in the following. We consider the molecules of the gas as solid spheres, their number is N_C. The initial distribution for the velocities of the molecules $f_0(V)$ is the Bobylev's distribution with parameter $\beta = 0.4$ [12]. We apply numerical scheme based on the majorant frequency principle which enables the linear dependence of the computer time on the number of the test particles. The value of the majorant M_C equals to 10. We consider the a model for chemical reaction based on the elastic collisions.

We evaluate the dependence of the normalized tenth moment of the molecule's velocity on the time variable. We evaluate the functional at time moments $T = 0, 1, \ldots, 10$.

For the simulation of statistical based algorithms PARMONC library [10] was used. For simulating large computations by the Monte-Carlo method, a computational core with an algorithm running on it was chosen as an elementary unit. That is, the same agent simulates the complete behavior of the computational core and the communication of the hardware. For this model two types of agents were designed:

- *MonteCarlo* agent involved in direct simulation of computations;
- *CentralMC* agent that simulates aggregation of calculation results.

The simulation model works according the following steps:

1) All functional agents are initialized and prepared for work.
2) Agents *MonteCarlo* start simulating calculations of realizations of the Monte Carlo algorithm. After completing each l realizations, data with results of these calculations are sent to the agent *CentralMC*.
3) Agents *CentralMC* collect data with results of calculations and process them periodically, saving resulting information to the disk (for analysis of simulation results).
4) The criterion for stopping the simulation is achieving the desired accuracy of calculations, in other words the number of realizations simulated by agents *MonteCarlo*.

Thus *CentralMC* counts the total number of performed operations and at achieving given threshold sends to AGNES a command to complete the simulation run.

ISBN: 978-1-61804-233-0

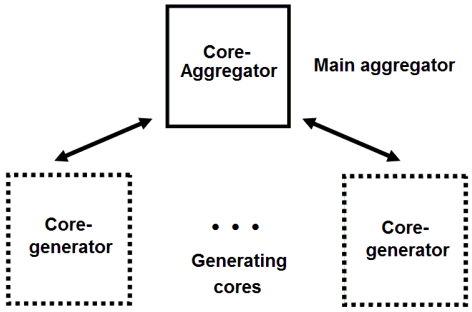


Figure 3. Basic simulation scheme of the Monte-Carlo method

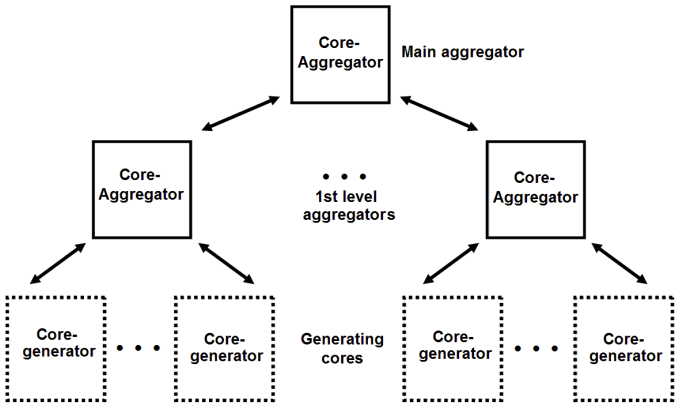


Figure 4. Modified simulation scheme of the Monte-Carlo method

The model was calibrated based on a real data obtained by profiler when running real computational program (in our case studies by the PARMONC). Important information that has been used for configuring the model includes a diagram of the computation algorithm and temporal characteristics of the stages of computation.

Results of the simulation run include the following reports generated by AGNES:

- General report contains information about the total model time spent for achieving given accuracy, resource usage, as well as service information.
- Report about realizations contains couples: a unique name for each agent *MonteCarlo* and a number of realizations which it had simulated.
- Report about times of realizations simulation. It allows obtaining and checking the statistics of calculations.
- Report on the intensity of the incoming data to the agent *CentralMC*. That is, agent *CentralMC* counts the number of messages received by it during regular intervals.

In the early stages of the simulation experiments the model consisted of a large number of agents *MonteCarlo* and one agent *CentralMC* (see Fig. 3). This allowed us to see the effect of overloading the agent *CentralMC* and its inability to handle the large amount of packets coming to it. In this regard, the model was modified so that one main *CentralMC* and several common *CentralMC* are created. Last agents collect information from *MonteCarlo* and in aggregated mode are transferred to main *CentralMC* (see Fig. 4).

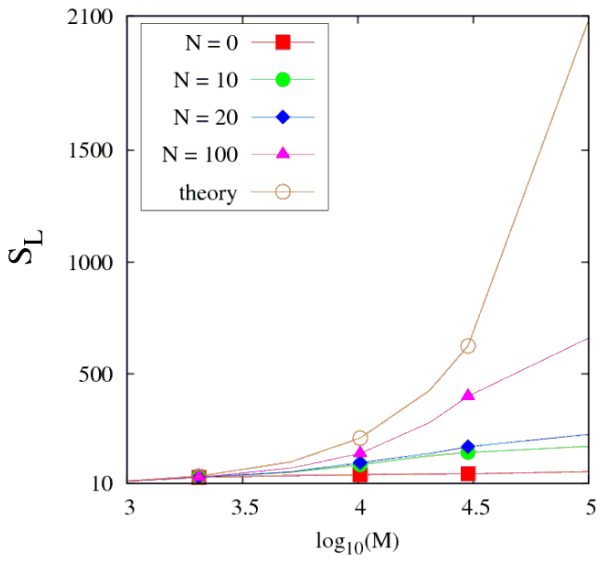


Figure 5. Comparison of modeled scalability of Monte-Carlo algorithm between theory and simulation modeling for 48 - 10000 cores range. N is a number of *CentralMC* functional agents, X-axis is in the logarithmic scale

A comparison of the speed-up

$$S_L(M) = T_L(M_{min})/T_L(M)$$

was made for 48-100000 modeled computational cores in AGNES while up to 480 real cores were used for simulation.

Fig. 5 shows the speed-up of modeled algorithm and impact of the *CentralMC* agents number on it. We can see that adding of several common *CentralMC* aggregators agents minimize an effect of main *CentralMC* agent overloading. One can say that the result, that is overloading of a single aggregator, is quite obvious. But by simulation we find the optimal number of aggregators and their distribution by levels. Fig. 6 shows dependence of the acceleration of calculations on the number of the aggregators on the intermediate level when 10^6 cores are used for calculations (simulation results).

IV. Conclusion

In this paper, we introduce AGNES (AGent NEtwork Simulator) software package as a tool for simulating supercomputers workload for different kinds of algorithms. The results of simulation of different algorithm types are in a good accordance with real run tests on supercomputer. We can state that such simulation is very useful for preliminary tuning and even modifying algorithms before their executing on real supercomputers so saving time and money. Such simulation is also very important in projects aiming on designing algorithms and software for non-existent supercomputers of future generations at their the design stage. Future development of AGNES concerns improvement of its interface abilities and implementing new algorithms for its own scalability, thus allowing high-performance simulation of really large systems. Another goal of the project is development of a number of object-oriented libraries of agents, specially designed for simulating promissory new supercomputer architectures.

ISBN: 978-1-61804-233-0

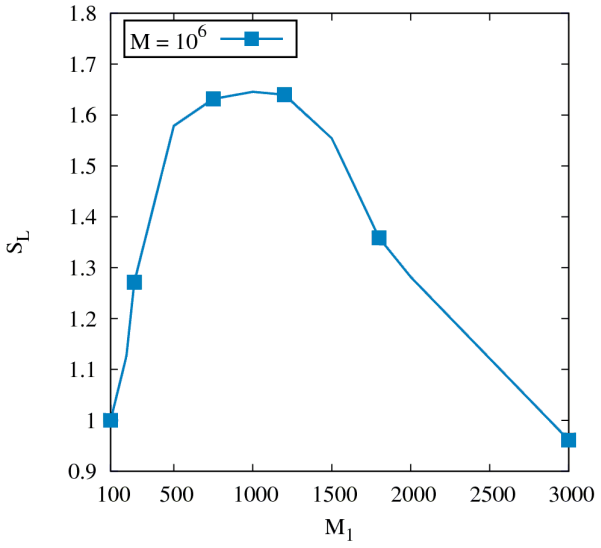


Figure 6. Dependence of the acceleration of calculations on the number of the aggregators on the intermediate level when 10^6 cores are used for calculations

REFERENCES

[1] A. Sivasubramaniam, A. Singla, U. Ramachandran, and H. Venkateswaran, *A Simulation-based Scalability Study of Parallel Systems,* Journal of Parallel and Distributed Computing, Vol 22(3) (1994) 411–426.

[2] G. Racherla, S. Killian, L. Fife, M. Lehmann, and R. Parekh, *Parsit: A parallel algorithm reconfiguration simulation tool,* Proc. of the Int. Conf. on High Performance Computing (1995).

[3] R.M. D'Souza, M. Lysenko, S. Marino, and D. Kirschner, *Data-parallel algorithms for agent-based model simulation of tuberculosis on graphics processing units,* Proc. of the 2009 Spring Simulation Multiconference (SpringSim '09, San Diego, CA, USA (2009).

[4] M.T. Goodrich, *Simulating Parallel Algorithms in the MapReduce Framework with Applications to Parallel Computational Geometry,* http://dblp.uni-trier.de/db/journals/corr/corr1004.html#abs-1004-4708 (2010)

[5] P. Kogge (Ed.), *ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems,* DARPA report. http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf. (2008)

[6] V. Ivannikov, A. Avetisyan, V. Padaryan, *Evaluation of dynamic characteristics of a parallel program on a model,* Programming, vol. 4, (in russian) (2006) 21–37.

[7] A.I. Avetisyan, S.S. Gaysaryan, V.P. Ivannikov, and V.A. Padaryan, *Productivity prediction of MPI programs based on models*, Automation and remote control, vol. 68, issue 5 (2007) 750–759.

[8] D. Podkorytov, A. Rodionov, O. Sokolova, and A. Yurgenson, *Using Agent-Oriented Simulation System AGNES for Evaluation of Sensor Networks,* Lecture Notes in Computer Science, vol. 6235, (2010) 247–250.

[9] D. Podkorytov, A. Rodionov, H. Choo, *Agent-based simulation system AGNES for networks modeling: review and researching,* Proc. of the 6th Int. Conf. on Ubiquitous Information Management and Communication (ACM ICUIMC 2012), ISBN 978-1-4503-1172-4, Paper 115 ACM (2012) 4 pages,.

[10] B. Glinsky, A. Rodionov, M. Marchenko, D. Podkorytov, and D. Weins, *Scaling the Distributed Stochastic Simulation to Exaflop Supercomputers,* High Performance Computing and Communication & 2012 IEEE 9th Int. Conf. on Embedded Software and Systems (HPCC-ICESS) IEEE (2012) 1131–1136.

[11] M. Marchenko, *PARMONC - A software library for massively parallel tochastic simulation*, Lecture Notes in Computer Science, vol. 6873, (2011) 302–315.

[12] C. Cercignani, *Rarefied Gas Dynamics.: From Basic Concepts to Actual Calculations*, Cambridge University Press, 2000.

[13] B. Glinskiy, D. Karavaev, V. Kovalevskiy, and V. Martynov, *Numerical simulation and experimental study of the mud volcano "Mount Karabetova" by vibroseismic methods,* Journal of Numerical Methods and programming, vol. 11, no. 1, (in russian) (2010) 99–108

[14] I. Kulikov, et. al, *Hydrodynamical code for numerical simulation of the gas components of colliding galaxies,* Astrophysical Journal: Suppl. series, vol. 194, no. 47 (2011)

[15] V. Snytnikov, T .Mischenko, Vl. Snytnikov, and I. Chernykh, *Physico-chemical processes in a flow reactor using laser radiation energy for heating reactants,* Chemical Engineering Research and Design, vol. 90, iss. 11 (2012) 1918–1922

[16] V. Snytnikov, T .Mischenko, Vl. Snytnikov, and I. Chernykh, *A reactor for the study of homogeneous processes using laser radiation energy,* Chemical Engineering Journal, vol. 150, iss. 1, (2009) 231–236

[17] O. Stoyanovskaya, I. Chernykh, O. Stadnichenko, *ChemPAK Software Package as an Environment for Kinetics Scheme Evaluation,* Chemical Product and Process Modeling, vol. 4, iss. 4 (2009)

[18] S. Souravlas, E. Kotsialos, A. Margaris, and Manos Roumeliotis *On Simulating Parallel Algorithms with VHDL,* 8th Hellenic European Research on Computer Mathematics and its Applications Conference, http://www.aueb.gr/pympe/hercma/proceedings2007/H07-FULL-PAPERS-1/SOURAVLAS-KOTSIALOS-MARGARIS-ROUMELIOTIS-1.pdf (2007)

Effects of speech codecs on a remote speaker recognition system using a new SAD.

Riadh AJGOU[(1)], Salim SBAA[(2)], Said GHENDIR[(1)],Ali CHAMSA [(1)] and A. TALEB-AHMED[(3)]
[(1)] Department of sciences and technology, faculty of sciences and technology, 'El-oued University.
[(2)] Electric engineering department, LESIA Laboratory, Med Khider University Biskra .Algeria
[(3)] LAMIH Laboratory Valenciennes University France.
E-mail:ajgou2007@yahoo.fr,s_sbaa@yahoo.fr, ghendir.emp@gmail.com, chemsadoct@yahoo.fr
abdelmalik.taleb-ahmed@univ-valenciennes.fr

***Abstract*—** This paper deals with the effects of speech codecs on remote text-independent speaker recognition performance in VoIP applications, considering three types of speech codec: PCM, DPCM and ADPCM conforming to International Telecommunications Union - Telecoms (ITU-T) recommendation used in telephony and VoIP (Voice over Internet Protocol). In order to improve the speaker recognition performance in noisy environment we propose a new robust speech activity detection algorithm (SAD) using "Adaptive Threshold", which can be simulated with speech wave files of TIMIT database submerged in an additive noise. Moreover, the speaker recognition system is based on Vector Quantization and Mel-Frequency Cepstral features extraction algorithm. Where, the feature extraction proceed after (for testing phase) and before (for training phase) the speech is sending over communication channel. Therefore, the digital channels can introduce several types of degradation. To overcome the degradation of channel, a convolutional code is used as error-control coding with the AWGN channel. Finely, the best overall performance of speech codecs was observed for the PCM code in terms of recognition rate accuracy and execution time.

***Keywords*—** PCM, DPCM, ADPCM, speaker recognition, SAD.

I. INTRODUCTION

Speaker recognition is the ability of recognizing a person from solely his voice. There are various techniques to resolve the automatic speaker recognition problem [1, 2, 3, 4, 5]. Where most published works where the speech does not under phone network and digital channels in noisy environment. Few published works on phone network and digital channels and degradation [6, 7, 8, 9, 10]. Our aim is to provide a comprehensive assessment of speech codecs, considering codecs conforming to ITU-T (International Telecommunications Union - Telecoms) and that are used in Internet telephony and internal IP network in a remote speaker recognition system. Figure 1 shows a general diagram of a remote information system using a remote speaker (or speech) recognition approach, where after recognition, the server provides and transmits the required information to the client [7]. Moreover, the speaker recognition system is based on Vector Quantization (VQ) [11, 12] and Mel - Frequency Cepstral Coefficients (MFCC) algorithm [12, 6]. There are various kinds of speech codecs available. The speech codecs generally complies industry standards like ITU-T. The main software of ITU standard are : G.711, G.722, G.723, G.726, G.728, G727, and G.729. We consider in this paper PCM, DPCM and ADPCM codecs and their effects on speaker recognition accuracy.

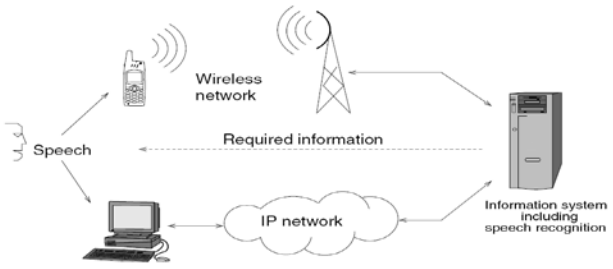


Fig. 1 General scheme of a speech-driven remote information system [7]

There are several possibilities for the implementation of a remote *speaker or speech recognition system over a digital channel*. In the first approach, usually known as *network speech recognition* (NSR), the recognition system resides in the network from the client's point of view [7]. In this case, the speech is compressed by a speech codec in order to allow a low bit rate transmission and/or to use an existing speech traffic channel (as in the case of mobile telephony). The recognition is usually performed over the features extracted from the decoded signal, although it is also possible to extract the recognition features directly from the codec parameters. Figure 2 shows a scheme of this system architecture. In the case where implementation is over an IP network, a VoIP codec can be employed [7]. The second approach known as *distributed speech (or speaker) recognition* (DSR) [7]. In this case, the client includes a local front end that processes the speech signal in order to directly obtain the specific features used by the remote server (back end) to perform recognition,

thus avoiding the coding/ decoding process required by NSR [7]. The conceptual scheme of DSR is shown in Figure 3.

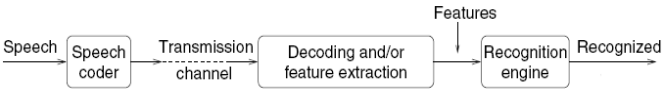


Fig 2. Scheme of a network speaker (or speech) recognition system [7].

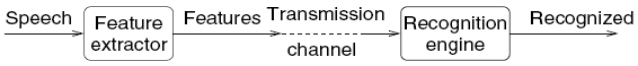


Fig 3. Scheme of a distributed speech recognition (DSR) system [7].

In our work we adopted network *speaker (speech) recognition* conforming to figure 4.

Our work is divided in five steps: in the first we illustrate speech codecs standard. In the second a system configuration is set including speaker recognition system over digital channel (AWGN). In the third we illustrate feature extraction used in recognition system (MFCC). In the forth, in order to improve the memory capacity and speaker recognition performance in noisy environment we propose a new robust speech activity detection algorithm (SAD) using "Adaptive Threshold". The key advantages of this algorithm is its simple implementation and its low computational complexity. This algorithm is based on energy and zero crossing rate. In the fifth, we introduce error correcting code methods that are necessary to improve immunity to noise of communication channels, in our work we considered convolutional code. In the sixth, a simulation results and discussion is done, where we started by the evaluation of speech activity detection algorithm (SAD), then a work is done about remote speaker identification accuracy using PCM, DPCM, and ADPCM with AWGN channel versus SNR. To illustrate the advantage of using channel code, we use remote speaker identification system with and without convolutional code. In term of execution time, we do a comparative study of PCM, DPCM, and ADPCM codecs.

II. Speech Codecs Standard

Access to a voice server is not only made through the conventional telephone network, but voice can also be transmitted through wireless networks or *IP* networks. The main factors that determine voice quality are choice of codec, packet loss, latency and jitter. The number of standard and proprietary codecs developed to compress speech data has been quickly increased. We present, only the codecs conforming to ITU-T.

A. PCM

G.711 [13,14] known as PCM codec used in VoIP and fixed telephony. VoIP standard describes two algorithms μ-law and A-law. The μ-law version is used primarily in North America. A-law version used in most other countries outside North America. Both algorithms code speech using 8 bits per sample which provides 50 % reduction in bandwidth use for original signal sampled with 16 bits at 8 kHz sample rate, i.e. reduction from 128 kb/s to 64 kb/s.

B. DPCM (G727)

Differential coding is a signal encoder that uses the baseline of PCM witch naturally appropriate in speech quantization. One of the first scalable speech codecs was embedded DPCM [15] Later in 1990, an embedded DPCM system was standardized by the ITU-T as G.727 [16]. The typical operating rate of systems using this technique is higher than 2 bits per sample resulting in rates of 16 Kbits/sec or higher [16].

C. ADPCM

G.726 [17, 14] algorithm provides conversion of 64 kb/s A-/μ-law encoded signal to and from a 40, 32, 24 or 16 kb/s signal using Adaptive Differential Pulse Code Modulation (ADPCM). The principle application of 40 kb/s is to carry data modem signals not speech. The most commonly used bitrates for speech compression is 32 kb/s which doubles the capacity compared to the G.711

III. System Configuration

Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification is the process of accepting or rejecting the identity claim of a speaker. The system that we will describe is classified as text independent speaker identification system.

The system we used for experiments included a remote text independent speaker recognition system which was set up according to the following block diagram in figure 4.

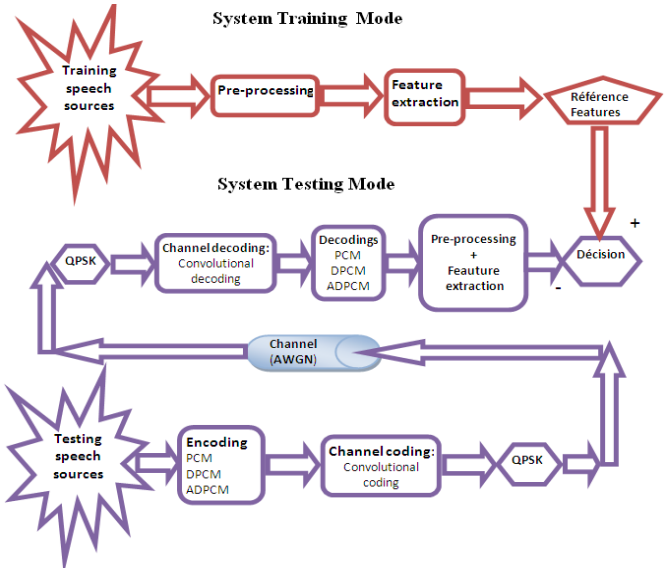


Fig. 4. Remote speaker recognition system.

A. TRAINING PHASE

In the training stage, pattern generation is the process of generating speaker specific models with collected data. The generative model used in speaker recognition is Vector Quantization [11].

The system was trained using speakers from the TIMIT database where we have used 30 speakers from different regions. Speech signal passed through pre-processing phase (emphases + speech activity detection). We emphasize the speech using a high pass filter. We usually use a digital filter with 6dB/ Octave. In formula (1), μ is a constant which is taken 0.97 usually [12] [6].

$$y(z) = 1 - \mu z^{-1} \tag{1}$$

After emphasising phase, silence segments are removed by the speech activity detection algorithm so that *twenty four* mel-frequency cepstral coefficients are extracted and form the characterization of the models using Vector Quantization.

B. TESTING PHASE

In this stage we have use speech codecs: PCM, DPCM and ADPCM therefore their coefficients are converted into a binary sequence. Before using QPSK modulation we introduce Convolutional code [7, 18] with a rate of ½. The coded signal is transmitted over AWGN channel. After demodulation (QPSK), convolutional decoding, and PCM, DPCM, or ADPCM decoding, the binary data is converted back to a synthesized speech file. Finally, cepstral coefficients are extracted (from the synthesized speech file) .

C. CECISION PHASE

Pattern matching is the task of calculating the matching scores between the input feature vectors (arrived from testing phase) and the given models. In our work we have used the Euclidean Distance (ED) classifier method. The Euclidean distance (ED) classifier has the advantage of simplicity and fast computational speed. The classification is done by calculation the minimum distance to decide which speaker out of all the training set and the most likely to be the test speaker.

Consider a class '***i***' with an m-component mean feature vector X, and a sample vector Y, given respectively by [19]:

$$\overline{X}_i = \left[\overline{X_{i1}}, \overline{X_{i2}} ..... \overline{X_{im}}\right]^t \tag{2}$$

and [19]:

(

$$Y = \left[y_1, y_2 .... y_m\right]^t \tag{3}$$

The ED between class i and vector Y is given by [19]:

$$d(i,Y) = \left\| \overline{X_{i1}} - Y \right\| = \sum_{k=1}^{m} (\overline{x_{ik}} - yk)^2 \tag{4}$$

For a number of classes C, the decision rule for the ED classifier is that Y be assigned to class ***j*** if [19]:

$$d(i,Y) = \min\{d(i,Y)\}, \forall i \in C \tag{5}$$

IV. FEATURES EXTRACTION

The speaker-dependent features of human speech can be extracted and represented using the mel-frequency cepstral coefficients MFCC [6]. These coefficients are calculated by taking the cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale [10]. After pre-emphasis and speech/silence detection the speech segments are windowed using a Hamming window. MFCC calculation is shown in figure 5.

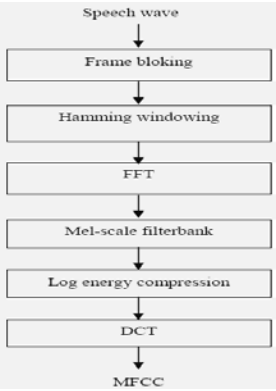


Fig 5. The process of calculating MFCC [20].

The TIMIT database files are sampled with a rate of 16000 samples/s, these files were downsampled to a rate of 8000 samples/s. In our speaker recognition experiments, the speech signal is segmented into frames of each 16 ms, generally it takes 128 points as a frame and the overlapped data is 64 points (8 ms). The Discrete Fourier Transform is taken of these windowed segments. The magnitude of the Fourier Transform is then passed into a filter-bank comprising of twenty four triangular filters (corresponds to MFCC coefficients).

V. NEW SPEECH ACTIVITY DETECTION (SAD) ALGORITHM WITH ADAPTIVE THRESHOLD

Speech presence detection in a background of noise is the important pre-processing step in Automatic Speaker Recognition (ASR) systems[21]. Removing nonspeech frames from the speaker recognition system input stream effectively reduces the insertion error rate of the system. SAD perform the speech/nonspeech classification and background noise reduction process on the basis of speech features extracted from the frame under consideration. There are many SAD algorithms available [21, 22, 23, 24, 25].

The new SAD is based on two original works [25, 26]. In [25] the author had used the LPC residual energy and zero crossing rate to detect speech activity using adaptive threshold where this threshold is calculated for every frame introduced in comparison with previous calculated features of frames which means a probable mistakes for firs frames (the algorithm is initiated and spans up to a few frames 0-15 frames, which is considered as non-speech). The second author [26] used energy and zero crossing rate ratios to voiced/non voiced classification of speech using a fixed threshold.

Our new SAD is based on Energy and Zero crossing Rate (EZR) ratios using an adaptive threshold to detect speech activity and remove the silent and noise intervals. SAD operating with a rectangular window of 8ms. The procedure of calculating threshold is as follows:

1- Segmenting the whole speech signal (speaker's signal) in frames of 8ms with rectangular window and without overlapping.

2- Calculating the energy (E[m]) and zero crossing rate (ZCR[m]) for each frame and calculating E[m]/ZCR[m].

3- Calculating the maximum and minimum of EZR.

4- Calculate Threshold (formula 10).

The principle of EZR application explains itself by the fact that the energy of the speech activity is important while the rate of zero crossing rates is weak; therefore the value of EZR is important.

If a frame have an EZR superior to a threshold, this frame classified as speech, if the opposite the frame considered as nonspeech (the recognition system does not extract features from this frame). The threshold determination is estimated by the SAD algorithm in automatic and adaptive way [26]:

$$EZR[m] = \frac{\overline{E}[m]}{ZCR[m]} \tag{6}$$

Where ZCR [m] and $\overline{E}[m]$ present respectively the zero crossing rate and the average energy of a frame [27]:

$$\overline{E}(m) = \sum_{n=0}^{N-1} x^2(n).w(m-n) \tag{7}$$

Where: w is a rectangular window of length N (length of a frame) and x(n) is the frame signal with N samples. ZCR is defined as [28]:

$$ZCR(m) = \sum_{n=0}^{N-1} \left|\text{sgn}[x(n)] - \text{sgn}[x(n-1)\right| w(m-n) \tag{8}$$

Where sgn(.) is the signum function which is defined as [28]:

$$\text{sgn}[x(n)] = \begin{cases} +1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases} \tag{9}$$

SAD algorithm calculates EZRs of all frames (for a speaker's signal) and estimate threshold:

$$Threshold = \min(EZR) + \alpha * [DELTA] \tag{10}$$

$$DELTA = \max(EZR) - \min(EZR) \tag{11}$$

α: is a real number in the interval of]0,1[. In our simulation we fixed: α=0,35. We can resume our algorithm of speech activity detection in figure 6.

First step: Segmenting the speech signal into frames of 8ms to calculate threshold where α=0.35

From 1 to "m" frames:
Calculate énergie 'E[m]'
Calculate 'ZCR[m]'
Calculate 'EZR[m]' .
End
max_r=max{EZR};
min_r=min {EZR};
delta=max_r-min_r;
threshold =min_r+ α *delta;

Second step: using calculated threshold to detect speech /nonspeech:

From 1 to "m" frames:
Calculate " EZR"
if EZR [m] > threshold
Speech frame.
else
non Speech frame
end
End

Fig. 6. Main algorithm of speech activity detection based on EZR .

VI. CHANNEL CODING TECHNIQUES

Channel coding is developed to maximize the recognition performance, There are different types of FEC (Forward Error Correction) techniques, namely Reed-Solomon and Convolutional codes [29]. The Viterbi algorithm is a method for decoding convolutional codes. A convolutional code with a code rate k/n also generates n output bits from every k input bits, as in the case of block codes. The difference is because the encoding of these k bits is not independent from the bits previously received but it has "memory." A general diagram of a convolutional encoder is shown in Figure 7. At each time unit, the encoder takes a k-bit input sequence, shifts it through a set of m registers, and generates an n-bit output by performing a linear combination (or convolution), in modulo-2 arithmetic, of the data stored in the registers. The integer m is called *constraint length*. When k = 1, we have a special case for which the input bitstream is continuously processed [7].

ISBN: 978-1-61804-233-0

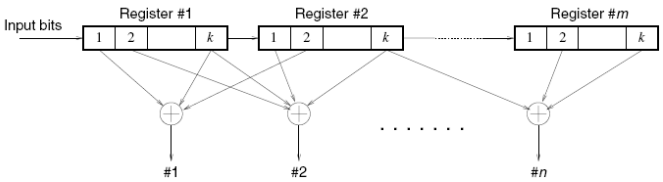


Fig. 7. Structure of a convolutional encoder[7].

In our work we use a convolutional code with ½ rate.

VII. SIMULATION RESUTS AND DISCUSSION

In our work, we are training and testing the same database (TIMIT). As mentioned above, the generative model used in speaker recognition is Vector Quantization using MFCC (24 features).

The first test we, evaluate our SAD algorithm, where the speech signal which is " *She had your dark suit in greasy wash water all year*" passed through the algorithm ($\alpha = 0.35$). The figure 8 represents the original speech signal. The figure 9 represents the speech signal after it has been passed through SAD algorithm. The figure 10 illustrates a speech signal (clean speech) and its SAD counter. From figures 9 and 10, we observe the efficiency of our algorithm where silent segments are eliminated, therefore the memory capacity and recognition accuracy improved. Figures 11, 12 and 13, represent the SAD counter as function of SNR for 10dB, 5dB, 0dB respectively where it is clear that the SAD counter is effective, where Silent and background noise segments are eliminated. Further, Speech activity detection is robust down to SNR=5 dB. To observe the effect of the SAD algorithm on speaker identification rate, we have used our system of speaker recognition with and without SAD (not over digital channel). Figure 14 shows an identification rate with and without speech activity detection algorithm as function of SNR. It is clearly shown that this figure represents an improvement of identification rate accuracy when using the SAD algorithm in noisy environment. Further, the detector functioned accurately in low SNR environments

The second test is about channel errors effect on a remote speaker recognition system. Therefore, we use original and reconstructed wave files after transmition over AWGN channel, furthermore we use these files with speaker identification system. Table 1 shows a simulation results of identification rate accuracy using original and reconstructed speech wave files, where we observe performance degradation of speaker identification accuracy when using reconstructed files.

The third test consists to do the identification rate of speaker recognition system using : PCM, DPM and ADPCM code used in our Remote Speaker Recognition system where the figure 15 illustrates this study using AWGN channel in noisy environment, where we can conclude the efficiency of PCM code.

The fourth test consists to the execution time of each codec used in our work. Table 2 shows a simulation results of: PCM, DPCM and ADPCM techniques in term of execution time using our remote speaker recognition system, where we can observe that DPCM require more time to execute than PCM and ADPCM, however PCM technique needed low time.

Even its clear, the use of channel coding gives good results, the fifth test is about effects of channel coding on speaker identification accuracy. Therefore, we evaluated the identification rate with and without Convolutional code as function of SNR (AWGN + code) considering PCM technique. Figure 16 shows a simulation results for identification rate as function of SNR with and without Convolutional code where we conclude the efficiency of channel code.

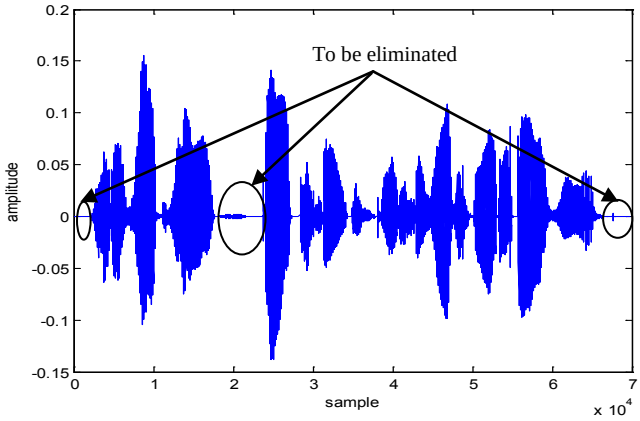


Fig.8. The original speech signal.

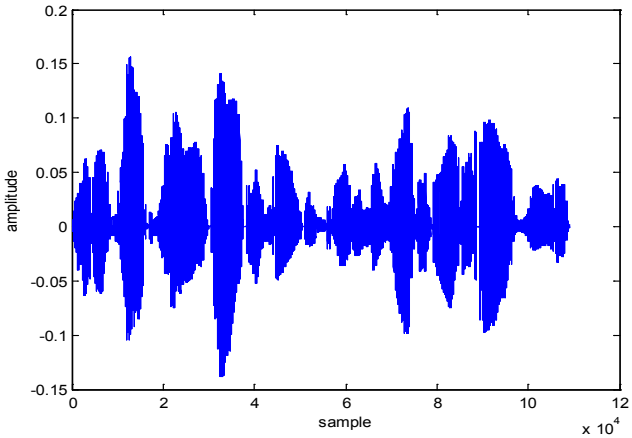


Fig. 9. Speech signal through activity detection algorithm (α=0.35).

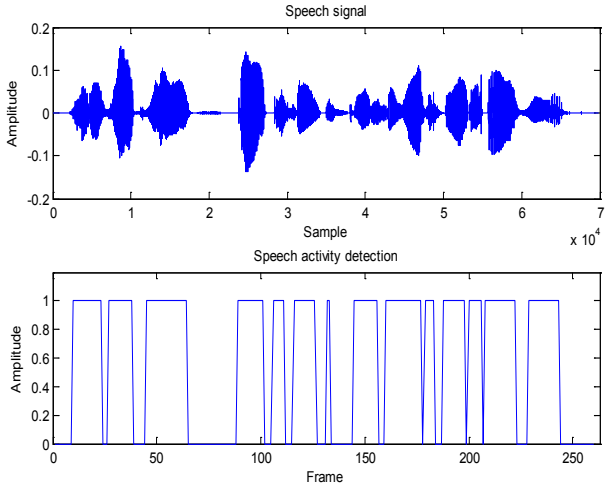


Fig. 10. Speech signal (clean signal) and SAD counter (below).

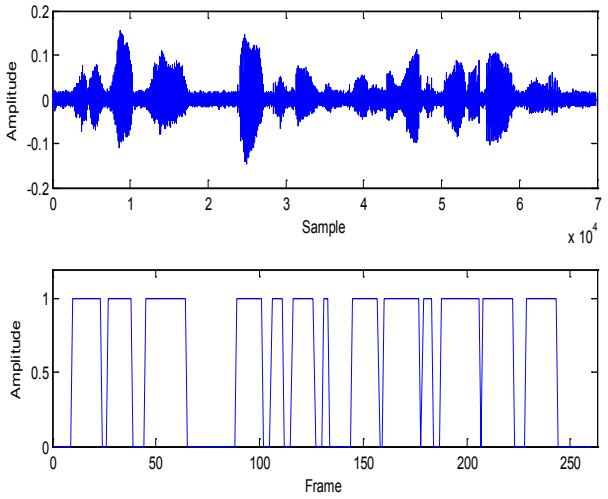


Fig. 11. Speech signal and SAD counter (below) at SNR=10dB.

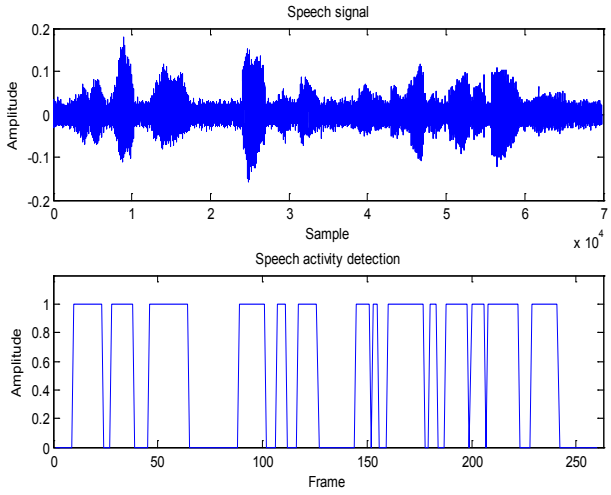


Fig. 12. Speech signal and SAD counter (below) at SNR=5dB.

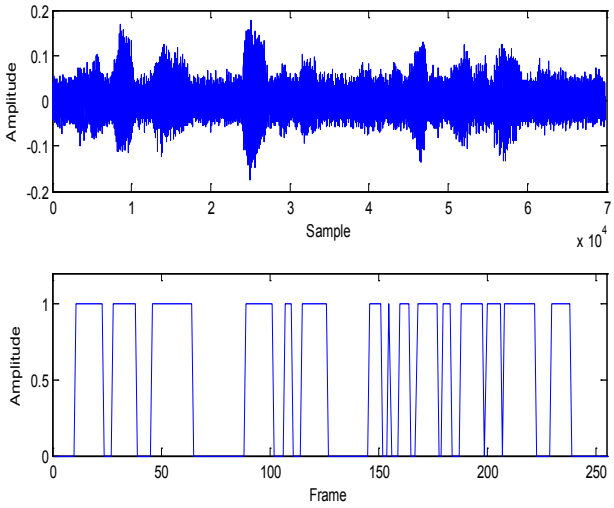


Fig. 13. Speech signal and SAD counter (below) at SNR=0dB.

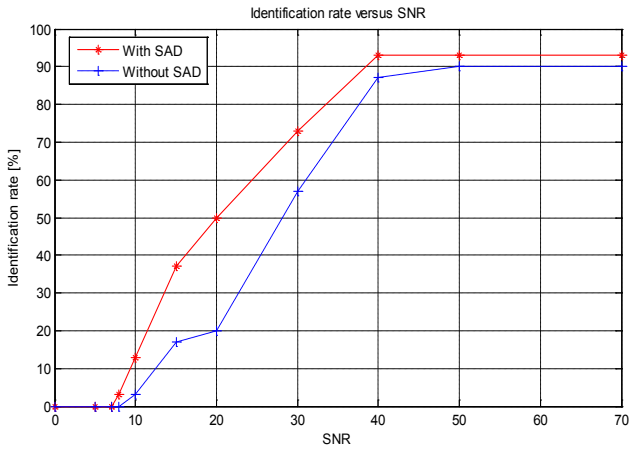


Fig 14. Identification rate with and without speech activity detection algorithm versus SNR.

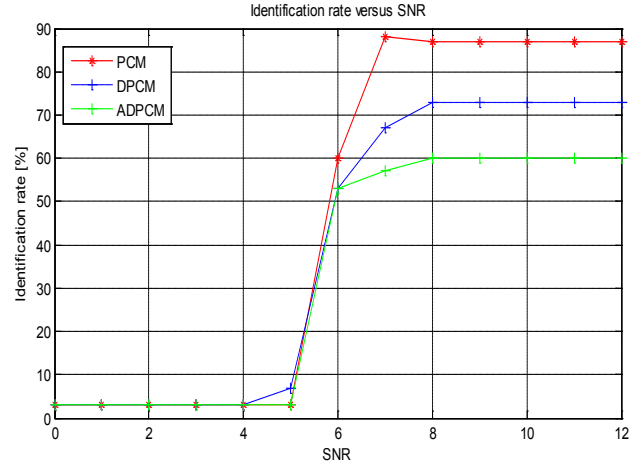


Fig. 15. Remote speaker identification accuracy using PCM, DPCM, and ADPCM versus SNR.

ISBN: 978-1-61804-233-0

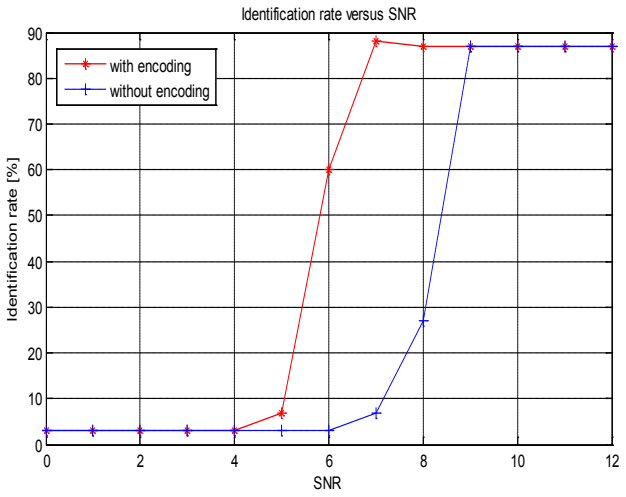


Fig. 16. Speaker identification accuracy with and without encoding (covolotional code).

	Speaker identification system	Speaker identification over AWGN Channel
Identification rate %	93	87

Table1. Identification rate accuracy using original speech waveform and reconstructed speech after transmition over AWGN channel (SNR=100 dB).

	PCM	DPCM	ADPCM
Elapsed time [sec]	100.84	764.24	554.32

Table2. Elapsed time of execution using: PCM, DPCM and ADPCM.

VIII. CONCLUSION

In this work we have done a comparative study of speech codec: PCM, DPCM and ADPCM in view of their effects on our remote automatic speaker recognition system performance in noisy environment. Therefore, a system configuration is set. Since, Speech activity detection (SAD) algorithm perform speech/nonspeech classification and background noise reduction process, we have developed a new (SAD) algorithm which improves memory capacity and identification rate accuracy.

Our SAD algorithm that is based on energy and zero crossing rate, performs suitable counter of speech activity. Furthermore, it functioned accurately in low SNR environments (down to SNR=5 dB) and leads to a good identification accuracy.

In order to improve identification rate accuracy, the use of channel coding is necessary to make the remote system more robust against channel errors, therefore we have chosen convolutional code. the best overall performance of speech codecs was observed for PCM code in terms of identification rate accuracy and execution time (Elapsed time of execution). Moreover, it's recommended using PCM technique as speech codec in remote speaker recognition system in VoIP applications.

REFERENCES

[1] LIU, Liwei, QIAN, Feng, et ZHANG, Yao. Application Research of HHT-IF Speech Feature Parameter in Speaker Recognition System. *Energy Procedia*, 2012, vol. 17, p. 1102-1108.

[2] FURUI, Sadaoki. Recent advances in speaker recognition. *Pattern Recognition Letters*, 1997, vol. 18, no 9, p. 859-872.

[3] LUNG, Shung-Yung. Feature extracted from wavelet eigenfunction estimation for text-independent speaker recognition. *Pattern recognition*, 2004, vol. 37, no 7, p. 1543-1544.

[4] RAMACHANDRAN, Ravi P., FARRELL, Kevin R., RAMACHANDRAN, Roopashri, *et al.* Speaker recognition—general classifier approaches and data fusion methods. *Pattern Recognition*, 2002, vol. 35, no 12, p. 2801-2821.

[5] Shung-Yung Lung "Multi-resolution form of SVD for text-independent speaker recognition" Department of Management Information Systems, Chung-Kuo Institute of Technology, 56, Section 3, Hsing-Lung Road, Wen-Shan District, Taipei, Taiwan, ROC. Pattern Recognition 35 (2002) 1637 – 1639.

[6] SAHIDULLAH, Md et SAHA, Goutam. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication*, 2012, vol. 54, no 4, p. 543-565.

[7] PEINADO, Antonio et SEGURA, Jose. *Speech Recognition Over Digital Channels: Robustness and Standards*. John Wiley & Sons, 2006.

[8] HECK, Larry P., KONIG, Yochai, SÖNMEZ, M. Kemal, *et al.* Robustness to telephone handset distortion in speaker recognition by discriminative feature design. *Speech Communication*, 2000, vol. 31, no 2, p. 181-192.

[9] ALEXANDER, Anil, BOTTI, Filippo, DESSIMOZ, D., *et al.* The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. *Forensic science international*, 2004, vol. 146, p. S95-S99.

[10] NEVILLE, Katrina, AL-QAHTANI, Fawaz, HUSSAIN, Zahir M., *et al.* Recognition of Modulated Speech over OFDMA. In : *TENCON 2006. 2006 IEEE Region 10 Conference*. IEEE, 2006. p. 1-3.

[11] HE, Jialong, LIU, Li, et PALM, Günther. A discriminative training algorithm for VQ-based speaker identification. *Speech and Audio Processing, IEEE Transactions on*, 1999, vol. 7, no 3, p. 353-356.

[12] XIE, Chuan, CAO, Xiaoli, et HE, Lingling. Algorithm of Abnormal Audio Recognition Based on Improved MFCC. *Procedia Engineering*, 2012, vol. 29, p. 731-737.

[13] RECOMMENDATION, G. 711:"Pulse Code Modulation (PCM) of voice frequencies". *ITU (November 1988)*, 1988.

[14] Silovsky, J., Cerva, P., & Zdansky, J. (2011, September). Assessment of speaker recognition on lossy codecs used for transmission of speech. In*ELMAR, 2011 Proceedings* (pp. 205-208). IEEE..

[15] N. S. Jayant, "Digital coding of speech waveforms-PCM, DPCM and DM quantizers," Proc. IEEE, vol. 62, pp. 621-624, May 1974.

[16] DONG, Hui, GIBSON, Jerry D., et KOKES, Mark G. SNR and bandwidth scalable speech coding. In : *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*. IEEE, 2002. p. II-859-II-862 vol. 2.

[17] 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), International Telecommunication Union Std. G.726 (12/90), Geneva 1990.

[18] MUÑOZ PORRAS, José María et IGLESIAS CURTO, José Ignacio. Classification of convolutional codes. *Linear Algebra and its Applications*, 2010, vol. 432, no 10, p. 2701-2725.

[19] AMADASUN, M. et KING, R. A. Improving the accuracy of the Euclidean distance classifier. *Electrical and Computer Engineering, Canadian Journal of*, 1990, vol. 15, no 1, p. 16-17.

[20] Snani Cherifa, Ramdani Messaoud. New Technique to use the GMM in Speaker Recognition System (SRS). Computer Applications Technology (ICCAT), 2013 International Conference 2013 IEEE.

[21] HATAMIAN, Shahin. Enhanced speech activity detection for mobile telephony. In : *Vehicular Technology Conference, 1992, IEEE 42nd*. IEEE, 1992. p. 159-162.

[22] PADRELL, Jaume, MACHO, Dušan, et NADEU, Climent. Robust speech activity detection using LDA applied to FF parameters. In : *Proc. ICASSP*. 2005.

[23] MACHO, Dusan, PADRELL, Jaume, ABAD, Alberto, *et al*. Automatic speech activity detection, source localization, and speech recognition on the CHIL seminar corpus. In : *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005. p. 876-879.

[24] LIANG, Zhang, YING-CHUN, Gao, ZHENG-ZHONG, Bian, *et al*. Voice activity detection algorithm improvement in adaptive multi-rate speech coding of 3GPP. In : *Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on*. IEEE, 2005. p. 1257-1260.

[25] HARSHA, B. V. A noise robust speech activity detection algorithm. In : Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on. IEEE, 2004. p. 322-325.

[26] KOTNIK, Bojan, HOGE, Harald, et KACIC, Zdravko. Evaluation of pitch detection algorithms in adverse conditions. In : Proc. 3rd international conference on speech prosody. 2006. p. 149-152.

[27] Charalampidis, D., & Kura, V. B. (2005, July). Novel wavelet-based pitch estimation and segmentation of non-stationary speech. In *Information Fusion, 2005 8th International Conference on* (Vol. 2, pp. 5-pp). IEEE.

[28] KATHIRVEL, P., MANIKANDAN, M. Sabarimalai, SENTHILKUMAR, S., *et al*. Noise robust zerocrossing rate computation for audio signal classification. In : *Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on*. IEEE, 2011. p. 65-69.

[29] LU, Jun, TJHUNG, Tjeng Thiang, ADACHI, Fumiyuki, *et al*. BER performance of OFDM-MDPSK system in frequency-selective Rician fading with diversity reception. *Vehicular Technology, IEEE Transactions on*, 2000, vol. 49, no 4, p. 1216-1225.

OWA – type Possibilistic Aggregations in a Decision Making Regarding Selection of Investments

Gia Sirbiladze and Gvantsa Tsulaia

Department of Computer Sciences
Iv. Javakhishvili Tbilisi State University
13, University st., 0186, Tbilisi, Georgia

***Abstract* —** In this work a new generalization of the OWA operator (introduced by R.R. Yager) is presented. Our focus is directed on the construction of the aggregation OWA operator – AsFPOWA in the possibilistic uncertainty environment. For the illustration of the applicability of the new aggregation operator - AsFPOWA an example of the fuzzy decision making problem regarding optimal selection of investment is considered. Several variants of the new aggregation operator are used for the comparing of decision making results.

***Keywords* —** *Decision-making system, expert evaluations, OWA operator, possibility uncertainty, selection of investments.*

I. New Possibilistic Aggregations in the OWA operator

It is well recognized that intelligent decision making systems (IDMS) and technologies have been playing an important role in improving almost every aspect of human society. In this type of problem the decision making person (DMP) has a collection $D = \{d_1, d_2, \ldots, d_n\}$ of possible uncertain alternatives from which he/her must select one or some rank decisions by some expert's preference relation values. Associated with this problem as a result is a variable of characteristics, activities, symptoms and so on, acts on the decision procedure. This variable normally called the state of nature, which affects the payoff, utilities, valuations and others to the DMP's preferences or subjective activities. This variable is assumed to take its values (states of nature) in the some set $S = \{s_1, s_2, \ldots, s_m\}$. As a result the DMP knows that if he/her selects d_i and the state of nature assumes the value s_j then his/her payoff (valuation, utility and soon) is a_{ij}. The objective of the decision is to select the "best" alternative, get the biggest payoff. But in IDMS the selection procedure becomes more difficult. In this case each alternative can be seen as corresponding to a row vector of possible payoffs. To make a choice the DMP must compare these vectors, a problem which generally doesn't lead to a compelling solution. Assume d_i and d_k are two alternatives such that for all $j, j = 1, 2, \ldots, m$ $a_{ij} \geq a_{kj}$. In this case there is no reason to select d_i. In this situation we shall say d_i dominates d_k $(d_i \succeq d_k)$. Furthermore if there exists one alternative (optimal decision) that dominates all the alternatives then it will be *optimal solution*. Faced with the general difficulty of comparing vector payoffs we must provide some means of comparing these vectors. Our focus in this work is on the construction of aggregation operator F that can take a collection of m values and convert it into a single value, $F : R^m \Rightarrow R^1$. In [5] R.R. Yager introduced a class of mean aggregation operators called Ordered Weighed Averaging (OWA) operator.

Definition 1: An OWA operator of dimension m is mapping $OWA : R^m \Rightarrow R^1$ that has an associated weighting vector W of dimension m with $w_j \in [0;1]$ and $\sum_{j=1}^{m} w_j = 1$, such that

$$OWA(a_1, \ldots, a_m) = \sum_{j=1}^{m} w_j b_j , \quad (1)$$

where b_j is the j-th largest of the $\{a_i\}, i = 1,2,\ldots,m$.

The Triangular Fuzzy Numbers (TFNs, denoted by Ψ) has been studied by many authors ([1] and others). It can represent in a more complete way as an imprecision variable of an expert knowledge. In the role of uncertainty measure a possibility distribution is taken. So, we consider possibilistic aggregations based on the OWA operator. Therefore we introduce the definition of a possibility measure [1]:

Definition 2: A possibility measure - Pos on 2^S can be uniquely determined by its possibility distribution function $\pi : S \to [0,1]$ via the formula:

$$\forall A \in 2^S, \; Pos(A) = \max_{s \in A} \pi(s).$$

ISBN: 978-1-61804-233-0

Let S_m be the set of all permutations of the set $\{1,2,..,m\}$, Let $\{P_\sigma\}_{\sigma\in S_m}$ be the associated probabilities class of a possibility measure - Pos [2-4]. Then, we have the following connections between $\{\pi_i\}$ and $\{P_\sigma\}_{\sigma\in S_m}$: $\forall\,\sigma\in S_m$,

$$P_\sigma\left(s_{\sigma(i)}\right)=\max_{\nu=1,i}\pi(s_{\sigma(\nu)})-\max_{\nu=1,i-1}\pi(s_{\sigma(\nu)}),$$

for each $\sigma=(\sigma(1),\sigma(2),...,\sigma(m))\in S_m$, which are called the associated probabilities.

Let $M:\Psi^k\Rightarrow\Psi\quad(k=m!)$ be some deterministic mean aggregation function [5].

Definition 3: An associated fuzzy-probabilistic OWA operator AsFPOWA of dimension m is mapping $AsFPOWA:\Psi^m\Rightarrow\Psi$, that has an associated objective weighted vector W of dimension m such that $w_j\in(0,1)$ and $\sum_{j=1}^{m}w_j=1$ some possibility measure $Pos:2^S\Rightarrow[0,1]$, according the following formula:

$$AsFPOWA(\tilde{a}_1,\tilde{a}_2,...,\tilde{a}_m)=\beta\cdot\sum_{j=1}^{m}w_j\tilde{b}_j+ \\ +(1-\beta)M\left(E_{P_{\sigma_1}}(\tilde{a}),E_{P_{\sigma_2}}(\tilde{a}),...,E_{P_{\sigma_k}}(\tilde{a})\right), \quad (2)$$

where $\tilde{b}_j$ is the j-th largest of the $\{\tilde{a}_i\},i=1,...,m$; $E_{P_{\sigma_i}}(\tilde{a})$ is a Mathematical Expectation of $\tilde{a}$ with respect to associated probability P_{σ_i}.

We will consider concrete AsFPOWA operators for concrete mean function M: AsFPOWAmin if M=Min, AsFPOWAmax, if M=Max and AsFPOWAmean if M=Mean.

II. The Problem of the Selection of Investments

We analyze an illustrative example of the using of the new AsFPOWA operator in a fuzzy decision-making problem regarding selection of investments. The main reason for using our new aggregation operator is that we are able to assess the decision making problem considering possibility distribution and the attitudinal characters of the DMPs.

In the following, we study a Company that wants to invest some money in a new market. They consider five alternatives: d_1: "Invest in the Asian market"; d_2: "Invest in the South American market"; d_3: "Invest in the African market"; d_4: "Invest in all three markets"; d_5: "Do not invest money in any market". In order to analyze these investments, the investor has brought together a group of experts. This group considers that the key factors are the economic situations of the world (external) and country (internal) economy for the next period. They consider 3 possible states of nature that in hole could occur in the future: s_1: "Bad economic situation"; s_2: "Regular economic situation"; s_3: "Good economic situation". As a result the group of experts gives us union one opinions and results. The results depending on the state of nature s_i and alternative d_k that the company selects, are presented in the Table 1.

Table 1: Expert's valuations in TFNs

D \ S	s_1	s_2	s_3
d_1	(60,70,80)	(40,50,60)	(50,60,70)
d_2	(30,40,50)	(60,70,80)	(70,80,90)
d_3	(50,60,70)	(50,60,70)	(60,70,80)
d_4	(70,80,90)	(40,50,60)	(40,50,60)
d_5	(60,70,80)	(70,80,90)	(50,60,70)

Following the expert's knowledge on the world economy for the next period, experts decided the objective weights (as an external factor) of states of nature must be $W=(0{,}5;0{,}3;0{,}2)$, which the country (in which is founded or works this Company) economy for the next period takes only some possibilities to occur presented states of nature in the country (as an internal factor). So, there exists some possibilities (internal levels), as an uncertainty measure, to occur states of nature in the country.

So, we can define subjective possibilities $\pi_i=Pos(s_i)$ based on the experts' knowledge. Let on the basis of some fuzzy term of internal factor – country economy experts define the possibility levels of states of nature: $poss(s_1)\equiv\pi_1=0{,}7;\quad poss(s_2)\equiv\pi_2=1;$ $poss(s_3)\equiv\pi_3=0{,}5.$

$$Pos(A)=\max_{s_i\in A}\pi_i,\ \forall\,A\subseteq S.$$

In this model weight $\beta\equiv 0{,}3$.

$W=(0{,}5;0{,}3;0{,}2)$;

For $\tilde{a}=(\tilde{a}_1,\tilde{a}_2,\tilde{a}_3)$ we have:

$$AsFPOWA(\tilde{a}_1,\tilde{a}_2,\tilde{a}_3)= \\ \beta\sum_{j=1}^{3}\tilde{b}_j w_j+(1-\beta)M\left(E_{P_{\sigma_1}}(\tilde{a}),E_{P_{\sigma_2}}(\tilde{a}),...,E_{P_{\sigma_6}}(\tilde{a})\right).$$

It is clear that k=m!=3!=6 and for calculation of the AsFPOWA operator we firstly define the associated probability class $\{P_\sigma{}^{Pos}\}_{\sigma\in S_3}$ for the $Pos:2^S\Rightarrow[0,1]$.

For every

$\sigma=\{\sigma(1),\sigma(2),\sigma(3)\}\in S_3$

$E_{P_\sigma}(d)=E_{P^{Pos}{}_\sigma}(\tilde{a})=\sum_{i=1}^{3}P_{\sigma(i)}\cdot\tilde{a}_{\sigma(i)}$.

The results presented in the Table 2.

Table 2: Associated Probability Class - $\{P_\sigma\}_{\sigma\in S_3}$

$\sigma=(\sigma(1),\sigma(2),\sigma(3))$	$P_{\sigma(1)}$	$P_{\sigma(2)}$	$P_{\sigma(3)}$
$(1,2,3)=\sigma_1$	$P_1=0{,}7$	$P_2=0{,}3$	$P_3=0$
$(1,3,2)=\sigma_2$	$P_1=0{,}7$	$P_3=0$	$P_2=0{,}3$
$(2,1,3)=\sigma_3$	$P_2=1$	$P_1=0$	$P_3=0$
$(2,3,1)=\sigma_4$	$P_2=1$	$P_3=0$	$P_1=0$
$(3,1,2)=\sigma_5$	$P_3=0{,}5$	$P_1=0{,}2$	$P_2=0{,}3$
$(3,2,1)=\sigma_6$	$P_3=0{,}5$	$P_2=0{,}5$	$P_1=0$

ISBN: 978-1-61804-233-0

Following the table 2 we calculate Mathematical Expectations - $\{E_{P_\sigma}(\cdot)\}_{\sigma\in S_3}$ (see Table 3).

Table 3: Mathematical Expectations - $\{E_{P_\sigma}(\cdot)\}_{\sigma\in S_3}$

$E_{P_\sigma}(\cdot)$ σ	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
$E_{P_\sigma}(d_1)$	(54,64,74)	(54,64,74)	(40,50,60)	(40,50,60)	(49,59,69)	(45,55,65)
$E_{P_\sigma}(d_2)$	(39,49,59)	(39,49,59)	(60,70,80)	(60,70,80)	(59,69,79)	(65,75,85)
$E_{P_\sigma}(d_3)$	(50,60,70)	(50,60,70)	(50,60,70)	(50,60,70)	(55,65,75)	(55,65,75)
$E_{P_\sigma}(d_4)$	(61,71,81)	(61,71,81)	(40,50,60)	(40,50,60)	(46,56,66)	(40,50,60)
$E_{P_\sigma}(d_5)$	(63,73,83)	(63,73,83)	(70,80,90)	(70,80,90)	(58,68,78)	(60,70,80)

Now we may calculate the values of different variants of the AsFPOWA operator with respect to different averaging operators M (Table 4):

Table 4: Aggregation results

D/Ag. Op.	$\text{AsFPOWA}_{\min}$	$\text{AsFPOWA}_{\max}$	$\text{AsFPOWA}_{\text{mean}}$
d_1	(44,54,64)	(54,64,74)	(49,59,69)
d_2	(45,55,65)	(64,74,84)	(57,66,75)
d_3	(52,62,72)	(56,66,76)	(53,63,73)
d_4	(45,55,65)	(60,70,80)	(51,61,71)
d_5	(60,70,80)	(68,78,88)	(64,74,84)

Calculating numerical values of AsFPOWAmin, AsFPOWAmax, AsFPOWAmean operators we can rank the alternatives from the most preformed to the less preformed. The results are shown in table 5.

Table 5: Ordering of the policies

N	*Aggreg. Operator*	*Ordering*
1	$\text{AsFPOWA}_{\min}$	$d_5 \succ d_3 \succ d_2 = d_4 \succ d_1$
2	$\text{AsFPOWA}_{\max}$	$d_5 \succ d_2 \succ d_4 \succ d_3 \succ d_1$
3	$\text{AsFPOWA}_{\text{mean}}$	$d_5 \succ d_2 \succ d_3 \succ d_4 \succ d_1$

It is clear that decision d_5 -"Do not invest money in any market" is an optimal solution (decision) in this problem.

III. Conclusion

In this work our focus is directed on the construction of a new generalization of the aggregation OWA operator – AsFPOWA in the possibilistic uncertainty environment. For the illustration of the applicability of the new aggregation operator - AsFPOWA an example of the fuzzy decision making problem regarding optimal selection of investments is considered, where we study a Company that is planning to invest some money in a new market.

References

[1] D. Dubois and H. Prade, Possibility Theory, New York: Plenum Press, 1988.

[2] GiaSirbiladze, Extremal Fuzzy Dynamic Systems: Theory and Applications. IFSR International Series on Systems Science and Engineering, 28, 1st Edtition, Springer, 2012.

[3] GiaSirbiladze and Anna Sikharulidze, Generalized Weighted Fuzzy Expected Values in Uncertainty Environment, Proceeding of the 9-th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, 2010, pp. 54-64.

[4] GiaSirbiladze, Anna Sikharulidze, Bezhan Ghvaberidze and Bidzina Matsaberidze, Fuzzy-probabilistic Aggregations in the Discrete Covering Problem, International Journal of General Systems,Vol.40, No. 2, 2011, pp. 169 - 196.

[5] R.R. Yager, On Ordered Weighted Averaging aggregation operators in multi-criteria decision making, IEEE Trans. On Systems, Man and Cybernetics, Vol. 18, No.1., 1988, pp. 183-190.

Feature Extracting Correlation Filter Trained by Perceptron Learning

Tan Loc Nguyen, Chan-Il Yoo, Jung-Ja Kim, Yonggwan Won

***Abstract*—** Although the indisputable strength of neural networks, some problems could not always be classified successfully in their original feature domains. Besides, datasets could actually be analyzed in many different feature domains depending on their characteristics. Also, a classification problem defined as a complicated one in a domain could be an easier one in other feature domains. Therefore, feature domain transformation or feature extraction by converting the feature set from the original domain in to new one could be useful for better classification results. This paper introduces a new approach to use neural network to train the correlation filter that transforms the input features into a new a new domain for better classification. In this scheme, the correlation filter implemented in frequency domain works as feature transformation layer produces new feature set of power spectrum. The classification system including the correlation filter can be then trained by the back-propagation learning algorithm. An experimental result shows that a problem which cannot be solved by the traditional multilayer perceptron neural network can be solved with a high classification rate by our approach.

***Keywords*—** Correlation filter, Feature extraction, Classification, Neural network.

I. Introduction

NEURAL network is currently one of the most powerful models for classification along its quite long development history. Since the first modern neural network definitions in the 1940s with the works of Warren McCulloch and Walter Pitts [1] in providing the artificial neurons abilities in computing any arithmetic or logical functions, a fundamental network had been provided with a fixed set of weights. Until 1949, Herb developed the first learning rule, which stated that the two activated neurons should provide the stronger connection between them [2]. In 1950s-1960s, perceptron theory, presented in Fig. 1 as a single later neural network, was applied to provide the weight adjustable learning algorithm. However, due to linear discriminant properties of the perceptron theory, the problems which is not linear separable could not be solved [3]. Until 1980s, the back propagation learning algorithm for the multilayer perceptron network (MLP) shown in Fig. 2 has been proven to be able to solve nonlinear separable problems [4], and it is currently considered as an essential and the most powerful system in classification problems.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Educations, Science and Technology (NRF-2013R1A2A2A04016782).

Tan Loc Nguyen and Yonggwan Won are with the School of Electronics and Computer Engineering, Chonanm National University, 77 Yongbong-ro, Buk-gu, Gwangju 500-757, Korea, (Telephone: +82-10-4513-2203; e-mail: nguyentanloc032003@gmail. com).

Chan-Il Yoo and Jung-Ja Kim are with the Department of Healthcare Engineering, Chonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do, Korea (e-mail: brubru113@naver.com).

The artificial neural network (ANN) or simply neural network (NN) is traditionally referred to the network of biological neurons in human brain or nervous system, which could transmit, learn and condition the input information according to the experiences or memories. The neural network is a set of artificial neurons or sometimes called perceptron which could be derived by mathematical or computational models, and its internal interconnections. The network can perform complicated global behaviors by changing its connections between the processing elements (neurons) and even the neuron structures. Therefore, artificial neural network is an adaptive system which could change its structures according to the external or internal information [5].

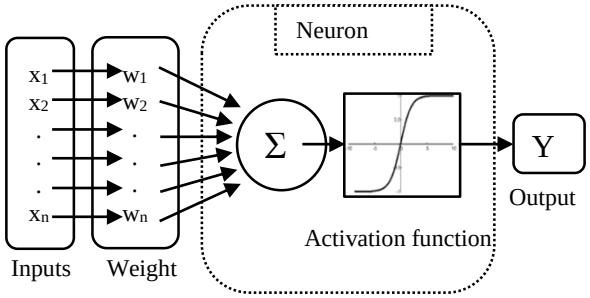


Fig.1Single layer perceptron neural network

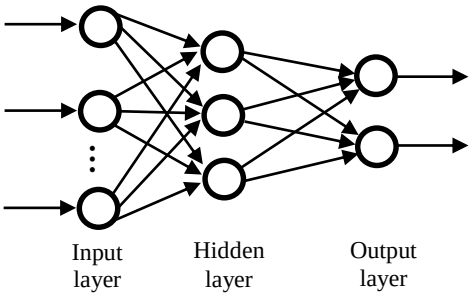


Fig.2 Multilayer perceptron (MLP) neural network

In a normal neural network, actually the more number of neurons per layer is, the better resolution for a particular feature domain is. Obviously, the resolution of feature domain could be helpful in analyzing the interesting feature properties and give more power in discriminating. In addition, the more

number of layers in network is, the more power in nonlinear classification ability the neural network can provide. Therefore, with the above definitions, the neural network could generally be able to solve most of the complex problems in reality.

However, despite of its advantages, neural network could not be able to solve some problems in which the classes are much overlapped together in the feature domain or high level of noise is added to the data. In general, the neural network performs a competitive learning process, in which the characteristics of the pattern data set would be analyzed and used to train the weights so that the largest difference between the class categories could be provided. Although increasing the size of the network and training time, sometimes, it could not provide any better result in classification rate. In other words, the classifier failed to solve such a complex problem because no competitive feature was provided in the given or intermediate feature domain. Besides, there are some signals which have similar characteristics in one particular feature domain but have big difference in another domain. A typical example is a frequency modulated (FM) signal which is difficult to distingue in time domain, but could be easy to process in frequency domain. Thus, it is possible to transform the input signal in the time domain to the frequency domain, and apply a classifier to the feature in frequency components.

This paper introduces a new multilayer feed-forward neural network which has a correlation filter layer that is followed by the output layer. The correlation filter layer performs the correlational filtering process implemented in the frequency domain, which can be considered as a feature mapping layer from the time domain features to the power spectrum component features. The filter can be trained by the back-propagation training approach in order to produce the new features with better discriminant ability by highlighting the characterized features in each class [6][7].

II. System Descriptions

A. Introduction

The classification system is composed of two cascaded layers: correlation filter and classification. The correlation filter layer actually plays as a hidden layer in ordinary neural network to pre-process the input signal. Its weights can be considered as the parameters for the filter which are the coefficients of the correlation filter and can be optimized by the iterative training process of the back-propagation algorithm [6].

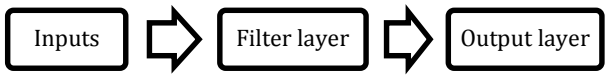
The correlation operation in time domain can be performed by the point-wise multiplication in the frequency domain. Since the correlation operation is processed in the complex domain, the output of this operation is also the complex number. This complex number output can be invertible into time domain if the correlation filter output is in conjugate symmetry. However, the iterative updating for the filter parameters cannot guarantee this requirement. Therefore, the output of the filter is transformed into its power spectrum which is a way of presenting the characteristics of the frequency components. Then, the power spectrum output is fed into the next layer which is the output layer or the next hidden layer of the normal multilayer feed-forward network.

In the back-propagation procedure for correcting the weights between layers in neural network as well as the coefficients in filter layer, error sensitivity δ and the amount of parameter correction Δ for each layer can be calculated and updated with a suitable learning rate η. We should note that the property for the correlation filter layer and the normal MLP network layer can be different, which implies the possibility of different learning rate.

B. Operations

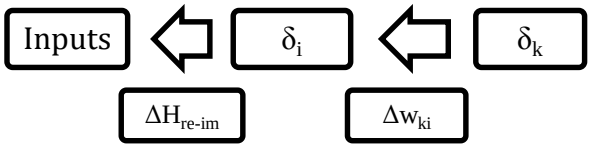
1) Correlational operation

Generally, the multilayer feed-forward neural network is composed of three layers: input layer, one or more hidden layer, and the output layer. From this normal network architecture, the first hidden layer can be substitute with the correlation filter operation layer which performs Fourier transformation of the input pattern vector, point-wise multiplication of this transformed input with the correlation filter, and computation of power spectrum [6][7]. Then, the power spectrum is fed into the next layer which is the normal feed-forward network.



2) Back-propagation

In order to train the system, the error back-propagation perceptron learning is used to correct the error in training for each of the pattern. Back propagation learning algorithm for the correlation filter layer can be derived by the chain rule as for the conventional feed-forward network layer [8][9].



In this scheme, the filter coefficients are considered as weights in normal hidden layer of neural network. However, note that the coefficients are the complex numbers.

C. Feed-forward operation

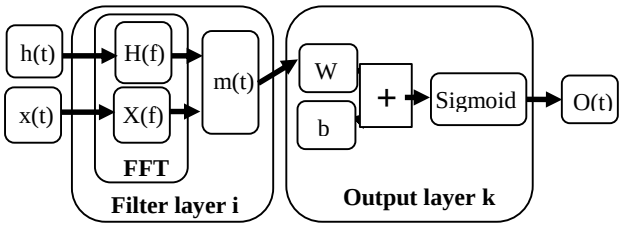


Fig.3 Filter - MLP diagram

Based on the theory of convolution and Discrete Fourier Transform properties, convolution in time domain will be point-wise multiplication in Frequency domain. In this stage, the filtering process can be done by global transform and calculation for all of the features at once.

$$x(t) * h(t) = \mathcal{F}^{-1}\{\mathcal{F}\{x(t)\} \times \mathcal{F}\{h(t)\}\} = \mathcal{F}^{-1}\{X(f) \times H(f)\} \quad (1)$$

where * denotes the correlation, $x(t)$ is the input pattern vector, $h(t)$ is the weights and their capitals are their corresponding Fourier transforms.

Power spectrum, which is the output of the correlation filter layer, of the correlation output can be computed by the following formula:

$$m_i(t) = \left(\sqrt{(X_r H_r - X_{im} H_{im})^2 + (X_r H_{im} + X_{im} H_r)^2}\right)^2 \quad (2)$$

where X_r denotes the real part of the Fourier transform of the input pattern, X_{im} does the imaginary part. In the same, H denotes the filter coefficients in the Fourier domain.

The next layer which can be another hidden layer or the output layer of the conventional feed-forward network performs the inner product (weighted sum) of the power spectrum output vector with the weight vector. The sum squared error defined in the equation 3 is then computed at the output layer, which should be minimized by adjusting the parameters.

$$error\ E = \sum_{k=1}^{c} (target_k - O_k)^2 \quad (3)$$

III. Experiments and Analysis

Several experiments have been carried out in this study to improve the current neural network architecture as well as the new network performance. The datasets were prepared to be different in each case. One of them is the one that cannot be solved by the conventional MLP, while the network with correlation filter layer produces higher classification results.

Based on the obvious difference in input dataset, the normal MLP could easily classify the pattern vectors with a short training time and a small number of neurons. In this case, because the time domain distribution of the input data can be clearly separable, the correlation filter network cannot shows any advantage for a successful classification comparing with the normal feed-forward neural network.

A. Experiment with a undistinguishable time domain data

In this experiment, dataset includes 400 samples equally divided into two groups of signals as shown in Fig.4 and the magnitudes of the signals in both cases are much similar together. The fundamental function to create the dataset is as below.

$$y(t) = \sin(\alpha t) \times \sin(2\pi f x t) + noise(t) \quad (4)$$

A sine function with various frequencies ranges x was modulated by another sine function with a constant frequency α. Each category (class) can contain a specific frequency range which is not totally overlapped together. Moreover, noise was added to the signals by directly adding a random noise with the same length of fundamental signals. This noise will ensure the complication with an overlap in time domain and will cause difficulty in normal classification.

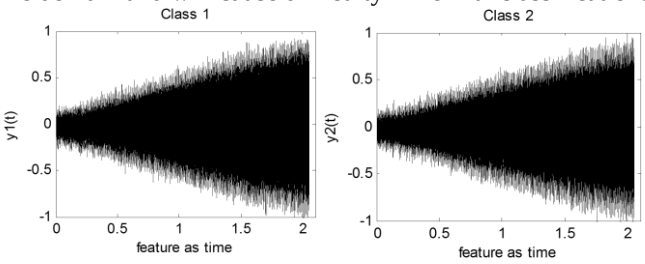


Fig.4 Input time domain representation for two classes

B. Results and Analysis

Two classification systems were used in this experimental study: the normal-MLP and correlation filter MLP. The filter MLP included 3 layers: input, filter, and output. The results are shown in the Table 1.

In this stage, the normal-MLP was failed to classify the problem. Even though the training error and classification rate were in acceptable level, the classification rate for the test data set was only below 0.6. For two class problem, the classification rate of 0.5 implies that the classifier tends to classify all patterns into one class and ignore the other, or randomly sort patterns into classes.

In fact, due to high signal to noise ratio (SNR) and high overlap between two classes in time domain, the normal MLP could not realize the competitive features to increase the discriminant power even though the size of the network was increased. The Fig.7 shows that the normal network was successful for training by reducing the sum squared error for the training data set. However, the sum squared error for the testing data which were not presented for training process was not reduced. This result shows that the normal MLP cannot be generalized for this type of classification problem.

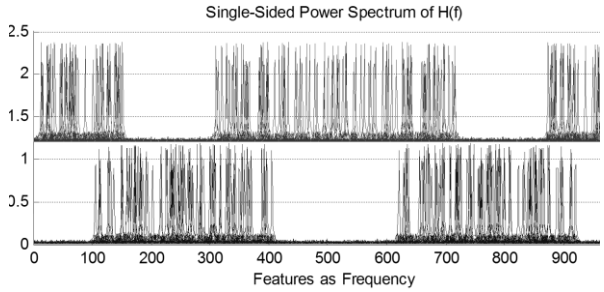


Fig.5 Input frequency domain representation for two classes

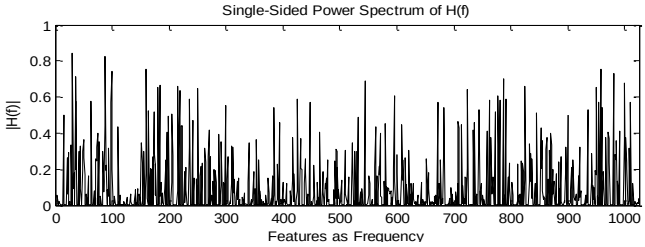


Fig.6 Single Sided Power Spectrum of the Filter Part

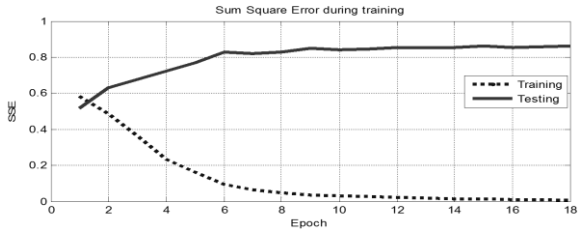


Fig7 Sum Square Error with epoch in Normal-MLP

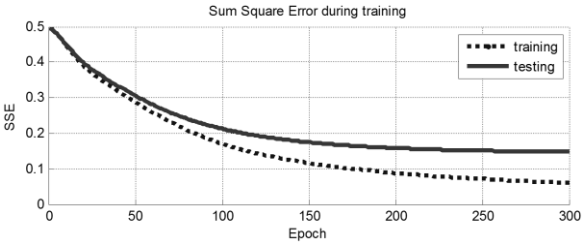


Fig.8 Sum Square Error with epoch in Filter-MLP

In contrast to the normal MLP, the correlation filter neural network produced the decreasing error values for the testing data set along with the decreasing error values for the training data set, as shown in the Fig. 8.Classification rate for testing data was achieved by 0.918 with the acceptably low SSE as shown in the Table 1.

Table 1: Result Comparison between Normal-MLP and Filter-MLP

Experiment of Time domain undistinguishing dataset		Normal MLP	Filter MLP
Epoch		18	300
Training	SSE	0.007895	0.041147
	Classification rate	0.997500	0.975000
Testing	SSE	0.863553	0.141505
	Classification rate	0.546512	0.918605

IV. Conclusion

Dataset usually contains many features that can be interesting in some domains and could be hidden in some others. Unsuccessful training in one domain may not yell to fail in solving that problem in other domain. In fact, the filter network that does not have impressive performance in time domain could be very powerful in classifying the problems in frequency domain by integrating Fourier Transform in neuron function. The network is acting as the nonlinear correlational filter that could pass all the competitive features and provide feature extraction for further processing. This is one of the most essential demonstrations for this new type of neural network.

V. Applications and Future Works

In reality, there are many kinds of signals and applications outside time domain could be done with this type of neural network. In zoology, cepstrum or frequency varying can also classify animal sounds. Besides, due to the less sensitive with noise property, frequency modulation or frequency encoding signals can be used for transmission or communication. These are some of the promising applications for this type of neural network.

In future, a comparison of this network's advantages to others can be done to analyze more details about how it works. At that time, the network definition can be used and open wider applications for this network.

References

[1] W. S. McCulloch and W. Pitts, "*A Logical Calculus of the Ideas Immanent in Nervous Activity*", Bulletin of Mathematical Biophysics, Vol.5, pp. 115-133, 1943.

[2] D. O. Hebb, *The Organization of Behavior.* Wiley & Sons, New York: 1949.

[3] F. Rosenblatt, "*The Perceptron: A probabilistic model for information storage and organization in the brain*," psychological Review, vol.65, Chap 1-5, pp. 386-408, 1958.

[4] D. E.Rumelhart and J. L. McClelland, eds, "*Parallel Distributed Processing: Explorations in the Microstructure of Cognition*", vol.1, Cambridge, MA:MIT Press, 1986, Chap 1, 11, 14, 19.

[5] M.T. Hagan, H.B. Demuth, M.Beale, "*Neural Network Design*", PWS Publishing Company, MA, 1996.

[6] Y.G. WON, "*Nonlinear Correlation Filter and Morphology Neural Networkd for Image Pattern and Automatic Target Recognition*", Ph.D. Dissertation, Faculty of the Graduate School, University of Missouri, Columbia, USA, 1995

[7] S.T. Bow, *Pattern Recognition and Image Preprocessing*, 2nd edition, Basel, NY, 2002, ISBN: 0-8247-0659-5, pp. 170-195.

[8] R. O. Duda, P. E. Hart and D. G.Stork, *Pattern Classification*, 2nd Edition, Wiley-Interscience Publication, 2000, Chap 5, 6. ISBN: 0-471-05669-3.

[9] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 2nd edition, Prentice Hall, 2002, ISBN: 0-130-94650-8, Chap 4.

[10] S. Haykin, *Neural Network: A comprehensive foundation*, 2nd edition, Prentice Hall, Chap 6.

[11] R.J. Schalkoff, *Artificial Neural Network*, Computer Science Series, MacGraw-Hill, Chap 6-7.

Mass segmentation in mammograms using energy minimization and proposed a method for classification and detection the lesion.

Khalid El Fahssi, Abdenbi Abenaou, Said Jai-andalouss, Abderrahim Sekkaki

***Abstract*—Breast mass segmentation in mammography plays a crucial role in Computer-Aided Diagnosis (CAD) systems. We propose in this article a method for the segmentation of mammography images. We use a combined solution of the two approaches; one based on levels set theory and the other based on the principle of the minimization of the energy of active contours. The elaborated algorithm which is an interactive approach permits to segment effectively the images from its coarse resolution to its refinement. The results of the segmentation are illustrated from the database of MIAS.Finally we propose a method for classification and detection of the lesion.**

***Index Terms*—segmentation; mammography; active contours; levels set;classification**

I. Introduction

Breast cancer is the second cause of death among women in the world [1], [2], [3]. Prevention of the disease is very difficult because the risk factors are either poorly understood or not influenced. Scientific studies have provided insight into the development of cancer, but it is not yet clear why such a person develops breast cancer. It should be noted that only 5-10% of breast cancers are hereditary origin related to the transmission of deleterious genes most commonly implicated are BRCA1 and BRCA2 (acronyms for breast cancer) associated with a predisposition to the disease[4]. Due to its delayed diagnosis, that often results into a heavy treatment, mutilating and costly which is accompanied by a high mortality rate. Various studies have confirmed that this is the early stage detection which can improve the prognosis and mammography in this case it is the best diagnostic technique [5]. However, it remains difficult for expert radiologists to provide a good interpretation of shots mammography due to low differences in densities of various tissues of the breast in the image. The interpretation of mammograms by radiologists is performed in the aim of finding the anomalies that indicate the changes associated with cancer. The indicators of probable presence of cancer are the microcalcifications and the opacities [6]. We analyse the mammography for the purpose of finding these indicators to characterize and classify Benign type and malignant, thus the importance of tools of aid to diagnostic (CAD) developed during these last years [7]. Generally, CAD systems of aid to diagnostic based on a range of approaches including the steps of pre-processing, segmentation, extraction of classification parameters (size, shape, density and grouping in hearth) and finally the classification of anomalies suspect [7], [8], [9] and [10]. As general the segmentation of image mammography is a very important step in solving the problem of detecting breast cancer. Currently there are many different methods of segmentation, which are: the segmentation based on regions [11], [12], [13], the segmentation based on contours [14], [15], [16] and the segmentation with thresholding [17], [18], [19]. The analysis of these diversity methods allows concluding that the application of such a method or another can influence on rates of certainty of diagnostic. Given the low contrast and the highly textured nature of the images mammography, all segmentation methods proposed depend on the values of selected parameters (thresholds, mean values, variances, etc..) And the models exploited by these methods (probability of density, function of belonging ...). Therefore, a low error of estimation of these parameters or these models can lead to segmentation results of poor qualities in terms of error rate at the pixel level. This bad estimation may also jeopardize the detection of small parts contained in the mammography images. So, this opens in a path to propose another method that can lead to a correct interpretation. In this article we propose a method for segmentation the image mammography based on theory of Level Set and the principle of the minimization of the energy of active contours.

This paper is organized as follows: Section II.A describes the database used for evaluation, section II.B we give a reminder on the theory of level sets, section II.C and section II.D present the formulation of problem and method proposed for minimizing the energy for the segmentation of mammography image.Section II.E present the segmentation algorithm and we proposed in section III method for classification and detection of the lesion. The result is presented in section IV and we end with a conclusion and perspective in section V.

II. Materials and Method

A. Database

The mini-MIAS [20] database contains a total of data 322 MLO view mammography images. This database is divided into categories such margin: speculated, circumscribed or poorly defined. The images have a resolution of 1024 1024 pixels. From this data set, a total of 111 lesions was selected. These include 60 benign and 51 malignant masses. An example of a series of the image is given by Figure 1.

Khalid El Fahssi, Said Jai-Andaloussi and Abderrahim Sekkaki are with LIAD Labs, Casablanca, Kingdom of Morocco.; elfahssi@etude.univcasa.ma

Khalid El Fahssi, Said Jai-Andaloussi and Abderrahim Sekkaki are with Faculty of science Ain-chok, Casablanca, Kingdom of Morocco.

Abdenbi Abenaou are with ENSA, Agadir, Kingdom of Morocco.

ISBN: 978-1-61804-233-0

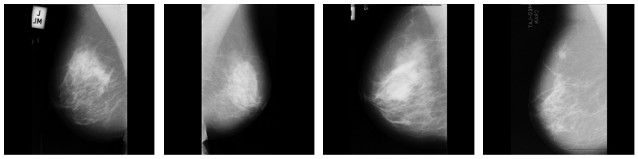


Fig. 1. Example of mammography study.

B. Level Set Theory

A level set is a vital category of deformable models. Level set theory, a formulation to implement active contours was proposed by Osher and Sethian [21]. They represent a contour implicitly via two dimensional Lipchitz continuous function $\varphi(x,y) : \Omega \rightarrow \mathbb{R}$, defined in the image plane. The function $\varphi(x,y)$ is called level set function, and a particular level, usually the zero level of $\varphi(x,y)$ is defined as the contour, such as

$$\mathcal{C} = \{(x,y) : \varphi(x,y) = 0\}, \forall (x,y) \in \Omega \tag{1}$$

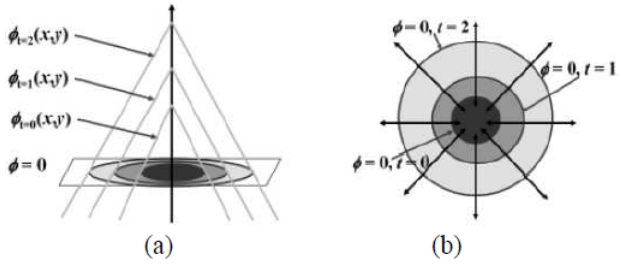


Fig. 2. Level set evolution and the corresponding contour propagation: (a) Topological view of level sets $\varphi(x,y)$ evolution (b) the changes on the zero level sets

Fig.2 (a) shows the evolution of level set function $\varphi(x,y)$ and Fig.2 (b) shows the propagation of the corresponding contour $\mathcal{C}$. As the level set function $\varphi(x,y)$ increases from its initial stage, the corresponding set of contours $\mathcal{C}$, propagate towards outside. With this definition, the evolution of the contour is equivalent to the evolution of the level set function,i.e. $\frac{\partial \mathcal{C}}{\partial t} = \frac{\partial \varphi(x,y)}{\partial t}$ the advantage of using the zero level set is that a contour can be defined as the border between positive and negative areas, so the contours can be identified by just checking the sign of $\varphi(x,y)$ the initial level set function $\varphi_0(x,y) : \Omega \rightarrow \mathbb{R}$ may be provided by the signed distance from the initial contour such as

$$\varphi(x,y) = \{\varphi(x,y) : t = 0\} = \pm D((x,y), N_{xy}(\mathcal{C}_0)) \tag{2}$$

Where $\pm D(a,b)$ denotes a signed distance between a and b, and $N_{xy}(\mathcal{C}_0)$ denotes the nearest neighboring pixel on initial contours $\mathcal{C}_0 = \mathcal{C}\ (t = 0)$ from (x,y) as a pixel (x,y) is located further inwards from the initial contours $\mathcal{C}_0$. The initial level set function is zero at the initial contour points given by $\varphi_0(x,y) = 0, \forall (x,y) \in 0$.

The deformation of the contour is generally represented in a numerical form as a partial differential equation. A formulation of contour evolution using the magnitude of the gradient of $\varphi(x,y)$ was initially proposed by Osher and Sethian [22] and is given by:

$$\frac{\partial \varphi(x,y)}{\partial t} = |\nabla \varphi(x,y)| = (\vartheta + \varepsilon k(\varphi(x,y))) \tag{3}$$

Where ϑ denotes a constant speed term to push or pull the contour,k denotes the mean curvature of the level set function $\varphi(x,y)$ and ε controls the balance between the regularity and robustness of the contour evolution.

C. Image Model and Problem Formulation

To cope with intensity inhomogeneity in image segmentation, we formulate our method based on an image model which describes the composition of images of the real-world, in which inhomogeneity of intensity is assigned to a component of an image. In this paper, we consider the following inhomogeneity intensity multiplicative model. From the physics of imaging in a variety of modalities (Mammography images), an observed image can be modeled as

$$I = bJ + n \tag{4}$$

Where J is the true image, b is the component that accounts for the intensity inhomogeneity, and n is additive noise. The component b is designated as a bias field (or shading image). The real image J measures an intrinsic physical property of the objects to be imaged, which is therefore assumed to be piecewise (approximately) constant. The bias field b is assumed to be slowly varying. The additive noise n can be considered to be zero-mean Gaussian noise.

In this paper, we consider the image $I : \Omega \rightarrow \mathbb{R}$ as defined on a continuous domain function. The assumptions about the true image and the bias field can be stated more precisely as follows:

- b Bias field varies slowly, which means that b can be well approximated by a constant in a neighborhood of each point in the image domain. (a1)
- The real image J approximately takes about N distinct values constant $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_N$ in disjoint regions $\Omega_1 ....\Omega_N$ respectively, where $\{\Omega_i\}_{i=1}^N$ forms a partition of the image domain, i.e $\Omega = \bigcup_{i=1}^N \Omega_i$ and $\Omega_i \bigcap \Omega_j = \emptyset$ for $i \neq j$.(a2)

Based on the model in (4) and the assumptions a1 and a2, we suggest a method for estimating regions $\{\Omega_i\}_{i=1}^N$, the constants $\{\mathcal{C}_i\}_{i=1}^N$, and the bias field $\hat{b}$. The resulting estimates of them are appointed by $\{\Omega_i\}_{i=1}^N$, the constants $\{\mathcal{C}_i\}_{i=1}^N$, and the bias field $\hat{b}$, respectively. The obtained bias field $\hat{b}$ should be slowly varying and the regions $\Omega_1 ...\Omega_N$ should satisfy certain regularity property to avoid spurious segmentation results caused by image noise. We will define a criterion for the search of this estimates based on the model and assumptions of the above image a1 and a2. This criterion will be defined in terms of the region, constants, and function, as energy in a variational framework, which is minimized to find the optimal regions $\{\Omega_i\}_{i=1}^N$, constants $\{\mathcal{C}_i\}_{i=1}^N$, and bias field $\hat{b}$. Consequently, the image segmentation and bias field estimation are simultaneously achieved.

ISBN: 978-1-61804-233-0

D. Energy Formulation

Let Ω be the image domain, and $I : \Omega \to \mathbb{R}$ be a gray level image. In [23], a segmentation of the image I is carried out by finding a contour $\mathcal{C}$, which divides the area of the image domain into disjoint regions $\Omega_i ... \Omega_N$, and a piecewise smooth function u that approximates the image I and is smooth within each region Ω_i. This can be formulated as a problem of minimizing the following Mumford-Shah functional

$$F^{MS}(u, \mathcal{C}) = \int_{\Omega} (I - u)^2 dx + u \int_{\Omega/\mathcal{C}} \nabla u^2 dx + \vartheta |\mathcal{C}| \quad (5)$$

Where $|\mathcal{C}|$ is the length of the contour $\mathcal{C}$.In the right part of (5), the first term is the term data, which requires u to be close to the image I, and the second term is the term of smoothness, which causes it to be smooth in each of the regions separated by the contour $\mathcal{C}$. The third term is introduced to regularize the contour $\mathcal{C}$.

To be written in a continuously the local intensity property combination indicates that the intensities in the neighborhood o_y can be classified into N clusters, with centers $m_i \cong b(y)\mathcal{C}_i$, $i = 1...N$, this allows to apply the K-means clustering standard to classify these local intensities.More specifically, the intensities $I(x)$ in the neighborhood o_y, the K-means algorithm is an iterative process to minimize the clustering criterion [24], which can form as

$$F_y = \sum_{i=1}^{N} \int_{o_y} |I(x) - m_i|^2 u_i(x) dx \quad (6)$$

We can rewrite as

$$F_y = \sum_{i=1}^{N} \int_{\Omega_i \bigcap \Omega_y} |I(x) - m_i|^2 dx \quad (7)$$

Therefore, we define the energy

$$\varepsilon \triangleq \int \varepsilon_y dy \quad (8)$$

i.e

$$\varepsilon = \int (\sum_{i=0}^{N} \int_{\Omega_i} k(y - x) |I(x) - b(y)\mathcal{C}_i|^2 dx) dy \quad (9)$$

In this paper, omit the domain Ω in the index integral symbol (as in the first integral above) if the integration is over the entire domain Ω the mammography Image segmentation by minimizing energy with respect to the region $\Omega_i ... \Omega_N$, constants $\mathcal{C}_i ... \mathcal{C}_N$ and bias field b. The kernel function K is chosen as a truncated Gaussian function defined by

$$K(u) = \{ \begin{matrix} \frac{1}{a} e^{-\frac{|u|^2}{2\sigma}} \\ 0, otherwise \end{matrix} \quad (10)$$

Where a is a normalization constant such that $\int K(u) = 1$, σ is the standard deviation (or the scale parameter) of the Gaussian function.

Our suggested energy in (9) is expressed in terms of the regions $\Omega_1 ... \Omega_N$. It is difficult to draw a solution to the energy minimization problem from this expression. In this section, the energy ε is converted to a level set formulation by representing the disjoint regions $\Omega_1 ... \Omega_N$ with a number of level set functions, with a regularization term on these level set functions. In the level set formulation, the energy minimization can be solved by using well-established variational methods [25].

Note that for a given $\mathcal{C}$, there is more than one possible level set representation: if φ is a level set function for Ω. φ Can only represent two regions, inside and outside the contour $\mathcal{C}$, as $\Omega_1 = inside(\mathcal{C}) = \{\varphi > 0\}$ and $\Omega_2 = outside(\mathcal{C}) = \{\varphi < 0\}$, respectively. In this case, a level set function is used to represent the two regions Ω_1 and Ω_2. The regions Ω_1 and Ω_2 can be represented with their membership functions defined by

$$TwoPhase\{ \begin{matrix} M_1(\varphi) = H(\varphi) \\ M_2(\varphi) = 1 - H(\varphi) \end{matrix} \quad (11)$$

Where $H(\bullet)$ is the Heaviside functional, φ represents the set of the level set functions such that $\varphi = \varphi$ for the Two-Phase model and $\varphi = \{\varphi_1, \varphi_2\}$ for the Four-Phase model.

$$FourPhase \begin{cases} M_1(\varphi) = H(\varphi_1)H(\varphi_2) \\ M_2(\varphi) = H(\varphi_1)(1 - H(\varphi_2)) \\ M_3(\varphi) = (1 - H(\varphi_1))H(\varphi_2) \\ M_4(\varphi) = (1 - H(\varphi_1))(1 - H(\varphi_2)) \end{cases} \quad (12)$$

Where

$$H(\varphi) = [\frac{1}{2}[1 + \frac{2}{\pi} arctan(\frac{\varphi}{\epsilon})]] \quad (13)$$

Thus, for the case of $N \geq 2$, the energy in (9) can be re-written as:

$$\varepsilon \triangleq \int (\sum_{i=0}^{N} k(y - x) |I(x) - b(y)\mathcal{C}_i|^2 M_i(\varphi(x)) dx) dy \quad (14)$$

E. Algorithm

Enter :

I Initial Image

μ Regularization parameter

$\triangle t$ the Pas in time

γ Proportionality constant

$IterNum$ Number of iterations

υ speed constant

Initialisation :

$$\varphi_0 = \begin{cases} +1 \ where \ (x, y) \ is \ inside \ C, \\ -1 \ where \ (x, y) \ is \ outside \ C. \end{cases}$$

estimate :

For n ranging from 1 to IterNum do

. - Calculate the contour carte by Algorithm of Canny

. - Calculate the gradient carte $\varphi(x, y)$

. $\varphi(x, y) = \sqrt{\sum_{s \in \eta_s} |I_p - I_s|^2}$

. - Calculate the gradient g

. For i ranging from 1 to nl do

. For j ranging from 1 to nc do

. $g(\varphi(x, y)) = \frac{1}{1 + \varphi(x, y)}$

. - Calculate the mean curvature term K

. For i ranging from 1 to nl do

. For j ranging from 1 to nc do

. $K(i, j) = div(\frac{\nabla \varphi(i,j)}{|\nabla \varphi(i,j)|})$

. - Calculate $\varphi^{n+1}(i,j)$

$$\varphi_{i,j}^{n+1} = \varphi_{i,j}^{n} + \triangle t \left[-\gamma\delta_\varepsilon(\varphi)(c_1 - c_2)\left(I(\varphi) - \frac{c_1+c_2}{2}\right)\right.$$

$$\left. + \mu\left(\triangle\varphi - div\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right)\right) + \upsilon\delta_\varepsilon(\varphi) div\left(g.\frac{\nabla\varphi}{|\nabla\varphi|}\right)\right]$$

. -end.

where $c_1 = \frac{K_\sigma * [H_\varepsilon(\varphi) I]}{K_\varepsilon * H_\varepsilon(\varphi)}$ and $c_2 = \frac{K_\sigma * I - K_\sigma * [H_\varepsilon(\varphi) I]}{K_\sigma * 1 - K_\sigma * H_\varepsilon(\varphi)}$
(1 is a function with a constant value of 1).
$H_\varepsilon(\varphi)$ is the Heaviside function and e_i functions given by

$$e_i = \int K_\sigma(y-\varphi)|I(\varphi) - c_i(y)|^2 dy \quad , \quad i = 1,2$$

σ_ε is the Dirac function smoothed given by

$$\sigma_\varepsilon(\varphi) = H'(\varphi) = \left[\frac{1}{2}[1 + \frac{2}{\pi} arctan(\frac{\varphi}{\varepsilon})]\right]' = \frac{1}{\pi}.\frac{\varepsilon}{\varepsilon^2 + \varphi^2}$$

III. The Proposed Method for Classification

The overall content of this paper was devoted to segmentation the mammograms image. The proposed method has facilitated extracting lesion areas of various classes of mammograms images that are considered an essential step in diagnosis. To complete the procedures in CAD system, the last step is to set up an automatic identification system of regions in different classes. The idea consisted, first time, calculating the statistical characteristics vectors (histogram or other) from the extracted regions during the segmentation. To ensure a higher identification rate, it is proposed thereafter to use a method of classification based on adaptive transformation [26] statistical characteristic vectors (given the size of the images and regions is important) into vectors informative characteristics of minimum dimension. This will facilitate the task thereafter to develop a decision rule with high certainty.

IV. Results and Discussion

This section presents the results of experiments performed on 111 selected mammography database mini-MIAS images. To evaluate the effectiveness of the proposed method, we used the information provided by the Mammographic Image Analysis Society database (MIAS) including the class of anomalous images and coordinates of their centers of regions of interest. To test our method, the algorithm is applied to each image containing a mass lesion, then verified if the algorithm detects a region of interest containing the abnormally and segment the mass, finally calculates the percentage of abnormally detection on all cases. The example for result is presented a figure 3.

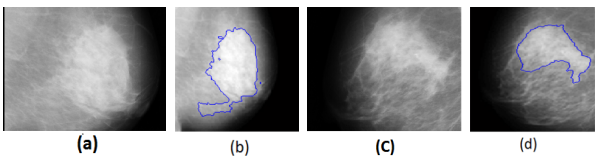


Fig. 3. (a) The original image with Calcification benign. (b) Detection the region of interest (contains the abnormally) (c) The original image with Calcification malignant. (d) Detection the region of interest (contains the abnormally)

The result of experiment carried on the 111 mammogram image to detect the region of interest that contains the disease and segment the mass gives the percentage of precision equal to 92.27%.

V. Conclusion and Perspective

In this work we presented an approach for segmentation of masses in digital mammograms. This approach consists of three main stages which are the (ROI) detection, edge extraction and mass segmentation. The experimentation gives a percentage of 92.27% for all cases studied. The results of the algorithm can contribute to solving the main problem in mammography image processing such as diagnostic and classification. The Efficiency of the proposed method confirms the possibility of its use in improving the computer-aided diagnosis.

References

[1] National Cancer Institue of Canada, canadian cancer statistics 2011, available at http://www.cancer.ca/~/media/cancer.ca/cw/cancer%20information/cancer%20101/canadian%20cancer%20statistics/canadian-cancer-statistics-2011-en.pdf

[2] Alireza Shirazi Noodeh, Hossein Rabbani , Alireza Mehri Dehnavi, Hossein Ahmadi Noubari , Detection of cancerous zones in mammograms using fractal modeling and classification by probabilistic neural network, proceedings of the 17th iranian conference of biomerucal engineering (icbme2010), 3-4 november 2010.

[3] Association Lalla Salma de lutte contre le cancer, guide de dtection prcoce des cancers du sein et du col de lutrus edition 2011, available at $http://srvweb.sante.gov.ma/documents/guide_pratique_v3.pdf$

[4] Pierre Kestener, Analyse Multifractale 2d et 3d a l'aide de la Transformation en Ondelettes : Application en Mammographie et en Turbulence Developpee soutenue le 21 novembre 2003, l'universite bordeaux i, ecole doctorale de sciences physiques et de lingenieur

[5] Songyang Yu and Ling Guan, A Cad System for the automatic detection of clustered microcalcifications in digitized mammogram films, ieee on medical imaging, vol. 19, n. 2, february 2000, pp. 115-126

[6] G.kom, A.tiedeu, M.kom, C.nguemgne, J.gonsu : Detection automatique des opacites dans les mammographies par la methode de minimisation de la somme de linertie, itbm-rbm 26 (2005) 347356,1297-9562/$ - see front matter 2005 elsevier sas

[7] Rangaraj M.Rangayyan, Fabio J.Ayres, J.E.Leo Desautels, A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs, sciencesdirect , journal of the franklin institute 344 (2007) 312348

[8] T.Balakumaran, Dr.Ila.Vennila, C.Gowri Shankar, Detection of microcalcification in mammograms using wavelet transform and fuzzy shell clustering, (ijcsis) international journal of computer science and information security,vol. 7, no. 1, 2010

[9] Arianna Mencattini, Marcello Salmeri, Giulia Rabottino, Simona Salicone, Metrological characterization of a cadx system forthe classification of breast masses in mammograms , ieee transactions on instrumentation and measurement, vol. 59, no. 11, november 2010.

[10] Leila Bahreini, Emad Fatemizadeh, Masoumeh Gity, Gradient vector flow snake segmentation of breast lesions in dynamic contrast-enhanced mr images, proceedings of the 17th iranian conference of biomedical engineering (icbme2010), 3-4 november 2010, 978-1-4244-7484-4110/ 2010 ieee.

[11] Arianna Mencattini, Giulia Rabottino, Marcello Salmeri, Roberto Lojacono, Emanuele Colini Breast mass segmentation in mammographic images by an effective region growing algorithm, acivs 2008,lncs 5259, pp 948-957, 2008 @ springer -verlag berlin heidelberg 2008.

[12] Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, Syed Khaleel Ahmed Automated breast profile segmentation for roi detection using digital mammograms ,ieee embs conference on biomedical engineering and sciences (iecbes 2010), kuala lumpur, malaysia, 30th november-2nd december 2010.

[13] Aminah Abdul Malek, Wan Eny Zarina, Wan Abdul Rahman, Arsmah Ibrahim, Rozi Mahmud, Siti Salmah Yasiran, Abdul Kadir Jumaat Region and boundary segmentation of microcalcifications using seed-based region growing and mathematical morpholog, procedia - social and behavioral sciences, volume 8, 2010, pages 634-639, 1877-0428 2010 published by elsevier ltd.

[14] Cascio, D.Fauci, F.Bagnasco Mammogram segmentation by contour searching and massive lesion classification with neural network nuclear science, ieee transactions on (volume:53 , issue: 5), 16 october 2006

[15] Tolga Berbera, Adil Alpkocak, Pinar Balci, Oguz Dicle Breast mass contour segmentation algorithm in digital mammograms. computer methods and programs in biomedicine, volume 110, issue 2, may 2013, pages 150-159, 0169-2607/$see front matter 2012 elsevier ireland ltd

[16] Qaisar Abbas, M.Emre Celebi, Irene Fondn Garcia Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system,biomedical signal processing and control, volume 8, issue 2, march 2013, pages 204-214. 1746-8094/$see front matter 2012 elsevier ireland ltd

[17] R. C. Gonzalez and R. E.Woods, Digital image processing. 2ed, prentice hall, 2002.

[18] P.S.Liao, B.C.Hsu, C.S.Lo, P.C.Chung, T.S.Chen, S.K.Lee, L.Cheng and C.I.Chang Automatic detection of microcalcifications in digital mammograms by entropy thresholding , engineering in medicine and biology society, bridging disciplines for biomedicine, vol. 3, pp.1075-1076, 1996.

[19] J.Mohanalin, P.K.Kalra and N.Kumar Tsallis entropy based contrast enhancement of microcalcifications, international conference on signal acquisition and processing, icsap, pp. 3-7, 2009.

[20] http://peipa.essex.ac.uk/info/mias.html

[21] S.Osher, J.A Sethian, Fronts propagating with curvature dependent speed: algorithms based on hamitonjacobi formulations, journal of computer physics. vol. 79 (1), 1988, pp. 12 49.

[22] J.Sethian, Theory, algorithms, and applications of level set methods for propagating interfaces acta numerica, 1996, pp. 309 395.

[23] D.Mumford and J.Shah, Optimal approximations by piecewise smooth functions and associated variational problems, commun. pure appl. math., vol. 42, no. 5, pp. 577685, 1989.

[24] S.Theodoridis and K.Koutroumbas, Pattern Recognition. Newyork, Academic, 2003.

[25] L.Evans, Partial Differential Equations. Providence, RI: Amer. Math. Soc., 1998.

[26] Abenaou A., Sadik M. 'Elaboration d' une méthode de compression des signaux aléatoires base d' une transformation orthogonale paramétrable avec algorithme rapide' 'WOTIC'2011, ENSEM,Casablanca.

A novel method for localization of rotor-stator rub in steam turbines

Jindrich Liska, Jan Jakl

***Abstract*—**The rotor-stator rub has become a pretty common phenomenon in terms of steam turbines operation in last couple of years. The reason is that clearances in flow paths of the turbines get smaller and smaller along efforts towards efficiency increase. This paper introduces a novel method based on vibration signal processing to localize rotor-stator rub in rotating machinery. The presented method utilizes the standard sensor installation in several measurement planes of the steam turbine. To investigate the behavior of the rub induced signals a Rayleigh-Lamb description of the stress waves propagation is discussed. This paper covers not only the introduction of procedures for rub signal pre-processing but also the method of automatic wave arrival time determination. The proposed localization method was tested during experiments on rotor stand and also in operation of steam turbine. Finally the accuracy of localization results was verified after an inspection of the steam turbine during the outage.

***Keywords*—** Rotor-stator rub, localization, steam turbine, monitoring.

I. Introduction

During the operation of steam as well as gas turbines, in some situations, such as during the start-up of a turbine, when the rotor is overcoming its natural frequencies and its vibrations are at the highest level, undesirable rubbing between the stator and the rotor may occur („rotor-stator rub“ or „rubbing“). In the first phase of this rubbing it is especially the seals arranged between the rotating and statical parts of the turbine that are abraded, and, as a result, the amount of leaking medium increases and the turbine efficiency decreases. In cases when rotor-stator rubbing is not detected in time and is not eliminated by an appropriate intervention of the machine operator, such as by changing the running speed during the start-up or run-down, or by changing parameters of the jacking oil during the operation on a turning gear, both the rotor and the stator may be heavily damaged, or an breakdown of the whole turbine leading to considerable economic losses may occur.

This article was supported by the European Regional Development Fund (ERDF), project "NTIS – New Technologies for the Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090.

The authors also would like to contribute to the support from the internal grant of the University of West Bohemia No. SGS-2013-041.

J. Liska is with the University of West Bohemia, NTIS - European Centre of Excellence, Pilsen, Univerzitni 22, 30614 Czech Republic (phone: +042-377632521; fax: +420-377632502; e-mail: jinliska@ kky.zcu.cz).

J. Liska is with the University of West Bohemia, NTIS - European Centre of Excellence, Pilsen, 30614 Czech Republic (e-mail: jjakl@kky.zcu.cz).

The rotor-stator rubbing may be either partial, when the rubbing between the rotor and the stator is brief, but at least several times repeated, or it is full annular rubbing, i.e. a continuous or almost continuous rubbing between the rotor and the stator. Nevertheless, the full annular rotor-stator rubbing is always preceded, at least for a short time, by the partial rotor-stator rubbing, see [1].

At present, detection of the partial rotor-stator rubbing is based especially on offline analysis of vibration signals, when during the measurements the machine operator monitors the overall level of vibrations as well as the phasor of the first harmonic component of the rotational frequency in vibration signals [2]. If a step change occurs in overall vibration, or in rotation of the phasor of the first harmonic component with variable or periodically varying amplitude, rotor-stator rubbing is detected, and after terminating the measurements a detailed data analysis is carried out with the purpose of excluding the possibility of false positive detection. Therefore this approach is not suitable for detecting rotor-stator rubbing during the real operation of the turbine and can only be used for laboratory research or experimental purposes.

Another approach [3] for detecting partial rotor-stator rub is based on the fact that apart from the change in the phasor of the first harmonic component of the vibration signals, the rubbing is also accompanied by formation of subharmonic spectral components, whose frequency corresponds to the frequency of impacts of the rotor on the stator. Nevertheless, the disadvantage of detecting rotor-stator rubbing based on monitoring these subharmonic components is the fact that the frequencies of some of them are too close to frequencies indicating other defects, such as the instability of the oil film in the slide bearing, etc., which may result in false positive detection of rotor-stator rubbing, or, on the contrary, in wrong interpretation of expresses of this rubbing as defects of different type. Authors of the paper have developed a monitoring system RAMS (Rub Advanced Monitoring System) which implements methods for detection of the rotor-stator rub and triggers the data for following localization procedure [4].

In this paper a novel method for localization of rotor-stator rub is proposed and verified in the operation of steam turbines. Currently, there is only few known and in practice applicable methods for localizinf of rubbing in steam turbines. Important is the method based on the measurement of acoustic emission (AE) on the bearing pedestals described in [5]. The authors use the replica of real labyrinth seal slices to induce

rubbing on experimental stand. A special holder allows pushing the seal against the rotor with a force up to 140 N. Within this experiment, the shaft was in contact throughout the revolution, which is a special type of rub (full annular rub). Measuring acoustic emission in the frequency range from 100 kHz to 1.2 MHz the authors concluded that the full rub expresses itself as amplitude modulation of the measured signal. The smoothed RMS values were afterthat calculated from the signal using Savitsky - Golay filter. Based on the course of the smoothed RMS values it was possible to estimate the rub origin and use the linear localization to determine the contact location.

II. Linear localization principle

Linear localization principle can be applied in cases where the excitation generated by the rub event spreads from the source of excitation on a straight line in both directions and at the same velocity. If the excited impulse is measured in at least two measuring points and we are able to determine the exact moment of the arrival time to the measurement position, then the knowledge of measurement points distance, spread velocity in the material, and the difference in wave arrival times make it possible to calculate the position of the excitation source relative to one of the measurement points. When measuring in more than two points it is possible to refine the existing results or to exclude the propagation velocity parameter from the calculation. Sample linear localization using two sensors is shown in Fig. 1 (top).

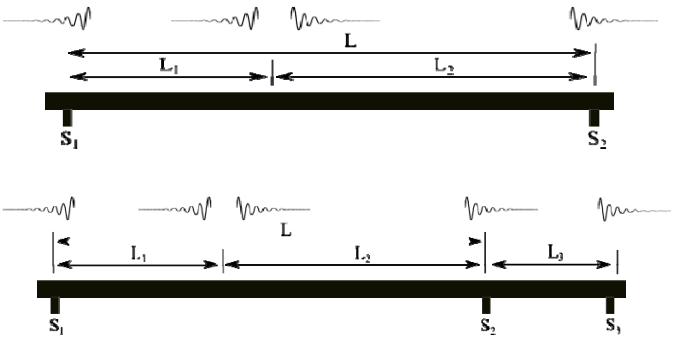


Fig. 1 Linear localization using two sensors (top) and three sensors (bottom).

Let's consider the total distance L between the sensors S_1 and S_2. Let the distances of sensors S_1 and S_2 from the excitation sources are L_1 and L_2. The time of arrival of the impulse to the individual sensors S_1, S_2 is t_1 and t_2. Then apply the following:

$$L = L_1 + L_2 \tag{1}$$

$$t_2 = t_1 + \Delta t \tag{2}$$

The equation (1) can be modified as follows:

$$L = vt_1 + v\left(t_1 + \Delta t\right), \tag{3}$$

where v is propagation velocity of the stress waves in rotor. From previous relations follows:

$$L_1 = \frac{L - v\Delta t}{2} \tag{4}$$

To determine the distance from the excitation source to one of the sensors is therefore necessary to know the distance between sensors, time difference of the event detection for both sensors and the propagation velocity in rotor material. The knowledge of the last parameter is very problematic and because the most of the rub impulses propagate through the rotor nonlinearly (dispersive waves) and the transmission path from the source to the sensor is complex, a linear localization in this case is mainly approximative. When the velocity of the rub wave propagation is properly selected and when a sufficient number of partial rub events are processed then the linear localization is successfully applicable based on a statistical evaluation of the results. The velocity evaluation of the impulse propagating from the rub origin to the sensor positions (4) depends on the method used to process the measured signals.

If we measure on the machine more than in two measuring planes, it is possible to exclude the propagation velocity from the calculation (assuming a constant impulse propagation velocity). This case is illustrated in Fig. 1 bottom.

Again, we assume that the source of excitation is located on the right side from the sensor S_1 and in the distance L_1. If it is possible to detect an event on all three sensors, then the propagation velocity could be determined from the time differences between rub impulse arrival times of sensors S_2 and S_3 (with knowledge of the distance).

$$v = \frac{L_3}{t_3 - t_2} \tag{5}$$

Thus, (4) could be modified to the following form:

$$L_1 = \frac{\left(t_3 - t_2\right)L - \Delta t L_3}{2\left(t_3 - t_2\right)} \tag{6}$$

Similar relations can be derived also for other relative positions of the excitation source and sensors. A more detailed analysis of this task is given in the following chapters.

III. Dispersive character of the stress waves propagation after rub impact

Measured rub signals mainly contain running vibrations of monitored machinery (rotor x stator). An impact pressure wave propagates from the place of origin through the wave fronts and, along its trajectory, is subject to various deformations, reflections and damping. Reflected waves and other impact waves further deform the signal. Superposition of the impact wave fronts and current running vibrations is

ISBN: 978-1-61804-233-0

measured and monitored at specific measurement planes of the turbine. The essential task of the diagnostic system based on vibration data analysis is to recognize the presence of impact wave fronts from the measurements.

An impact signal is similar to the impulse response of a slightly damped mechanical system, the so called burst shape signal. Signal reaches its maximum relatively fast (in milliseconds) but its overall length is in tens of milliseconds. With increasing distance of impact from sensor, the burst leading edge increases, signal maximum decreases and the signal-noise ratio decreases.

Vibrations propagate from the impact origin as a multimode waveform, whose damping and velocity are dependent on the dominant wave mode. Propagation velocity of a particular waveform mode is described by the so called dispersive curves (see [7] or [8]), which describe velocity dependence on frequency. It is possible to analytically describe the particular wave modes solving the Rayleigh-Lamb equations. Hence, the resulting waves are called Lamb waves. Analytical model of dispersive curves (Lamb curves) is the following:

$$\frac{\tan(qh)}{\tan(ph)} = -\left[\frac{4\kappa^2 pq}{(q^2-\kappa^2)^2}\right]^{\pm 1} \tag{7}$$

where $2h$ is the material thickness, κ is the wave number, power of +1 is used for symmetric Lamb waves and power of -1 for antisymetric ones. Variables p and q are defined as follows:

$$p^2 = \left(\frac{\omega}{c_L}\right)^2 - \kappa^2,\ q^2 = \left(\frac{\omega}{c_T}\right)^2 - \kappa^2 \text{ and } \kappa = \frac{2\pi}{\lambda}, \tag{8}$$

where ω is frequency, λ is wavelength, c_L and c_T are longitudinal and transversal wave propagation velocities for the current material. Wave number κ is thus the number of oscillation periods per a unit of length. Equation (7) can be transformed into the following form for symmetric modes

$$\frac{\tan(qh)}{q} + \frac{4\kappa^2 p \tan(ph)}{(q^2-\kappa^2)^2} = 0 \tag{9}$$

and the following form for antisymmetric modes

$$q\tan(qh) + \frac{(q^2-\kappa^2)\tan(ph)}{4\kappa^2 p} = 0. \tag{10}$$

Real roots of the above mentioned equations are especially significant. They represent the undamped propagation of the Lamb waves in the material structure. Resulting dependency of the wave number κ on the frequency f ($\omega/2\pi$) is displayed in Fig. 2. It shows the solution of (3) and (4) which was computed for a steel plate (c_L = 5900 m/s; c_T = 3100 m/s) with thickness of 10 cm. Individual modal groups are distinguished by color: blue – symmetrical; red – antisymmetrical mode.

In reality, waves in the material propagate in the form of packets (groups). Velocity of these packets is given by the so-called group velocity c_g, which is described by the following formula:

$$c_g = \frac{d\omega}{d\kappa}. \tag{11}$$

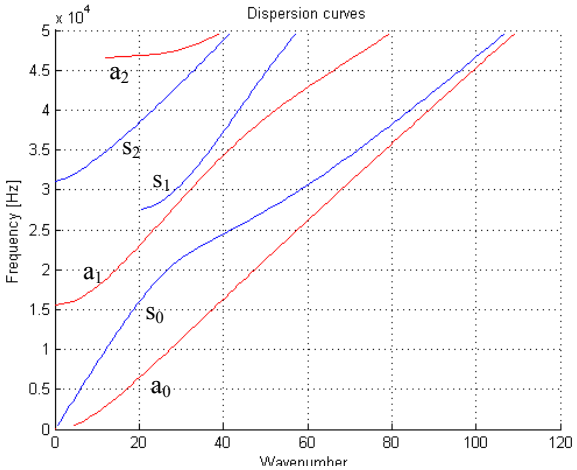


Fig. 2 Solution of (9) and (10) in wave number – frequency domain.

The dependency of the group velocity c_g on the frequency f is displayed in Fig. 3.

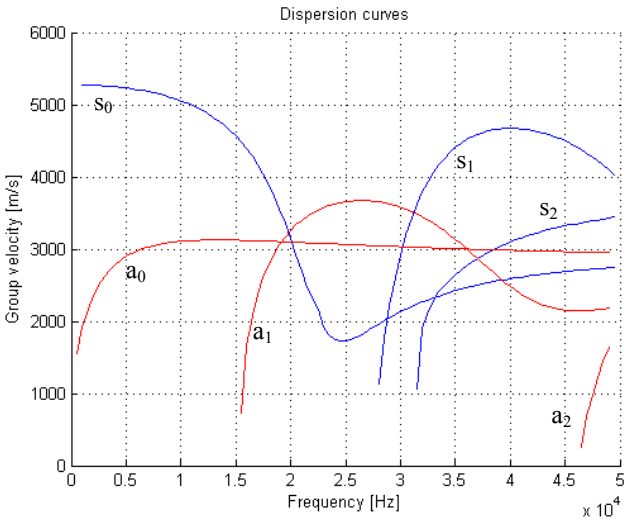


Fig. 3 Dispersion curves - representation of group velocity cg depending on frequency f.

IV. Signal Processing for Rub Localization

In the case when the contact between the rotor and stator does not occur, there are only fundamental oscilations of the rotor monitored by the sensor. These vibrations should be considered as stationary when considering constant rotation speed.

A. A method based on computation of moving variance

One possibility to determine the rub wave time of arrival is

to monitor the signal variance. Due to the nonstationary signal character during the rubbing phase an appropriate recursive algorithm for signal mean and variance estimation should be applied. If we choose a recursive least squares method with exponential forgetting, then the equations for estimated mean and variance have the following form:

$$m(t) = \lambda m(t - T_s) + (1 - \lambda) x(k) \tag{12}$$

$$s(t) = \lambda s(t - T_s) + (1 - \lambda)\left(x(t) - m(t)\right)^2, \tag{13}$$

where *m(t)* and *s(t)* denotes the estimated mean value, and variance in time *t*. The parameter λ is the forgetting factor. The relation between the forgetting coefficient and time constant of the filter is as follows

$$\lambda = 1 - \frac{1}{\tau f_s}, \tag{14}$$

where f_s denotes the sampling frequency.

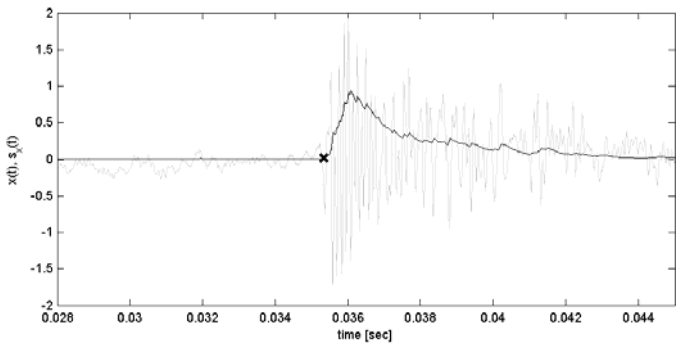


Fig. 4. An estimated variance of the measured rub signal.

In Fig. 4 there is time behaviour of the measured rub vibration signal (gray) together with the result from recursive estimation of moving variance (black). Automatically detected arrival time of the wave (the beginning of the leading edge) is shown with a cross. It can be seen that the position of the cross corresponds with the beginning of the impulse leading edge. The method of impulse arrival time detection will be presented in the following text. In cases when the effective signal of the rub impulse is masked by the machinery running vibrations, the suitable waveform filter should be applied (windowed sinc filter was succesfully tested on operational signals from steam turbine).

B. A method based on instantaneous amplitude computation

Another characteristic that describes the changes of signal oscilations is the envelope. Let's consider a compex signal *z(t)* in the following form:

$$z(t) = z_r(t) + j z_i(t) \tag{15}$$

The envelope or instantaneous amplitude is defined as the modulus of the complex signal (16)

$$A_z(t) = \left|z(t)\right|. \tag{16}$$

Definition of signal instantaneous amplitude is also important for multi-component signals (unlike the instantaneous frequency). Firstly, it is necessary to calculate the complex signal imaginary component when applied to real signals. Useful method to do this is the Hilbert transform defined as follows:

$$H\left[z_r(t)\right] = \frac{1}{P}\int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \tag{17}$$

where *P* denotes the Cauchy principal value. The real part of the analytic signal *z(t)* is formed by the measured signal $z_r(t)$ and the imaginary component is the Hilbert transform:

$$z(t) = z_r(t) + jH\left[z_r(t)\right] \tag{18}$$

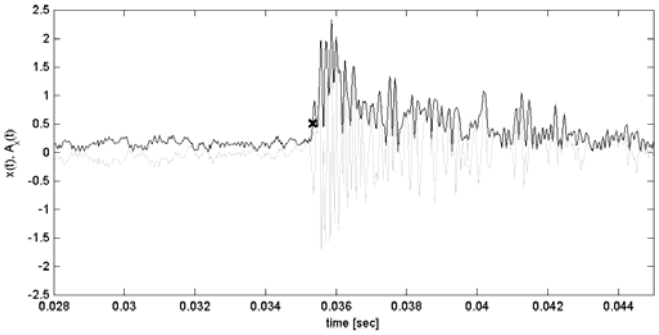


Fig. 5. Envelope of the signal based on Hilbert transform computation.

In Fig. 5, the vibration signal (gray) together with time behaviour of instantaneous amplitude is displayed. Again, the cross sign indicates the detected time of rub wave arrival.

C. A method based on signal stochastic normalization

The principle of stochastic normalization was described in [6] and the method is based on normalization of each frequency line of the amplitude spectrogram by its estimated mean and standard deviation, which are determined recursively. Originally, this method was applied to a spectrogram calculated using a discrete short-time Fourier transform. However, this method of normalization is able to normalize any of the time-frequency representations of the vibration signal. The results reported in this paper have been obtained by application of stochastic normalization on approximation of continuous Gabor transform [9]. Gabor transform is a special case of continuous short-time Fourier transform, where the Gaussian function is used as a window. Gabor transform of a signal x (t) is defined as

$$G(t, f) = \int_{-\infty}^{\infty} x(\tau) e^{-\frac{(\tau - t)^2}{\sigma^2}} e^{-j2\pi f \tau} d\tau. \tag{19}$$

Changing the parametr σ the resolution of the time-frequency spectrogram could be changed. The result of the stochastic normalization is the function $G_n(t,f)$.

$$G_n(t,f)=\frac{G(t,f)-M(t,f)}{\sqrt{S(t,f)}} \tag{20}$$

The functions $M(t,f)$ and $S(t,f)$ are determined from the following recursive computation formula:

$$\begin{aligned} M(t,f)&=\lambda M(t-T_s,f)+(1-\lambda)G(t,f)\\ S(t,f)&=\lambda S(t-T_s,f)+(1-\lambda)\left(G(t,f)-M(t,f)\right)^2 \end{aligned}, \tag{21}$$

where the parameter λ has the same meaning of forgetting factor as in the section A.

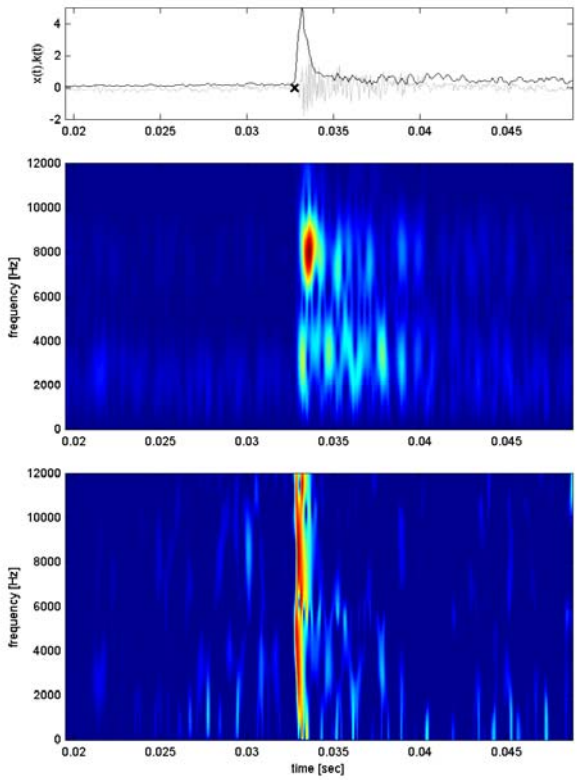


Fig. 6 Stochastic normalization of signal x(t) - time domain (top), amplitude spectrogram (middle) and normalized amplitude spectrogram (bottom).

This method uses a similar principle for detecting the impact in the signal, such as the method based on calculation of the signal moving variance, i.e. the suppression of the stationary signal components and on the contrary highlighting the nonstationary components. The difference is that the method of stochastic normalization is applied not in the time domain, but onto the time-frequency representation. The origin of the nonstationary components in the signal can be detected using the resulting k-value. It describes the overall change of the amplitude normalized spectrogram in a certain frequency band and can be defined as follows:

$$k(t)=\frac{\int_{f_1}^{f_2} G_n(t,f)\,df}{f_2-f_1} \tag{22}$$

The original rub signal with the corresponding $k(t)$ value are shown in the top part of the Fig. 6. In the middle of the Fig. 6, there is an amplitude spectrogram, where the broadband excitation associated with the rub impact is evident. In the bottom part of the figure, the normalized amplitude spectrogram calculated with parameter λ=0.995 is shown. Due to the normalization, the impact can be detected also in frequency bands where it was not significant in the former amplitude spectrogram.

D. A method based on spectral variance computation

It is also possible to use function $S(t,f)$, which was defined in (21) to detect impacts in signal. This function carries the information about origination of nonstationarity in signal. In the same manner as in section *C*, the characteristic k-variance value k_s can be introduced. This value describes the transient changes of the signal in summary.

$$k_S(t)=\frac{\int_{f_1}^{f_2} S(t,f)\,df}{f_2-f_1} \tag{23}$$

The resulting k_s value is shown in top part of the figure 7. The spectral variance representation is displayd in the bottom part of the figure, i.e recursively estimated variance behaviour on each frequency of the spectrogram.

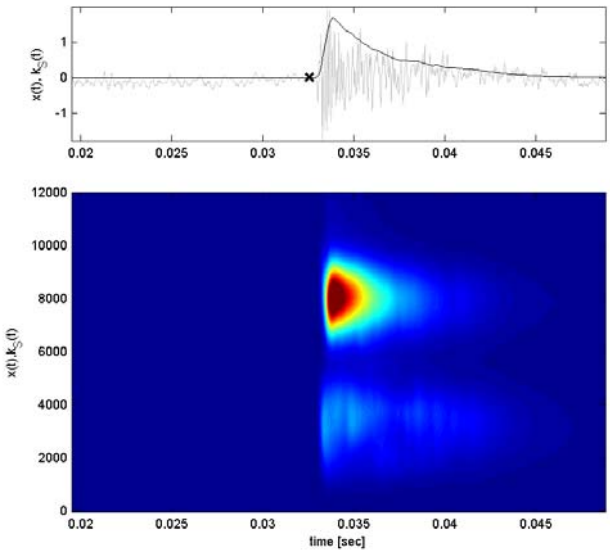


Fig.7 Spectral variance of the signal.

In above sections, there were described and discussed some of the proposed methods for the arrival time determination of the rub induced impacts covered in the measured vibration

signals. The method based on variance and envelope computation operate only in time domain and their use is conditional upon the rub impulse amplitude exceeds the value of the overall vibrations. The remaining two methods are based on the signal processing in the time- frequency domain, which increases their usefulness also in the case of arrival time detection of weak impacts. In general, these methods process changes and attributes of signals that can be used to localize the source of impacts. As allready discussed, all these methods filter the signal by the certain manner and hence the estimated propragation velocity of the stress waves in the rotor can slightly vary according to the used method. Using the measured signal from only two sensors is then necessary to compute the velocity analyticaly from the Rayleigh-Lamb equations described in chapter III. When three or more sensors with sufficient amplitude increase are available, the propagation velocity could be calculated directly from the measured signals (see chapter II.)

E. Determination of the wave arrival time

To identify the wave arrival time in the predefined frequency bands, a method based on a comparison of signal characteristics in two consecutive sliding windows of length N is used. It is necessary to identify the local signal change, therefore two windows were selected, whose ratio shows changes in local properties. Consider a signal *s* and its *n-th* sample. A quantity *w*, which characterizes changes in the signal, is determined by dividing signal part between samples *n* and *n+N-1* to the previous signal part from *n-N* to *n*. The wave arrival time of rub impacts are characterized by an increase in the signal amplitude, therefore the absolute value of the signal difference was chosen as an indicator of the increase in amplitude.

$$w(n) = \frac{\sum_{i=n}^{n+N-1} |s(i) - s(i-1)|}{\sum_{i=n-N}^{n-1} |s(i) - s(i-1)|} \tag{24}$$

Quantity *w* thus displays the ratio of signal amplitude changes for two consecutive signal parts.

The whole localization algorithm is used for quasi offline data processing after event detection, therefore noncausality of the wave arrival time detection method is not a limitation for the localization algorithm. Fig. 8 plots the vibration amplitude in a selected frequency band (black curve) and the time behavior of the *w*-value (red curve). At the time of the *w*-value maxima, the position of both running windows is marked.

Already from the calculation of the structure of the *w*-value in (24), it is clear, that at the time of wave arrival the *w*-value will take its local maximum. Due to signal stationarity before the rub wave arrival and its relatively low damping, the appropriate choice of window length *N* can also guarantee that the wave front peak yields the global maximum in the *w*-value. In practice, the window length of 1 ms (i.e. $N = 100$) was proved for the sampling frequency of 50 kHz. The advantage of this method is its adaptive nature in determining the beginning of the rub event, which is successfully used in determining the wave arrival time at different frequencies in the time-frequency domain.

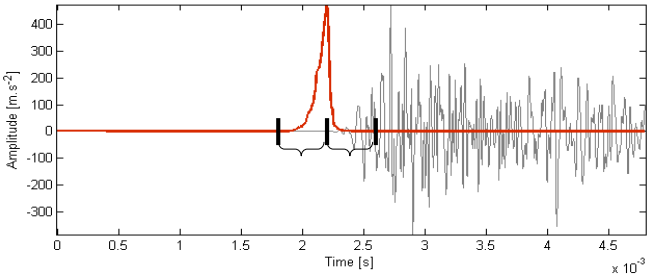


Fig. 8 Time behavior of w-value (red) - wave arrival time detection in selected frequency band vibrations (black).

V. Experimental Rotor Stand and Test Verification

The experiments on the rotor stand have been part of the research and development of the rub localization method. The rotor kit RK4 by Bently Nevada was utilized to carry out several rub experiments, where we took the chance to change the position of the rub source, which was primarily a teflon seal. The copper and bronze seals were further also covered into the experiments. The vibration signals were measured on the bearing pedestals where the accelerometers measuring absolute vibration speed were installed. The photo of the experimental stand with the sensors of relative and absolute vibration is displayed in the following figure.

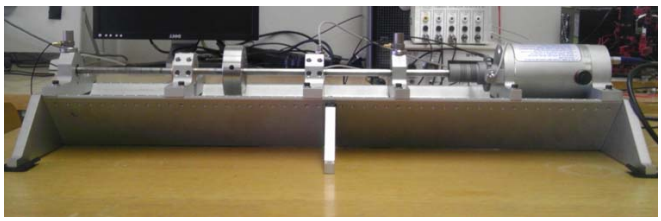


Fig. 9 Experimental rotor kit RK4.

The sampling frequency of the measured signals was 51.2 kHz (based on configuration of the measurement chain). The distance between the two measured planes was 0.48m. Let's label the measured position near the motor as bearing 2 (right bearing in Fig. 9) and the farther position as bearing 1 (left end of the shaft).

For the purpose of rub localization methods verification, various manifestations of rub were analyzed in time and time-frequency domain. Based on knowledge of the seal position, the actual propagation velocity was also analyzed.

In Fig. 10 (top), there is time behaviour of a signal containing rub impacts for the signal measured on the bearing 1. There are three impacts in the time view resulting from the contact of the shaft with the copper seal by the rotation speed of 2160 rpm. In the bottom part of the figure, the amplitude spectrogram in frequency band 0-10 kHz is displayed. The spectrogram was computed as approximation

ISBN: 978-1-61804-233-0

of the Gabor transform. This time-frequency representation serves more accurate resolution than the commonly used discrete short-time Fourier transform. In Fig. 11 the same signal analysis but for the farther sensor on bearing 2 is displayed.

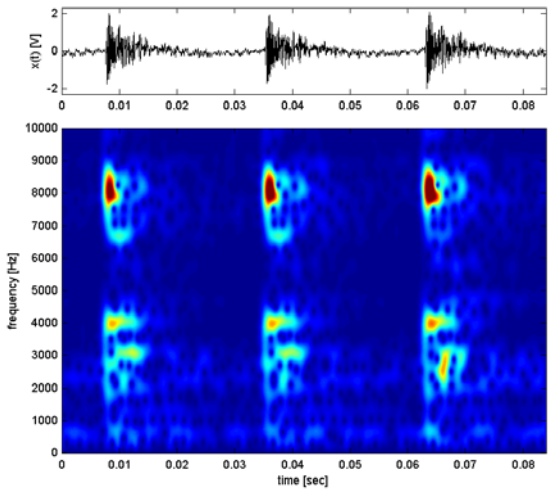


Fig. 10 Expression of rub impacts in time and time-frquency domain for the measured position near the seal (bearing 2).

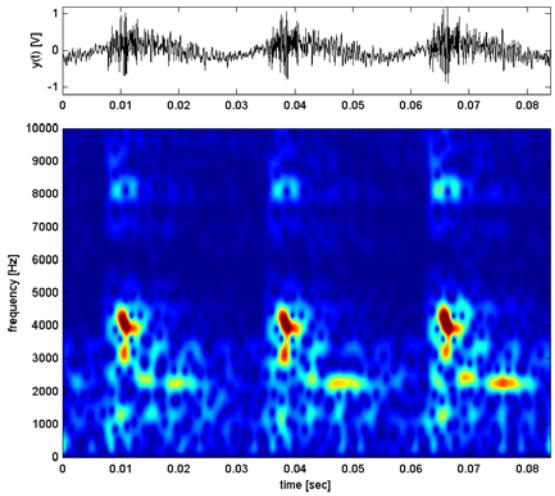


Fig. 11 Expression of rub impacts in time and time-frquency domain for the measured position farther from the seal (bearing 1).

VI. Localization of Rub in Machine Operation

An example of localization results for the seal position of described experimental rotor stand configuration is shown in the following Fig. 12. The true position of the seal is shown as a blue vertical line. The recursive variance computation of the vibration signal was used for the processing and in the section E chapter IV described method was applied for the wave arrival time detection. On the basis of the performed experiments the mean propagation velocity of 560 m/s was evaluated. The resulting histogram displays the rate of the localization outputs depending on the localized distance to the bearing 1.

The set up of the histogram column width covers a localization error of 5 mm. The results show relative good match between the true and evaluated seal position, especially if we accept the fact, that the seal width is 30 mm.

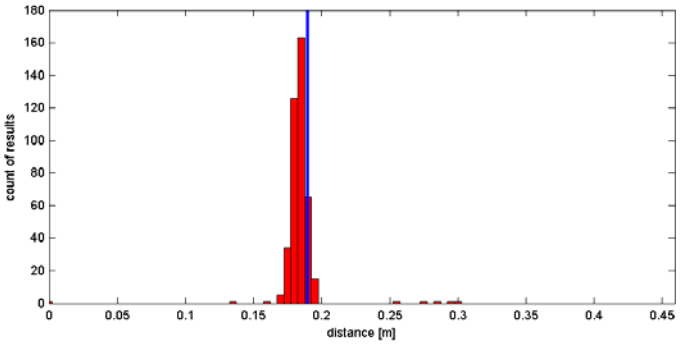


Fig. 12 Localization of the seal position in the experimental rotor stand, distance of 0.192 m to bearing 1.

The symptoms of shaft and blade rubbing were succesfully detected and measured by the RAMS monitoring system during the steam turbine operation in the year 2012 (see Fig. 13). The symptoms of the rub were significant and simultaneously detected in three measured planes. Consequently, the rub origin localization was performed.

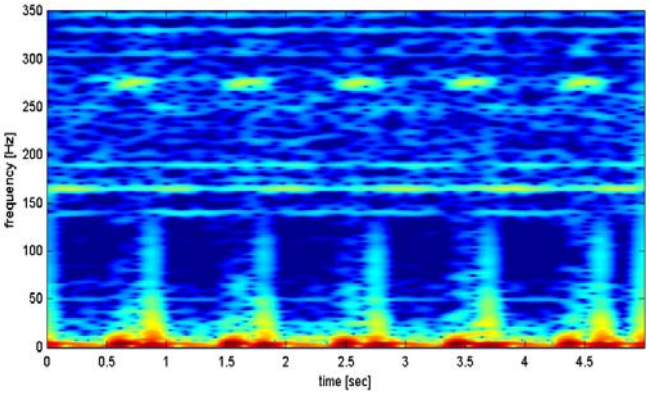


Fig. 13 Time-frequency representation of the shaft and blade rubbing measured by the operation of steam turbine.

The blade rub was identified by the excited frequency (276 Hz) which agrees with the frequency of shrouds crossing the swirl brakes installed in stator part above the blades. The excitation of the low frequency band in the signal was due to the shaft rubbing - rub of the rotor and stator labyrinth seals.

It is really difficult to determine the trasmission path and the propagation velocity of the rub impact signals throught the rotor and stator parts of the machine. Therefore it is appropriate to simplify the determination of the probable rub place by the use of linear localization and direct distances between the measured planes (bearing pedestals). In this case, the localization principle processes three measurement planes, and that is why there is no need to know the wave propagation velocity. A signal in frequency band of 270-280 Hz was used for analysis of the blade rub and the frequency band of 1-6Hz was determined as a shaft rub region. Dependency of the amplitude on time in the selected frequency bands was

ISBN: 978-1-61804-233-0

calculated using the Gabor transform. To estimate the arrival time of the rub impact wave, the method implementing the (24) was used.

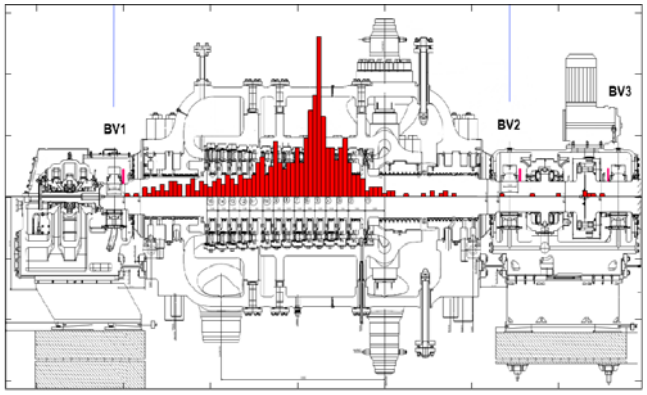


Fig. 14 Axial localization of the shaft rubbing in the HP part of the steam turbine.

In Fig. 14 there is the resulting axial localization of the shaft rubbing in form of histogram. The column width is 50mm. It can be seen the scheme of the HP part of the turbine in the backgroung and also several rotating blades stages are noticable in the figure. The measurement planes are also highlighted and marked as BV1 - BV3 (Bearing Vibrations measured on bearing pans).

According to the obtained results, the highest occurence of the rubbing impacts is between the 4th and the 6th bladed wheel. The variance of the obtained results is caused by the noise in measured signals and also (as it was found by the turbine inspection) the rub occurs on slighty different places depending on the operational deflection shape of the rotor.

Except the axial localization, there is also need to localize the contact place on the circumference of the turbine stator. An approximate tangential localization is possible when using signals from both absolute and relative vibration sensors. When fusing the information from the sensors of absolute and relative vibrations the angle of the maximum rub intensity could be obtained. The method is based on the representation of the rotor motion in the measured plane. From the ralative shaft vibrations an orbit of the rotor motion is reconstructed. The time behaviour of mean amplitudes in selected frequency bands, ie rub characteristic values, can be taken into account when rendering the filtered orbit of the rotor motion. Time synchronization is ensured by the keyphasor.

In Fig. 16, the result of tangential localization for blade rub is shown. The upper part shows the waveform of average amplitudes versus time for the signals from the front and rear bearing block of the HP steam turbine part, i.e. BV1 and BV2. At the bottom the filtered 1X orbits corresponding to the signals of shaft vibration SV1 and SV2 are displayed. The amplitude of the characteristic frequencies is scaled in colors, where blue corresponds to the minimum amplitude and red is the maximum value. From this figure it is evident when the highest intensity of the contact between the rotor and sator occurs and it is only a matter of processing to automatically evaluate the angle of maximum.

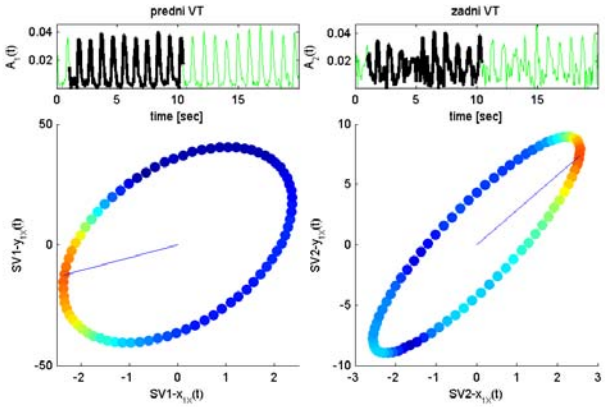


Fig. 15 Tangential localization of the blade rubbing for the measured possitions BV1 and BV2.

VII. Conclusions

In summer of 2012 the HP part of the turbine was opened and the eastimated places of rub were inspected. The results from the linear localization presented above (axial and tangential part) were in accordance with the results of inspection. Between the 4th and 6th bladed wheel, there were the most significant damages of the seals.

Acknowledgment

The authors Howould like to thank to the company Doosan Skoda Power for the possibility to do the research in the area of industrial interest and to have an opportunity to solve the actual problems on the field of steam turbine troubleshooting.

References

[1] A. Muszynska, *Rotordynamics*, Taylor&Francis, London, 2005.
[2] P. Goldman, A. Muszynska, and D. E. Bently, “Thermal Bending of the Rotor Due to Rotor-to-Stator Rub”, *International Journal of Rotating Machinery*, vol. 6(2), pp. 91-100, 2000.
[3] J. Liska, J. Jakl, and V. Cerny, “The use of time-frequency methods in rotor/stator impact-rubbing detection”, in *Proc. Asme 2011 International Design Engineering Technical Confereces & Computers and Information in Engineering Conference, IDETC/CIE 2011*, Washington, DC, 2011, pp. 1-10.
[4] J. Liska, J. Jakl, and E. Janecek, “Steam turbine rotor/stator impact and rubbing detection”, in *Proc. 9th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, London, 2012, pp. 1-12.
[5] L. D. Hall, D. Mba, “Diagnosis of continuous rotor-stator rubbing in large scale turbine units using acoustic emissions”, *Ultrasonics*, vol. 41, pp. 765-773, 2004.
[6] J. Liska, P. Kodet, and E. Janecek, "Source Location in Loose Parts Monitoring Using Time-Frequency Analysis", *Proceedings of 9th International Carpathian Control Conference (ICCC‘ 2008)*, Sinaia, Romania, 2008.
[7] E. Dieulesaint and D. Royer, *Elastic Waves in Solids – Applications to Signal Processing*, John Wiley & Sons, Chichester, 1980.
[8] H. J. Pain, *The Physics of Vibrations and Waves – 6th edition*, John Wiley & Sons, Chichester, 2005.
[9] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, New Jersey, 1995.

An Overview on Cryptography and Watermarking

Med Karim ABDMOULEH, Ali KHALFALLAH and Med Salim BOUHLEL
Research Unit: Sciences and Technologies of Image and Telecommunications
Higher Institute of Biotechnology
Sfax, Tunisia
medkarim.abdmouleh@isggb.rnu.tn, khalfallah.ali@laposte.net, medsalim.bouhlel@enis.rnu.tn

***Abstract*—**Nowadays, the transmission of information in all its forms, reaches a revolution. In certain areas it is necessary to protect the confidential information to make it. This requirement touches different areas such as military, industry, medicine and politics. In the literature, there exist several techniques that aim to information security. In this sense, we can cite the watermarking, the first aim is the protection of copyright. Also the cryptography, which aims to make the exchange of information incomprehensible by all unauthorized persons.

***Keywords*—**Chaos, Cryptosystem, Image encryption, Image Watermarking.

I. CRYPTOGRAPHY

A. Definition

The cryptography presents the basic science to secure information. It uses mathematical processes to transform this information and make it secret and to all those who are not permitted. Cryptography has two methods:

- The first is called encryption, it is to encode the information and make it secret in order to have an encrypted message or ciphertext.
- The second is called decryption is the inverse method of retrieving the original message.

The encryption is to apply mathematical methods to transform a clear text. The encryption algorithm or cryptosystem uses a set of parameters called encryption key to make the message incomprehensible called ciphertext or cryptogram. In order to retrieve the original text, we must carry out inverse transformations called decryption based on some parameters called the decryption key. During the exchange of data an attacker can intercept the encrypted data exchanged. This hacker will try to recover the clear text by applying mathematical transformations on the ciphertext. This process is called cryptanalysis.

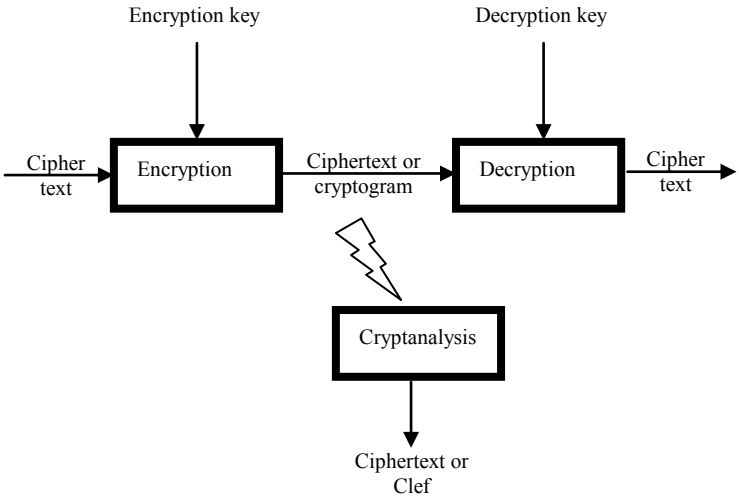


Fig. 1 Principle of cryptography

B. The purpose of a cryptosystem

The initial objective of encryption is to ensure the confidentiality of the transmitted message. Taking advantage of the properties of encryption, other applications are possible which are: authentication, integrity control and non-repudiation.

- Confidentiality: it ensures that only authorized people can read the information.
- Authentication: It ensures that the origin of the information is a source assured. Many methods can be used such as the signature and digital certificate.
- Data integrity: a cryptosystem is very sensitive to attacks, so any changes to the ciphertext leads to the loss of the original text. Thus, after the successful decryption of the text, the recipient ensures the handling of non-document.
- Non-repudiation: it is to prevent the message sender or the receiver to deny the content of the message.

C. Various cryptographic techniques

In this section we are interested in studying the different

encryption techniques to ensure the confidentiality of information and make it incomprehensible by anyone who are not qualified to access during the transmission process. In order to have the original receiver must perform the decryption operation that requires the possession of the decryption key. This key may be the same as that used when encrypting the clear text is talking about symmetric key encryption. Moreover, the two keys can be different, we are talking asymmetric key encryption. Classification according to the type of key is not the only one. Cryptosystems can be classified according to the procedure of message encryption.

In the following we present the different types of algorithms used in cryptography:

- Symmetric-key algorithm: Symmetric cryptography is also known by secret key cryptography. Like any cryptosystem, it is to use an encryption algorithm to transform a clear text is using an encryption key to a ciphertext. To decrypt the information, the receiver must use the same private key used by the sender [1] [2] [3].
- Public-key cryptography: This type of encryption is also called asymmetric encryption. It uses a pair of different keys for encryption and decryption in order to solve the problem of key exchange caused by the symmetric encryption [4] [5].
- Hybrid cryptography: The concept of hybrid cryptography uses the advantages of the two previous techniques. Hybrid cryptography is based on the combination of the speed of private key cryptography and security of asymmetric cryptography [6].
- Stream cipher: In this type of algorithm the plaintext is divided into blocks of fixed size. Each block is encrypted independently. The block size is between 64 and 256 bits [7].
- Block cipher: In contrast to stream cipher which divides the message to be encrypted into several blocks, the stream cipher encryption proceeds symbols one by one. The idea is to generate a sequence of key (key stream) the same size as the plaintext from a secret key [8].

II. Performance Analysis of Encryption Image Scheme

Image information differs from the textual one and has very interesting properties such as its large capacity, redundancy and strong correlation strength between neighboring pixels. According to these properties, the security of a cryptosystem image is evaluated by a statistical analysis, the sensitivity analysis and the differential analysis [9] [10].

A good cryptosystem should ideally be robust against all kinds of attacks (cryptanalytic and statistical attacks).

A. Key Space Analysis

For a system of secure encryption, the key space must be extended to resist brute-force attack (Brute-Force Attack). The size of the key range depends firstly of all parameters constituting a key, and you must specify the nature of the format of each parameter (real, integer, ...), then all possible values. The size of the key range is expressed in bits, and is defined by the number of possible combinations of these parameters are calculated by a power of 2. To calculate the range of key used frequently enough to avoid a brute force attack is of 128 bits or more.

B. Histogram Analysis

The encrypted image should be totally different from the clear image and without any resemblance static. The histogram of the encrypted image should be uniform. In fact, the similarity between the histogram of the clear and the encrypted image can lead to breaking the cryptosystem.

In order to have a perfect encrypted image, the histogram of the image must have a uniform distribution.

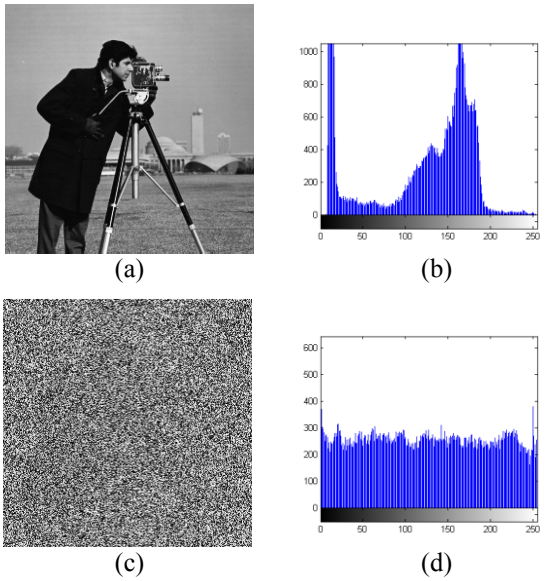


Fig. 2 Histogram analysis: (a) plain image, (b) plain image histogram, (c) encrypted image, (d) encrypted image histogram

A. Information Entropy

As the histogram of the encrypted image is almost uniform. The entropy was founded by Shannon in 1948 [11] [12] and is given in the following equation:

$$H(m) = \sum_i P(m_i) \log_2 \frac{1}{P(m_i)} \quad (1)$$

Where $P(m_i)$ represents the probability of symbol m+. The entropy H(m) is expressed in bits.

The entropy of encrypted image should be very close to 8 bits / pixel [13].

C. Correlation Coefficient Analysis

Another type of statistical analysis is the correlation coefficient analysis [14] [15] [16].

An image is often characterized by the strong correlation between a pixel and its neighboring pixels and in particular the pixels on the same row or the same column or the same

diagonal.

We calculate the correlation coefficient for a sequence of adjacent pixels using the following formula:

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}} \tag{2}$$

Here, x and y are the intensity values of two adjacent pixels in the image. r_{xy} is the correlation coefficient. The cov(x,y), E(x) and D(x) are given as follows:

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

$$D(x) = \frac{1}{N}\sum_{i=1}^{N} [x_i - E(x_i)]^2 \tag{4}$$

$$\operatorname{cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N} [(x_i - E(x_i))(y_i - E(y_i))] \tag{5}$$

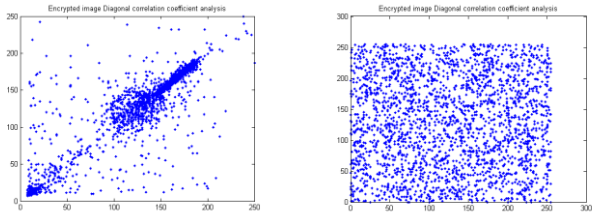
N is the number of adjacent pixels selected from the image to calculate the correlation.

The table below presents the correlation between two adjacent pixels for the plain and the encrypted Camera image.

To test the correlation coefficient, we have chosen 2500 pairs of two adjacent pixels which are selected randomly from both the plain and encrypted image.

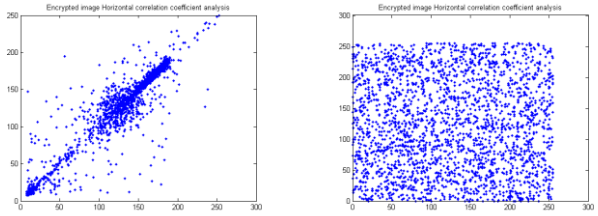
TABLE I. CORRELATION COEFFICIENTS OF TWO ADJACENT PIXELS

Direction	Correlation plain image	Correlation encrypted image
Diagonal	0.9084	0.0081
Horizontal	0.9335	0.0047
Vertical	0.9591	-0.0201



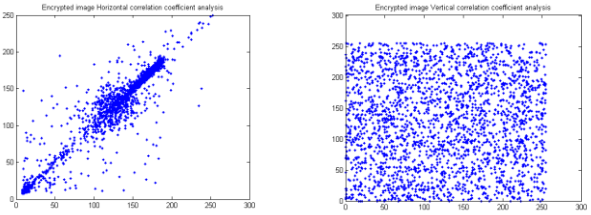
(a) (b)

Fig. 3 Correlation between two diagonally adjacent pixels: (a) in the plain image, (b) in the encrypted image



(a) (b)

Fig. 4 Correlation between two horizontally adjacent pixels: (a) in the plain image, (b) in the encrypted image



(a) (b)

Fig. 5 Correlation between two vertically adjacent pixels: (a) in the plain image, (b) in the encrypted image

Fig. 3-5 represents the correlation between two diagonally, horizontally and vertically adjacent pixels of the plain and encrypted image. We see that the neighboring pixels in the plain image have a high correlation, while in the encrypted one there is a little correlation.

D. Differential Analysis

The study of the differential analysis is carried out through both measures NPCR defined by equation (7) and UACI given by equation (8). The objective of this analysis is to show that a small change in the image clearly introduces a major change on the image encrypted [1] [17].

The theoretical values for an encryption algorithm ideal image are greater than 99% to the value of NPCR while UACI value must be very close to 33%.

The NPCR means the Number of Pixels Change Rate of ciphered image while a pixel of the plain image is changed.

Let A_2 be the changed plain image on one pixel. C_1 and C_2 are the ciphered images of the plain image and A_2. D is a matrix having the same size as the image figures C_1 and C_2. D (i, j) is determined as follows:

$$D(i, j) = \begin{cases} 1 & if\ C_1(i, j) \neq C_2(i, j) \\ 0 & else \end{cases} \tag{6}$$

The NPCR is defined by:

$$NPCR = \frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} D(i, j)}{M \times N} \times 100 \tag{7}$$

M and N are the height and width of encrypted images C_1 and C_2.

The Unified Average Changing Intensity (UACI) between these two images is defined by the following formula:

$$UACI = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} \frac{|C_1(i, j) - C_2(i, j)|}{255} \times 100 \tag{8}$$

The average of the two measures based on the original image with the corresponding encrypted image is presented in the following table.

TABLE II. VALUES RESULTS OF NPCR AND UACI

	NPCR	UACI
Camera	98.9975	31.3237

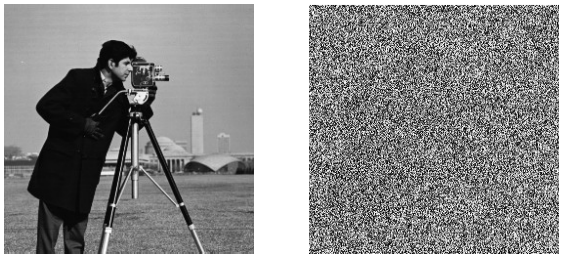
Experimentally measured value of NPCR is 98.9975 % and UACI is 31. 3237 % for Camera image. These results indicate that a small change in the plain image introduces a high alteration on the ciphered one.

E. Cryptanalysis

While cryptography is the discipline that ensures the security of confidential information, cryptanalysis [18] is the discipline that studies and validates the robustness of cryptosystems against attacks. According to Kerckhoffs, the security of a cryptosystem depends on the secrecy of the key and not on that of the encryption algorithm. To study the cryptosystem security, we can utilize the Kerckhoffs principle. In this case, the cryptanalyst must be unable to find the key even if he has access to the plaintext and its corresponding ciphertext. He tries to apply more attacks such as:

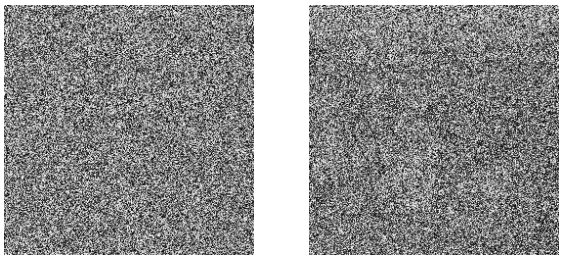
- Ciphertext only attack
- Chosen plaintext attack
- Known plaintext attack
- Chosen ciphertext attack

Known-plaintext attack is one of the attacks where the cryptanalyst owns a plaintext image and its corresponding encrypted image. This frequently used attack utilizes the known clear image and the ciphered image to extract the decryption key or to decrypt another ciphered image.



(a) (b)

(c) (d)



(e) (f)

(g)

Fig. 6 Failed crack attempt:
(a) plain image Camera, (b) encrypted Camera image, (c) extracted key stream, (d) decrypted Camera image, (e) encrypted Bird image, (f) failed attempt to crack the cipher image of Bird, (g) plain image Bird

III. WATERMARKING

A. Definition

Digital watermarking is a technique of dissimulation which is to integrate, inside of a digital document (audio, video, image, text), an invisible "watermark" using a computer algorithm in order to deal with any attack to change the digital document [19].

This mark should be transparent with respect to a human observer, however, retain the original data visible.

B. Different types of watermarking

This mark should be transparent with respect to a human observer, however, retain the original data visible.

There are two types of watermarking: visible and invisible.

1) Visible watermarking

This type of watermarking is simple, leads to change the original document in a manner perceptible, adding a mark to distinguish it from others. This operation is equivalent to stamping a watermark on the paper, what is called in the digital domain 'digital shading". In this context, we recall the example of agencies that insert a picture watermark visible in the form of copyright ("©") on the low-resolution photos to ensure that they do not change those by high-resolution charge. The efficacy of visible watermarking is a topic for discussion, which on the one hand it can be easily attacked, and on the other hand we find highly used as logos for companies in the television programs.

2) Invisible watermarking

In contrast to previous type, invisible watermarking is much more complex, which is the mark of an integrated imperceptibly.

Consider the previous example of the photo agency, in this case, it inserts a watermark invisible to high-resolution photos sold. The majority of techniques for the protection of intellectual property following the branch of invisible watermarking, and this is what we will explore in the following.

C. Diagram of watermarking

Tattooing image is to insert a signature him. To do this, you must select the area of insertion. Thus, the original image will

be transformed in this area, and then we, the selection of the holders of the brand. These carriers will be marked by the signature using one of the functions of integration. Then we return to space for the watermarked image (Fig. 8).

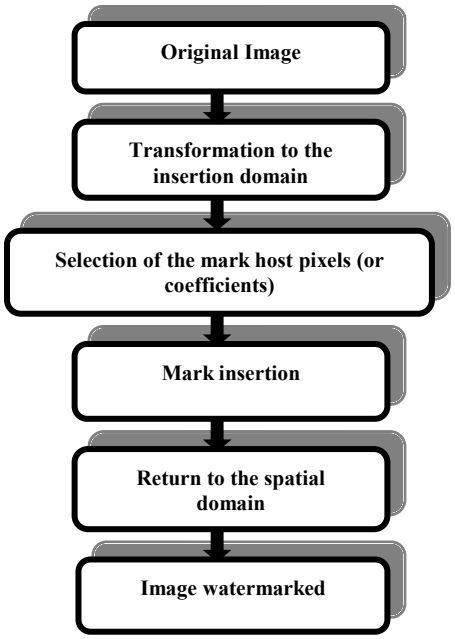


Fig. 8 Insertion scheme

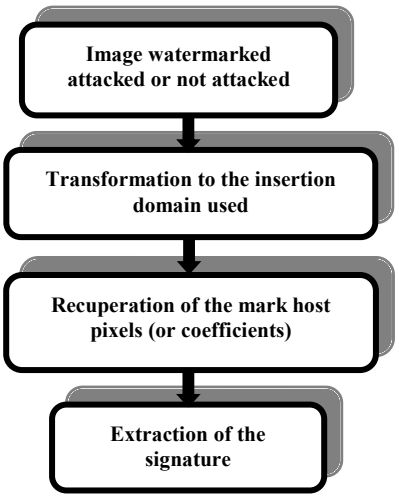


Fig. 9 Extraction scheme

Fig. 9 shows the path to extract the signature or to detect its presence. This phase requires the transformation of the watermarked image to the area of insertion, and then relocate the trademark holders to recover eventually signing inserted.

D. Constraint watermarking

The watermarking technique is faced with a compromise between several key parameters such as imperceptibility, robustness and security.

1) The imperceptibility

The principle here is that the watermark should be invisible to a human observer. In addition, we must maintain the fidelity between the original image and the watermarked.

2) Robustness and fragility

The watermarked image can be attacked by malicious to remove the watermark. It must be sufficiently robust and resistant against these attacks to remain legible. Fragility may also be more favored than the strength in some cases, such as checking the integrity of a document.

3) The ability to integrate

View the large image size, we try to insert a maximum of information to facilitate the recognition of the true owner of the signed document. As the size of the brand is more great recognition is easy. But insert a signature of large volume can impair image quality signed.

E. Categories of watermarking

The watermark can be classified into three categories (Fig. 10) based on data available at the detecting:

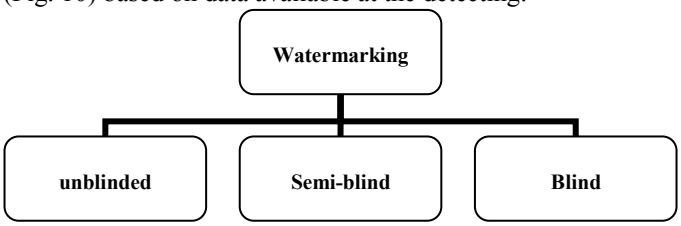


Fig. 10 Diagram of watermarking

1) Watermarking unblinded

This type of marking requires the presence of the original image, the watermarked image and the signature of the author in the extraction phase. In this case the extracted signature is simply the difference between the original image and the watermarked.

2) Semi-blind watermarking

In this type, you will not need the presence of the original image. Présance to check the signature, just to have the image signed and the signature pretending document owner. This type of marking recognizes a difficult relocation carriers signature.

3) Watermarking blind

This type requires only the presence of the image tattooed. It is very susceptible to attack. In this category of watermarking we often use binary signatures.

F. Areas of insertion

Like any signal, the image can have multiple presentations. Thus, the insertion of the signature can be done in one of these areas of image presentation. The areas most commonly used are [20]:

- Frequency domain: The move to this area is done by applying the Discrete Cosine Transform (DCT) or the Discrete Fourier Transform (DFT). This field shows the frequency distribution of the image.
- Domain multiresolution: The wavelet domain to allow the passage multiresoltuion. This area highlights the simultaneous spatial and frequency

representation of the image.

- Spatial Domain: Direct manipulation of pixel luminance [21] and in the case of our application.

G. Watermarking applications

- Copyright: The first role is to ensure the tattoo copyright. To define the owner, the copyright applies as a digital signature that can be inserted into the medium, in the works...
- Medical ethics: the patient information is in the form of a signature that is inserted into the document. The watermarking ensures the preservation of secret medical and economy Storage patient record.
- Monitoring program: The watermarking is used here to avoid hacker attacks, and alert viewers by inserting information about programs.
- Indexing: You can insert images into information that describes the content of the document so that it can be classified automatically in a database.
- Integrity checking: A signature must be inserted into sensitive image. Recovery will be impossible with any image manipulation signed. Modification or not the received image can be inferred after comparing the signature inserted and extracted it.

IV. Conclusion

In the multimedia encryption area, we need an algorithm with excellent performance. Also, the proposed encryption scheme should be secure and efficient.

In the area of copyright protection for multimedia data, the watermarking algorithms developed should be robust and fragile watermarking schemes and preserving copyright protection.

References

[1] G. CHEN, Y. MAO and C. K. CHUI, "A symmetric image encryption scheme based on 3D chaotic cat maps", *Chaos, Solitons and Fractals*, Vol. 21, n° 3, pp. 749-761, 2004.

[2] M. K. ABDMOULEH, A. KHALFALLAH and M. S BOUHLEL, "Image Encryption With Dynamic Chaotic Look-Up Table", *the 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'12)*, Sousse, Tunisia, March 2012

[3] M. K. ABDMOULEH, A. KHALFALLAH and M. S BOUHLEL, "Dynamic Chaotic Look-Up Table for MRI Medical Image Encryption", *the International Conference On Systems, Control, Signal Processing And Informatics (SCSI 2013)*, Rhodes Island, Greece, July 2013

[4] Z. KOTULSKI, "Discrete chaotic cryptography (DCC): New method for secure communication", *the 11th Workshop on nonlinear Evolution Equations and Dynamical Systems NEEDS'97*, OAK, Kolymbari near Chania, Crete Greece, June 1997.

[5] A. ROYO, J. MORN and J. C. LÓPEZ, "Design and Implementation of a Coprocessor for Cryptography Applications", *The European Design and Test Conference*, Paris -France, March 1997,

[6] S. GARFINKEL and G. SPAFFORD, *Practical UNIX and Internet security*, O'Reilly, Cambridge, 2. ed., 1996.

[7] A. CANTEAUT, Cryptographic Functions and Design Criteria for Block Ciphers, *The Second International Conference on Cryptology in India: Progress in Cryptology*, Springer-Verlag, London, UK, pp. 1-16, 2001.

[8] A. J. MENEZES, P. C. VAN OORSCHOT and S. A. VANSTONE, Handbook of Applied Cryptography , CRC Press , 2001

[9] A. KHALFALLAH, N. MBAYA, M. S. BOUHLEL, "An easy adapted image chaotic cipher to AWGN canal transmission", *Revue d'Information Scientifique Technique*, vol. 17, n°. 1-2, pp. 153-170, 2007.

[10] L. KOCAREV, "Chaos-based cryptography: a brief overview", *IEEE Circuits and Systems Magazine*, vol. 1, n°. 3, pp. 6-21, 2001.

[11] R. DEVANEY, *An Introduction to Chaotic Dynamical Systems*, 2nd ed. Redwood City, CA: Addison-Wesley, 1989.

[12] D. SOCEK, Li. SHUJUN, S.S. MAGLIVERAs and B. FURHT, "Enhanced 1-D chaotic key-based algorithm for image encryption", *IEEE, Security and Privacy for Emerging Areas in Communications Networks*, 2005.

[13] C. Q. LI, S. J. LI, G. ALVAREZ, G. R. CHEN, and K. T. LO, "Cryptanalysis of two chaotic encryption schemes based on circular bit shift and XOR operations", *Physics Letters A*, vol. 369, no.1-2, pp. 23-30, 2007.

[14] L.M. PECORA and T.L. CARROLL, "Synchronization in Chaotic Systems", *Physical Review Letters*, vol. 64, n°. 8, pp. 821-824, February 1990.

[15] T. PARASKEVI, N. KLIMIS and K. STEFANOS, Security of Human Video Objects by Incorporating a Chaos-Based Feedback Cryptographic Scheme, *The 12th annual ACM (Association for Computing Machinery) international conference on Multimedia*, New York, NY, USA, October 10 - 16, 2004.

[16] G. CHEN, Y. MAO, and C.K. CHUI, "A symmetric image encryption based on 3D chaotic maps", *Chaos Solitions Fractals*, vol. 21, pp. 749-761, 2004.

[17] Y. MAO, G. CHEN and C.K. CHUI, "A novel fast image encryption scheme based on 3D chaotic Baker maps", *Int. J. Bifurcation Chaos*, vol. 14, pp. 3613-3624, 2004.

[18] S. DOUGLAS, *Crypography: Theory and practice*, 2nd edition, Chapman & Hall, 2002.

[19] F. KAMMOUN, A. KHALFALLAH and M. S BOUHLEL, "New Scheme of Digital Watermarking Using an Adaptive Embedding Strength Applied on Multiresolution Filed by 9/7 Wavelet", *Journal of Imaging Systems and Technolgy*, Vol. 16, Issue 6, pp. 249-257, 2006.

[20] A. KHALFALLAH and M. S BOUHLEL, "The Impact of the Concept of the Family Relative Signatures on the Non-Blind Watermarking in the Multiresolution Domain using 9/7 and 5/3 Wavelets", *International Journal on Information and Communication Technologies*, Vol.4, N. 3-4, pp. 111-118, 2011.

[21] W. BENDER, D. GRUHL, N. MORIMOTO and A. LU, "Techniques for data hiding", IBM Systems Journal, Vol 35, n° 3&4, pp.313-336, 1996.

Analysis for the Pattern of the Lower Limbs Disease Using Decision Tree Model

J. K. Choi, K. H. Jeon, Y. G. Won and J. J. Kim

***Abstract*—**Datamining, which search for useful pattern and rule systematically and automatically in a large-scale data, is method to find out unknown and meaningful knowledge. Decision-tree algorithm, which is one of representative technique in the datamining, have advantage of that researcher is able to understand and explain analysis process easily compared with other methods. The purpose of this study was to find out significant knowledge from developing prediction model that classified category of lower limbs disease based on clinical data stored in Chungnam National University Hospital Foot Clinic. Sample data of 80 patients diagnosed with a disease in data of 1267 patients except data of 18 patients with missing value was used for analysis. Dependent variable was consisted of total 8 items. Independent variables were selected with 28 variables of the whole 37 attributes closely related to disease. The factors related closely with 8 disease were RCSP, right pelvis elevation and pelvis rotation condition. In the case of both pes planus and left pes planus, it were shown same result that left RCSP was higher than right RCSP. However, we could confirm the left pes planus would accompany the body unbalance by pelvis rotation and pelvis elevation difference. Pes cavus was always shown if any feet had above 1° for RCSP, and gastro-soleus muscle tightness was appeared if there was pelvis elevation difference with similar RCSP on both feet. In conclusion, we were able to conclude that factor of each node, including RCSP, which formed decision-tree, major diagnosis for distinction lower limbs disease and obtain the 8 diagnosis rules. On the basis of result in this study, datamining for the whole data will proceed from now on.

***Keywords*—**Datamining, Decision-tree, Foot, Disease, RCSP

I. Introduction

THE most important element in the standing posture and bipedalism of the human body.[1] When a human walks for 1 km, there is about a 15 t weight increase on the foot, and the pressure that occurs by weight or the push-off exercise causes stress or soft tissue strain. [2], [3] In this condition, there are close connections between form of the abnormal foot and cause of the various diseases in lower limbs as biomechanical concept.

Generally, the foot shape is classified into three types such as pes planus, normal foot, and pes cavus by visual inspection, radiographic examination, resting calcaneal stance position (RCSP) and footprint examination. [4] Pes planus is form of the foot touched excessively on the floor in the midfoot by the lower medial longitudinal arch (MLA) aberrantly and caused by the loose spring ligament. However, most of cases are idiopathic. [5] Pes cavus is a medical condition, in which the MLA has the height of the foot raised, and accepted to be rigid structurally. As a consequence, the contact area of the foot on the floor is narrowed, while the ankle or heel is tilted out exteriorly. [6] This foot condition is known to be commonly caused by constricted the Achilles' tendon or high-heel shoes in woman. This abnormal the foot shape have a bad influence on balance of the knee, pelvis and spine besides the ankle and the unstable body cause excessive and unnatural movement in the joints. [7]

The various field containing medicine have a large-scale data by development of information technology. The result analyzed on this data is applied index or information extracted on the basis of experience. Datamining, which is drawing attention as the technology of information modeling on account of search for useful pattern and rule systematically and automatically in this data, is method to find out new and meaningful knowledge with utilizing pattern recognition technology, statistics technique or mathematic algorithm. [8], [9] Decision-tree, which is one of representative technique in the datamining, is an analysis method to predict or classify from group of interested object to a couple of subgroup by observing relationship and modeling regulation in data, and non-parametric method that linearity, normality and homoscedasticity supposition are unnecessary. [10] The decision-tree have advantage of that researcher is able to understand and explain analysis process easily compared with neural network, discriminant analysis and regression analysis because it show result of learning as tree structure shape. [11] The typical algorithm to generate decision-making are C5.0 based on ID3 algorithm, C&RT, Quest and CHAID. [12], [13] This datamining methods are utilized in several medical field such as research, diagnosis, hospital management and customer management in the present and used for analysis of clinical data as useful tool, especially. [14], [15] According to previous studies, the

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1B3003952, NRF-2013R1A2A2A04016782).

J. K. Choi is with the Department of Healthcare Engineering, Chonbuk National University, 664-14, Deokjin-dong 1, Deokjin-gu, Jeonju, Jeolabuk-do, Republic of Korea (e-mail: chaosjoozak@nate.com).

K. H. Jeon is with the Department of Occupational therapy, Kunjang College, 608-8, Doamri, Sungsan-myun, Gunsan, Jeollabuk-do, Republic of Korea (e-mail: khjeon@kunjang.ac.kr).

Y. G. Won is with the School of Electronics and Computer Engineering, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea (e-mail: ykwon@chonnom.ac.kr).

J. J. Kim is with the Division of Biomedical Engineering and Research Center of Healthcare & Welfare Instrument for the Aged, Chonbuk National University, 664-14, Deokjin-dong 1, Deokjin-gu, Jeonju, Jeolabuk-do, republic of Korea (corresponding author to provide phone: 82-63-270-4102; fax: 82-63-270-2247; e-mail: jungjakim@jbnu.ac.kr).

ISBN: 978-1-61804-233-0

evolutionary-driven support vector machine was used to anticipate stage of hepatic fibrosis that determine hardness degree of liver or operation in chronic hepatitis C, and the logistic regression and neural network are utilized to extract attribute and perform learning based on widespread clinical data of acute myocardial infarction for forecast short-term relapse mortality of ST-segment elevation myocardial infarction (SEMI) patients. Through this study, the model to foresee short-term mortality of SEMI patients is suggested. [16], [17] In addition, age, associated disease, pathology scale, course of hospitalization, respiratory failure and congestive heart failure were came out to be danger factors on death of pneumonia by using the decision-tree model for analysis of death factor on pneumonia patients. [18] More intelligent and exquisite analysis is necessary because most existing studies just figured out simple correlation from only quantitative analysis based on statistical method despite it is very important in disease diagnosis to grasp various interconnection of several parameters measured with reference to disease on analysis of medical data.

Accordingly, the purpose of this study was to find out significant knowledge from developing prediction model that distinguished category of lower limbs disease and deducting relevant rule based on clinical data stored in the Foot Clinic of Chungnam National University Hospital.

II. Method

A. Subjects

The first medical examination clinical data of total 1285 patients in the Foot Clinic of Chungnam National University Hospital were utilized in this study. The data was composed of parameters of 37 attribute measured and diagnosed by a podiatrist besides gender and age. Sample data of 80 patients diagnosed with a singular disease in data of 1267 patients except data of 18 patients with missing value was used for analysis.

B. Variables

Dependent variable in this study was single disease and

Table I Dependent variable

Variable	Description
Pathologies	(L) Pes cavus
	(L) Pes planus
	(R) Pes cavus
	(R) Pes planus
	Gastro-Soleus m.
	Pes cavus
	Pes planus
	Scoliosis

consisted of total 8 items such as (L)Pes cavus (left pes cavus), (L)Pes planus (left pes planus), (R)Pes cavus (right pes cavus), (R)Pes planus (right pes planus), Gastro-soleus m. (gastro-soleus muscle tightness), Pes cavus, Pes planus and Scoliosis (Table I).

Independent variables were selected with 28 variables of the whole 37 attributes closely related to disease through opinion of a podiatrist (Table II).

C. Study Procedure

In this study, combination of variables that explained effectively 8 diseases, nominal dependent variable, were analyzed by insert the 28 independent variables in algorithm at the same time. Data analysis was performed by IBM SPSS Modeler 14.2 (SPSS Inc., Chicago, IL, USA). It was efficient to make a number of predictive model from one data and carry out comparison analysis for producing ideal model. [19] The whole data, therefore, was divided into training data (70%) and test data (30%) by using partition node. Prediction rate was verified by analysis node after model of each data was developed by C5.0 algorithm of datamining (Fig. 1). Tree-structured classification model, which is knowledge discovery technique for using purpose of classification, is comprised of organization as 'If A, then B. Else B2'. [20]

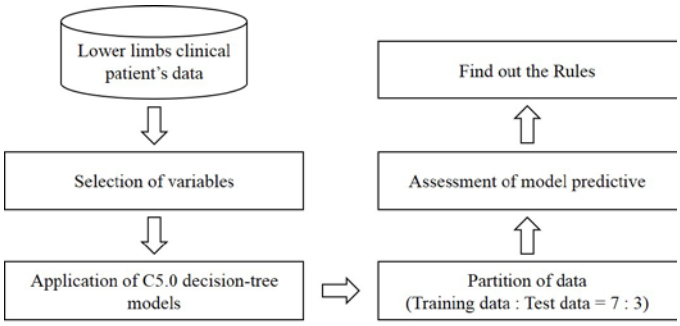


Fig. 1. Datamining procedure

III. Result

Data of 80 patients (Male : 31, Female : 49) diagnosed 8 disease in the whole 1267 patients data was used for experiment. The decision-tree model was generated for analysis of disease category by applying the C5.0 algorithm (Fig. 2). The measured prediction rate is Correct : 94.64 % and Wrong : 5.36% in the training data and Correct : 80.77 % and Wrong : 19.23% in the test data (Table III).

Table III Predictive according to Modeling

	Training data		Test data	
	Correct	Wrong	Correct	Wrong
C5.0	94.64	5.36	80.77	19.23

As a result of analysis on disease by using C5.0 algorithm, significant prediction factor was shown in the order : (L)RCSP (0.58), (R)RCSP (0.21), (R)PELVIS_Elva (0.16), PELVIS_Rot (0.05) (Fig. 3). In addition, total 8 rules were verified about training data (70%) of the whole data : 1) If '(R)RCSP' is below 1°, '(L)RCSP' is below -2°, 'PELVIS_Rot' is left-lateral rotation (L_Lt-R_N) and '(R)PELVIS_Elva' is below 2°, then '(L)Pes planus', 2) If '(R)RCSP' is below 1°, '(L)RCSP' is below -2° and 'PEVIS_Rot' is normal, then 'Pes planus', 3) If '(R)RCS'P below -2° and '(L)RCSP' is above -2° ~ below 1°, then '(R)Pes planus', 4) '(R)RCSP' is below 1° and '(L)RCSP' is above 1°, then '(L)Pes cavus', 5) '(R)RCSP is above 1° and

Table II Independent variables

Variables	Type	Description
Sex	Nominal	'Male', 'Female'
Age	Nominal	'Adolescent', 'Adult', 'Child', 'Infants'
(R) TIBIA_TMA = (Right) Tibia rotation angle	Numeric	-18° ~ 10°
(L) TIBIA_TMA = (Left) Tibia rotation angle	Numeric	-29° ~ 10°
Forefoot Adductus = Forefoot Adducted conditions	Nominal	'L_Mi-R_Mi', 'L_Mo-R_Mo', 'L_N-R_Mo', 'N' (L=Left, R=Right, Mi=Mild, Mo=Moderate, N=Normal)
Hip_ExtRot = Hip joint External Rotation conditions on knee flexed	Nominal	'L_Mi-R_Mi', 'L_Mo-R_Mo', 'N' (L=Left, R=Right, Mi=Mild, Mo=Moderate, N=Normal)
Hip_IntRot = Hip joint Internal Rotation conditions on knee flexed	Nominal	'L_Mi-R_Mi', 'L_Mi-R_N', 'L_Mo-R_Mo', 'L_N-R_Mi', 'N' (L=Left, R=Right, Mi=Mild, Mo=Moderate, N=Normal)
Ankle_Dorsi_Kneeflex = (Left) Ankle joint Dorsiflexed conditions on Knee flexed	Nominal	'L_B-R_B', 'L_B-R_Mi', 'L_B-R_Mo', 'L_B-R_S', 'L_Mi-R_B', 'L_Mi-R_Mi', 'L_Mi-R_Mo', 'L_Mo-R_B', 'L_Mo-R_Mi', 'L_Mo-R_Mo', 'L_S-R_B', 'L_S-R_Mi' (L=Left, R=Right, Mi=Mild, Mo=Moderate, B=Bad, S=Severe)
Ankle_Dorsi_KneeExten = Ankle joint Dorsiflexed conditions on knee extended	Nominal	'L_B-R_B', 'L_B-R_Mi', 'L_B-R_Mo', 'L_B-R_S', 'L_Mi-R_B', 'L_Mi-R_Mi', 'L_Mi-R_Mo', 'L_Mo-R_B', 'L_Mo-R_Mi', 'L_Mo-R_Mo', 'L_Mo-R_S', 'L_S-R_B', 'L_S-R_Mi', 'L_S-R_Mo', 'L_S-R_S' (L=Left, R=Right, Mi=Mild, Mo=Moderate, B=Bad, S=Severe)
(L) STJ_Inversion = (Left) SubTalar Joint Inversion maximum angle	Numeric	2° ~ 69°
(R) STJ_Inversion = (Right) SubTalar Joint Inversion maximum angle	Numeric	2° ~ 70°
(L) STJ_Eversion = (Left) SubTalar Joint Eversion maximum angle	Numeric	-6° ~ 43°
(R) STJ_Eversion = (Right) SubTalar Joint Eversion maximum angle	Numeric	-10° ~ 39°
(L) STJ_ROM = (Left) STJ_Inversion + STJ_Eversion	Numeric	-4° ~ 99°
(R) STJ_ROM = (Right) STJ_Inversion + STJ_Eversion	Numeric	8° ~ 101°
(L) FFtoRF = (Left) torsion angle of ForeFoot to RearFoot(0°)	Numeric	-31° ~ 24°
(R) FFtoRF = (Right) torsion angle of ForeFoot to RearFoot(0°)	Numeric	-18° ~ 49°
(L) RCSP = (Left) Resting Calcaneal Stance Position angle	Numeric	-20° ~ 16°
(R) RCSP = (Right) Resting Calcaneal Stance Position angle	Numeric	-36° ~ 13°
Pelvis_Tilting = Pelvis Tilting conditions	Nominal	'L_A-R_N', 'L_N-R_A', 'N' (L=Left, R=Right, A=Anterior tilting, N=Normal)
Pelvis_Rot = Pelvis Rotation conditions	Nominal	'L_Lt-R_N', 'L_N-R_Lt', 'N' (L=Left, R=Right, Lt=Lateral rotation, N=Normal)
(L) Pelvis_Trend = (Left) Pelvis angle of the maximum height on each one leg	Numeric	-8° ~ 20°
(R) Pelvis_Trend = (Right) Pelvis angle of the maximum height on each one leg	Numeric	-8° ~ 16°
(L) Pelvis_Eleva = (Left) Pelvis angle of height difference between left/right sides	Numeric	-1° ~ 12°
(R) Pelvis_Eleva = (Right) Pelvis angle of height difference between left/right sides	Numeric	-1° ~ 10°

'(L)RCSP' is below 1°, then '(R)Pes cavus, 6) '(R)RCSP' is

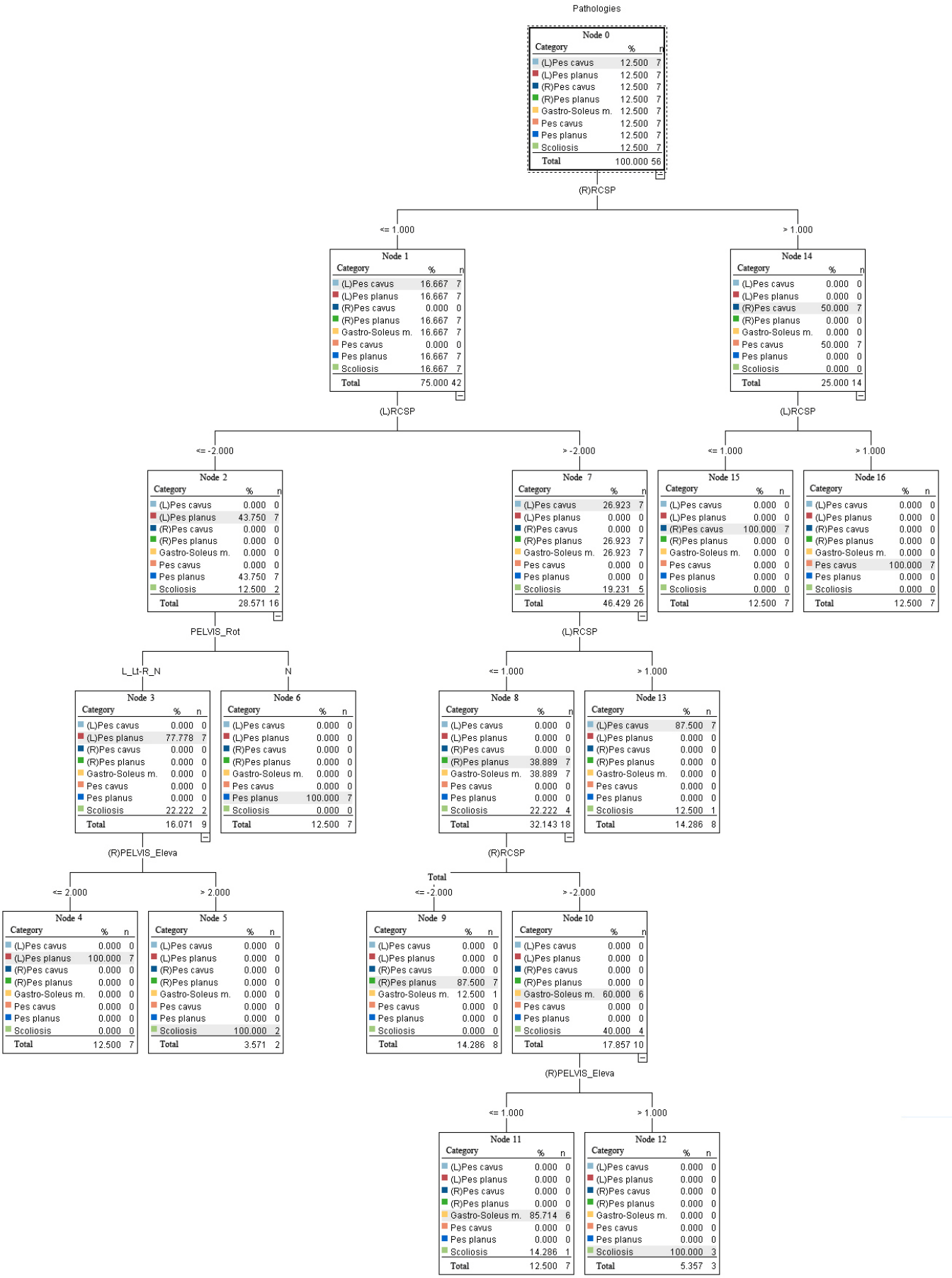


Fig. 2. The result of C5.0 decision-tree model

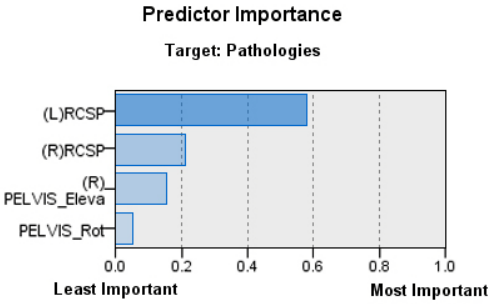


Fig. 3. Predictor importance of Independent variables

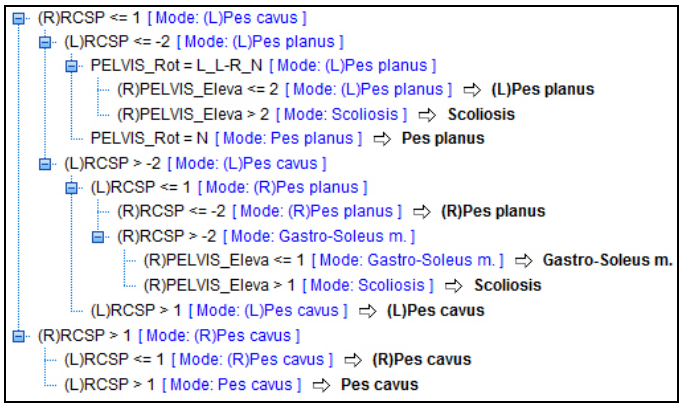


Fig. 4. The result of discoverd rule

above 1° and '(L)RCSP' is above 1°, then 'Pes cavus', 7) '(R)RCSP' and '(L)RCSP' are above -2° ~ below 1° and '(R)PELVIS_Eleva is below 1°, then 'Gastro-Soleus m., 8) '(R)RCSP' is below 1°, '(L)RCSP' is below -2°, 'PELVIS_Rot' is left-lateral rotation (L_Lt-R_N) and '(R)PELVIS_Elva' is above 2° or '(R)RCSP' and '(L)RCSP' are above -2° ~ below 1° and '(R)PELVIS_Eleva' is above 1°, then 'Scoliosis' (Fig. 4).

IV. Discussion and Conclusion

In this study, the object was to develop prediction model that classified category of lower limbs disease and draw diagnosis rule from learning data measured on patient based on foot clinical data by utilizing the decision-tree model of datamining technique.

The sample data of 80 patients diagnosed with 8 disease for the first medical examination clinical data of 1267 patients in the Foot Clinic of Chungnam National University Hospital was used for analysis. Dependent variable was consisted of total 8 disease and independent variables were selected with 28 variables. The data was divided into training data (70%) and test data (30%) for producing ideal model and prediction rate was confirmed after model of each data was developed by C5.0 decision-tree. As a result, the measured prediction rate is Correct: 94.64 % and Wrong: 5.36% in the training data and Correct: 80.77 % and Wrong: 19.23% in the test data.

As the result, the factors related closely with 8 disease were RCSP, right pelvis elevation and pelvis rotation condition. In the case of both pes planus and left pes planus, it were shown same result that left RCSP was higher than right RCSP. However, we could confirm the left pes planus would accompany the body unbalance by pelvis rotation and pelvis elevation difference. Pes cavus was always shown if any feet had above 1° for RCSP, and gastro-soleus muscle tightness was appeared if there was pelvis elevation difference with similar RCSP on both feet. In conclusion, we were able to conclude that factor of each node, including RCSP, which formed decision-tree, major diagnosis for distinction lower limbs disease and obtain the 8 diagnosis rules.

This study shows a similar result to previous studies classified shape of the foot according to RCSP. [21], [22] A clue, especially, which the body unbalance would come together, was figured out if there was gastro-soleus muscle tightness or pes planus on one side. In the study, the error rate of test data was relatively high due to extract 80 patients diagnosed with 8 diseases from the whole data, and predicted to decrease if decision-tree model learn more data. On the basis of result in this study, datamining analysis for the whole data will proceed from now on.

Acknowledgment

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Techn ology (NRF-2012R1A1B3003952, NRF-2013R1A2A2A04016 782).

References

[1] A. K. Ramanathan, P. Kiran, G. P. Arnold, W. Wang, and R. J. Abboud, "Repeatability of the pedar-x in-shoe pressure measuring system," *Foot and Ankle Surgery*, vol. 16, pp. 70-73, June 2010.

[2] S. B. Choi, and W. J. Lee, "Influence of Shoe Shape and Gait Characteristics on feet Discomforts according to Women's Foot Type," *The Costume Culture Association*, vol. 10, pp. 306-317, June 2002.

[3] D. J. Lott, M. K. Hastings, P. K. Commean, K. E. , Smith, and M. J. Muller, "Effect of footwear and orthotic devices on stress reduction and soft tissue strain of the neuropathic foot," *Clinical Biomechanics*, vol. 22, 352–359, Mar. 2007

[4] W. H. Lee, and S. W. Lee, "A study of the Relationship between Normal Adults Resting Calcaneal Stance Position and Postural Sway," *J Korean Soc Phys Ther*, vol. 11, pp. 5-17, June 2004.

[5] Y. T. Wang, C. K. Kim, and V. Farrar & T. Ford "The effects of heel height on ground reaction forces in female's normal walking," *Journal of Sport and Leisure Studies*, vol. 2, pp. 187-212, May 1994.

[6] S. L. Turek, *Turek's Orthopaedics: Principles and their application 5thed*. Baltimore, USA: Lippincott Williams and Wilkins, 1984.

[7] M. G. Benedetti, F. Catani, F. Ceccarelli, L. Simoncini, S. Giannini, and A. Leardini, "Gait analysis in pes cavus," *Gait & Posture*, vol. 5, pp. 169, April 1997.

[8] K. Kwak, J. Yoo, and J. Kim, "The investigation of protective facotrs and risk factors in child poverty : On the 3-yesr-old children," *The Korean Journal of Developmental Psychology*, vol. 20, pp. 1-19, Feb. 2007.

[9] H. C. Park, "Decision Process for right association rule generation," *Journal of the Korean Data & Information science Society*, vol. 21, pp. 263-270, Mar. 2010.

[10] J. H. Choi, H. C. Kang, E. S. Kim, S. K. Lee, S. T. Han, and M. K. Kim, *Prediction and excess of data mining using decision tree analysis*. Seoul, KOR: SPSS Academy, 2002.

[11] J. Y. Song and H. K. Kim, "A Study of Decision Tree in Detecting Intrusions," *Journal of the Korean Data Analysis Society*, vol. 12, pp. 983-996, Apr. 2010.
[12] J. A. Hartigan, *Clustering Algorithms*. New York, USA: John Wiley & Son Inc., 1975
[13] J. R. Quinlan, *C4.5 Programs for Machine Learning*, San Francisco, USA: Morgan Kaufmann Publishers, 1993.
[14] M. O. Suhn, Y. M. Chae, H. J. Lee, S. H. Lee, S. H. Kang, and S. H. Ho, "An application of datamining approach to CQI using the discharge summary," *Proceedings of the Korea Inteligent Information System*, vol. 6, pp. 1225-8903, Dec. 2000.
[15] W. S. Bae, D. H. Cho, K. H. Seok, B. S. Kim, K. L. Choi, and J. E. Lee, *Data mining using SAS enterprise miner*. Seoul, KOR: Kyowoosa, 2004.
[16] R. Stoean, C. Stoean, M. Lupsor, H. Stefanescu, and R. Badea, "Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis C," *Artificial Intelligence in Medicine*, vol. 51, pp. 53-65, Jan. 2011.
[17] K. H. Lim, K. S. Ryu, S. H. Park, H. S. Shon, and K. H. Ryu, "Short-term Mortality Prediction of Recurrence Patients with ST-segment Elevation Myocardial Infarction", *Journal of The Korea Society of Computer and Information*, vol. 17, pp. 145-154, Oct. 2012.
[18] Y. M. Kim, "A study on analysis of factors on in-hospital mortality for community-acquired pneumonia," *Journal of the Korean Data & Information Science Society*, vol. 22, pp. 389-400, Mar. 2011.
[19] M. Park, S. Choi, A. M. Shin, and H. Chul, "Analysis of the Characteristics of the Older Adults with Depression Using Data Mining Decision Tree Analysis," *J Korean Acad Nurs*, vol.43, pp. 1-10, Feb. 2013.
[20] M. H. Huh, and Y. G. Lee, *Data mining modeling and case 2nded*. Seoul, KOR: Hannarae, 2008.
[21] L. K. Dahle, M. Mueller, and A. Delitto, "*Visual assessment of foot type and relationship of foot type to lower extremity injury*," J. Ortho. Sports Phys. Ther., vol. 14, pp.70-74, Aug. 1991
[22] M. L. Root, W. P. Orien, and J. H. Weed, "Normal and abnormal function of the foot," *Clinical Biomechanicsm*, vol. 3, pp. 1977,

Jung-Kyu Choi He received the B. S., M. S. degree from the Chonbuk National University, Republic of Korea, in 2011 and 2013, respectively. He is currently pursuing a doctorate in foot biomechanics at the university. His major research interest is the podiatry, medical informatics and biomechanics.

Keun-Hwan Jeon He received the B. S. in Computer Science from Kunsan National University in 1993, and M. S. and Ph. D. degree in Computer Science from Chonbuk National University in 1995 and 2002, respectively. He is currently an assistant professor at Department of Occupational Therapy, Kunjang College and in master course in biomedical informatics at Chonbuk National University. His major research interest is the biomechanics, medical informatics and computer science.

Yonggwon Won He received the B. S. in Electronics Engineering from Hanyang Univesity in 1987, and M. S. and Ph. D. degrees in Electrical and Computer Engineering from University of Missouri-Columbia in 1991 and 1995, respectively. He worked with Electronics, and Telecommunication Research Institute (ETRI) from 1995 to 1996, and Korea Telecomm (KT) from 1996 to 1999. He is currently a professor in Chonnam National University in Korea, and the director of Korea Bio-IT Foundary Center at Gwangju. His major research interest is the computational intelligence for image analysis, pattern recognition, network and communication security, bio and medical data analysis.

Jung-Ja Kim She received the B.S., M.S. degree in 1985, 1988 and Ph.D. degree from 1997 to 2002, in Computer Science from Chonnam National University respectively. She worked with electronic telecommunication Laboratory at Chonnam National University from 2002 to 2004, and Korea Bio-IT Foundry Center at Gwangju from 2004 to 2006. She is currently an associate professor at Chonbuk National University. Her major research interest is the bio and medical data analysis, database security, and pattern recognition.

Applying "ABCD Rule of Dermatoscopy" using cognitive systems

Ionut Taranu, Iunia Iacovici

***Abstract*—** Difficulties in accurately assessing pigmented skin lesions (nevi, melanoma) are always present in practice. ABCD rule of dermatoscopy (skin microscopy magnification XIO), based on the criteria of asymmetry (A), edges (border-B), color (C) and differential structure (D), improved diagnostic accuracy when it was applied retrospective on clinical images. This paper is a prospective study to assess the value of ABCD rule for dermatoscopy on melanocytic lesions using cognitive systems.

***Keywords*—** computer-aided diagnosis, skin cancer, melanoma, dermoscopy, segmentation, border detection

I. Introduction

Stolz et al. introduced ABCD rule of dermatoscopy, based on multivariate analysis of four criteria (asymmetry, pigment pattern sudden stop at the border, color variation, and different structure) with a semi-quantitative scoring system. .

The rule was developed based on a retrospective analysis of color prints of dermoscopy images in the same size as those seen using a Dermatoscope. This study tests the ABCD rule prospectively to see if the criteria determining when examining an injury accurately predicts the correct diagnosis.[1]

Steps	Criteria for separation
Easily identifiable melanocytic lesions	Pigment network, regular streaks, brown globules, steel-blue areas (blue nevus)
Angiomatous tumors and hemorrhages	Red lakes, homogeneous reddish or reddish black areas, well-defined outline, lack of network, regular streaks and globules
Pigmented seborrheic keratoses	Yellowish or white "comedo-like" openings, brownish gray appearance, well-defined outline, lack of network, regular streaks and brown globules, presence of teleangiectasias
Basal cell carcinomas	Maple leaf-like areas, telangiectases, lack of network, regular streaks and globules
Remaining	Benign and malignant melanocytic lesions

Tabel I. Algorithm for distinguishing melanocytic and nonmelanocytic pigmented lesions by dermatoscopy Kreusch J, Rassner G. Hautarzt 1991;42:77-83.

For the calculation of the score ABCD asymmetry criteria (A), the sudden interruption of the pattern at the border pigment (B), various colors (C), and various structural components (D) were measured to obtain a semi-quantitative score.

The asymmetry was assessed in terms of color and structure of mutually perpendicular axes selected in such a way as to result in the lowest possible asymmetry score. A score of 0 was given if the were two axis of asymmetry, a score of 1, if the asymmetry was in one axis, and a score of 2 in the case of asymmetry in both axes.

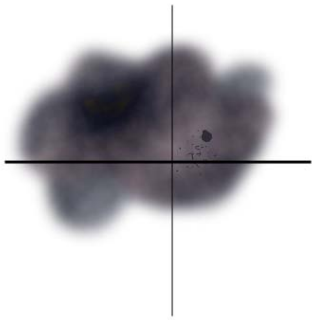


Fig1. Evaluation of melanocytic lesions asymmetry

To calculate the border, the lesions were divided into 8 segments as shown in Fig. 2.

For each segment in which a pattern of sudden stop of the model of border pigment was present, a point was added to the score.

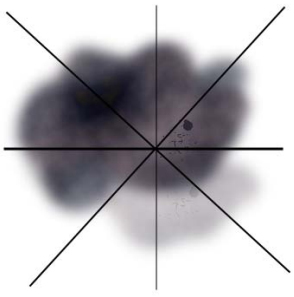


Fig.2 Evaluation of melanocytic lesions border

ISBN: 978-1-61804-233-0

For the color score the number of colors in the lesion was counted (possible colors: white, red, dark brown, blue-gray and black), resulting in a minimum score of 1 and a maximum of 6 (Fig 3.). Possible differential structural components were: network, homogeneous areas, dots, globules, and streaks (Fig. 4). Thus, the differential structure score (D) was multiplied by different weight factors derived from multivariate analysis

Fig3. Evaluation of color in melanocytic lesions.

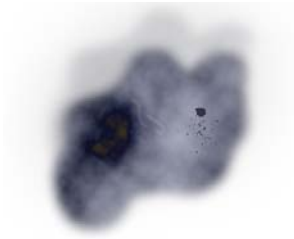


Fig4. The evaluation of the various structural components

II. RESEARCH

A. *The calculation method*

ABCD score was calculated by adding the four individual scores (A to D), resulting in a final score of minimum 1,0 (A = 0, B = 0, C = 1, D = 1) and a maximum of 8.9 (A = 2, B = 8, C = 6, D = 5) (Tabel II).

Criteria	**Score X**	**Factor =**	**Result**
Asymmetry **(Asimetrie)** Perpendicular axes: shape, color and structure	0 - 2	1.3	0 - 2.6
Borders (**M**argini) 8 segments: irregular edges of the pigmented model	0 - 8	0.1	0 - 0.8
Color White, red, light brown, dark brown, gray, blue, black	1 - 6	0.5	0.5 - 3.0
Different structural components (pigmented network, points, areas without structure, etc..)	1 - 5	0.5	0.5 - 2.5
Total	Benign		<4.76
	Suspect		4.76-5.45
	Melanoma		>5.45

Tabel II. Calculation of ABCD score in dermatoscopy

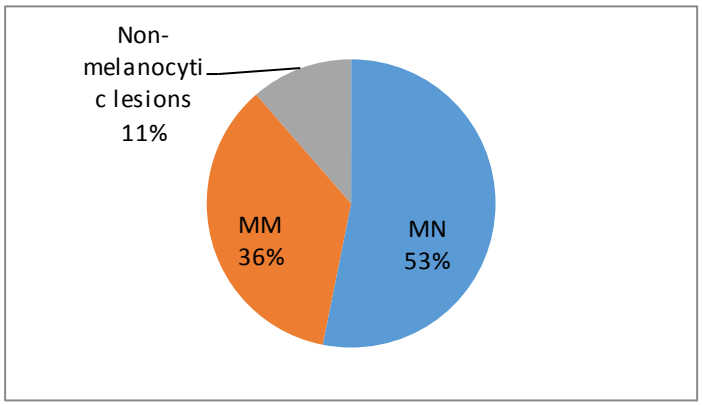
B. *Materials, Methods, Results*

In preparing this paper we used, on condition of anonymity, 79 dermoscopy images taken by Dr. Daniel BODA by using an experimental assembly consisting of a dermatoscope attached to a mobile phone. Image capture was immediately followed by storing it in a database of images, transmission being made through a wireless connection.

Received images were then processed by applying a series of processing methods - normalization, segmentation, restoration, extracting color, edge detection and various structures in the image.

In parallel, Dr. Daniel Boda put at our disposal the results previously obtained by using clinical diagnosis in order to make a comparative study of the results.

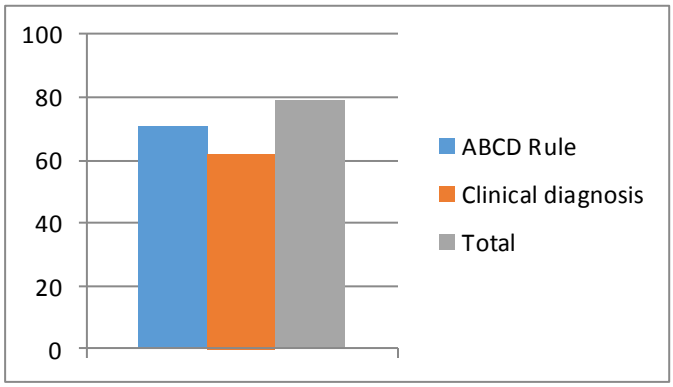
Of the 79 cases included, the investigation revealed 28 MM, 42 MN and nine non-melanocytic lesions.



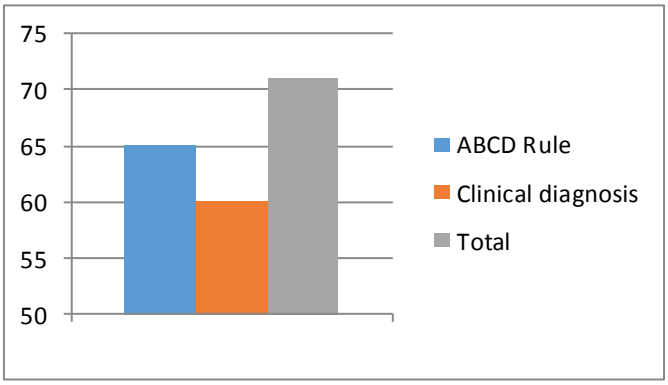
MM group consisted of 19 superficial melanoma (SSM) (66.7% of MM), 5 MM nodular (NMM) (17.4% of MM) and four with a different classifications (15.9% of MM). Breslow depth was 1.11 mm (± 1.03 SD, range, 0.15 - 4.10 mm).

Of the 79 lesions, 70 were melanocytic and 9 were non-melanocytic histologically pigmented lesions. In total, 71 pigmented lesions (90.2%) were diagnosed correctly with the

ABCD rule (melanocytic lesions or non-melanocytic lesions) compared with 62 (79.4%) correct clinical diagnoses.



Using the ABCD rule, 65 of the 71 melanocytic lesions histologically proven were correctly classified (91.3%), while the clinical diagnosis was correct in 60 cases (85.5%).



C. Methods

By the nature of the activities involved in the objective of our study, this article actually heralds the objectives and activities of the next stages of the research project. Therefore, this work was carried out in several directions of study.

1. Image normalization methods
2. Image segmentation methods
3. Identification of the outline methods
4. Structural differences identification methods

In the first theme we treated the illumination normalization methods, which, together with noise or other disturbances that may occur during the image acquisition, can negatively influence the performance of analysis of an automated classification of such data.

The five methods that were investigated were: Ratio Method (MR), The morphology ratio method (MRMD), The dynamic ratio morphology (MRMD), Homomorphic filter (FH) [3] and Recombined Multiscale Retinex method (MRMR) [4].

Ratio Method (MR) requires only a single image for extracting illumination invariant features in an image. And further, unlike the other methods, MR does not require a phase alignment of the images. MR is based on modeling Lambertian reflections of light on an object. This model considers that any light source falling on a 3D object can be decomposed into three factors: surface normal, "albedo" factor and point source of light. Auto-aspect ratio is defined as the ratio between the original image and the processed image where the processed image is a filtered version of the original image. Typically, the filtration is performed using a Gaussian filter weighted.

The morphology ratio method (MRMD). The basic idea of this method is given by morphological operations of the image to estimate the illumination component of an image.

More specifically it's about using morphological closing operator. An operator of the closure (CL) is represented by a non-linear filter consists of two stages: a dilation followed by erosion. This filter tends to eliminate noise characteristics and dashed edges in the image while retaining well-defined edges (large scale features). Unlike the classical Gaussian filter that smoothes a noisy image, the morphological filter type distort the closure to a lesser extent, the image quality (particularly observed in the vicinity of the edges that remain sharp).

The dynamic ratio morphology (MRMD). A disadvantage of this filter is that it uses a window of constant size convolution morphological representative image. Large scale features are normalized effective but on the other hand is not effective illumination compensation in dark areas (locally). For this reason, a variation of the method proposed by the MRM, called dynamic MRM (MRMD), which uses three different sized windows. The dimensions will vary depending on three parameters determined experimentally.

Homomorphic filter (FH) considers that an image can be decomposed as the product of a lighting model and one of reflection. By applying a logarithmic function, the multiplicative relationship turns in an additive relationship so that the two components can be separated. Applying further a Fourier transform, using a high-pass filter,

low frequency components (usually associated lighting component) are removed. By determining an optimal cutting frequency method gives satisfactory results.

Recombined Multiscale Retinex method (MRMR) consists in using a set of Gaussian filters to the original image and recombination results in well-defined proportions on the basis of the parameter recombination.

As a preprocessing step, we used the method of decomposition into components and for the classification technique we used proximity support vector method.

III. Conclusion

Tests have shown that the method that led to the greatest increase in the rate of recognition of color, edges and structures (from 46.11% to 91.67%) is MR.

In the next stage, we will analyze five color recognition methods to identify the algorithm with the highest success rate

ISBN: 978-1-61804-233-0

in order to improve the algorithm of ABCD method according to each parameter involved.
Also, for the next stage of the research we will introduce and present "parameter E" – evolution, to determine to what extent we can automatically determine the degree of evolution of the lesions using image processing algorithms and the impact on the accuracy of the results obtained in compared with the conventional methods of diagnosis.

IV. References

[1]Nachbar F, Stolz W, Merkle T, et al. The ABCD rule of dermatoscopy. J Am Acad Dermatol. 1994;30:551-559.

[2] Medical Image Processing, Reconstruction and Restoration: Concepts and Methods, Jiri Jan. CRC Press 2006

[3] http://en.wikipedia.org/wiki/Homomorphic_filtering

[4] Zia-ur Rahman, Daniel J. Jobson, Glenn A. Woodell. Multi-scale retinex for color image enhancement. ICIP (3), 1996

Study on EEG Steady State Alpha Brain Wave Signals Based on Visual Stimulation for FES

I.S. Isa, Z. Hussain, S.N. Sulaiman, N.H.Hamzah

***Abstract*—**This study is conducted to determine the steady state condition of EEG alpha brainwave signal between male and female subject. The brainwave signals measured from scalp using surface electrodes and EEG (Electroencephalograph) to detect the brain activities based on visual stimulation. From the analyzing results, it indicates that alpha waves during periods of eyes closed are the strongest EEG brain signal.Alpha waves is the most suitable signals to be integrated for FES stimulation as it consistency in producing electrical brainwave signals The findings of this conceptual study can be recommended to be implemented to control amount of FES induced based on physical activities specifically for stroke rehabilitation therapy.

***Keywords*—**Alpha wave, steady state response, brainwaves activities, EEG.

I. Introduction

THEnumber of existing neurons in human brain cortex which being activated in synchronize patterns are resulting in certain rhythmic behavior. The potential stimulation produced in brain cortex is recordable with the recommended electrodes position on the skull. The potential patterns occur due to electrical rhythms and transient discharge represents in electroencephalogram (EEG). EEG signals can be classified based on skull positions, frequency ranges, amplitudes, signal waveforms, periods and signal-induced actions. Basically, the EEGs signals are synchronize when the external stimulated has been measured. The EEGs are affected due to different degree of alertness as example the separate sleeping periods result in different EEG characteristics. There are several techniques proposed by Do *et al.*[1], Domino *et al.* [2], and Jacobs & Friedman [3],in integration the EEG and FES and the challenging of EEG signals mainly due to its small amplitude. The EEG signal passes through dura, cerebrospinal fluid and skull to scalp will produces peak-to-peak amplitude is only about 1 ~ 100μV with frequency range 0.5 ~ 100 Hz. In addition, the electrode material, contact tightness and electrode paste may even affect the recordings due to some unpredictable noise which interfere with EEG detection.

Brain waves are measured in cycles per second or Hertz (Hz) also known as frequency of brain wave activity. The lower number of frequency indicates that the slower the brain activity or the slower the frequency of the activity. The study from Teplan [8]has been categorized brain waves signal into four basic groups as shown in Table 1.

TABLE 1
BRAINWAVES SIGNALS WITH DOMINANT FREQUENCY BAND

Brainwaves	Frequency (Hz)
Beta (β)	>13
Alpha (α)	8 – 13
Theta (θ)	4 – 8
Delta (δ)	0.5 – 4

The objective of this study is to measure an alpha wave electrical activity and observed the effect on visual stimulation of alpha wave for healthy normal brain signal. This is the motivation to conduct this research since the best-known and most extensively studied rhythm of the human brain is the normal alpha rhythm as proposed by Teplan [8]. According to Rahman [7], the normal alpha rhythm is the best-known and most extensively studied rhythm of the human brain because alpha wave can be usually observed better in the posterior and occipital regions with typical amplitude about 50 μV. Previous established research by Klimesch [4] has reviewed and analyzed evidences of alpha and theta oscillations to determine the performance of cognitive and memory based on power contribution. From the reviewed findings, the study proposed that the encoding of new information is reflected by theta oscillations in hippocampo-cortical feedback loops, whereas search and retrieval processes in (semantic) long-term memory are reflected by upper alpha oscillations in thalamo-cortical feedback loops. A study conducted by Jacobs and Friedman [3] analyzed the central nervous system effects of relaxation technique using spectral analysis of alpha and theta EEG activity in all cortical regions. The findings indicate that relaxation technique produced significantly greater increases in theta activity in multiple cortical regions compared to the music condition. A study by Domino *et al.* [2]to determine either tobacco smoking in humans produces localized or widespread neocortical dominant alpha EEG frequency of the brainstem activating system. From the study demonstrates the consistency of tobacco smoking produces widespread bilateral neocortical increases in dominant alpha EEG frequencies.

II. Methodology

A. EEG Circuit Principle

In the experiment, the electrodes place on the positions of frontal, Occipital and A1 as potential ground as shown in Figure 1. These positions had been chosen since the presences

of hair on the scalp make it difficult to place electrodes on the proper region. The isolation concept was incorporated in circuit designing to avoid electrical shock caused by power supply or measuring instrument leakage.

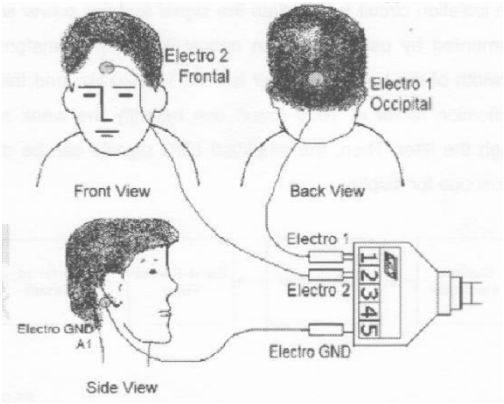


Fig. 1Position of electrodes

Figure 2 shows the block diagram of EEG measurement circuit. The surface electrodes were used to measure the alpha wave that was induced by stimulation the eye with light. Arrhythmic alpha waves were responsive if eye opened or closed. The instrumentation amplifier with gain 50 was used as the preamplifier for picking up the unipolar component of EEG signals. The isolation circuit functions to isolate the signal and line power source. The bandwidth of the band-pass filter is from 1 to 20 Hz and the amplifier with amplification factor of 1000 can magnify the weak signal that passed through the filter. An amplified EEG signals directly sent to oscilloscope for display.

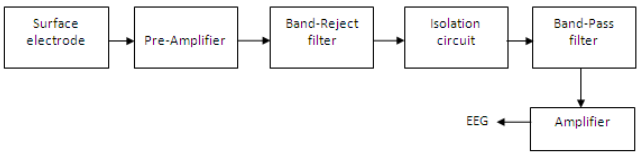


Fig. 2Block diagram of EEG measurement circuit

Preamplifier circuit is composed of an instrumentation amplifier with gain determination using (1). Figure 3 is showing preamplifier circuit where OP1 is the reference terminal and Z_2 is the compensated potential that can adjust the drift level of output voltage.

$$A_v = \frac{49.4k\Omega}{Z_1} + 1 \qquad (1)$$

Band-reject filter composed of RC networks including OP2A where the center frequency can be determined using (2). Figure 4 shows a twin-T band-reject filter.

$$f = \frac{1}{2\pi Z_3 Z_5} \qquad (2)$$

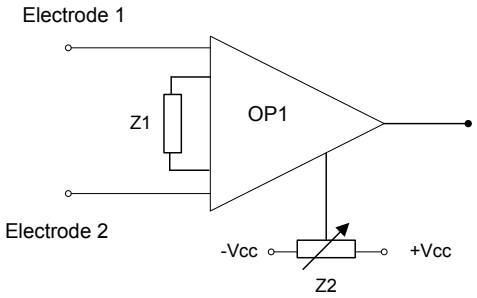


Fig. 3Preamplifier circuit

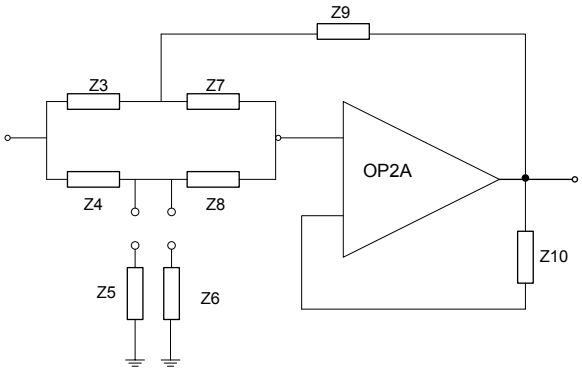


Fig. 4Band-Reject filter circuit

B. EEG Measurement Technique

Three subject volunteers in this study were selected from a group of 2 male and 1 female students. The inclusion criteria were adult, mentally and physically healthy males 21 to 25 years of age. The subjects were installed with EEG electrodes on the scalp and connected to computer to display brain wave signals. Figure 8 shows the measurement procedure during lab session. The subjects were asked to calm and relax with open eye. After a minute, subjects were asked to closed eye for five times and the signals were recorded. The step repeated two or three time to obtain smooth and clear waves. The same procedures were conducted for all three subjects.

Fig. 5Male subject

ISBN: 978-1-61804-233-0

Fig. 6Female subject

III. Results And Discussion

The EEG recordings even though focusing only on alpha waves, there were no significant differences between both subjects on frontal alpha waves. However the amplitudes during eye closing are higher compare to eye opening.Based on visual analysis, EEG alpha brainwaves of male subjects are showing much higher frequency compare to female subject.This can explain that female subject is much relaxed and calm compare to male subject.Table 2 tabulates the comparison results in terms of minimum, maximum and standard deviation of alpha wave activities. These results are provides strong evidences that alpha activity is induced by closing the eyes and by relaxation, and abolished by eye opening or alerting by any mechanism (i.e: thinking or calculating) [7]. All Fig 7 to Fig 13 are results recorded of all three subjects during experimental session. Fig 14 is shown the steady state response of alpha signal where comparison between male and female subject is having a little bit different. Male subject tend to have higher amplitude (>8µvolt) compare to female (<8µvolt).

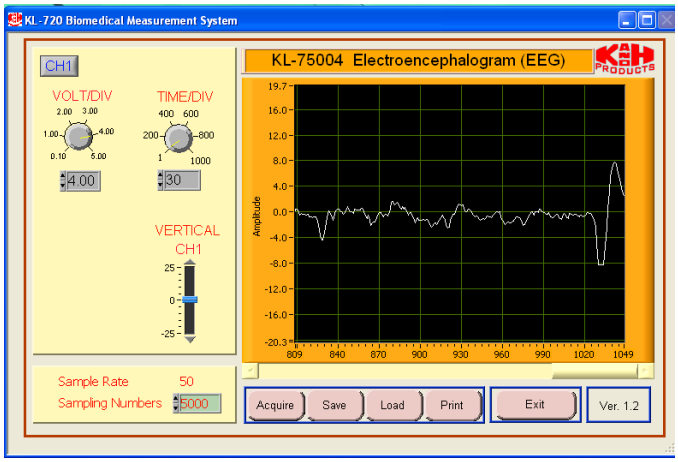


Fig. 7Rest / No activity(Male student 1)

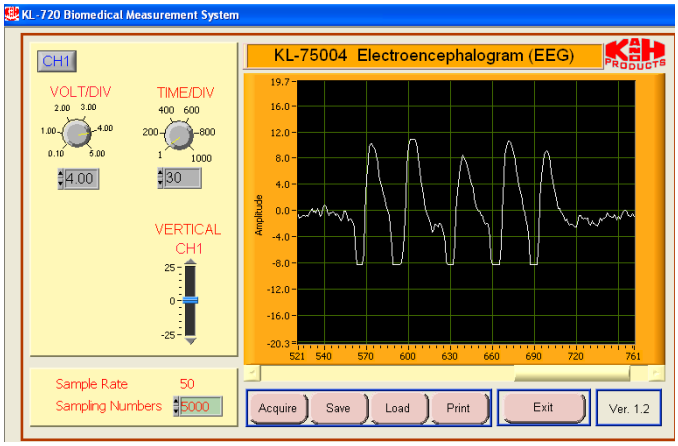


Fig. 8Eye Opening and Closing (Male student 1)

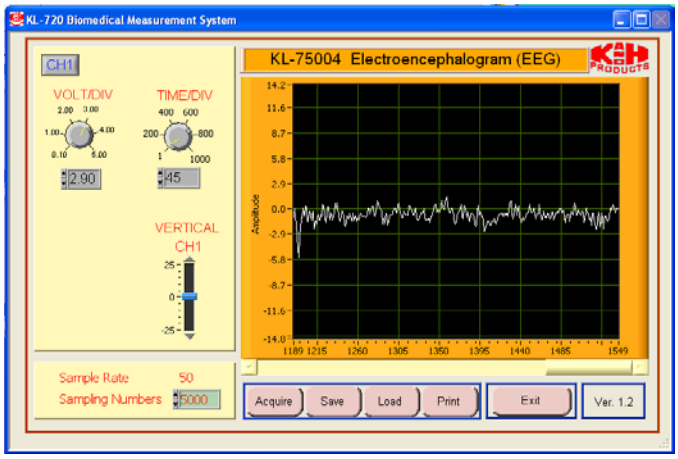


Fig. 9Rest / No activity (Female student)

ISBN: 978-1-61804-233-0

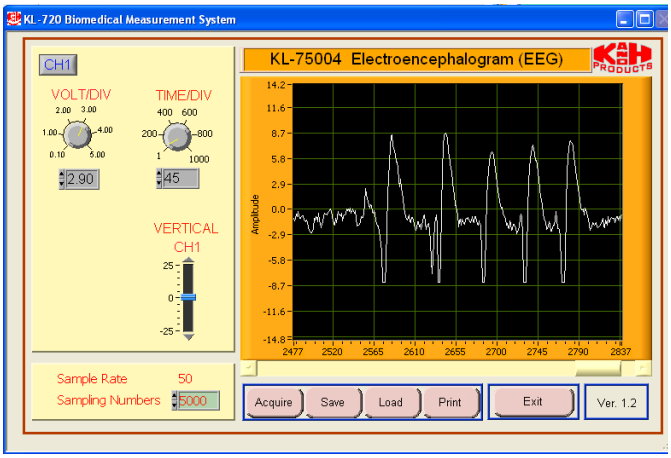


Fig. 10Eye Opening and Closing (Female student)

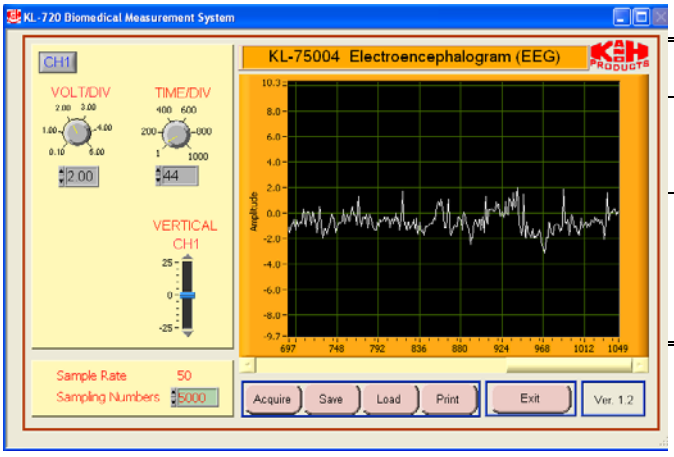


Fig. 11Rest / No activity (Male student 2)

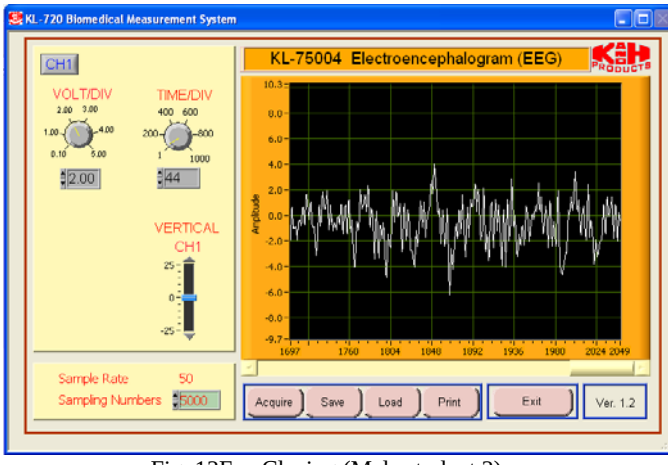


Fig. 12Eye Closing (Male student 2)

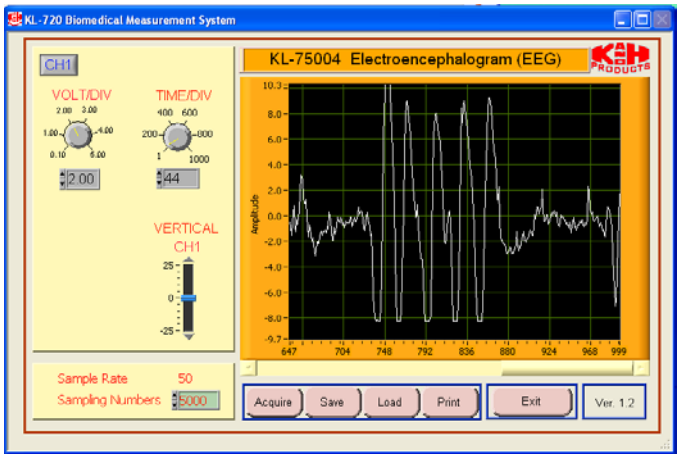


Fig. 13Eye Opening and Closing (Male student 2)

TABLE 2
STEADY STATE RESPONSE OF ALPHA SIGNAL (MALE STUDENTS)

	Blinking Eye			Non-Blinking Eye		
	Average Max Value	Average Min Value	Standard Deviation	Average Max Value	Average Min Value	Standard Deviation
Male Student 1	8.389	- 8.32	0.070	0.210	- 0.799	0.017
Male Student 2	7.823	- 8.23	0.123	0.378	- 0.959	0.020

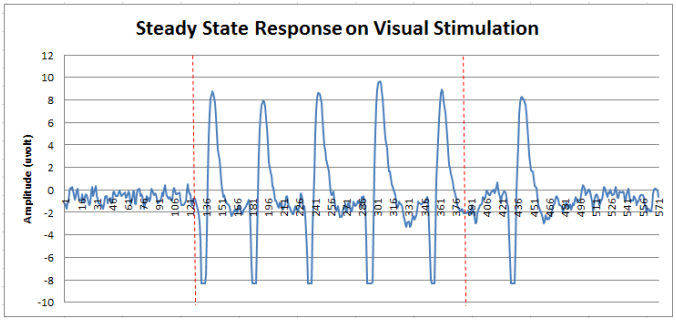


Fig. 14Steady state response of alpha signal

IV. CONCLUSION

This paper investigated the activities of alpha wave brain signal on visual stimulating. Alpha waves are reduced with open eyes and during periods of eyes closed are the strongest EEG brain signals. Alpha waves is the most suitable signals to be integrated for FES stimulation as it consistency in producing electrical brainwave signals as recommended by [7] that each action will produce its own signal and this is the easy way to recognize the action done by subject by refer to the brain signal pattern. This fundamental study is further experiment to determine the characteristics of alpha waves during body activities movement.

ISBN: 978-1-61804-233-0

ACKNOWLEDGMENT

This research is funded and supported byResearch Management Unit (RMU), UniversitiTeknologi MARA and Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS 2012) for project 600-RMI/FRGS 5/3 (71/2012).

REFERENCES

[1] Do, A.H., Wang, P.T., King, C.E., Abiri, A. &Nenadic, Z. 2011. "Brain-computer interface controlled functional electrical stimulation system for ankle movement". *Journal of neuroengineering and rehabilitation*, 8, 49.

[2] Domino, E.F., Ni, L., Thompson, M., Zhang, H., Shikata, H., Fukai, H., Sakaki, T. &Ohya, I. 2009. "Tobacco smoking produces widespread dominant brain wave alpha frequency increases". *International Journal of Psychophysiology*, 74, 192-198.

[3] Jacobs, G.D. and Friedman, R. 2004. EEG spectral analysis of relaxation techniques. *Applied psychophysiology and biofeedback*, 29, 245-254.

[4] Klimesch, W. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29, 169-195.

[5] Moreno, J.C., Del Ama, A.J., De Los Reyes-Guzman, A., Gil-Agudo, A., Ceres, R. and Pons, J.L. 2011. Neurorobotic and hybrid management of lower limb motor disorders: a review. *Medical & biological engineering & computing*, 49, 1119-1130.

[6] Pfurtscheller, G., Muller-Putz, G.R., Scherer, R. and Neuper, C. 2008. Rehabilitation with brain-computer interface systems. *Computer*, 41, 58-65.

[7] Rahman, K., Ibrahim, B., Leman, A. &Jamil, M. 2012. "Fundamental study on brain signal for BCI-FES system development", Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on, 2012. IEEE, 195-198.

[8] Teplan, M. 2002. Fundamentals of EEG measurement. *Measurement science review*, 2, 1-11.

Semantic Analysis: An Approach to Improve Spotted Words Recognition System

Mohamed FEZARI, Ali Al-Dahoud

Abstract**—**The technique adopted in this work is based on : combining pertinent features used in automatic speech recognition system such as Crossing Zero, energy, Mel frequency cepstral coefficients (MFCC) and derivatives; combining efficient classifiers such as Dynamic time warping and HMM; and finally , improving the speech recognition by including syntactic and semantic processing .
The results show that by adding more computation (combination of DTW and HMM) decrease the rate of error from 17.37% and 12.22 to 09.22%, adding the syntactic agent the error rate is reduced to 8.20 %, finally adding the semantic agent led to error rate less than 6.5 %. To implement the approach on a real-time application, a Personal Computer interface was designed to control the movements of a five degree of freedom (DOF) robot arm by transmitting the orders via radio frequency circuits. The voice command system for the robot arm is designed. The main parts of the robot are a microcontroller from Microchip PIC16F628 and a radio frequency wireless module. The possibilities to control other parts of the automatic robot arm are investigated by inserting sensors.

Keywords**—** Continuous Speech recognition, Robot Arm, syntactic and semantic processing, speech features

I. Introduction

Human-robot voice communication interface has a key role in many applications. Robot arms, or *manipulators*, comprise a 2 billion dollar industry. Bolted at its shoulder to a specific position in the assembly line, the robot arm can move with great speed and accuracy to perform repetitive tasks such as spot welding, manipulating chemical products and painting. In the electronics industry, manipulators place surface-mounted components with superhuman precision, making the portable telephone and laptop computer conception possible [1].
Yet, for all of their successes, these commercial robots suffer from a fundamental disadvantage: lack of human voice control. A fixed manipulator arm has a limited range of commands provided by an operator. Mainly using : a keyboard, joystick, mouse or pre-programmed sequences.

This paper proposes a new approach to the problem of the robot arm command. Based on the recognition of spotted words within a sentence, using a set of traditional pattern recognition in speech recognition and a combined discrimination approach based on test results of classical methods [2][5] and [7] in order to increase the rate of recognition. The increase in complexity, by adding more agents such as syntactic and semantic processing, as compared to the use of only traditional approach is negligible, but the system achieves considerable improvement in the recognition phase, thus facilitating the final decision and reducing the number of errors in decision taken by the voice command system.

Moreover, speech recognition constitutes the focus of a large research effort in Artificial Intelligence (AI), which has led to a large number of new theories and new techniques. However, it is only recently that the field of robot arm control and AGV navigation have started to import some of the existing techniques developed in AI for dealing with uncertain information.
Hybrid method is a simple, robust technique developed to allow the grouping of some basic techniques advantages. It therefore reduces the rate of error in pattern recognition . The selected features are : Zero Crossing and Extremes (CZEXM), Linear Predictive Coefficient (LPC) parameters , Energy Segments (ES), and cepstral coefficients (MFCC: Mel Frequency Cepstral coefficients). The combination of two classifiers DTW and HMM is used. Then, to improve the recognition by eliminating ambiguity, two agents are added namely: syntactic process and semantic process of the sentence. This study is part of a specific application concerning robot control by simple sentence commands. The sentence is composed of a set of spotted Arabic words in order to control the directions of five DOF robot arm. It has to be implemented on a DSP [8] and has to be robust to any background noise generated by the environment of the voice command system.
The aim of this paper is therefore the recognition of spotted words from a limited vocabulary in the presence of background noise. The application is speaker-dependent. Therefore, it needs a training phase. It should, however, be pointed out that this limit does not depend on the overall approach but only on the method with which the reference patterns were chosen. So by leaving the approach unaltered and choosing the reference patterns appropriately, this application can be made speaker-independent [9].

As application, a vocal command for a five DOF robot arm is chosen. The robot arm is “TERGANE -TR45”. There have been many research projects dealing with robot arm control, among these projects, there are some projects that build intelligent systems [10-12]. Since we have seen human-like robots in science fiction movies such as in “I ROBOT”, making intelligent robots or intelligent systems became an obsession within the research group. Voice command needs the recognition of spotted words from a limited vocabulary

used in Automatic Robot Arm Control system (RACS)[13] . the paper is presented as follow: in section 2, we present the application , in section 3 the speech recognition agent is detailed, the hardware part is presented in section 4, finally in section 5 and 6 we present the tests and results [2]followed by a discussion and a conclusion.

II. Description of Application

The application is based on the voice command for a set of degree of freedom for a robot arm T45. It therefore involves the recognition of isolated words from a limited vocabulary used to control the movement of selected parts of the arm.

The vocabulary is limited to nine words to select the arm part (upper limb, limb, hand and forceps or grip) and to command the selected part (up, down, left right and stop). These commands are necessary to control the movement of the T45, Up movement, Down movement, stop, turn left and turn right. The number of words in the vocabulary was kept to a minimum both to make the application simpler and easier for the user.

The user selects the robot arm part by its name then gives the movement order on a microphone, connected to sound card of the PC. A speech recognition agent based on hybrid technique recognises the words and send to the parallel port of the PC an appropriate binary code. This code is then transmitted to the robot T45 via a radio frequency emitter.

The application is first simulated on PC. It includes two phases: the training phase, where a reference pattern file is created based on pre-recorded data base, and the recognition phase where the decision to generate an accurate action is taken. The action is shown in real-time on parallel port interface card that includes a set of LED's, showing what command is taken, and a radio Frequency emitter.

III. SPEECH RECOGNITION MODULE

The speech recognition agent is based on a traditional pattern recognition approach. The main elements are shown in the block diagram of Figure 1. The pre-processing block is used to adapt the characteristics of the input signal to the recognition system. It is essentially a set of filters, whose task is to enhance the characteristics of the speech signal and minimize the effects of the background noise produced by the external conditions and the motor. A Kalman filter is used to enhance the word command regarding the other words [14].

The Speech locator (SL) block detects the beginning and end of the word pronounced by the user, thus eliminating silence. It processes the samples of the filtered input waveform, comprising useful information (the word pronounced) and any noise surrounding the PC. Its output is a vector of samples of the word (i.e.: those included between the endpoints detected).

The SL implemented is based on analysis of crossing zero points and energy of the signal, the linear prediction mean square error computation helps in limiting the beginning and the end of a word; this makes it computationally quite simple.

A. *Features extraction:*

Features extraction means finding good data that hepls to categorize the healthy status of patient, features selection make a boundary between each class.

The Features extraction block analyses the signal, extracting a set of parameters with which to perform the recognition process. First, the signal is analysed as a block, the signal is analysed over 20-mili seconds frames, at 256 samples per frame. Five types of parameters are extracted: Normalized Extremes Rate with Normalized Zero Crossing Rate (CZEXM), LPC coefficients (Ai), Energy Segments (ES) and cepstral parameters MFCC's [4]

The Mel-Filter Cepstral Coefficients are extracted from the speech signal . The speech signal is pre-emphasized, framed and then windowed, usually with a Hamming window (PE-FB-W). Mel-spaced filter banks are then utilized to get the Mel spectrum. The natural Logarithm is then taken to transform into the Cepstral domain and the Discrete Cosine Transform is finally computed to get the MFCCs. as shown in the block diagram of Figure 3.2

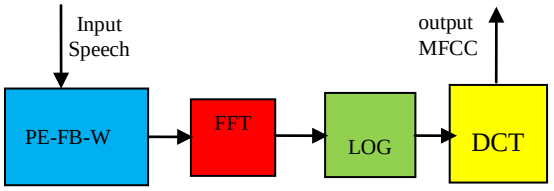


Fig. 3.2 : *MFCC block diagram*

Mel frequency cepstral coefficients are given by:

$$\check{C} = \sum_{k=1}^{K} \log(\tilde{S}) \cos[n(k-\tfrac{1}{2})]\frac{\Pi}{K} \qquad (1)$$

These parameters are extracted by 32 filter bank applied on 10 ms (256 points) Hamming windowed frames at 50% of overlap.

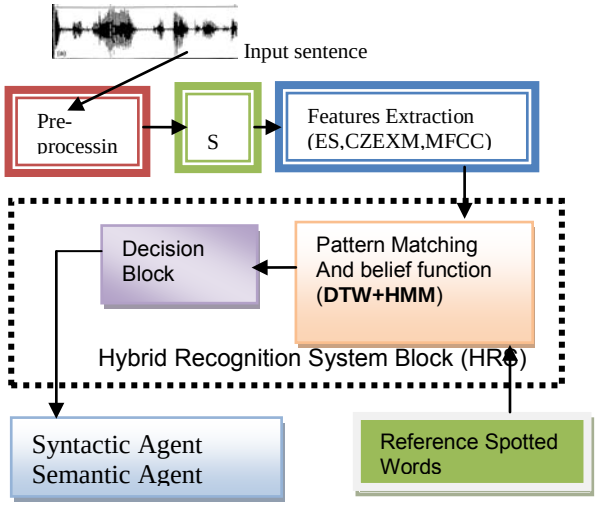


Fig. 1: BLOCK DIAGRAM

These parameters were chosen for computational simplicity reasons (CZEXM, ES), robustness to background noise (12 Cepstral parameters) and robustness to speaker rhythm variation (DTWE) [7].

The reference pattern block is created during the training phase of the application, where the user is asked to enter ten times each command word. For each word and based on the ten repetitions, ten vectors of parameters are extracted from each segment and stored.

B. Hybrid Recognition System Block (HRS)

This block main job is to test the referenced spotted words with incoming sentence word. It designed using two classifiers in parallel then based on response of each one and based on belief function the block select the corresponding word from reference

B.1 Dynamic Time Warping (DTW)

Dynamic Time Warping algorithm (DTW) [19] is an algorithm that calculates an optimal warping path between two time series. The algorithm calculates both warping path values between the two series and the distance between them.

Suppose we have two numerical sequences (a_1,a_2, ..., a_n) and (b_1, b_2, ..., b_m). As we can see, the length of the two sequences can be different. The algorithm starts with local distances calculation between the elements of the two sequences using different types of distances. The most frequent used method for distance calculation is the absolute distance between the values of the two elements (Euclidian distance). That results in a matrix of distances having *n* lines and *m* columns of general term:

$$d_{ij} = \left| a_i - b_j \right| \; i=1..n \text{ and } j=1..m \qquad (2)$$

Starting with local distances matrix, then the minimal distance matrix between sequences is determined using a dynamic programming algorithm and the following optimization criterion:

$$a_{ij} = d_{ij} + min(a_{i-1,j-1}, a_{i-1,j}, a_{i,j-1}), \qquad (3)$$

where a_{ij} is the minimal distance between the subsequences (a_1,a_2, ..., a_i) and (b_1, b_2, ..., b_j).

A warping path is a path through minimal distance matrix from a_{11} element to a_{nm} element consisting of those a_{ij} elements that have formed the a_{nm} distance.

The global warp cost of the two sequences is defined as shown below:

$$GC = 1/P \sum_{i=1}^{P} Wi \qquad (4)$$

where w_i are those elements that belong to warping path, and *p* is the number of them . There are three conditions imposed on DTW algorithm that ensure them a quick convergence:

- monotony – the path never returns, that means that both indices i and j used for crossing through sequences never decrease.
- continuity – the path advances gradually, step by step; indices i and j increase by maximum 1 unit on a step.
- boundary –the path starts in left-down corner and ends in right-up corner.

Because optimal principle in dynamic programming is applied using "backward" technique, identifying the warp path uses a certain type of dynamic structure called "stack". Like any dynamic programming algorithm, the DTW one has a polynomial complexity. When sequences have a very large number of elements, at least two inconveniences show up:

- memorizing large matrices of numbers;
- performing large numbers of distances calculations.

Words identification can be performed by straight comparison of the numeric forms of the signals or by signals spectrogram comparison.

The comparison process in both cases must compensate for both the different length of the sequences and non-linear nature of the sound. The DTW Algorithm succeeds in sorting out these problems by finding the warp path corresponding to the optimal distances between two series of different lengths.

Figure 1.a illustrates the DTW comparison between reference word and test word:

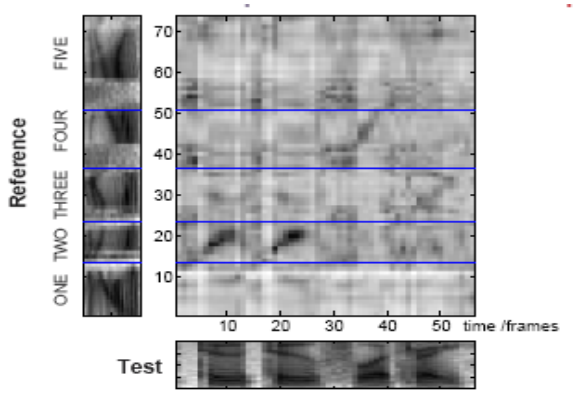


Figure 3.a Dynamic Time-Warping on two 1D signals

B2. HMM for isolated words

A Hidden Markov Model (HMM) as shown if figure 3.b is a type of stochastic model appropriate for non stationary stochastic sequences, with statistical properties that undergo distinct random transitions among a set of different stationary processes. In other; words, the HMM models a sequence of observations as a piecewise stationary process. Over the past years, Hidden Markov Models have been widely applied in several models like pattern [6], or speech recognition [6][9] . The HMMs are suitable for the classification from one or two dimensional signals and can be used when the information is incomplete or uncertain. To use a HMM, we need a training phase and a test phase. For the training stage, we usually work with the Baum-Welch algorithm to estimate the parameters (Π_i ,*A*,*B*) for the HMM [7, 9]. This method is based on the maximum likelihood criterion. To compute the most probable state sequence, the Viterbi algorithm is the most suitable.

A HMM model is basically a stochastic finite state automaton, which generates an observation string, that is, the sequence of observation vectors, $\boldsymbol{O}=\boldsymbol{O}_1,..\boldsymbol{O}_t$,… ,$\boldsymbol{O}_T$. Thus, a HMM model consists of a number of N states $S=\{S_i\}$ and of the observation string produced as a result of emitting a vector $\boldsymbol{O}_t$ for each successive transitions from one state S_i to a state S_j. $\boldsymbol{O}_t$ is d dimension and in the discrete case takes its values in a library of M symbols.

The state transition probability distribution between state S_i to S_j is $A=\{a_{ij}\}$, and the observation probability distribution of emitting any vector $\boldsymbol{O}_t$ at state *Sj* is given by $B=\{b_j(\boldsymbol{O}_t)\}$. The probability distribution of initial state is

$\Pi=\{\pi_i\}$.

$$a_{ij} = P(q_{t+1} = S_j / q_t = S_i) \quad (3.1)$$

$$B=\{b_j(\boldsymbol{O}_t)\} \quad (3.2)$$

$$\pi_i = P(q_0 = S_i) \quad (3.3)$$

Given an observation O and a HMM model $\lambda=(A,B,\Pi)$, the probability of the observed sequence by the forward-backward procedure $P(O/\lambda)$ can be computed [10]. Consequently, the forward variable is defined as the probability of the partial observation sequence $O_1O_2,....O_t$ (until time t) and the state S at time t, with the model λ as $\alpha(i)$. and the backward variable is defined as the probability of the partial observation sequence from t+1 to the end, given state S at time t and the model λ as $\beta(i)$. the probability of the observation sequence is computed as follow:

$$P(O/\lambda) = \sum_{i=1}^{N} \alpha_t(i) * \beta_t(i) = \sum_{i=1}^{N} \alpha_T(i) \quad (3.4)$$

and the probability of being in state I at time t, given the observation sequence $\boldsymbol{O}$ and the model λ is computed as follow:

$$\pi_i = P(q_0 = S_i) \quad (3.5)$$

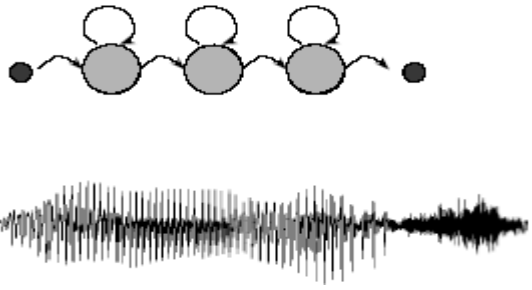


Fig. 3.b: ***Presentation of left-right (Bakis) HMM***

The matching block compares the reference patterns and those extracted from the input signal. The matching decision integrates: a hybrid recognition block based on two methods, and a belief function to decide in case of a conflict between both classifiers (DTW and HMM). The hybrid module is built on applying both methods in parallel, then combines their results based on belief function which works as a weighting vector , in other word the results of each method applied independently were used to generate a Weight vector.

Tests were made using each method separately. From the results obtained, a weighting vector is extracted based on the rate of recognition for each method. Figure 1 shows the elements making up the main blocks for the hybrid recognition system (HRS).

Table 1. The meaning of spotted words used in command

Word	Meaning
1) Kaiida	Base motor(M0) rotate (left,right)
2) Diraa	Upper limb motor (M1)
3) Saad	Limb motor (M2)
4) Meassam	Wrist (hand) motor (M3)
5) Mikbath	Grip motor (M4)
6) yamine	Right (M0)
7) Yassare	Left (M0)
8) Fawk	Up movement M1, M2 and M3
9) Tahta	Down movement M1, M2 and M3
10) Eftah	Open Grip, action on M4
11) Eghlak	Close grip, action on M4
12) Kif	Stop the movement, stops M1, M2, M3 or M4
13) Yade	Key word means "Arm"
14) Tabeek	Execute for all motors

C . Syntactic agent

Based on some syntactic language rules created and saved in a dictionary the agent checks for the order of spotted words within the uttered phrase. The syntax for each sentence should follow the natural language rule i.e. "key-word + Robot-name + action". The key-word is necessary as a security so that the system would not replay to any generated phrase that might contain the spotted word and spoken by another person than the operator. If there is any syntactic error or non defined syntax in dictionary then the agent will not provide any action to the following agent. As example, if the operator says "you red go forward", in this case the key-word "robot" has not been detected n the sentence therefore the command is refused.

The rules are:

Phrase= key-word + nominal-group + verb + complement.

Nominal-group= determinant + name [+ preposition+ nominal-group].

D. Semantic agent

The uttered phrase should have a meaning, because we can construct correct phrases using the spotted words however they have no meaning and therefore the system can detect errors and hence it will not generate commands (i.e. "robot eats red carrots" or "robot go forward next to red table", or "robot blue go to the table in its right"). This luck of precision leads the agent to not classify the sentence as a correct one. This type error is due in general to some phonemes of different successive words in the phrase. The semantic agent has a rule semantic table, in which accepted set of words are stored in the table called semantic table.

E. Vocabulary (Spotted words)

A voice command system and an interface to the robot arm TR-45 are implemented. The application is based on the voice command for a set of five stepper motors. It therefore involves the recognition of spotted words from a limited vocabulary used to recognise the elements and control the movement of a robot arm.

The vocabulary is limited to fourteen words divided into two subsets: object name subset necessary to select the part of the robot arm to move and command subset necessary to control the movement of the arm example: turn left, turn right and

stop for the base (shoulder), Open close and stop for the gripper. The number of words in the vocabulary was kept to a minimum both to make the application simpler and easier for the user.

The user selects the robot arm part by its name then gives the movement order on a microphone, connected to sound card of the PC. The user can give the order in a natural language phrase as example: "**Yade**, **gripper open execute**". A speech recognition agent based on HMM technique detects the spotted words within the phrase, recognises the main word "Yade" witch is used as a keyword in the phrase, it recognises the spotted words, then the system will generate a byte where the four most significant bits represent a code for the part of the robot arm (gripper in this case) and the four less significant bits represent the action to be taken by the robot arm (open in this case). The sentence should finish with the spotted word execute (it is and end of the sentence and an order to be executed: like return key in keyboard). Finally, the byte is sent to the parallel port of the PC and then it is transmitted to the robots via a radio frequency emitter.

The application is first simulated on PC. It includes three phases: the training phase, where a reference pattern file is created, the recognition phase where the decision to generate an accurate action is taken and the appropriate code generation, where the system generates a code of 8 bits on parallel port. In this code, four higher bits are used to codify the object names and four lower bits are sued to codify the actions. The action is shown in real-time on parallel port interface card that includes a set four stepper motors to show what command is taken and a radio Frequency emitter.

IV. Robot Arm Control System

A. Robot Arm Description

As in Figure 3, the structures of the mechanical hardware and the computer board of the four DOF robot arm in this paper are similar to those in [5-6].

The system is composed of:

- A robot arm reference T45 from TEREL, France.
- A command module.
- A power supply box.

Four movements for the robot (base, shoulder, limb and hand) can be controlled. The command of each motor is realised by a moto-reductor block (1/500) powered by + and – 12 volts.

The recopy voltage is provided by a linear rotated potentiometer fixed on the moto-reductor block.

The grip, however is in open loop. The motors used in this manipulator arm are cc/2.5 watts/CCL potentiometers.

Base displacement is 290°.

Shoulder displacement is 107°.

Limb displacement is 280°

And Hand displacement is 360°

B. The New Design for Tele-Guiding The TR-45

A parallel port interface was designed to show the real-time commands. It is based on two TTL 74HCT573 Latchs and 10 light emitting diodes (LED), 5 green LED to indicate each recognized word for robot arm parts ("kaiida", "Dirrar", "Saad", "Meassam", "mekbadh") and 7 yelow LED to indicate each recognised command word ("fawk", "tahta", "Yamine", "Yassar","eftah", "eghlak", "Kif"), respectively and a red LED to indicate wrong or no recognized word. Other LED's can be added for future insertion of new command word in the vocabulary example the words: "Iftah", "Ighlak" for the grip as shown in Figure 2.b. The order to be executed by the TR45 is transmitted by the wireless transmission circuit TXM-433-10, which provide a transmission rate of 10kb/s.

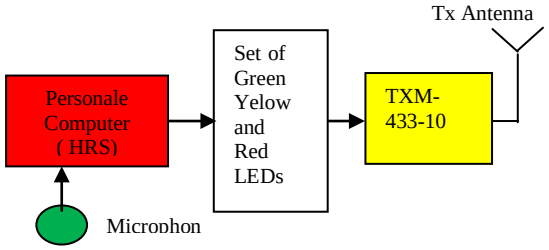


Fig. 2.b : Block Diagram of Parallel Port Interface.

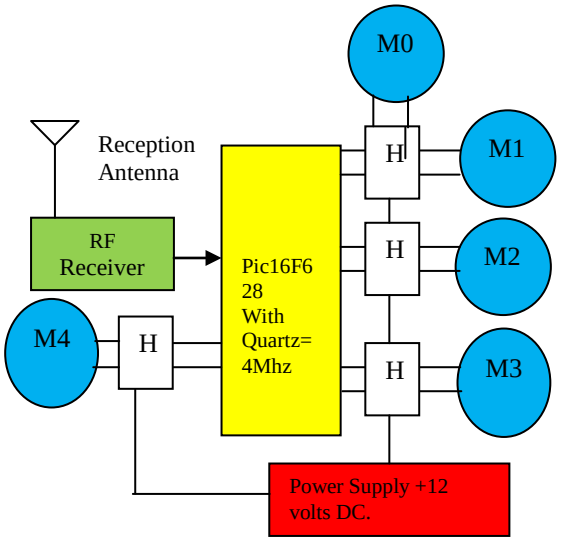


Fig. 3. Robot Arm Control System block diagram

C. TR45 Interface Description

However, since the robot arm in this paper needs to perform simpler tasks than those in [5-6] do, robot arm electronic control system can be designed simply. The computer board consists of a PIC16F628 , with 1K-instruction EEPROM (Electrically Programmable Read Only Memory) some analogue inputs[15], five H bridges drivers using BD134 and BD133 transistors for DC motors, a RF receiver module from RADIOMETRIX which is the SILRX-433-10 (modulation frequency 433MHz and transmission rate is 10 Kbs) [16]. In order to protect the microcontroller from power feedback signals, a set of opto-transistors were added. Each motor within the robot arm T45 is designed by a name and performs the corresponding task to a received command as in Table 1. Commands and their corresponding tasks affected to robot arm motors may be changed in order to enhance or adapt the application.

In the recognition phase, the application gets the word to be processed, treats the word, then takes a decision by setting the corresponding bit on the parallel port data register and hence

the corresponding LED is on. The code is also transmitted in serial mode to the TXM-433-10.

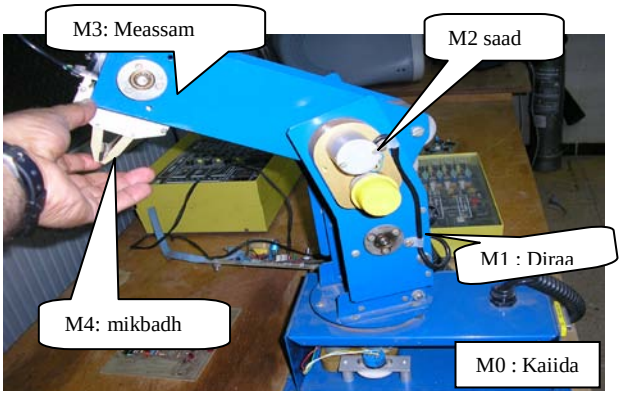


Fig.3.c. Photo of the Robot Arm T45.

V. Experiments Results and Discussion

The developed system has been tested within the laboratory of L.A.S.A. The tests were' done only on the eight command words. Three different scenarios were tested:

1. Combining Features with every Classifier and the results are presented in table2, with columns:
   a) ES+CZEM with DTW, b) MFCC with DTW
   c) ES+MFCC with DTW, d) ES+CZEM with HMM.
   e) MFCC with HMM, f) ES+MFCC with HMM.
2. Combining classifiers with different features, results are presented in table 3.
3. Adding Syntactic agent.
4. Adding Semantic agent to the best Features and combining (DTW+HMM).

For all tests, each command word is uttered 25 times. The recognition rate for each word is presented in table2.

From this table 2, we concluded that best features are the combination of Energy segments with MFCC and their first and second derivatives as shown in column c and column f of this table. We have selected these feature (ES and MFCC with first and second derivatives) then, we have combined the two classifiers DTW and HMM, the results are presented in column 3 of the table 3.

The tests gave better results by combining DTW and HMM as shown if column 4 and 5, the main difference is that in column 5 we applied the belief function based on results obtained in column 2 and 3 of that table however in column four we just computed the mean.

We observed that the word "kif" which means stop has high rate of recognition compared to the other words, combining DTW and HMM by taking into account the belief function of each one gave the better results as shown in column 5.

Table 2. Results of Different Features with DTW and HMM Classifiers.

Word	ES+MFCC & DTW	ES+MFCC & HMM	ES+MFCC & (DTW+HMM)/2	ES+MFCC & DTW+HMM
Ammame	75	83	79	88
Wara	84	86	85	91
Yamine	82	86	84	87
Yassar	78	89	83.5	88
Kif	83	90	86.5	94
Yade	87	89	88	93
Mikbadh	85	89	88	94
Eftah	87	91	89	92
Average	82.63	87.88	85.5	90.88

Table 3. Best Features with DTW and HMM

Word	a	b	c	d	e	f
Ammame	60	70	75	73	82	83
Wara	65	71	84	75	85	86
Yamine	70	75	82	78	86	86
Yassar	68	72	78	75	88	89
Kif	75	78	83	80	89	90
Yade	75	80	87	81	87	89
Mikbadh	78	82	85	83	90	89
Eftah	79	80	87	82	87	91
Average	71.25	76	82.63	78.38	86.75	87.88

Table 4. Adding Syntactic and Semantic Agents

Word	*Adding Syntactic Agent*	*Adding syntactic & Semantic Agents*
Ammame	90	90.5
Wara	93	93.5
Yamine	89	89.5
Yassar	92	92.8
Kif	94.1	94.6
Yade	93.5	93.5
Mikbadh	95	96.2
Eftah	93.5	94
Average	91.80	94

VI. Conclusion And Future Directions

Improvement in recognition rate was reached by combining pertinent features and interesting classifiers, more over the rates were improved by adding syntactic and semantic agents(as hearing impairment persons can understand the hole sentence based on syntactic and semantic sentence formulation). The results of the tests shows that a better recognition rate can be achieved inside the laboratory and especially if the phonemes of the selected word for voice

ISBN: 978-1-61804-233-0

command are quite different. However, a good position of the microphone and additional filtering may enhance the recognition rate. Several interesting applications of the proposed system different from previous ones are possible as mentioned in section 5.

Since the designed robot consists of a microcontroller and other low-cost components namely RF transmitters, the hardware design can easily be carried out.

The use of hybrid technique based on classical recognition methods makes it easier to separate the class represented by the various words, thus simplifying the task of the final decision block. Tests carried out have shown an improvement in performance, in terms of misclassification of the words pronounced by the user.

Beside the designed application, a hybrid approach to the implementation of an isolated word recognition agent HSR was used. This approach can be implemented easily within a DSP or a CMOS DSP microcontroller or even FPGA modules.

Finally we notice that by simply changing the set of command words, we can use this system to control other objects by voice command such as an electric wheelchair movements or a set of home appliances(TV ,curtain, lights on/off) . One line of research is under way is the implementation of feedback force sensors for the grip and adapted algorithms to get smooth objects and investigate the Intelligent Haptic Sensor System for Robotic manipulator[20].Another line of research will be investigated in the future relates to natural language recognition based on semantic and large data base.

REFERENCES

[1] M Mokhtari., Ghorbal M. and Kadouche R., "Mobilité et Services : application aux aides technologiques pour les personnes handicapées », in procedings 7th ISPSAlgiers, may, 2005, pp. 3948, 2005.

[2] Nguyen L., Belouchrani A., Abed-Meraim K. and Boa-shash B., "Separating more sources than sensors using time- frequency distributions." in Proc. ISSPA, Malysia, 2001.

[3] P. Arbeille, J Ruiz, M. Chevillet," Teleleoperated robotic Arm for Echographic Diagnosis in Obstetrics and Gynecology", Untrasound In Obstetrics and Gynecololy Vol 24. No.3 pp. 242-243, August 2004. .

[4] Gu L. and Rose K., "Perceptual Harmonic Cepstral Coefficients for Speech Recognition in Noisy Environment," *Proc ICASSP 200*1, Salt Lake City, Utah, May 2001.

[5] M. Fezari , and M. Bedda, "Hybrid technique to enhance voice command system for a wheelchair", ACIT'05, Al_Isra University, Jordan, 2005.

[6] Mohamed Fezari" HMM and Kalman Filters to improve voice command system for manipulator arm", paper accepted in proc. IDT'09 ,IEEE conference Riyadh, KSA

[7] Wang T. and Cuperman V., "Robust Voicing Estimation with Dynamic Time Warping," *Proceedings IEEE ICASSP'98*, pp. 533-536, May 1998.

[8] Hongyu L., Zhao Y., , Dai Y. and Wang Z., " a secure Voice Communication System Based on DSP", *IEEE 8th International Conf. on Cont. Atau. Robotc and Vision*, Kunming, China, 2004, pp. 132-137, 2004.

[9] Ahmed-Foitih Z., Nouibat W., Belarbi A.T. and Benkouider M., "Commande dynamique du robot Puma 762 en environnement virtuel" CIP 2001, juin 2001.Alger, Algérie, pp 38-43, 2001.

[10] Vitrani M., Morel G., and Ortmaier T., "Automatic guidance of a surgical instrument with ultrasound based visual servoing," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, April 2005.

[11] Reynolds D. **,** **"**Automatic Speaker Recognition, Acoustics and Beyond"**,** Senior Member of Technical Staff, MIT Lincoln Laboratory, JHU CLSP, 10 July 2002.

[12] Kim J. H. , et al, "Cooperative Multi-Agent robotic systems: From the Robot-Soccer perspective", *1997 Micro-Robot World Cup Soccer Tournament Proceedings*, Taejon, Korea, June 1997, pp. 3-14.

[13] Kwee H., Stanger C.A., "The Manus robot arm", Rehabilitation Robotics News letter, Vol. 5, No 2, 1993, H. G. Evers, E. Beugels and G. Peters, "MANUS toward new decade", Proc. ICORR 2001, Vol. 9, pp. 155-161, 2001

[14] Evers H. G., Beugels E. and Peters G., "MANUS toward new decade", Proc. ICORR 2001, Vol. 9, pp. 155-161, 2001

[15] Data sheet PIC16F628 from Microchip inc. User's Manual, 2003, Available at: http://www.microchip.com.

[16] Radiometrix components, TXm-433 and SILRX-433 Manual, HF Electronics Company. Available at: http://www.radiometrix.com.

[17] Laila ABDOUN and Mohamed FEZARI," Detection and Cmlassification of Abnormal sounds in a habitat', in Proc STA'2011 Sousse, Tunisia 2011.

[18] Mohamed FEZARI and Ali Aldahoud," Embedded Multiprocessors Vocal Command System: A way to Improve the tele-control of a Didactic manipulator Arm", in Proc. ICIT 5th int confetrence on IT IEEE Jordan section, ISBN:9957-8583-0-0 Al-Zaytoonah Univ. Jordan 2011.

[19] T. Wang and V. , "Robust Voicing Estimation with Dynamic Time Warping," Proceedings IEEE ICASSP'98, pp. 533-536, May 1998.

[20] C. pasca, P. Payeur, E.M. Petriu, A.M. Cretu, "Intelligent Haptic Sensor System for Robotic manipulator", in Porc. IMTC/2004, IEEE Instrum. Meas. Technol. Conf. pp. 279-284, Italy 2004.

[21] Kwee Hok, (1997). Intelligent control of Manus Wheelchair. *In proceedings Conference on Reabilitation Robotics, ICORR'97*, Bath 1997, pp. 91-94.

[22] M. Fezari, Attoui Hamza, Mouldi BEDDA "Arabic Spotted Words Recognition System Based on HMM Approach to control a didactic Manipulator Arm", In Proc. MS'08, nt. Conf. On Modelling and Simulation, PETRA/ Jordan, Vol. 2008.

[23] Djemili R. , Bedda M. , Bourouba H. Recognition Of Spoken Arabic Digits Using Neural Predictive Hidden Markov Models. International Arab Journal on Information Technology, IAJIT, 2004, **Vol.1, N°2**, pp. 226-233.

Mohamed FEZARI (MS'87-PHD'2006, Habilitation in directing Research: 2010) is associate Prof. at Badji Mokhtar Annaba University BMAU Faculty of Engineering Dept. Electronics , he is currently in LASA laboratory, he has many publications in different journals and participated in many conferences as author , reviewer and TPC.

Ali AL-DAHOUD is Professor at Zaytoonah University of Amman Jordan, Senior Member of IEEE , conference chair of ICIT'09 ICIt'11 and ICIT'13, has many Publications in different journals well indexed and actually he is in the faculty of IT.

Severe accidents management in PWRs

J. Rajzrová, J.Jiřičková

***Abstract*—** According to a new trend in safety upgrades in PWRs, the nuclear power plants have started to adopt strategies to mitigate events beyond the design envelope for nuclear energy facilities. This paper deals with implementation of some measures to existing nuclear power plants which do not include severe accident mitigation measures in their original design projects. It discusses the contribution of severe accident mitigation measures on the limitation of radiological consequences.

***Keywords*—** Severe accident, Severe accident management, Mitigation measures.

I. Introduction

In case of pressurized water reactors, severe accidents are the accidents associated with significant core damage. The phenomena occurring during severe accidents are as follows:

- Cladding damage
- Fuel melting and core degradation
- Fuel-coolant interaction in the reactor vessel
- High-pressure core melt ejection from the reactor vessel
- Slow reactor vessel melting through
- Hydrogen combustion
- Containment overpressurization
- Core-concrete interaction
- Containment by-pass

Cladding damage is caused by excessive heat up and pressure. These two factors acting on cladding can lead to loss of its integrity with gap release of fission products and also fuel fragmentation can occur in case of high burn-up. Fission products accumulated in the fuel matrix release if the core damage is not terminated and once fuel melting has started. The possible fuel-coolant interaction can lead to steam explosion with potential generation of missiles. High-pressure core melt ejection from the reactor vessel is followed by direct containment heating. It leads to rapid increase of containment temperature and pressure, challenging the containment integrity. If the high-pressure core melt injection is not the case, the reactor vessel will be slowly melt through with a possibility of ex-vessel steam explosion and generation of missiles. Another severe accident phenomenon is hydrogen combustion. The hydrogen deflagration or detonation could lead to fast loading with possible early containment failure due to loss of its integrity. Containment overpressurization can be caused by generation of steam or non-condensable gases from decomposition of the containment concrete and combustion of other combustible gases. Another possible loss of containment integrity is due to basement melt-through in case of core-concrete interaction. Containment bypass is an event in which the function of the containment as the final barrier is lost without its failure [3].

Consideration of severe accidents currently represents the main trend in design of new reactors and also in safety upgrading of existing nuclear power plants. It requires addressing a number of specific challenges to barriers against releases of radioactive substances into the environments. In accordance with defense in depth philosophy, occurrence of severe accidents shall be considered in the plant design independently of any measures implemented for prevention of such accidents. A set of measures and actions taken during the evolution of a severe accident are called severe accident management and its objectives are to:

- Terminate the progress of core damage once it has started
- Maintain the capability of the containment as long as possible
- Minimize on-site and off-site releases
- Return the plant to a controlled safe state

Severe accident management should cover all states of plant operation as well as selected external events, such as fires, floods, earthquakes and extreme weather conditions that could damage significant part of the plant [1], [6].

The current trend in nuclear industry is to implement new safety requirements to existing reactors and simultaneously extend their operation significantly beyond the originally projected life time. Nevertheless implementation of severe mitigation strategy in existing nuclear power plant is a complicated task due to a several basic principle, which should be taken into account during development of the strategy. Selected severe accident mitigation strategy shall be fully in compliance with the relevant legislations and its technical and organization means shall be reasonably applicable, sufficiently proven in similar facility and feasible for implementation in short period to avoid production losses due to the shutdown [2].

II. Identification of Severe Accident Management Strategies

A. Depressurization of the primary circuit

Depressurization of the primary circuit is a strategy which

J. Rajzrová is with the University of West Bohemia, Pilsen, 306 14 Czech Republic (phone: +420-377-634-385; e-mail: jrajzrov@ kee.zcu.cz).

J.Jiřičková is with the University of West Bohemia, Pilsen, 306 14 Czech Republic (phone: ++420-377-634-385; e-mail: jjiricko@ kee.zcu.cz).

provides preventive and as well as mitigative effect. It initiates primary system feed and bleed procedure and use of low pressure water sources to avoid the core melt. It also ensures that induced high-pressure core melt ejection or direct containment heating cannot lead to containment failure. The depressurization strategy is well demonstrated and it is commonly implemented in severe accident management in modern power plants by a dedicated depressurization station or by opening pressurizer relief valve. In existing nuclear power plants the reactor coolant system depressurization strategy could be relatively well adopted and implemented within their safety upgrading programs (e.g. by means of the pressurizer power operated relief valve if capacity of the valve is sufficient to depressurize the primary system below before the bottom head could fail) [2], [7].

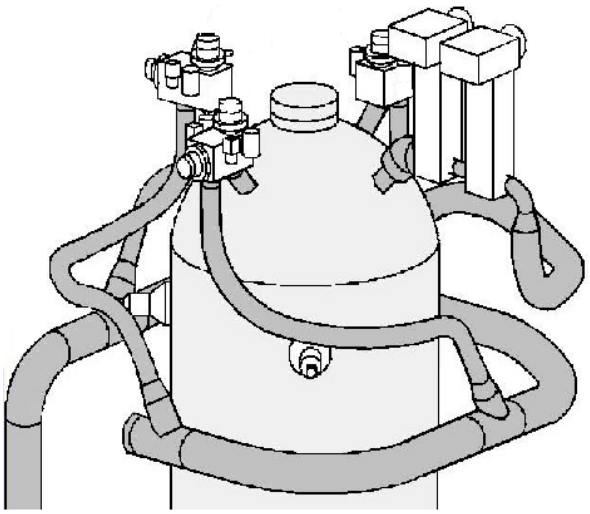


Figure II-1 Depressurization system of the primary circuit, EPR AREVA [2]

B. In-vessel retention of molten corium

The re-establishment of water injection into the reactor pressure vessel can stop the progression of core degradation and avoid reactor vessel failure. Nevertheless the injection could also cause several potentially negative effects, depending on time when it is initiated. However, it can be concluded that positive cooling effects prevail over potential negative ones and it is recommended to restart water injection into the reactor pressure vessel as soon as possible from any available borated water source. This strategy is ensured by emergency core cooling system recoveries or a modified emergency core cooling function [2], [7].

C. In-vessel retention of molten corium through external cooling of the reactor pressure vessel

External cooling of the reactor pressure vessel is ensured by the cavity filled by water. The reactor pressure vessel is submerged in water and the decay heat from the corium pool is transferred through the vessel wall into the water-filled cavity. Flooding of the cavity should be done passively, using condenser as a source of water. This strategy for halting the progress of the severe accident is used in design of modern reactors and it has already been implemented in several existing power plants in order to proceed their safety upgrading. Nevertheless implementation of this strategy requires significant hardware modifications, because the following issues must be addressed:

- Coolant inventory required and available for reactor cavity flooding
- Time window available for cavity flooding as required for successful in-vessel retention
- Critical heat flux on the external vessel surface
- Behavior of molten corium inside the vessel
- Thermal-hydraulic conditions of coolant in the reactor cavity determining heat removal, boric acid mixing steam venting from the cavity
- Impact of molten corium on the reactor pressure vessel wall and resistance of the vessel against failure [2], [7].

D. Ex-vessel cooling of molten corium or core debris

There are a number of strategies in new designs of nuclear power plants to mitigate ex-vessel phenomena, such as modified cavity configuration or installation of core catcher to improve corium spreading. The current focus is on the issue of corium release into a dry or a wet cavity. Water would be obviously required to enhance corium cooling also in originally dry cavity, but conditions for coolability of corium in the reactor cavity and its pressurization due to possible steam explosions are still under study. Due to the complexity of ex-vessel phenomena and significant hardware modification needs for implementation of measures to mitigate the phenomena, in-vessel corium retention should be given a preference in existing designs [2], [7].

E. Containment overpressure protection

Under certain conditions pressure buildup inside the containment during a core melt accident may exceed the containment design pressure. The containment could possibly fail from overpressure if there are no dedicated measures for the containment overpressure protection. A filtered venting or spraying of the containment atmosphere can prevent such a failure. The filtered venting is primary used to overpressure protection, but it can significantly reduce also source term to the environment. In general, the filtered venting equipment are expensive, complex and large and in existing designs it may be therefore advantageous to use for the given objective ventilation systems with appropriate modifications. In the case of source term reduction, it seems more appropriate to use containment spraying than forced venting of the containment atmosphere with very large required capacity of filters [2], [7].

F. Hydrogen mitigation

Hydrogen combustion is very complicated issue, because it may occur during the early part of the accidents. The containment vulnerability mainly depends on the low design pressure, the shape and generally bed conditions for internal mixing of the containment atmosphere. The new nuclear power plants take into account these factors in their design projects in order to reduce the hydrogen mitigation as much as possible. The current focus of hydrogen combustion research

ISBN: 978-1-61804-233-0

is on the issue of transition to detonation and for what geometrical conditions and hydrogen concentrations this phenomenon can occur.

There are three hydrogen control strategies have been found as the more effective and applicable also in existing designs:

- Catalytic recombiners
- Combination of CO2 inerting with recombiners
- Combination of severe accident spray, igniters and upgraded recombiners

The most promising option seems to be combination of combination of passive autocatalytic recombiners and igniters. Another feasible but more expensive option is intertization of the containment atmosphere. The both should satisfy the requirements to avoid hydrogen risk, but it is obvious that the choice of appropriate strategy mainly depends on a given containment [2], [7].

The global state of implementation of severe accident mitigation strategies shows that this trend is broadened worldwide and implementation into operation of existing plants is practiced even if this action is not formally required by the regulatory body. The highest priority in worldwide practice is given to fast reactor coolant systems depressurization, use of passive autocatalytic recombiners and in-vessel retention strategy as an efficient way for halting severe accident and reduction of possible consequences [4], [5], [7], [8].

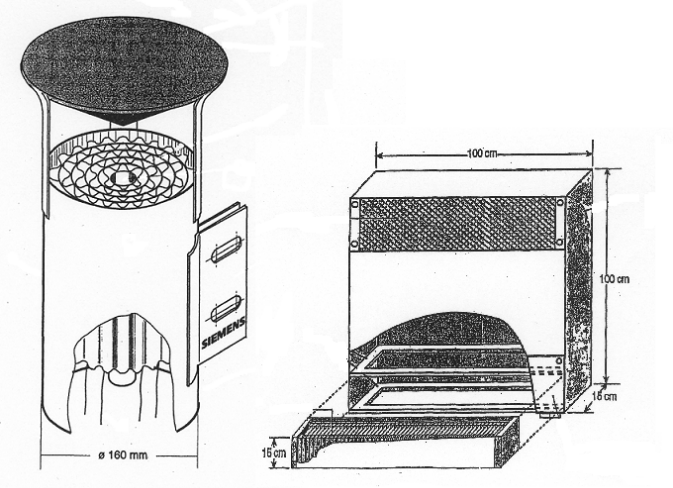


Figure II-2 Igniter and PAR, Siemens [2]

III. Acceptance criteria for severe accidents

The acceptance criteria addressing severe accidents shall be specified in national legislation or any guidance on applicability of design basis criteria to severe accidents shall be available. In general, besides probabilistic acceptance criteria, deterministic limitation of radiological consequences of severe accidents is the main criterion. There are three objectives specified in EUR related to radiological impact on the public:

- Minimal emergency protection action beyond 800 m from the reactor during releases from the containment (criterion: 50 mSv effective dose)
- No delayed action at any time beyond about 3 km from the reactor (criterion: 30 mSv effective dose)
- No long term action at any distance beyond 800 m from the reactor (criterion: 100 mSv effective dose)

In addition, restrictions on the consumption of foodstuff and crops shall be limited in terms of timescale and ground area. The quantitative dose targets for an individual from external radiation and simultaneous intake of radioactive materials can be obtained by using EUR values as guidance. The dose targets are set up as follows:

- For emergency protection actions effective dose < 50 mSv beyond the exclusion zone
- For delayed actions effective dose < 30 mSv beyond the exclusion zone
- For long term actions effective dose < 100 mSv beyond the exclusion zone

In addition in compliance with EUR it was required to address the long-term effects of the accident by limiting releases of Cs 137 (which is the most significant radioisotope for long-term effects) to 100 TBq.

The radiological design targets can only be met if the containment integrity is ensured (in addition to actions towards minimization of containment releases and reduction of the in-containment source term). Neither early nor late loss of containment integrity is accepted with sufficiently high probability. Therefore, special design targets were established for ensuring the containment integrity and they are recommended to use also for safety upgrading of existing units. The design targets take into account containment integrity acceptance criterion (maximum internal overpressure), reactor coolant system pressure, coolability of molten corium, survivability of the equipment important for the containment performance, etc [2],[8], [9].

IV. Effect of severe accident management measures on radiological consequences

A study has been performed in order to estimate the contribution of severe accident mitigation measures on the limitation of radiological consequences. The main purpose was to answer a question whether the implementation of severe accidents measures to a reactor, originally projected without these measures, is able to reduce radiological impact of severe accidents in compliance with acceptance criteria specified in EUR (as described above). The nuclear reactor system analyzed during this study was a WWER/ V 320 reactor. This reactor is originally projected with a simple containment, but it was assumed that it is additionally equipped with the severe accident mitigation measures as follows:

- fast reactor coolant systems depressurization,
- passive autocatalytic recombiners for hydrogen management,
- reactor cavity flooding for halting accident progression, special ventilation system surrounding the containment.

ISBN: 978-1-61804-233-0

The reactor system is analyzed during the first 24 hours of LOCA (analyzed by Melcor 1.8.5).The LOCA accident is associated with complete loss of all power sources leading to a complete loss of all active emergency cooling system. After 12 hours of the accident, one loop of the containment spray system and the activity of a special ventilation system in areas surrounding the containment are assumed to be restored. For the first twelve hours a containment leakage is expected (the containment leak rate 0.1% vol at the design pressure of 0.5 MPa).

The trend of pressure in containment shows a significant increase of pressure when the coolant was boiling in the reactor pit and its decrease when the spray systems starts. However, the sufficient reserve time is demonstrated and the design pressure is not exceeded until the containment spray system has been restored.

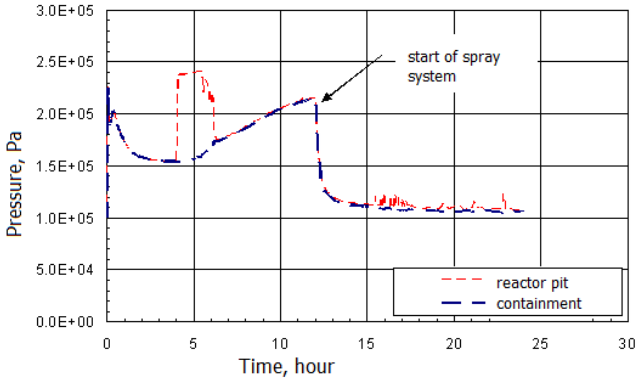


Figure IV-1 Pressure in containment and reactor pit [2]

It can be concluded that the containment integrity is ensured and radioactive releases to the environment will be caused by containment leaks (0.1 vol.% per day at design pressure) during the first 12 hours when the ventilation was considered not operable. The investigation of release from containment leakage has been done according to the Report NUREG -1465 Accident Source Terms for Light-Water Power Nuclear Plants and the EUR document as guidance.
The study has verified that the radioactive release will be in compliance with the following acceptance criteria specified in EUR:

- No emergency Action beyond 800 m from the reactor (criterion: 50 mSv effective dose)
- Limited Impact for economic impact (criterion: the sum of ground and elevated releases during the entire release time shall be compared with each of the reference values as follows: I 131 – 4000TBq, Cs 137 – 30TBq, Sr 90 – 400TBq)

The first target shall be verified according to the following methodology:

- the releases from the plant to the atmosphere are broken down into the 9 reference isotope groups (Xe 131, I 131, Cs 137, Te 131, Sr 90, Ru 103, La 140, Ce 141, Ba 140),
- these releases are combined and compared with one criterion according to the following formula :

$$\sum_{i=1}^{9} R_{ig} C_{ig} + \sum_{i=1}^{9} R_{ie} C_{ie} < 5.10\text{-}2 \qquad (9)$$

where

- Rig and Rie are the total releases (at ground and elevated level) of the nine reference isotopes during the related release period from the containment
- Cig and Cie are the coefficients related to environmental effects of unitary releases.

The release of radioactive products to the environment is presented in Table 1. For calculation the total activity released to environment we used data on activity during the 24 hours after reactor shutdown. Reactor was operated at a nominal power of 5 cycles with duration of campaign of 300 days. After the 5th cycle the fuel achieved burnout of 44505 MWd,

Table 1 Release of radioactive products [2]

Release of radioactive products			
Isotopes	*Fraction released to environment*	*Activity during the 24 hours after reactor shutdown [Bq]*	*Activity released to environment [Bq]*
Xe 131	1,278E-03	5,77E+18	7,374E+15
I 131	2,350E-05	2,74E+18	6,439E+13
Cs 137	1,932E-05	3,20E+17	6,182E+12
Te 131	2,369E-05	3,15E+17	7,462E+12
Sr 90	4,392E-06	2,49E+17	1,094E+12
Ru 103	6,306E-07	4,25E+18	2,680E+12
La 140	7,803E-07	5,52E+18	4,307E+12
Ce 141	6,940E-07	5,00E+18	3,470E+12
Ba 140	4,392E-06	5,11E+18	2,244E+13

For verification of EUR impact target: No emergency action beyond 800 m from the reactor, we used the formula as explained above takes into account the activity of the reference isotopes released to the environment and the coefficients related to environmental effects and finally we obtained

$$\sum_{i=1}^{9} R_{ig} C_{ig} + \sum_{i=1}^{9} R_{ie} C_{ie} = 1{,}42\ 10^{-2} < 5.10^{-2} \qquad (2)$$

For verification of EUR impact target: Limited impact for economic impact , we directly compared the activity of 3 reference isotopes released to the environment as follows:

ISBN: 978-1-61804-233-0

Release of Cs 137 = 6,182 TBq < 30 TBq
Release of Sr 90 = 1,094 TBq < 400 TBq
Release of I 131 = 64,390 TBq < 4000 TBq

The analysis results show that radiological impact of this accident is in compliance with acceptance criteria specified in EUR. [2].

V. Conclusion

All current evolutionary nuclear power plants are equipped with dedicated systems for severe accident management allowing reaching safety objectives. The means are following:

- The reactor is equipped either with a single containment with a shielding wall, or with a double containment with a thin or thick secondary containment.
- Depressurization of the reactor coolant system is made either by a dedicated depressurization station or by opening pressurizer relief valve.
- Molten corium stabilization is either in vessel by flooding the reactor cavity, or ex-vessel, in the core catcher or in corium spreading compartment.
- Hydrogen management is ensured either by igniters, by combination of igniters and recombiners, or by recombiners only.
- Long term containment heat removal is ensured either fully passively, or actively, by means of a dedicated containment spray system.
- None of the designs uses filtered venting of the containment as an ultimate overpressure protection.
- Initiation of actions can either be passive or active by the operator action.

Severe accident mitigation measures were not included in the original design of existing nuclear power plants. The current trend is to eliminate this design deficiency within their safety upgrades. Although the safety and cost effective severe accident management strategy for existing reactors in compliance with the new safety requirements is a complicated issue, some of the measures developed for new reactors are also applicable for existing designs. Our analysis showed that the implementation of some feasible measures is able to reduce radiological impact of severe accidents in compliance with acceptance criteria specified in EUR. Nevertheless more detailed safety analyses are still needed [2].

ACKNOWLEDGEMENT

The work has been supported by The Technology Agency of the Czech Republic (TA CR), under contact No. TE01020455 and The Student Project SGS- 2012 – 047.

References

[1] International Atomic Energy Agency, *Approaches and Tools for Severe Accident Analysis for NPPs*, Safety Report No. 56, Vienna 2008

[2] J. Rajzrova,"Assessment of efficiency of SAM in evolutionary reactors", 2011

[3] B. Kujal , "Complex phenomena and processes characteristic for severe accidents", Technical Report UJV -12017-T , 2004

[4] J. Dienstbier, B. Kujal, *Review of SAM programmes for Czech NPPs*, Technical Report UJV-12407-T, 2006

[5] P. Krakora, J.Zatloukal, L.Tetula," Safety Assessment of selected NPPs", Technical Report UJV Z 2521 T, 2009

[6] International Atomic Energy Agency, *Severe Accident Management Programmes for NPPs*, Safety Guide No. NS-G-2.15, Vienna 2009

[7] J. Dienstbier, J. Jakab, B. Kujal, a A. Rydl , "Severe Accident Management Strategies and Technical Means", Technical Report UJV-12116-T, 2004

[8] Misak, J, *Study on effects of new safety requirements during safety assessment of NPPs*, Technical Report UJV- Z13479T , 2010

[9] European Utility Requirements for LWR Nuclear Power Plants, Safety Requrements, Volume 2.

Optimization of Fuzzy Metagraph Based Stock Market DSS Using Genetic Algorithm

A. Thirunavukarasu, S.Uma Maheswari

***Abstract*—** Investing in stock market is not an easy task. It needs careful decision making skills to maintain profit in the long run. Though many research works are available to guide the investors, they are not reliable and do not guaranty maximum profit possible. In this paper a fuzzy metagraph based stock market decision support system is proposed for short term investors. Technical indicators like RSI and William-%r are used to support the system. Membership functions are optimized using genetic algorithm. The proposed system reduces the risks involved in stock investing by significantly. The genetic algorithm optimization improves overall profit by 30% to 40%.

***Keywords*—** DSS, Fuzzy Metagraph, Genetic Algorithm, RSI, Will %R, stock market.

I. Introduction

Stock market is one of the most popular investments due to its high returns. High return does not come without risk. If proper care was not taken while investment it may lead to adverse effects. There are numerous stock prediction models to help investors to predict a stock direction. Stock prediction models usually belong to two categories. They are fundamental analysis based model and technical analysis based model. Fundamental analysis involves global economic conditions, whether conditions, annual budget and company information. Technical analysis involves predicting stock price using technical indicators like RSI, MACD and William -%R. Most of the proposed prediction models are based on technical analysis. Classifiers like SVM, Fuzzy logic and neural networks are used to classify stock status.

Recently research work in stock market predictions are gaining momentum since stock investment is still a mystery and difficult to correctly predict. The predictions so far been based on portfolio management, fundamental analysis and technical analysis. Fuzzy metagraph is an emerging technique used in the design of many information processing systems like transaction processing systems, decision support systems, and workflow Systems. In our work a fuzzy metagraph based fuzzy rule management and genetic algorithm based optimization are proposed.

The rest of the paper is organized as follows. Section 2 gives the related work. Section 3 deals with metagraph technique and also two technical indicators of the stock market. Proposed model DSS is presented in section 4. The membership function optimization using GA is presented in section 5. In section 6 the system results and discussion have been discussed. Section 7 concludes the paper.

Thirunavukarasu is with the Computer Science and Engineering Department, Anna University, University college of Engineering, Ramanathapuram, Tamilnadu, India (mobile number;919865749460, e-mail: thsa07@ gmail.com).

Dr.S.Uma maheswari is now with the Department of Electronics and Communication Engineering , Coimbatore Institute of technology, Coimbatore, tamilnadu, india (e-mail: umamaheswari.cit@gmail.com).

II. Related Work

Ahmed A. Gamil, and Raafat.S have proposed a multi agent and fuzzy logic model for stock market decision making based on technical analysis. They used short term to long term Moving Average indicators to develop a decision making system. The model is tuned and modified using genetic algorithms. The DSS has been integrated into an agent based frame work to enhance the stock information retrieval process, and to be accessible through the Internet [2].

They developed a novel hybrid model based on a chaotic firefly algorithm and support vector regression (SVR) for stock market price forecasting. Integration of chaotic motion with a firefly algorithm as a simple and novel optimization method [3]. An-Sing Chen et al [7] developed a system to predict the direction of return on the Taiwan Stock Exchange Index. The probabilistic neural network is used to forecast the direction of index return. The performance of the PNN based system is compared with that of the generalized methods of moments (GMM) with Kalman filter and random walk. The results showed that PNN outperforms other models.

L. J. Cao et al [11] proposed a SVM based stock market prediction system for futures contracts collated from the Chicago Mercantile Market are used as the data sets. The performance of the system is investigated by varying kernel parameters. They have proposed a multiple-kernel support vector regression approach for stock market price forecasting. A two-stage multiple kernel learning algorithm is developed to optimally combine multiple-kernel matrices for support vector regression. The learning algorithm applies sequential minimal optimization and gradient projection iteratively to obtain Lagrange multipliers and optimal kernel weights [13]. Jacek Ma_ndziuk et al[23] developed a ANN and GA based system for short term stock index prediction. Technical variables are

taken as neural network inputs. GA is used to find an optimal set of input variables. The system performs well still it is time consuming and complex.

Kasemsan.K, and Radeerom.M have proposed a decision-making model based on the application of Neuro fuzzy systems. The model has been applied in order to make a one-step forward decision, considering historical data of daily stock returns. They have used RSI, MACD and other indicators to support the trading system o formulate a trading strategy which achieves more stable results and higher profits when compared with Neural Networks and the Buy and Hold strategy [24]. Kyoung-jae Kim et al [25] proposed a support vector machine based stock prediction system. They used SVM to predict the direction of daily stock price change in the Korea composite stock price index. For this study they selected 12 technical indicators to make up the initial attributes.

Simone Bova, and Pietro Codara have proposed to analyze Mamdani type fuzzy control systems in logical terms, with special emphasis on the fuzzy inference process Mamdani type implementation of the fuzzy system. They have used RSI and MACD to analyze stock market. They have developed an alternative inference system to Mamdani system [37]. Takashi Kimoto et al [43] proposed a index based prediction of Tokyo stock exchange prices indexes using ANN. They used both technical indexes and economic indexes to predict buy and sell timing for one month in the future. They developed a new high-speed learning method called supplementary learning to train neural networks. The buying and selling signal were accurate but were not tested for selected stocks.

Financial markets are highly volatile and generate huge amount of data on a day to day basis. The present study applied the popular data mining tool of k-NN for the task of prediction and classification of the stock index values of BSE-SENSEX and NSE-NIFTY. The results of k-NN classifier are compared with the Logistic regression model and it is observed that the k-NN classifier outperforms the traditional logistic regression method as it classifies the future movement of the BSE-SENSEX and NSE-NIFTY more accurately [42]. SVM is a promising type of tool for financial forecasting. They proposed a combining model by integrating SVM with other classification methods. The weakness of one method can be balanced by the strengths of another by achieving a systematic effect. The combining model performs best among all the forecasting methods [47].

Tiffany Hui-Kuang et al [44] used neural networks to forecast Taiwan stock index. They also developed a new fuzzy time series model to improve the forecasting. Vaidehi.V, and Monica.S have proposed a subtractive clustering based fuzzy system identification method to model a prediction system that can predict future movement of stock prices by taking samples of past events. When recent data are given to the trained system, it gives the possibility of a rise or a fall along with the next possible value of data. The prediction model is trained by daily market price data. It can also be used as a weekly or a monthly predictor [46]. Yakup Kara et al [49] proposed a neural network and support vector machine based stock prediction system to predict securities of Istanbul stock exchange. The models use both artificial neural networks (ANN) and support vector machines (SVM) for classification. Ten technical indicators including RSI, MACD simple MA and will %R were selected as inputs of the proposed models. The proposed model performs well when neural network is used as a classifier. Zheng-Hua has proposed a Fuzzy Metagraph based knowledge representation. The FM has been applied to fuzzy rule-based systems for knowledge representation and reasoning. In the format of algebraic representation and FM closure matrix [50]. In our work a fuzzy metagraph based fuzzy rule management and genetic algorithm based optimization are proposed.

III. METHODOLOGY AND TECHNICAL INDICATORS

A. *Metagraph and its adjacency matrix*

A metagraph S = <X, E> is a graphical representation consisting of two tuples X and E. Here X is its generating set and E is the set of edges defined on generating sets. The generating set X of the metagraph S i.e. the set of elements X = $\{x_1, x_2, x_3, \ldots\ldots x_n\}$ represents variables and occurs in the edges of the metagraph. An edge e in a metagraph is a pair e = $<V_e, W_e> \in E$ (where E is the set of edges) consisting of an invertex $V_e \subseteq X$ and an outvertex $W_e \subseteq X$. A simple path h(x, y) from an element x to an element y is a sequence of edges $<e_1, e_2, \ldots, e_n>$such that $x \in$ invertex(e_1), $y \in$ outvertex(e_n),and for all e_i,i=1,2,…,n-1. outvertex(e_i) ∩ invertex(e_i) ≠∅. The coinput of x in the path (denoted coinput(x)) is the set of all other invertex elements in the path's edges that are not also in the outvertex of any edges in the path, and the cooutput of y (denoted cooutput(y)) is the set of all outvertex elements other than y. The length of a simple path is the number of edges in the path [9].

The graphical structure can be represented by the adjacency matrix of a metagraph. The adjacency matrix A of a metagraph is a square matrix with one row and one column for each element in the generating set X. The ijth element of A, denoted a_{ij} , is a set of triples, one for each edge e connecting x_i to x_j. Each triple is of the form < CI_e, CO_e, e>, in which CI_e is the coinput of x_i in e and CO_e is the cooutput of x_j in e. Table 1 represents adjacency matrix A of given metagraph. For example in the figure 1, $a_{ll} = a_{31} = \emptyset$, since there are no edges connecting x_1 to itself or connecting x_3 to x_1. On the other hand, a_{13} contains one triple, since there is one edge connecting x_1 to x_3. The first component of the triple is the coinput of x_1 for this edge, the second component is the cooutput of x_3, and the third component is the edge. Since x_1 has coinput x_2, while x_5 has no cooutput, and the edge is e1, we have $a_{13} = \{< \{x_2\}, \emptyset , < e_1 >>\}$.Similarly, since e1 connects x_2 to x_4 with no coinputs or cooutputs and no other edges connect x_2 to x_4 we have $a_{24} = \{< \emptyset, \emptyset, < e_2 >>\}$. The metagraph construct is unable to tackle the issue of uncertainty and imprecision. Originally, metagraph focused on structural aspects, i.e., the connectivity relationships among components of systems. As an outcome, this approach cannot support uncertain knowledge representation and approximate reasoning.

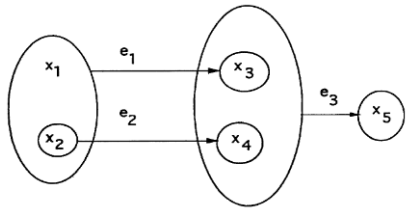


Fig. 1 example of metagraph

	X_3	X_4	X_5
X_1	$<x_2, \emptyset, e_1>$	$\emptyset$	$\emptyset$
X_2	$<x_1, \emptyset, e_1>$	$<\emptyset, \emptyset, e_2>$	$\emptyset$
X_3	$\emptyset$	$\emptyset$	$<x_4, \emptyset, e_3>$
X_4	$\emptyset$	$\emptyset$	$<x_3, \emptyset, e_3>$

Table. 1 adjacency matrix for given metagraph

B. *Fuzzy Graph*

Fuzzy set theory is primarily concerned with quantifying the vagueness in human thought and perception, where linguistic terms can be properly represented by the approximate reasoning of fuzzy sets. Linguistic variables are used in fuzzy logic. In the fuzzy theory, fuzzy set *A* of universe *X* is defined by function μ_A called the membership function of set *A*. Membership Functions are used to convert crisp inputs into fuzzy values. $\mu_A : X \rightarrow [0, 1]$, where $\mu_A(x) = 1$ if x is totally in A, $\mu_A(x) = 0$ if x is not in A, $0 < \mu_A(x) < 1$ if x is partly in A. A graph is a symmetric binary relation on a nonempty set V. Similarly, a fuzzy graph is a symmetric binary fuzzy relation on a fuzzy subset. Let V *be a*\ non empty set. A fuzzy graph is a pair of functions G = (V, σ, μ) Where σ is a fuzzy subset of V and μ is a symmetric fuzzy relation on σ. i.e. $\sigma : V \rightarrow [0,1]$ and $\mu : V \times V \rightarrow [0,1]$, a fuzzy edge set of G such that for all x , $y \in V$, $\mu(x, y) \leq \sigma(x) \wedge \sigma(y)$.

C. *Technical Indicators and Its Calculation*

Technical indicators are numerical values calculated using mathematical formulas on historical stock information. There are numerous technical indicators calculated by well known mathematicians. Each indicator can be used independently or in combination with other indicators. Among them relative strength index (RSI) and William %r are well known and widely used indicators. In our work Relative strength index (RSI) and William-%r are used.

Relative strength index: It is a momentum oscillator that measures the speed and direction at which prices are moving. RSI oscillate within the band of zero to 100. The formula used to calculate RSI is

RSI= 100-(100/1+RS)

Where, RS=Average gain/Average loss

Simple 14 period averages are used as initial value for average gain and average loss. For subsequent values the following formula is used

Average Gain = [(previous Average Gain) x 13 +current Gain] / 14.

Average Loss = [(previous Average Loss) x 13 + current Loss] / 14.

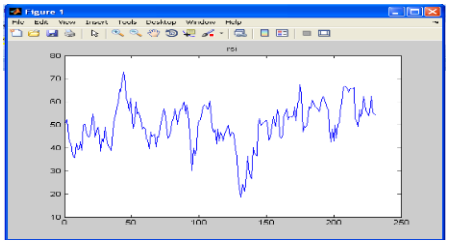


Fig. 2 RSI indicators for TCS, India

A stock is considered overbought when the value of RSI is above 70 and oversold when below 30. Signals can also be generated by looking for divergences, failure swings and centerline crossovers. RSI can also be used to identify the general trend. Buy signal is generated when RSI is near 30 and sell signal is generated when RSI is near 70. The value between 35 and 65 is considered as normal [45]. The RSI indicator for TCS, BSE and India is shown in figure 2.

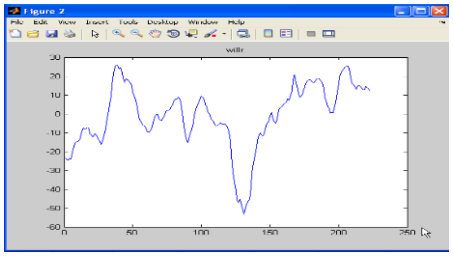


Fig. 3 Will %R indicators for TCS, India

William -%R: Developed by Larry Williams, William-%R is a momentum indicator which reflects the level of the close relative to the highest high and lowest low for the look-back period.

William- %R oscillates from 0 to -100. A stock is considered overbought when William-%R Reads from 0 to -20. Readings from -80 to -100 are considered oversold. The default setting for William- %R is 14 periods, which can be days, weeks, months or an intraday timeframe. Buy signal is generated when William -%R is near -80.Sell signal is generated when indicator is near -20. The will %R indicator for TCS, BSE and India is shown in figure 3.

IV. PROPOSED MODEL OF THE DSS

The structure of the proposed model is shown in figure 4. The proposed stock market decision support system consists of several stages. Initially historical stock information is collected from stock market. This historical information of selected stocks is collected. Technical indicators are calculated using historical stock information. In our work RSI and William -%r

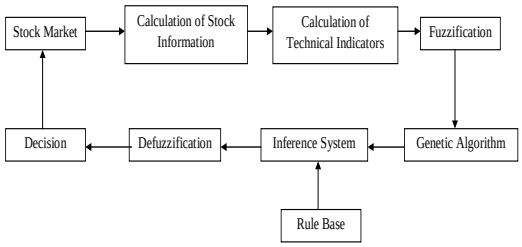


Fig. 4 Block diagram of the proposed DSS model.

are used.

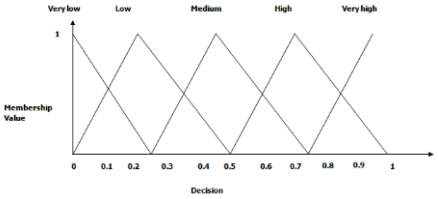


Fig. 5 Memberships function of decision system.

Decision	Value for triangular membership function
Very low	(0, 0, 0.25)
Low	(0, 0.25, 0.5)
Medium	(0.25, 0.5, 0.75)
High	(0.5, 0.75, 1)
Very high	(0.75, 1, 1)

Table. 2 Fuzzy set representation for each linguistic term

In fuzzification stage crisp values of input variables are converted into linguistic fuzzy variables. Then genetic algorithm is used to optimize membership function in order to maximize stock investment return. In defuzzification stage fuzzy variables are converted to crisp variables. These crisp values are used in decision making process. In coming sections some stages of the metagraph based fuzzy expert system will be elaborated.

Fuzzification and Fuzzy metagraph:

Fuzzification is the process of changing a crisp value into a linguistic fuzzy value. Fuzzification is achieved by using different types of membership functionsTriangular and trapezoidal membership functions are used for the propose system. Table 2 represents linguistic variables have five terms and value for triangular membership function. Decision is expressed in terms of Very low, low, Medium, high and Very high. For example fuzzification for a decision of a system is shown in the figure 5 [3].

Fuzzy graphs are widely applied in the modeling and analysis of systems. $\tilde{G} = \{\tilde{X}, \tilde{E}\}$ is a fuzzy graph where both vertices and edges have membership values. Often, the membership value of an edge is also called certainty factor (CF) of the edge. For simplicity, assign $\tilde{X}$ i denoting (x_i, $\mu(x_i)$) and $\tilde{e}_k$ denoting (e_k, CF_k). Fuzzy hypergraph describes any fuzzy relationship as a set of fuzzy elements, but it is not able to distinguish the input variables from the outputs. Over the past decades, a number of fuzzy graphs have been proposed to represent uncertain relationships between fuzzy elements or sets of fuzzy elements. However, existing fuzzy graphs are not capable of effectively modeling the directed relationships between sets of fuzzy elements.

A fuzzy metagraph is a triple $\tilde{S}=\{X, \tilde{X}, \tilde{E}\}$ in which $\tilde{X}$ is a fuzzy set on X and $\tilde{E}$ is a fuzzy relation on X×X. A fuzzy set X on X is completely characterized by its membership function μ:X→ [0, 1] for each X×X, μ(x) is the truth value of the statement of "x belongs to $\tilde{X}$ ". $\tilde{E}$ is a fuzzy edge set $\{\tilde{e}_m$,m=1, 2, 3,.. m}. Each component $\tilde{e}$ in $\tilde{E}$ is characterized by an ordered pair $<\tilde{V}_m, \tilde{W}_m>$. In the pair $\tilde{V}_m$ subset of $\tilde{X}$ is the in-vertex of $\tilde{e}_m$ and $\tilde{W}_m$ subset of $\tilde{W}$ is the out-vertex. The membership value of an edge is also called certainty factor of the edge [50].

Fuzzy Inference System and Rulebase

A Fuzzy Inference System is used to map an input space to an output space using fuzzy logic. Fuzzy inference system uses a collection of membership function and rules to drive output from a crisp input.A fuzzy inference system implements a nonlinear mapping from its input space to output space through a number of fuzzy if-then rules. The three most popular inference systems used in fuzzy logic are the: Mamdani fuzzy model, Tagaki-Sugeno fuzzy model, and Tsukamoto model. In this work Mamdani Fuzzy model is used.

Rule Base for Fuzzy Metagraph Based System

A rule base is an ordered pair T =< P, R > in which P is a set of propositions and R is a set of rules. Given a generating set X and a metagraph S =< X, E > on X with E and a rule base T=<P, R>, S corresponds to T if there are bijective mappings between X and P and between E and R. If S corresponds to T, then X corresponds to P and E corresponds to R. The metagraph notation to refer the rule base system, that is, xi will denote the i^{th} proposition and e_k will denote the k^{th}

rule. For example, in the metagraph of Figure 1, e_1 represents the rule $x_1 \hat{} x_2 \rightarrow x_3$, where "^" denotes conjunction and "→ " denotes implication. e_2: $x_2 \rightarrow x_4$, e_3: $x_3 \hat{} x_4 \rightarrow x_5$ .Since M_2 is a metapath from {x1, x2} to {x5}, we can conclude that x5 can be inferred from x1 and x2-that is $x_1 \hat{} x_2 \rightarrow x_5$. Fuzzy rules for fuzzy metagraphs are formed based on fuzzy Metagraph shown in figure 5. Technical indicators RSI,WILLIAM-%R are used as input vector to the fuzzy Metagraph system. Decision vector (BUY, HOLD, and SELL) is output of the system. Where, RSI ={x1, x2, x3}, WILLIAM-%R={x4, x5, x6} and DECISION={x10, x11, x12}. Elements Proposition For generating set from figure 6. X1: RSI <30,

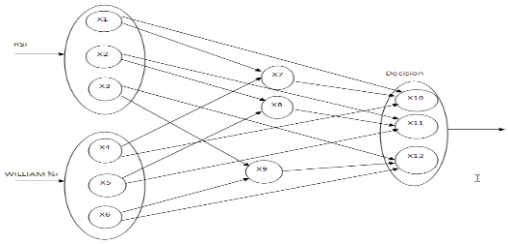


Fig. 6 Block diagram of the proposed DSS model.

X2: 30<=RSI<=70, X3: RSI>70,
X4: WILLIAM - %R <-80
X5: -80<=WILLIAM - %R<=-20 ,
X6: WILLIAM - %R> -80
X7: X1^ X4→ X7, X8: X2 ^ X5→ X8,
X9: X3 ^ X6→ X9,
X10: BUY, X11: HOLD, X12: SELL

Defuzzification and Decision

Defuzzification is the process of finding one single crisp value that summarizes the fuzzy set that enters it from the inference block. The output crisp value is used to make decisions. Here any value between 0 and 35 can be used as a BUY signal. Value between 35 and 65 is used as STAY signal. Any value between 65 and 100 is used as SELL signal.

V. Membership Function Optimization Using Genetic Algorithm

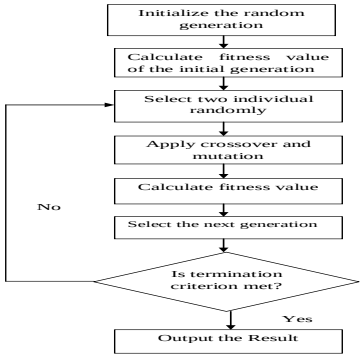


Fig. 7 Flow diagram of genetic algorithm based membership function optimization.

The genetic algorithm is a model of machine learning algorithm based on the evolutionary ideas of natural selection and genetic. The steps of genetic algorithm includes

1. Randomly generate an initial population
2. Compute and save the fitness value for each individual in the current population
3. Define selection probabilities for each individual
4. Generate a new population set by probabilistically selecting individuals from random population to produce offspring via genetic operators
5. Repeat step 2 until satisfying solution is obtained.

The decision making ability or profit depends on the membership function of the fuzzy system. If the membership function was properly chosen the profit earned can be increased. One of the ways to find out the suitable membership function parameters is optimization of the membership function. Genetic algorithm is used to optimize the membership functions. The results of optimized membership functions are compared with traditional membership functions. Optimization increased the total profit by 30% -40%. The flow diagram for genetic algorithm based membership function optimization is shown in figure 7. In this work both trapezoidal and triangular membership functions are used to define RSI and William-%R. The parameters of input membership function are taken as gene of chromosomes. The parameters that are fixed are not taken. Initially five individuals are randomly generated with some constraint. Total profit for given time period is taken as fitness value of the membership function parameters.

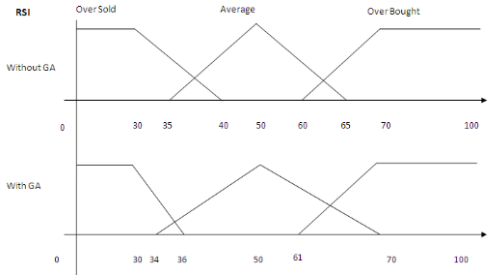


Fig. 8 Membership functions of RSI before optimization and after optimization.

T

he profit is calculated by assigning these values to the membership functions. Two individuals are chosen randomly and applied crossover and mutation. 30th bit from the initial point is taken as crossover point and mutation is done by complementing a randomly chosen bit. Profit for these two individual are calculated. The individual with maximum profit is taken as next generation. The randomly generated individual with minimum profit is replaced by the next generation individual. This process is repeated until maximum profit is achieved. The membership function of RSI and William-%r

before and after optimization are given in figure 8 and figure 9.

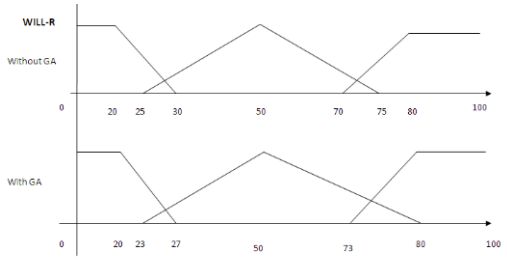


Fig. 9 Membership functions of William-%r before optimization and after optimization.

VI. RESULTS AND DISCUSSIONS

To test the performance of the proposed system two stocks of Bombay stock exchange, India are considered. They are Tata consultancy service (TCS) and Reliance Industry Limited (RIL). Both companies are popular and well performing companies in India. The year 2011-2012 had been very challenging year for Indian share market. In this study stock information for that period is taken to analyze the performance of the system at hard times. Table 3 represents testing results of TCS for one year .Table 4 represents testing results of RIL for one year. The system is tested with last one year historical trading data of the respective companies. In order to test the effectiveness of the system the results obtained from genetic algorithm optimized system is compared with that of non optimized system and BUY and HOLD method. The findings are listed in the table 5 given below. From the experience the proposed system outperforms both fuzzy logic based stock market decision support system and traditional BUY and HOLD method. With our system one can expect 30% to 40% more profit than any other methods.

VII. CONCLUSION

Fuzzy metagraph based decision support system for stock market investing is proposed. Technical indicators are used to aid the decision making. Genetic algorithm is used to improve the overall profit of the system. The proposed system can be used in short term investing. The system performance is tested using stocks listed in Bombay stock exchange. The results are satisfactory. This method reduces the risk factor involved in stock investing. Future works may concentrate on optimization techniques applied for tuning the input parameters to further enhance the performance of the system.

Date(DD-MM-YY)	**Price**	**Signal**	**Buy price**	**Profit**	**Total profit**
09-02-2011	1098	BUY	1098	-	-
31-03-2011	1182	SELL	1182	84	-
20-06-2011	1069	BUY	1069	-	-
14-10-2011	1134	SELL	1134	64	148

Table .3 Testing results of TCS for one year

Date(DD-MM-YY)	**Price**	**Signal**	**Buy price**	**Profit**	**Total profit**
27-01- 2011	943	BUY	943	-	-
01-03-2011	988	SELL	988	45	-
20-06-2011	834	BUY	834	-	-
25-07-2011	882	SELL	882	48	-
04-08-2011	812	BUY	812	-	-
08-09-2011	853	SELL	853	48	-
16-11-2011	786	BUY	786	-	-
07-12-2011	809	SELL	809	23	164

.

Table .4 Testing results of RIL for one year

Name	Price at the beginning of the investment period in Rs	Price at the end of the investment period in Rs	Profit per share without applying GA in Rs	Profit per share with GA in Rs	Profit using BUY and HOLD method in Rs
TCS	1178	1170	108	148	-8
RIL	1079	698	93	164	-381

Table .5 Comparative analyses of the various results.

ISBN: 978-1-61804-233-0

REFERENCES

[1] Aditya Gupta and Bhuwan Dhingra, "Stock Market Prediction Using Hidden Markov Models," pp. 1-4, 2011.

[2] Ahmed A. Gamil and Raafat S, "Stock Technical Analysis using Multi Agent and Fuzzy Logic,"Proceedings of the World Congress on Engineering, pp. 142-147, 2007.

[3] Ahmad Kazem, Ebrahim Sharifi, Farookh Khadeer Hussain, Morteza Saberi, "Support vector regression with chaos-based firefly algorithm for stock market price forecasting," *Elsevier, Applied Soft Computing*, vol.13, pp. 947-958, 2013.

[4] Ajith Abraham , "Rule-Based Expert Systems," Handbook of Measuring System Design, Oklahoma State University, Stillwater, USA, pp. 909-919,2005.

[5] A. Alaoui, "On Fuzzification of Some Concepts of Graphs," Fuzzy Sets and Systems, vol. 101, pp. 363-389, 1999.

[6] R.A Aliev and R.R Aliev "Soft Computing and Its Applications," World Scientific Publishing, Singapore, London, 2001.

[7] An-Sing Chen, Mark T. Leung, Hazem Daouk, "Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index",Computers & Operations Research 30, pp. 901–923, 2003.

[8] A. Basu, "A Metagraph View-Based Approach to Multi-firm Process Coordination", Proceedings of the CAiSE'05 Forum, ISBN 972-752-078-2, pp.69-74, 2005.

[9] A. Basu and R.W. Blanning, "Metagraphs and their applications,"Springer Science and Business Media, LLC, New York, NY 10013, USA, 2007.

[10] Binoy B. Nair, V.P Mohandas, N.R. Sakthivel, "A Genetic Algorithm Optimized Decision Tree- SVM based Stock Market Trend Prediction System", International Journal on Computer Science and Engineering, vol. 02, no. 09, pp. 2981-2988, 2010.

[11] L. J. Cao, Francis E. H. Tay, "Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting", *IEEE Transactions on Neural Networks*,vol.14,no.6, pp. 1506–1518,2003.

[12] Chin-Yuan Fan, Pei-Shu Fan, Chia-Huei Wu, Wen-Wei Yen, "Stock Turning Points Detection by Particle Swarm Optimization clustering and Support vector Machine decision system," pp. 1-5,2009.

[13] Chi-Yuan Yeh, Chi-Wei Huang, Shie-Jue Lee, "A multiple-kernel support vector regression approach for stock market price forecasting," *Elsevier, Expert Systems with Applications*, pp. 1-10,2010.

[14] Dattatray P. Gandhmal, Ranjeetsingh B. Parihar, "An Optimized Approach to Analyze Stock market using Data Mining Technique," Proceedings published by International Journal of Computer Applications, International Conference on Emerging Technology Trends, pp. 38-42, 2011.

[15] Deepti Gaur, AdityaShastri, Ranjit Biswas,"Metagraph a new hierarchical data structured as a decision tree," the journal of computer science and information technology, vol.6, no.1, pp.1-5, July-Dec. 2007.

[16] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Metagraph: A new model of Data Structure," *IEEE International Conference on Computer Science& Information Technology*, pp. 729-733, 2008.

[17] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Fuzzy Meta Node Fuzzy Metagraph and its Cluster Analysis," Journal of Computer Science, vol.4 , pp. 922-927, 2008, India.

[18] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Vague Metagraph," International Journal of Computer Theory and Engineering, vol.1, vo.2, pp.126-130, 2009, India.

[19] Deng-Yiv Chiu and Kun-Po Chuang, "Applying Artificial Neural Network and Chinese News Classification Techniques to Taiwan Stock Market," Tamsui Oxford Journal of Mathematical Sciences, vol.19, no.2, pp. 201-215, 2003.

[20] Diego J.Bodas-Sagi, Pablo.Fernandez, J. Ignacio Hidalgo, "Multi objective Optimization of Technical Market Indicators," GECCO'09, *ACM*, pp. 1999- 2004, 2009.

[21] P.Fernandez Blanco, D.Bodas-Sagi, F.Soltero, J.I.Hidalgo, "Technical market indicators optimization Using evolutionary algorithms,"*ACM*, pp.1851-1857, 2008.

[22] Festus Oluseyi Oderanti, "Fuzzy inference game approach to uncertainty in business decisions and market competitions," Oderanti SpringerPlus, vol.2, no.484, pp. 1-16, 2013.

[23] Jacek Ma_ndziuk, Marcin Jaruszewicz, "Neuro-genetic system for stock index prediction,"Journal of Intelligent & Fuzzy Systems, vol.22, no.2, pp. 93-123, 2011.

[24] M.L. K. Kasemsan, M. Radeerom, "Intelligence Trading System for Thai Stock Index Based on Fuzzy Logic and Neuro fuzzy System," Proceedings of the World Congress on Engineering and Computer Science, Oct 2011, San Francisco, USA.

[25] Kim, K., "Financial time series forecasting using support vector machines," Neuro computing, vol.55, pp.307–319, 2003.

[26] R.J. Kuo, C.H. Chen, Y.C. Hwang, "An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network, " Fuzzy Sets and Systems, vol.118 ,pp. 21-45, 2001.

[27] Nassim Homayouni ,Ali Amiri, "Stock price prediction using a fusion model of wavelet, fuzzy logic and ANN," 2011 International Conference on E-business, Management and Economics IPEDR , vol.25, IACSIT Press, Singapore, pp. 277-281, 2011.

[28] PankajDashore and Suresh Jain, "Fuzzy Rule Based Expert System to Represent Uncertain Knowledge of E-commerce," International Journal of Computer Theory and Engineering, vol.2, no.6, pp. 882-886, 2010.

[29] PankajDashore and Suresh Jain, "Fuzzy Metagraph and Rule Based System for Decision Making in Share Market," International Journal of Computer Applications, vol.6, no.2, pp. 10-13, 2010.

[30] G. Preethi and B. Santhi, "Stock market forecasting techniques: a survey," Journal of theoretical and applied information technology, ISSN: 1992-8645, pp. 24-30, 2012.

[31] Punam Varghade and Rahila Patel, "Comparison of SVR and Decision Trees for Financial Series Prediction," International Journal on Advanced Computer Theory and Engineering ,ISSN : 2319 – 2526, vo.1, no.1, pp. 101-105, 2012.

[32] Rama Bharath Kumar, Bangari Shravan Kumar, Chandragiri Shiva Sai Prasad , " Finical news classification using SVM," International Journal of Scientific and Research Publications, vol.2, no.3, pp.1-6,2012.

[33] Ramin Rajabioun and Ashkan Rahimi-Kian, "A Genetic Programming Based Stock Price Predictor together with Mean-Variance Based Sell/Buy Actions," ISBN: 978-988-17012-3-7, Proceedings of the World Congress on Engineering 2008, vol.2, London, U.K.

[34] Sneha Soni, "Applications of ANNs in Stock Market Prediction: A Survey", International Journal of Computer Science & Engineering Technology, ISSN: 2229-3345 Vol. 2, No. 3, pp.71-83, 2010.

[35] Sneha Soni, "Classification of Indian Stock Market Data Using Machine Learning Algorithms," International Journal on Computer Science and Engineering, ISSN: 0975-3397,vol. 02, no. 9, pp. 2942-2946, 2010.

[36] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik , "Financial Stock Market Forecast using Data Mining Techniques, IMECS 2010, pg 1-5, 2010.

[37] Simone. Bova, P. Codara, "A Logical Analysis of Mamdani type Fuzzy Inference- Theoretical Bases," *FUZZY-IEEE International Conference on Fuzzy Systems*, 2010.

[38] Sivanandam and S.N. Deepa , "Principles of Soft Computing". 1st Edn., Wiley India, New Delhi, ISBN: 9788126510757,2007.

[39] S. N. Sivanandam, S. Sumathi and S. N. Deepa, "Introduction to Fuzzy Logic using MATLAB", Springer, 2007.

[40] Sheta, "Software Effort Estimation and Stock Market Prediction Using Takagi Sugeno Fuzzy Models," *IEEE International Conference on Fuzzy System*, pp.171-178, 2006.

[41] Srinivasan Vaiyapuri and Rajenderan Govind, "A fuzzy fast classification for share market Database with lower and upper bounds," science publication, American journal of applied sciences, vol.9, no.12, ISSN: 1546-9239, pp.1934-1939, 2012.

[42] M. V. Subha, s. Thirupparkadal nimbi, "classification of stock index movement usingK-nearest neighbours algorithm", SEAS transactions on Information science and applications, E-ISSN: 2224-3402, vol.9, no.9, pp. 261-270, September 2012.

[43] Takashi Kimoto, Kazuo Asakawa, Morio Yoda, Masakazu Takeoka, "Stock Market Prediction System with Modular Neural Networks", in Proceedings of the International Joint Conference on Neural Networks, pp. 1-6.

[44] Tiffany Hui-Kuang yu and Kun-Huang Huarng, "A Neural network-based fuzzy time series model to improve forecasting", *Elsevier*, pp: 3366-3372, 2010.

[45] Thirunavukarasu,Uma Maheshwari, "Technical analysis of Fuzzy Metagraph based decision Support system for capital market," ISSN: 1549-3636, Science Publications, Journal of Computer Science, vol.9, no. 9, pp.1146-1155, 2013.
[46] Vaidehi .V ,Monica .S , Mohamed Sheik Safeer .S, Deepika .M, " A Prediction System Based on Fuzzy Logic," Proceedings of the World Congress on Engineering and Computer Science 2008, San Francisco, USA
[47] Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang, "Forecasting stock market movement direction with support vector machine," *Elsevier, Computers & Operations Research*, vol.32, pp. 2513–2522, 2005.
[48] Yuan Luo, Kecheng Li,Darryl N. Davis, " A Multi-Agent Decision Support System for Stock Trading," *IEEE Network* , pp.1-9,2002.
[49] Yakup Kara, Melek Acar Boyacioglu, omer Kaan Baykan , "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Elsevier, Expert Systems with Applications*, vol.38 ,pp. 5311–5319,2011.
[50] Zheng Hua Tan, "Fuzzy Metagraph and Its Combination with the Indexing Approach in Rule-Based Systems," *IEEE transactions on knowledge and data Engineering*, vol. 18, no. 6, pp. 829-841, June, 2006.

About the Authors

Mr.A.Thirunavukarasu completed his B.E Degree in Computer Science and Engineering from Coimbatore Institute of Technology, Coimbatore in the year 2006 and M.E. Degree in Computer Science and Engineering from Anna University of Technology, Coimbatore in the year 2009. Currently he is pursuing PhD degree from Anna University, Chennai. He is working as a Teaching Fellow, Department of Computer Science and Engineering in Anna University, University College of Engineering, Ramanathapuram Campus. He is having more than 5 years of teaching and research experience. He has published technical papers in National / International conferences/Journals. His areas of specialization include Data Structures and Algorithms, Compilers, Theory of computation, Data mining, Database Security and Metagraph.

Dr. S. Uma Maheswari received her B.E Degree in Electronics and Communication Engineering from Government College of Technology, Coimbatore in the year 1985 and M.E (Applied Electronics) from Bharathiar University in 1991. She received her Ph.D degree in the area of Biometrics from Bharathiar University, Coimbatore in the year 2009. She is Associate Professor of Electronics and Communication Engineering department in Coimbatore Institute of Technology. She is having more than 27 years of teaching experience. She has published technical papers in National /International Conferences/ Journals. Her special fields of interest are Digital Image Processing and Digital Signal Processing. She is a Member of IE (India), Life Member in Indian Society for Technical Education (India), Life Member in Systems Society of India, and Life Member in Council of Engineers (India).

Character Segmentation in Overlapped Script using Benchmark Database

Tanzila Saba[1], Amjad Rehman[3]
[1] Faculty of Computing Universiti Teknologi Malaysia, Malaysia
[3]MIS Department CBA Salman bin Abdul Aziz University Alkharj KSA

Saleh Al-Zahrani[2]
College of Computer and Information Sciences
Al-Imam M. Saud Islamic University Riyadh KSA

***Abstract*—**Character segmentation of highly slanted and overlapped characters is an important issue of the day. This paper presents a novel approach for character segmentation of such overlapped cursive words without using any slant correction technique. Hence, a new concept of core-zone is introduced for segmenting such overlapped cursive script. Following core-zone detection, the character boundaries are detected in the core-zone area only. However, due to inherent nature of cursive words, few characters are over-segmented. A threshold is selected heuristically to overcome this problem. For fair comparison overlapped words are extracted from CEDAR benchmark database. Experiments thus performed exhibit promising result and high speed.

Keywords—character segmentation, analytical approach, core-zone, slant correction, overlapped script.

I. INTRODUCTION

Character segmentation of cursive script words is a long standing problem at the heart of systems for conversion of handwritten information to electronic form [1,21]. This problem becomes more serious, when we are dealing with highly slanted and horizontally overlapped cursive script words. Nonetheless, the state of the art is well mature for recognition of disjoint handwritten numerals and characters [2,22]. However, the research into the segmentation and recognition of characters extracted from cursive, touching and horizontally overlapped characters is still fresh [3, 4]. Accurate character segmentation is the overlapped phase for correct and speedy recognition, particularly in context of segmentation-based, word recognition [4]. Still there are several reasons why we may find advantage in more accurate and reliable word image segmentation. Firstly, for reason of efficiency, we would like to keep the number of segmentations as less as possible. Secondly, each possible segmentation point in the image requires some resolution of target character, which can extend the time for recognition of the character image significantly [5].

Therefore, higher the segmentation accuracy, the more beneficial it is to the recognition rates [6]. Hence the segmentation is the backbone of the recognition process and is still active research topic. Researchers are agreed that high accuracy of character segmentation is guarantee for high accuracy in handwriting recognition process [7-9]. Hence, it is mandatory to adopt new and accurate character segmentation techniques particularly for touched, overlapped and broken script. Additionally, their results should be compared in the state of art on benchmark databases.

The script with highly slanted and overlapped characters could not segment accurately into individual characters despite of slant correction techniques [10, 11]. This paper presents a simple but effective approach for segmenting such overlapped cursive words. In proposed approach, rather performing complex slant correction preprocessing techniques, word core-zone height is detected. Segmentation performed in the core zone yields better results with high speed. Once character boundaries are found, overlapped slanted and horizontally overlapped characters can be detected by following connected pixels.

Rest of the paper is organized in three sections. Sections 2 presents proposed character segmentation approach for overlapped cursive script words. In section 3 experimental results are discussed and conclusion is presented in section 4.

II. PROPOSED SEGMENTATION APPROACH

In this section, preprocessing steps and proposed segmentation algorithm is presented. An overview of the proposed segmentation algorithm is shown in figure 1

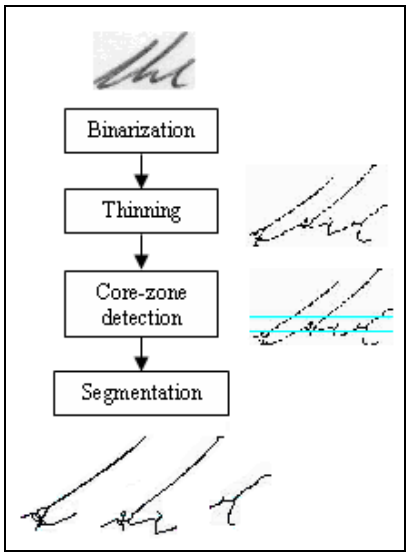


Figure 1. Overview of segmentation algorithm

A. *Biniraziation*

The first step removes noise using Otsu algorithm to change original image into binary form [12].The output binary image have values 1 (for white pixels) and values 0 (for black pixels).

B. *Thinning*

Following binarization, image is converted to skeleton image to allow users verity of writing device, pen tilt and to suppress extra data.

C. *Core-zone detection*

For overlapped handwritten words, slant correction techniques do not yield good results [10,11,14]. Likewise to segment horizontally overlapped characters, we introduced a simple concept of word core-zone height. Core-zone is the area that lies between lower base line and upper baseline of the word. Again a very simple and fast technique is proposed to calculate lower and upper baselines.

From top left point, count number of white pixels row by row. Calculate average of those rows for which there is a change in count between current and preceding row until the first big change occurs. Average row thus found is upper base line. The same procedure is adopted for the lower base line but from bottom to top. Results of the core zone detection for the overlapped words are shown in figure2.

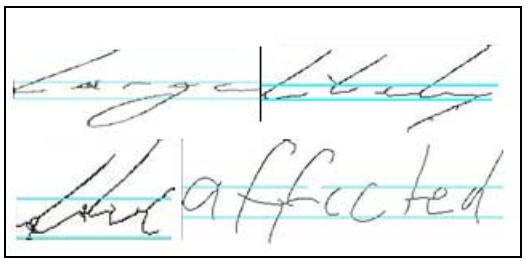


Figure2. Core-zone height detection

D. *Proposed Segmentation approach*

Proposed character segmentation algorithm for overlapped cursive handwritten words is presented in fig.3.

Step 1: Select overlapped script images from CEDAR benchmark database.
Step 2: Perform pre-processing.
Step 3: Determine core-zone region.
Step 4: In core zone, calculate sum of foreground pixels (white pixels) for each column. Save those columns as candidate segment column (CSC) for which sum is 0 or 1.
Step 5: By previous step, we have more candidate segmentation columns than actual required. Threshold is selected empirically from candidate segment columns to come out with real segment columns.

Figure 3. Proposed segmentation algorithm

Due to the simplicity of the proposed segmentation technique, it is very fast and performs well in most of the cases. For few characters such as w, u, v, m, n etc over segmentation occurs and this technique fails to find accurate character boundaries. However this problem is solved heuristically by selecting a threshold.

III. RESULTS

A. *Handwriting database.*

To exhibit efficiency of the proposed approach, slanted and horizontally overlapped words are selected from a benchmark database CEDAR [3]. A few samples are shown in figure 4.

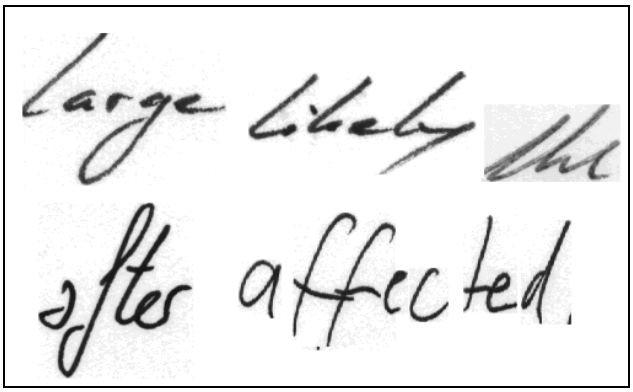


Figure 4. Overlapped word samples from CEDAR database.

Two sets (100 words each) of experiments are performed. Firstly, for normal words secondly, for overlapped words extracted from benchmark database [3]. For normal words, segmentation is accurate without core-zone concept while for overlapped words we introduced core-zone processing. Few segmentation results, for normal

ISBN: 978-1-61804-233-0

words and for overlapped words (with core-zone concept) are presented in figure 5 and figure 6 respectively.

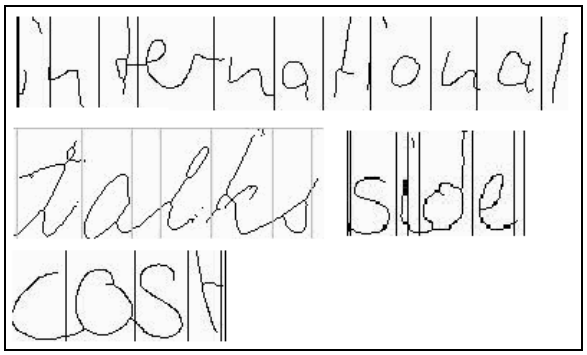


Figure5. Character segmentation results for normal words

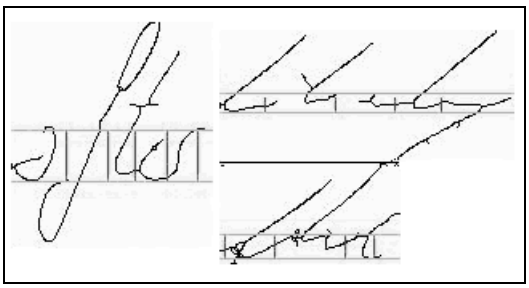


Figure 6. Character segmentation results for overlapped words

To evaluate performance of the proposed approach, valid, invalid and over-segmented characters are counted. There is valid detection when the algorithm finds correct character boundaries. Invalid segmentation, when algorithm finds inaccurate character boundaries and over-segmentation when it cut a character more than desired boundaries.

Although, character segmentation approach without introducing core zone concept, performed well for normal handwritten cursive words but segmentation rate decreased significantly for the overlapped words. Hence for overlapped words, core-zone concept is introduced. Following, detection of character boundaries in the core-zone, script word is segregated into individual characters.

For 100 (487 characters) overlapped words, 451 characters were segmented correctly. Few characters such as u, v, w, m, n are over-segmented and miss segmentation occurs for touching/ badly overlapped characters. Segmentation results are presented in Table 1.

Table 1: Segmentation results

Valid segmented words	92.6%
Miss segmentation	1.64 %
Over-segmentation	1.87%

IV. ANALYSIS AND COMPARISION

For the sake of space , a brief comparison of achievement for segmentation rate is presented. Verma and Gader [15] obtained 91% segmentation rate using neuro-feature based approach on words taken from CEDAR, however the number of words was not mentioned. Likewise, Blumenstien and Verma[16] claimed 78.85% segmentation results using neuro-feature based approach for words taken from CEDAR without mentioning number of words. In the same way, Verma[17] claimed 84.87 % accuracy for segmenting 300 CEDAR words based on neuro feature based approach. Similarly, Chen et al [18] used neuro-feature based approach and got 95.27 % segmentation rate from 317 CEDAR words. Finally Cheng and Blumenstein [19] using neuro-enhanced feature based approach obtained 84.19 % segmentation rate for 317 CEDAR words. Character segmentation results avaliable in the literature since year 2000 is presented in Table 2.

Table 2: Segmentation results

Author	**Technique**	**Success Rate**	**Database**
Verma and Gader [15]	Feature based + ANN	91	CEDAR
Blumenstein and Verma [16]	Feature based+ ANN	78.85%	CEDAR
Verma [17]	Feature based + ANN	84.87	CEDAR 300 words
Cheng et al [18]	Feature based + ANN	95.27	CEDAR 317 words
Cheng Blumenstein [19]	Enhanced feature based+ ANN	84.19	CEDAR 317 words
Blumenstein [20]	Fused features + ANN	95.27%	CEDAR 317 words

V. CONCLUSION

Slant correction techniques do not yield good results for highly slanted and horizontally overlapped characters in script [14]. A simple and effective core-zone based character segmentation approach is introduced without employing any slant correction to avoid complexity and to enhance speed. For experiments, 100 overlapped handwritten words are selected from CEDAR benchmark database. Character segmentation results thus obtained are encouraging and exhibit high speed because of simple approach. In future major experiment with significantly large data set will be performed.

ISBN: 978-1-61804-233-0

References

[1] T. Saba, and A. Rehman. "Effects of Artificially Intelligent Tools on Pattern Recognition", International Journal of Machine Learning and Cybernetics, 2012. DOI: 10.1007/s13042-012-0082-z.

[2] A.Rehman, and T. Saba. "Neural Network for Document Image Preprocessing" Artificial Intelligence Review, 2012, DOI: 10.1007/s10462-012-9337-z.

[3] J.J. Hull. "A Database for Handwritten Text Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence", vol. 16, pp.550–554, 1994.

[4] A. Rehman and T. Saba. "Features extraction for soccer video semantic analysis: current achievements and remaining issues" Artificial Intelligence Review, 2012, DOI: 10.1007/s10462-012-9319-1

[5] A. Rehman and T. Saba. "Off-line Cursive Script Recognition: Current Advances, Comparisons and Remaining Problems". Artificial Intelligence Review Springer, vol. 37(4), pp.261-268. 2012, DOI: 10.1007/s10462-011-9229-7.

[6] T.Saba, A. Rehman, and M. Elarbi-Boudihir, "Methods and Strategies on off-line Cursive Touched Characters Segmentation: A Directional Review". Artificial Intelligence Review, 2011, DOI 10.1007/s10462-011-9271-5.

[7] A. Rehman and D. Mohamad (2008). "A Simple Segmentation Approach for Unconstrained Cursive Handwritten Words in Conjunction of Neural Network". International Journal of Image Processing, vol. 2(3), pp. 29-35.

[8] A. Rehman, and T. Saba, "Performance Analysis of Segmentation Approach for Cursive Handwritten Word Recognition on Benchmark Database". Digital Signal Processing, vol. 21, pp. 486-490,2011.

[9] T. Saba, Amjad Rehman and Ghazali Sulong (2011) "Improved Statistical Features for Cursive Character Recognition" International Journal of Innovative Computing, Information and Control (IJICIC) vol. 7(9), pp. 5211-5224.

[10] A. Rehman, D. Mohamad, G. Sulong and T. Saba, "Simple and Effective Techniques for Core Zone Detection and Slant Correction in Script Recognition". The IEEE International Conference on Signal and Image Processing Applications (ICSIPA'09), pp. 15-20, 2009.

[11] M. Harouni, M. S. M. Rahim, D. Mohamad, A.Rehman, T. Saba (2012) Online Cursive Persian/Arabic Character Recognition by Detecting Critical Points , International Journal of Academic Research, Vol. 4(2). 209-214.

[12] S.Bekhti, A. Rehman, M. Al-Harbi (2011). "AQuASys An Arabic Question-Answering System based on Extensive Question Analysis and Answer Relevance Scoring" Information and Computation International Journal of Academic Research, Vol. 3(4), pp.45-54.

[13] A. Rehman and T. Saba (2011). "Document Skew Estimation and Correction: Analysis of Techniques, Common problems and Possible Solutions" Applied Artificial Intelligence, Vol. 25(9), pp. 769-787.

[14] G.Sulong, T. Saba, and A. Rehman, "Dynamic Programming Based Hybrid Strategy for Offline Cursive Script Recognition", IEEE Second International Conference on Computer and Engineering, vol. 2, pp. 580-584, 2010.

[15] B. Verma, P.D. Gader, P.D., and Chen, W. Fusion of Multiple Handwritten Word Recognition Techniques. Pattern Recognition Letters, vol. 22 (9), 991-998.2001.

[16] M. Blumenstein, and B. Verma, "Analysis of Segmentation Performance on the CEDAR Benchmark Database", Proceedings of Sixth International Conference on Document Analysis and Recognition. 1142-1146,2001.

[17] B. Verma, "A Contour Code Feature Based Segmentation for Handwriting Recognition. Proceedings of Seventh International Conference on Document Analysis and Recognition (ICDAR'03), pp. 1203-1207.2003.

[18] C.K. Cheng, X.Y. Liu, M. Blumenstein, and V.Muthukkumarasamy, "Enhancing Neural Confidence-Based Segmentation for Cursive Handwriting Recognition", 5th International Conference on Simulated Evolution and Learning, Busan, Korea, SWA-8,2004.

[19] C.K. Cheng, and M. Blumenstein, "The Neural Based Segmentation of Cursive Words Using Enhanced Heuristics", Proceedings of the Eighth International conference on Document analysis and recognition, vol. 2, 650-654,2005.

[20] M. Blumenstein, "Cursive Character Segmentation Using Neural Network Techniques". Machine Learning in Document Analysis and Recognition, vol. 90, 259-275, 2008.

[21] Saba, T. and Alqahtani, F.A. (2013) Semantic Analysis Based Forms Information Retrieval and Classification, 3 D Research, vol. 4(4).DOI. 10.1007/3DRes.03(2013)4.

[22] K.Neamah,D.Mohamad,T.Saba and A. Rehman (2014) Discriminative Features Mining for Offline Handwritten Signature Verification, 3D Research vol. 5(1), 2014.10.1007/s13319-013-0002-3

Conducting effective penetration testing

Stanislav Zitta, Josef Horalek, Ondrej Marik, Sona Neradova
Faculty of electrical engineering and informatics
University of Pardubice
Pardubice, Czech Republic
stanislav.zitta@student.upce.cz, josef.horalek@upce.cz, ondrej.marik@student.upce.cz, sona.neradova@upce.cz

***Abstract* — Nowadays, information security is very important, because more and more confidential information like medical reports is being stored electronically on computer systems and those systems are often connected to computer networks. This introduces new challenges to people working in information technology. They have to ensure that those systems are as much secure as possible and confidential information won't be revealed. One possible way how to prove systems security is to conduct regular penetration tests – that is simulate attacker's malicious activity. This paper introduces good practices that can help conduct penetration test that is effective and provides valuable results. Software tools and techniques described in this work are also valid and applicable for SCADA systems and moreover in any other field where are computer networks used.**

Keywords—Penetration testing methodology; security; OSSTMM

I. INTRODUCTION

Todays society is often called information society. That is because people who have right information have power and value of some information is immeasurable. If such information is stored in computer system, somehow must be proven this system is safe and not vulnerable. Way how to do so is called penetration testing. Although penetration testing is approach how to increase and strengthen information system security, it definitely doesn't prove that system is completely safe and not prone to hacker attacks. Penetration testing is able to detect publicly known security issues [1] that have been previously revealed and published.

In quite short history of computers and computer networks, there was plethora of incidents related to security and computers in general. Although these incidents differ in impact and danger, none of them was insignificant. In publication dealing with basics of hacking and penetration testing [2] is noted one such example of harmful security incident. Administrator of medium sized serverfarm was suffering from random server restarts. Problem, he wasn't aware of was, restart is only collateral damage and is caused by attacker ending his remote shell session he shouldn't have access to, of course.

This paper outlines what tools and aids can be adopted for better and sometimes faster workflow of penetration test procedures and how to collect and show results in understandable and presentable form to involved audience.

II. AIDS FOR PENETRATION TESTING

First aid which can be used is penetration testing methodology. Some of penetration testing methodologies aim to lead penetration testers through penetration test of web application and other are better suitable for penetration test of operating system, software installed in that operating system and associated network infrastructure.

A. Methodologies and benefits stemming from their usage

As there are many penetration testing methodologies available, it could be sometimes difficult to choose the right one. The more experienced penetration tester is and better knowledge about tested environment he has, the more accurate choice of appropriate methodology he does. There are plenty of penetration testing methodologies penetration tester can choose from. Possible representatives of nowadays modern penetration testing methodologies are NIST SP800-115 created by National Institute of Standards and Technology, Open Source Security Testing Methodology Manual created and maintained by ISECOM and last but not least – Open Web Application Security Project. Each of abovementioned methodologies is suitable for different penetration tests.

OWASP project is created and maintained by OWASP foundation – it is non-profit organization and its members are inter alia experts from web application development industry or software experts. On OWASP project webpage [4] can be found numerous details about OWASP target audience and project targets.

OWASP foundation released numerous publications and one of them is A Guide to Building Secure Web Applications and Web Services [5], where can be found what should be done in each phase of software development if resulting product targets to be as much secure as possible. OWASP project adopted Microsoft's threat modeling because of its simplicity and effectivity. Flowchart for Microsoft's threat modeling is shown on figure 1.

There are more interesting publications stemming from OWASP project. The OWASP Application Security Code of

ISBN: 978-1-61804-233-0

Conduct for Educational Institutions [6] is list of recommendations for educational institutions what to teach if they want to teach programming science in accordance with nowadays modern trends and with emphasis on application security. Book that penetration testers can get most benefit from, is OWASP testing guide [7]. Tremendous amount of valuable hints and information can be found here. Organizations and companies developing software products can establish their own software testing methodology during its life cycle which is based on this publication. Rest of this book contains details about techniques used for penetration testing of web applications.

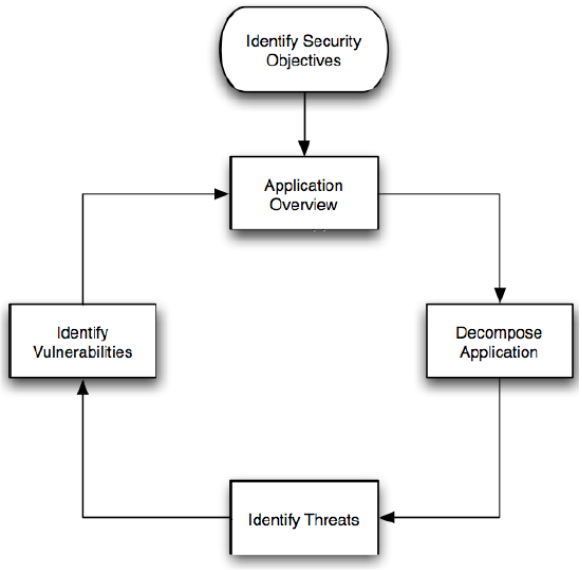


Fig. 1. Threat model flow [5]

NIST SP 800-115 is methodology created by National Institute of Standards and technology – agency funded and governed by US Department of Commerce and can be found at NIST website [8]. There are three ways how system security can be judged according to NIST SP800-115 methodology:

- Examination,
- testing,
- interviewing.

Examination method is least invasive and relies on studying internal directives of organization (possible security vulnerabilities are searched in here), logfiles of critical infrastructure elements are analyzed and if any anomalies are found here, it could lead to finding of security problem. It's fairly questionable how accurate and how valuable information does that kind of testing provide. Contrast to that approach is testing method. Although it gives most accurate results, caution is must. In testing method, one simulates malicious hacker activity against infrastructure like network or servers. Problem is, some of exploit-vulnerability activities are often fatal to target system. Therefore if testing is done in live environment, one must be aware of danger of unintentional Denial of Service. Last method contained in NIST SP800-115 is interviewing. It relies on dialogues about security measures in tested environment with people responsible for infrastructure. Similarly as in examination method, outcomes stemming from this method should be verified or evaluated by testing method.

NIST SP800-115 also provides basic workflow of penetration test procedure which can be used when conducting penetration test. This workflow is shown on figure 2.

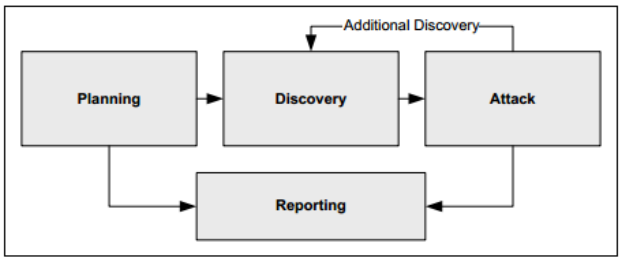


Fig. 2. Penetration testing workflow according to NIST[8]

This methodology isn't updated anymore, therefore it's better to look for another methodology when conducting penetration test against network or servers. One such methodology is OSSTMM.

OSSTMM was first introduced in 2000 and is maintained in current version 3.0 by Institute for Security and Open Methodologies. This methodology is heavily reputable in penetration testers community. This methodology targets to be up to date with modern software trends and technologies. Therefore is updated with half-year periodicity. Although this methodology defines lot of new terminology and theory, it's practical outcomes are highly valuable. Philosophy of this methodology is as follows: OSSTMM claims, that infrastructure elements interact. Approach how to increase security is seen as controlling and restricting those interactions. This methodology has the broadest area of activity. Security is here discussed in five areas:

- Physical security,
- human factor,
- wireless communication.
- telecommunication,
- data networks and operating systems.

Like other two methodologies, this methodology also establishes penetration test workflow, that helps to better organize penetration test procedures. Phases of penetration test, in accordance with OSSTMM, are:

- Induction phase,
- interaction phase,
- inquest phase,
- intervention phase

Main tasks of each phase can be easily deducted from its name, however brief introduction of each phase follows.

ISBN: 978-1-61804-233-0

Induction phase should clarify time range for penetration test and types of test to be used. Interaction phase determines which targets are in scope of particular penetration test. Inquest phase finds out as much data as possible about target systems and last but not least – intervention phase verifies functionality of security and alerting mechanisms.

After penetration test has been finished, reporting phase, where results are processed, begins. Compared to other methodologies, OSSTMM has one immense advantage. It has toolset to process results effectively – it is named Security Test Audit and Reporting. Usage of this toolset is simple – it is predefined spreadsheet where inputs are identified interactions and controls of that interactions, output is numerical value called rav. Rav expresses if there is need for additional controls (value below 100) or everyting is properly balanced (value equal 100), or if there are more controls of interactions than necessary (value above 100). Therefore, if one environment is tested twice, it can be easily seen if security of this environment has been improved or has deteriorated.

Truth is, OSSTMM is quite comprehensive. It takes considerable amount of time to get familiar with all the details and new terms defined there. Despite that fact OSSTMM offers complete guideline for penetration testing to non-experienced penetration testers. More information about OSSTMM can be found on [9].

All of abovementioned methodologies are suitable for different purposes and penetration tests. Although there has to be done some initial effort to understand terms and procedures defined in methodologies, this effort is often worth the result, especially for testers who have little or no experience with commercial penetration testing. It is often shown what needs to be done before starting penetration test, what should contract contain and what steps are needed to be taken during penetration test. Some of methodologies even give some tips regarding final reporting. There are often different audiences, thus different kinds of reports have to be created – more comprehensive for technical staff and less detailed for managers. Methodologies often contain hints what should each kind of report contain and how it should be structured.

B. Software tools for penetration testing

Some of the methodologies outlined in preceding paragraphs often give hints how to structure penetration test. Moreover they often suggest to divide test into phases, where output from each phase servers as input for next phase. Consequences of that approach are that different kinds of tools are suitable for different phases of penetration test. Following paragraphs outline and introduce these tools briefly.

Different approaches can be taken when choosing what toolsets can be used for carrying out phases of penetration test. There are plenty of tools and toolsets for penetration testing and can be used for testing various types of products and conducting diverse types of attacks. Following paragraphs introduce briefly some of these toolsets and discuss what benefits they bring into penetration testing.

Backtrack is Linux Debian-based distro. Nowadays its development has been discontinued and is superseded by Kali Linux, which has same purpose and way of use. Kali‘s or Backtrack’s feature is that it can run on tester’s computer without installation, in so-called live mode. On the other hand, changes in operating system made in live session are lost after reboot. Abovementioned distros are pre-loaded with large amount of software for different areas of penetration testing. Kali contains numerous tools and frameworks for penetration testing. These tools are usually used for:

- Database security audit,
- SQL injection techniques,
- network traffic eavesdropping or tampering,
- network infrastructure attack,
- network stress testing,
- DoS attacks,
- manipulating user data,
- web application penetration testing,
- and much more tasks.

Further info about Kali linux can be found at Kali project webpage [10].

First goal of each penetration test is to find systems on network that reply for network communication and thus are interesting from pentetration‘s tester perspective in further phases of penetration test, because there could be vulnerable software. This can be done using network scanner. Most known network scanner is Nmap.

Name of this network scanner is composed of beginnings of words network and mapper. Originally, Nmap was small toolkit with minimal capabilities, but during the years, developer community began to grow. Thus, amount of Nmap‘s capabilities grew and grew until todays state, when Nmap is most comprehensive and complex network scanner available with highest amount of options for computer network exploration. Community of Nmap developers has also clear vision of direction Nmap future development should follow [11]. What is also pleasant, is amount of documentation and available information on Nmap, most of it can be found on Nmap website [12]. Nmap allows advanced users to script routine tasks with Lua scripting language. There is a graphical extension for Nmap called Zenmap, which is suitable for users who aren't comfortable with command line interface.

Nmap is able to detect hosts in networks in different ways and some of them are able to detect hosts that are undetectable with classical approaches. Matter of course is ICMP ping scan that verifies host status with ICMP echo-request and ICMP echo-reply. Another technique to detect hosts on network segment is called ACK ping. This method exploits some features of IP stack in operating system designed and developed in compliance with RFC 793 standard. This standard defines how to reply when host receives TCP segment that isn't part of any existing TCP connection and has ACK flag in header. RFC 793 standard says host should reply

with segment having RST flag turned on. This method reveals whether systems that have communication with ICMP protocol forbidden, are up and running. Truth is, state firewalls often drop traffic when network is explored with ACK ping, because TCP segments sent out by tester are usually not part of existing TCP connection. Further info can be found in [13]. If either ICMP or ACK ping isn't enough, there is another technique for finding which hosts are alive on network subnet - UDP ping. As its name suggests, it's based on UDP protocol. If host receives UDP datagram on port that is closed, it usually responds with ICMP - port unreachable and de facto reveals its existence.

After all available hosts have been found, it's necessary to discover which services are running at which host. Port scanning techniques are intended for these purposes. First option how to determine which ports are open on target system, is TCP connect scan. This technique is based on TCP protocol internals. When scanning with TCP connect scan technique, Nmap tries to establish connection with three way handshake in defined port range. After receiving a reply, Nmap evaluates it and if there are ACK and SYN flags set, port is definitely open. On the other side, when segment with RST flag set is received, port is closed for sure. TCP SYN scan is technique similar to TCP connect scan. One difference is - TCP SYN scan does not complete connection establishment in order to be less noticeable for firewalls and similar network protection systems. UDP scan, again - as its name suggests, strives to find whether particular host is running any services on UDP protocol. This information is often valuable, because in history there has been plenty of vulnerable services running on UDP protocol - e.g. DNS service running on BIND daemon. In book written by Gordon Fyodor - author of Nmap scanner [13] - can be found further info about port scanning.

Last but not least, Nmap has one technique that other network scanning tools don't have. It is called idle scan. If scanning techniques would be evaluated acording to secrecy, idlescan would be winner. Idea of idlescan is simple - there should be intermediary host, so-called zombie host, available and traffic is sent to scanned system on behalf of this zombie host. Before scan is initiated, details about TCP segment numbers are detected, scan is conducted and TCP segment numbers are evaluated again. Either Nmap or Metasploit contain plugin which is capable of determining whether particular host is appropriate candidate for zombie host role or not.

Software vulnerabilities are mostly caused by errors that have been incorporated info software intentionally or by mistake. These vulnerabilities allow attackers to perform actions that couldn't be done without vulnerabilities presence. Source code for vulnerability exploitation is different from vulnerability to vulnerability. Thus, it is necessary to know what vulnerabilities are present on target systems. This information can be provided by vulnerability scanning systems like OpenVAS or Nessus.

Nessus can be controlled via web interface. There are default policies that can be used for scanning right after install or one can implement and fine-tune own policies. If second option is chosen, it must be defined which techniques will be used for port scanning or what plugins will be used for vulnerability site survey. Once test is finished, summary of results follows. Individual vulnerabilities are shown and each vulnerability is categorized depending on its severity. Obviously, highest attention should be payed to vulnerabilities whose category is critical. Nessus can be extended with plugins that are written in NASL - Nessus attack scripting language. There is also possibility to buy professional edition which has some extended capabilities like SCADA systems scanning compared to free edition. Accurate comparsion between Nessus versions can be found at official Nessus website [14]. Professional license costs are in thousands dollars per year. Free alternative to Nessus is OpenVAS. Years ago, Nessus was free and had available source code. OpenVAS developers got inspired by Nessus' source code, thus OpenVAS has similarities compared to Nessus.

Responsibility of Nessus and OpenVAS is to determine if there are any exploitable vulnerabilities in tested system. Next step when performing penetration test is to attempt to exploit vulnerabilities. For these purposes, snippet of source code - so-called exploit - is needed. After successful exploitation, sequence of steps is usually performed. As a result of vulnerability exploitation, tester often gets permission to execute arbitrary commands on target system. Tasks like exploitation of vulnerability or post-exploitation exploration of target system can be automated to a certain extent. For these purposes is available Metasploit project toolkit. Its development began in 2003. Since 2003, Metasploit went through rapid development and in 2009 was acquired by Rapid7 company. Nowadays, Rapid7 is in charge of funding and development of Metasploit. Metasploit is designed with an emphasis on simple extension. Metasploit's architecture aims to be modular and it's depicted on figure 3.

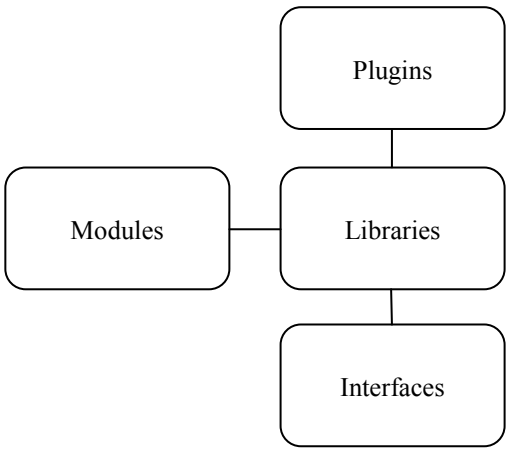


Fig. 3. Metasploit architecture overview

Libraries provide basic services like networking and files manipulation to neighboring components, thus, developer doesn't need to take care of routine tasks like communicating via network - it is already programmed in libraries. Via interfaces, Metasploit core is controlled by users. There are

many interfaces and following paragraphs introduce them briefly.

Msfcli is used directly from command line. It is suitable for beginners and newcomers to Metasploit, because it helps to understand how Metasploit internals work. Msfcli is controlled via command line parameters and arguments. There are detailed tutorials for Msfcli usage on Metasploit unleashed website [15].

Msfconsole is another interface available for Metasploit interaction. Compared to Msfcli, Msfconsole is more robust, scalable and easier to use. It allows to define global variables, perform lookups in exploit database and more Meterpreter sessions can be maintained in single Msfconsole - practical usage is shown in Metasploit unleashed tutorial.

There is also GUI for Metasploit - it is called Armitage. It's suitable for users who don't have so much experience with command-line usage and displays connection topology of tested sysytems.

One can benefit from metasploit usage in several ways. There is effective exploit management (lookup, update, documentation), plethora of payloads (tasks that are performed after successful exploit). Payloads which provide commandline of target system are used mostly. Figure 4 shows sequence of steps that are necessary to establish two-way communication channel between tester‘s and tested system.

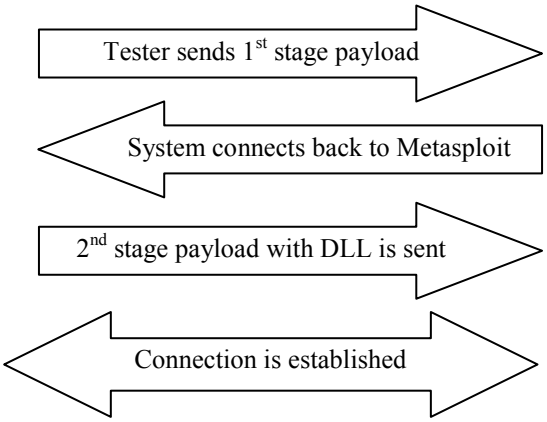


Fig. 4. Metasploit connection establishment

If two-way communication channel is established between testers's and target system, there are almost unlimited options what can be done with target system. This is so because of special DLL library, which is sent to target system. This library is loaded via DLL injection technique into RAM of target system and here resides in order to receive and execute commands from tester's special shell. This shell is called Meterpreter shell and is used for target system controlling, one can execute arbitrary commands, browse files or monitor user's activity with it. Communication protocol between target and tester's machine is built on type-length-value model. This is so because Meterpreter and Metasploit are developed with scalability in mind. There are other applications of type-length-value model, too. One such example is EIGRP routing protocol developed by Cisco company. If protocol based on type-length-value model needs to be extended new type is defined and current source code implementing current protocol behavior doesn't need to be changed, only extended. Traffic between Meterpreter client (tester) and server (target) is encrypted, thus, privacy is ensured.

Metasploit Framework is also updated on regular basis – new malicious code used for exploiting system vulnerabilities is added when update is triggered. When penetration tester gets access to tested system after vulnerability has been exploited, there is often need for remote control of that system. Metasploit offers remote shell called Meterpreter, where various tasks like uploading or downloading files, starting or terminating processes, taking snapshots on webcam can be done without difficult programing on penetration tester’s side.

Despite its initial complexity for newcomers to penetration testing, tools like Metasploit Project offer many valuable tools to conduct penetration testing and simplify some tasks, that would have to be solved programatically without framework.

III. PROCESSING RESULTS EFFECTIVELY

In every phase of penetration test are some results provided. It's a must to work with these results quickly and effectively, because during penetration test, time is an important matter. One must be able to perform a lookup quickly. Software products introduced above are able to work with results in several different ways. For example, Nmap offers output in four different formats:

- Classical,
- XML,
- greppable,
- script kiddiez.

Classical format only saves console output to text file, XML format presents results in XML language. Greppable format is suitable for analysis in Linux command line. Its intended for grep command, which allows to perform lookups in text and is able to use complex regular expressions. Benefit of script kiddiez output is questionable. It can be understood as some kind of easter egg, because output is similar to textual, but some parts are deliberately mispelled. One example of script kiddiez line output is shown below and says that on target systém is port 445 in open state:

```
445/tCp 0p3n  M|croSOft-dS
```

Metasploit also offers advanced options for working with scan and exploitation results. This is so because Metasploits’ core closely cooperates with PostgreSQL database engine. Msfconsole contains command db_nmap that is used for calling nmap from Msfconsole environment. This command has completely same way of usage as classic Nmap from Linux shell. Results of scan launched by db_nmap command

aren't stored as a text file. Instead of that, results are put into database. Immense advantage of that approach is that one can perform queries with parameters against that database. Moreover syntax of this kind of queries is easy to understand. Despite Metasploits' versatility and its complexity, storing Nmaps' output in database has some limitations. Following overview briefly outlines what can be stored in Metasploit's database when performing network and host scanning:

- IP addresses,
- open ports,
- services running on open ports,
- vulnerabilities of those services,
- operating systems of scanned hosts,
- hashdumps.

Not even conduction of penetration test requires rigorous and careful planning, but also results processing should be taken into consideration, because fast orientation in collected data is essential and sometimes critical for successful penetration test.

IV. Conclusion

In nowadays modern world, there is need for proactive approach to information security in order to avoid potential security breaches. It is so, because security breach and consequent data loss or data tampering usually costs huge amount of money and causes loss of reputation for company. There are often increased costs for security measures, but advantages brought by these measures worth it. Level of security can be determined by discipline called penetration testing. There are many toolsets and utilities which can be used – from methodologies to software utilities introduced in this article. Usually, it is desirable to spend some time with learning and getting familiar with methodologies and software tools, because cumbersome task are automated and done in few clicks or with few commands. Thoughts about complex and rigorous preparation for penetration test can be emphasized by quotation of former US president, Abraham Lincoln, who said:"If I had eight hours to chop down a tree, I'd spend six hours sharpening my axe".

ACKNOWLEDGMENT

This work was supported by the University of Pardubice in Student's Grant contest project.

References

[1] Towards a practical and effective security testing methodology. Proceedings - IEEE Symposium on Computers and Communications [online]. 2010, č. 1, s. 320-325 [cit. 2013-04-03]. DOI: 10.1109/ISCC.2010.5546813. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5546813f49b

[2] Mati Aharoni ([2011]). Penetration testing with Backtrack . 3rd ed. Cornelius, NC: Offensive Security. 12. SP800-115. Technical Guide to Information Security Testing and Assessment.

[3] Gaithersburg: [NIST Computer Security Division], 2008. Available: http://csrc.nist.gov/publications/nistpubs/800-115/SP800-115.pdf

[4] Owasp foundation. (2014). Owasp project web page. Available: https://www.owasp.org/index.php/Main_Page. Last accessed 14th Jan 2014. OWASP foundation. (2011). A Guide to Building Secure Web Applications and Web Services [Online]. Available: http://sourceforge.net/projects/owasp/postdownload?source=dlp

[5] Owasp foundation. (2005). A Guide to Building Secure Web Applications and Web Services. Available: prdownloads.sourceforge.net/owasp/OWASPGuide2.0.1.pdf?download. Last accessed 14th Jan 2014.

[6] Offensive Security. (2013). A Guide to Building Secure Web Applications and Web Services. Available: https://www.owasp.org/images/6/6b/OWASP_Blue_Book-Educational_Institutions.pdf. Last accessed 14th Jan 2014.

[7] Owasp foundation. (2008). OWASP Testing Guide. Available: https://www.owasp.org/index.php/OWASP_Testing_Guide_v3_Table_of_Contents. Last accessed 14th Jan 2014.

[8] Karen Scarfone, Murugiah Souppaya, Amanda Cody, Angela Orebaugh (2008). Technical Guide to Information Security Testing and Assessment. Gaithersburg: NIST. 1-80.

[9] ISECOMM. 2010. Open Source Security Testing Methodology Manual (OSSTMM). [ONLINE] Available at: http://www.isecom.org/mirror/OSSTMM.3.pdf. [Accessed 26 January 14].

[10] Kali Linux. 2014. Kali Linux | Rebirth of BackTrack. [ONLINE] Available at: http://www.kali.org/. [Accessed 26 January 14].

[11] NMAP. 2013. The history and future of Nmap. [ONLINE] Available at: http://nmap.org/book/history-future.html. [Accessed 26 January 14].

[12] NMAP. 2014. Nmap - Free Security Scanner For Network Exploration & Security Audits. [ONLINE] Available at: http://nmap.org/. [Accessed 26 January 14].

[13] Gordon Fyodor Lyon, 2009. Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning. Edition. Nmap Project.

[14] Tenable network security. 2014. Nessus editions | Tenable network security. [ONLINE] Available at: http://www.tenable.com/products/nessus/editions. [Accessed 26 January 14].

[15] Metasploit Unleashed. 2014. Metasploit Unleashed. [ONLINE] Available at: http://www.offensive-security.com/metasploit-unleashed/Main_Page. [Accessed 26 January 14].

Authors Index

ISBN: 978-1-61804-233-0