David C. Wyld
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

Second International Conference on Computer Science, Information
Technology and Applications
Zurich, Switzerland, January 2~3, 2017

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Dhinaharan Nagamalai,
Wireilla Net Solutions, Australia
E-mail: dhinthia@yahoo.com

Preface

The Second International Conference on Computer Science, Information Technology and Applications was held in Zurich, Switzerland, during January 2~3, 2017. The Second International Conference on Image and Signal Processing, The Second International Conference on Artificial Intelligence, The Second International Conference on Data Mining and Applications, The Seventh International conference on Computer Science and Information Technology, The Fifth International Conference on Artificial Intelligence, Soft Computing, The Fifth International Conference on Signal, Image Processing and Pattern Recognition, The Sixth International Conference on Parallel, Distributed Computing Technologies and Applications and The Second International Conference on Software Engineering was collocated with The Second International Conference on Computer Science, Information Technology and Applications. The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld
Dhinaharan Nagamalai

Organization

General Chair

Natarajan Meghanathan,	Jackson State University, USA
Brajesh Kumar Kaushik,	Indian Institute of Technology - Roorkee, India

Program Committee Members

Aalya Alajaji	Prince Sultan University, Saudi Arabia
Abd El-Aziz Ahmed	Cairo University, Egypt
Abraham Gizaw	Hawasa University, Ethiopia
Addi AIT-MLOUK	Cadi Ayyad university, Morocco
Ahmad T. Al-Taani	Yarmouk University, Jordan
Ahmed Mohamed Khedr	Sharjah university, UAE
Akhil	Shantou University, China
Ali Salem	University of Sfax, Tunisia
Ammar Al-Masri	Albalqa Applied University, Jordan
Anis Zouaghi	University of Sousse, Tunisia
Bartholomew Placzek	University of Silesia, Poland
Bernardo Almada-Lobo	University of Porto, Portugal
Christina Kluever	University of Duisburg-Essen, Germany
Chunshien Li	National Central University, Taiwan
Daniel Gomes	Estácio de Sá, Brasil
Daniele Codetta-Raiteri	University of Eastern Piedmont, Italy
Dinyo Omosehinmi	Colossus Technology, Nigeria
Dongchen Li	Peking University, China
Emad Al-Shawakfa	Yarmouk University, Jordan
Emilio Serrano	Technical University of Madrid, Spain
Wael abou el-wafa	Minia University, Egypt
Eyad M. Hassan ALazam	Yarmouk University, Jordan
Fatih korkmaz	Cankiri Karatekin University, Turkey
Fernando Tello Gamarra	Federal University of Santa Maria, Brazil
Gammoudi Aymen	University of Tunis, Tunisia
Gelenbe	Imperial College, UK
Goreti Marreiros	Polytechnic of Porto, Portugal
Guilherme Oliveira	Univates University Center, Brazil
Isa Maleki	Islamic Azad University, Iran
Ismail Shahin	University of Sharjah, UAE
Jan Lindstrom	MariaDB Corporation, Finland
Joao Gama	University of Porto, Portugal
John Tass	University of Patras, Greece
Jose Antonio Correia	University of Porto, Portugal
Jose Raniery	University of Sao Paulo, Brazil
Khaled Almakadmeh	Hashemite University, Jordan
Khalid M. Oqla Nahar	Yarmouk University,Jordan

Lim Seng Poh	Wawasan Open University, Malaysia
Liviu Octavian Mafteiu-Scai	West University of Timisoara, Romania
Liyakath Unisa	Prince Sultan University, Saudi Arabia
Liyakathunisa Syed	Prince Sultan University, Saudi Arabia
Mahdi Salarian	University of Illinois, USA
Mahmoud R. Delavar	University of Tehran, Iran
Mamoun Abu Helou	Al Istiqlal University, Palestine
Manish Mishra	University of Gondar, Ethiopia
Marco Anisetti	University of Milan, Itlay
Mario Henrique Souza Pardo	University of São Paulo, Brazil
Marta Ruiz	Polytechnic University of Catalonia, Spain
Masoumeh Javanbakht	Hakim Sabzevari University, Iran
Mastaneh Mokayef	UCSI University, Malaysia
Maysam Toghraee	Yasouj Science and Research Branch, Islamic
Mehrdad Jalali	Mashhad Azad University, Iran
Meriem bensouyad	Université Mentouri de Constantine, Algeria
Mohamad Heidari	Islamic Azad University, Iran
Mohamed Arezki Mellal	M'Hamed Bougara University, Algeria
Mohamed Tounsi	Prince Sultan University, Saudi Arabia
Mohammad Arshi Saloot	University of Malaya, Malaysia
Mohammad Ashraf OTTOM	Yarmouk University, Jordan
Mohammad Zarour	Prince Sultan University, Saudi Arabia
Mohammed AbouBakr Elashiri	Beni Suef University, Egypt
Mohammed Al-Sarem	Taibah University, KSA
Mohammed Ghazi Al-Zamel	Yarmouk University,Jordan
Mudassir Khan	King Khalid University, Saudi Arabia
Nahlah M. Ameen Shatnawi	Yarmouk University,Jordan
Nicolas H. Younan	Mississippi State University, USA
Pasi Luukka	Lappeenranta University of Technology Finland
Paulo Roberto Martins de Andrade	University de Regina, Canada
Quang Hung Do	University of Transport Technology, Vietnam
Rafah M. Almuttairi	University of Babylon, Iraq
Rim Haddad	Innov'com Laboratory, Tunisia
Ronghuo Zheng	University of Texas, United States
Saeed Tavakoli	University of Sistan and Baluchestan, Iran
Samad Kolahi	Unitec Institute of Technology, New Zealand
Samadhiya	National Chiao Tung University, Taiwan
Sanjay Sharma	University of london, UK
Seyed Ziaeddin Alborzi	Universite de Lorraine, France
Siuly Siuly	Victoria University, Australia
Soha Mohamed	Harbin Institute of Technology, Egypt
Soumaya Chaffar	Prince Sultan University, Saudi Arabia
Suleyman Al-Showarah	Isra University, Jordan
Tahmid Rahman Laskar	Leading University, Bangladesh
Thiago Pinheiro	Federal University of Amapá, Brazil
Tranos Zuva	Tshwane University of Technology, South Africa
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Wei Cai	University of California, USA
Wenjuan (Wendy) Xu	Frostburg State University, USA

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Digital Signal & Image Processing Community (DSIPC)

Software Engineering & Security Community (SESC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

Second International Conference on Computer Science, Information Technology and Applications

Second International Conference on Image and Signal Processing

Sixth International Conference on Parallel, Distributed Computing Technologies and Applications

Second International Conference on Software Engineering

Invited Paper

THE SUSCEPTIBLE-INFECTIOUS MODEL OF DISEASE EXPANSION ANALYZED UNDER THE SCOPE OF CONNECTIVITY AND NEIGHBOR RULES

Maria Teresa Signes Pont, Higinio Mora Mora and Antonio Cortés Castillo

Departamento de Tecnología Informática y Computación,
Universidad de Alicante, Spain
teresa@dtic.ua.es; hmora@dtic.ua.es;
antoniocortescastillo@gmail.com

ABSTRACT

This paper presents a model to approach the dynamics of infectious diseases expansion. Our model aims to establish a link between traditional simulation of the Susceptible-Infectious (SI) model of disease expansion based on ordinary differential equations (ODE), and a very simple approach based on both connectivity between people and elementary binary rules that define the result of these contacts. The SI deterministic compartmental model has been analysed and successfully modelled by our method, in the case of 4-connected neighbourhood.

KEYWORDS

Infectious disease expansion, deterministic compartmental models, ODE, Neighbour binary rules, Connectivity.

1. INTRODUCTION

The outbreaks of infectious disease pandemics have shaped nations and civilizations through the ages. History tells us the terrible impact of biblical plagues in Ancient Egypt, bubonic plague in Europe in the Middle Age, influenza at the beginning of the twentieth century or AIDS the more recently emerging pandemics [1]. Mathematical modeling is playing a very important role to assess and control the potential outbreaks [2]. The first paper presenting a model for an infectious disease appears in 1760. The author, Bernouilli, a swiss mathematician and physicist, dealt with a statistical problem involving censored data in order to analyze smallpox morbidity and mortality that aimed to demonstrate the efficacy of vaccination [3]. At the beginning of the twentieth century two pioneering works can be mentioned. W.H. Hamer [4] published a discrete time epidemic model for the transmission of measles in 1906. The model assumes that the number of cases per unit time (incidence) depends on the product of densities of the susceptibles and infectives. In 1911 R. Ross [5] demonstrated that malaria is produced by the bite of a mosquito.

His mathematical model of expansion is based on a set of equations to approximate the discrete-time dynamics of malaria and asserts is possible to control the disease whenever the population of mosquitos is reduced below a threshold. This was a new and crucial idea. Between 1927 and 1939 Kermack and McKendrick [6, 7] published papers on epidemic models and obtained the epidemic threshold that the density of susceptibles must exceed for an epidemic outbreak to occur. This model includes three states, the S (susceptible), I (Infectious) and R (Recovered) instead of the two, S and I, of the Bernouilli's model. From the mid-twentieth century a great variety of epidemiological models have been developed after the recognition of the importance of modeling in public health decision making [8]. In the nineties, when the scientists began to pay attention to complex systems new paradigms spread out in order to better understand and model the impact of numerous variables that go beyond the micro host–pathogen level, such as ecological, social, economic, and demographic factors. Many scientists coming from such different fields as medicine, molecular biology, computer science and applied mathematics or economy have teamed up for rapid assessment of urgent situations of contagious diseases by means of a multidisciplinary approach. The case of HIV/AIDS pandemic [9-12] is a good example.

This paper presents a model to approach the dynamics of infectious diseases expansion by means of a set of neighbour rules between elements located in a lattice that represents the whole population. Following the introduction, Section 2 provides a brief summary of the deterministic compartmental models and highlights the Susceptible-Infectious (SI) model which has traditionally been solved by ODE. Section 3 presents our model which considers the population confined in a lattice. The contacts between people are performed by neighbour binary rules, that are tailored to model different situations such as Susceptible, Infected, with or without capability to infect further. The neighborhood is also defined depending of connectivity. We consider 4-connection, 8-connection and horse jumping chess connection. The results are compared with those of the simulation of ODE. Section 4 presents a discussion upon the suitability of the model and proposes futures research. Section 5 summarises the work and presents concluding remarks.

2. MATHEMATICAL MODELLING

Three are the main categories encompassing mathematical modeling [1]. The statistical methods deal with real epidemics. They identify their spatial patterns and allow surveillance of outbreaks. The empirical models are based on machine learning methods such as data mining that allow the forecasting of the evolution of an ongoing epidemic spread. The mathematical or state-space methods provide quantitative predictions that have to be validated to forecast the evolution of a hypothetical or real epidemic spread. These methods also redefine our understanding of underlying mechanisms.

2.1. The Deterministic Compartmental Models

The description used in epidemiologic compartmental models is composed of standard categories represented by the variables that model the main characteristics of the system. These compartments, in the simplest case, divide the population into two health states: susceptible to the infection (denoted by S) and infected (denoted by I) [13]. The way that these compartments interact is often based upon phenomenological assumptions, and the model is built up from there. Usually these models are depicted by ODE, which are deterministic, but can also be viewed in more realistic stochastic framework [14]. To achieve more realism, other compartments are often included, namely the recovered (or removed or immune) compartment labelled by R, or the

exposed compartment, labelled by E. The stratification of these compartments lead to well-known models such as SIS [15], SIR [16], SEIR [17, 18],…or more complex ones [19]. The number of variables to be incorporated to the model depends on the particular disease being studied as well as on the desired complexity of the model. Other variables incorporated into the equation represent fundamental quantities such as birth rate, rate of transmission of infectious agent, death rate, and so forth, and are constants that can be changed.

2.2. The Traditional Susceptible-Infectious (SI) Model

In the SI model the two groups are the susceptible hosts, S, that are not infected by the pathogen but can get infected, and the infected hosts, I, who are infected by the pathogen. Assuming the mass-action model, the rate at which susceptible hosts become infected is a product of the number of contacts each host has per unit time, r, the probability of transmission of infection per contact, β, and the proportion of the host population that is infectious, I/N, where N = S + I is the total population size. This model is suitable to represent the case of the human immune deficiency virus (HIV) where there is no recovery. A schematic of the model is shown in Figure 1.

Figure 1. SI Model

Equations (1) for the SI model are as follows:

$$\frac{dS}{dt} = -\beta r S \frac{I}{N}$$
$$\frac{dI}{dt} = \beta r S \frac{I}{N} \quad (1)$$

Since the population size is fixed, we can reduce the system to one dimension with the substitution S = N – I to provide the logistic Equation (2).

$$\frac{dI}{dt} = \beta r (N - I) \frac{I}{N} \quad (2)$$

We can analytically solve Equation (2) with the initial condition $I(0) = I_0$, so

$$I(t) = \frac{I_0 N}{(N - I_0) e^{-\beta r t} + I_0} \quad (3)$$

The simulation of the SI model is shown in Figure 2 with the initial value $I(0) = I_0 = 1$
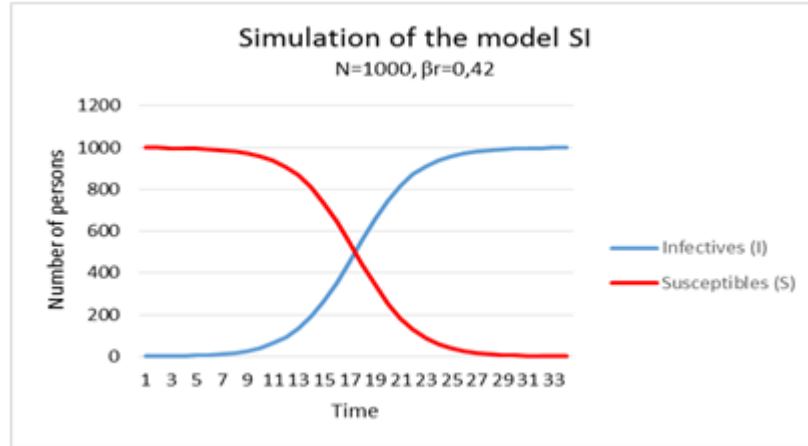
Figure 2. Simulation of the traditional SI Model

3. OUR PROPOSAL

Our proposal is based on a set of elementary binary rules that have the capability to model interactions between two individuals [20-23]. Without loss of generality we consider a two-dimensional square lattice, every cell represents a susceptible person except the one at the center which locates an infected one. When the infected person contacts with his/her neighbors he/she spreads the disease. The new infected people have then the capability to infect other people.

3.1. Binary neighbor rules

Equation (4) defines a generic binary neighbor rule denoted $\otimes$.

$$\otimes : (0,1) \rightarrow (0,1)$$
$$(x, y) \rightarrow z = x \otimes y \tag{4}$$

The $\otimes$ rule can be represented by a two input table that defines concretely the operation, as shown in Figure 3.

$\otimes$	0	1
0	a_3	a_1
1	a_2	a_0

Figure 3. Generic neighbor rule represented by a table

Let m stand for the number of the rule. The number is represented by the four bits stored in the cells; $m = a_3 a_2 a_1 a_0, \in [0, 2^4-1]$; $a_i \in (0, 1)$; $i \in [0, 3]$. As an example, we consider $m = 7$, that is to say $a_3 = 0$; $a_2 = 1$; $a_1 = 1$ and $a_0 = 1$.

$\otimes$	0	1
0	0	1
1	1	1

Figure 4. The rule n° 7 represented by a table ($a_3 = 0$; $a_2 = 1$; $a_1 = 1$; $a_0 = 1$)

The table defines concretely the operation as follows:
$0 \otimes 0 = 0$; $0 \otimes 1 = 1$; $1 \otimes 0 = 1$ and $1 \otimes 1 = 1$

This operation is suitable to model the interaction between infected people and susceptible people by identifying "0" as susceptible and "1" as infected. The previous operation means that infected people can transmit the disease to susceptible people ($1 \otimes 0 = 1$), when infected people contact other infected people, all them remain infected ($1 \otimes 1 = 1$), susceptible people have no effects upon people ($0 \otimes 0 = 0$ and $0 \otimes 1 = 1$).

3.2. Neighborhood rules modelling the spreading of a disease

In the following two-dimensional square lattices we present the spreading of a disease by a unique Infected ("1") located at the center of the lattice. All the empty cells are considered to be Susceptible ("0"). The red numbers stand for the generation number (time unit) the spreading occurs. Figure 5 shows the case of a 5x5 lattice with a contagion rate $\rho = 4$ per generation (4-connected cells are neighbors to every cell that touches one of their edges, following the Von Neumann neighbourhood).
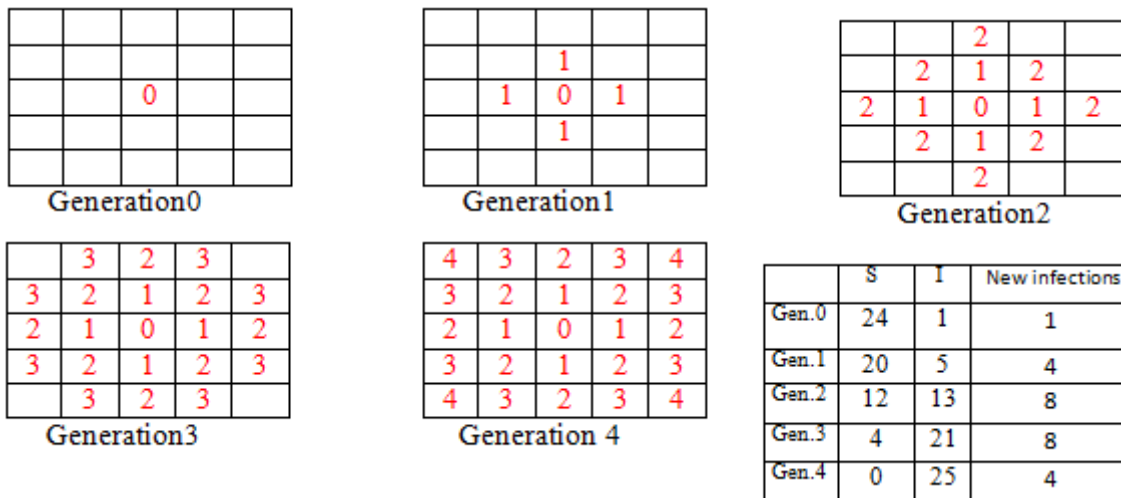
Figure 5. The spreading of a disease in a 5x5 lattice with $\rho = 4$ (4-connected cells)

	S	I	New infections
Gen.0	24	1	1
Gen.1	20	5	4
Gen.2	12	13	8
Gen.3	4	21	8
Gen.4	0	25	4

The spreading of the disease results in a diamond-shaped region shown for rate = 4 in Figure 5. The evolution of the infected people can be carried out by means of the equation $1+2\rho(\rho+1)$, where ρ stands for the rate.

The same example is presented for 8-connected cells (with horizontal, vertical, and diagonal connection) following the Moore neighborhood, and for the jumping chess neighbourhood. See Figures 6 and 7 respectively.

		0		

Generation0

	1	1	1	
	1	0	1	
	1	1	1	

Generation1

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

Generation2

	S	I	New Infections
Gen.0	24	1	1
Gen.1	16	9	8
Gen.2	0	25	16
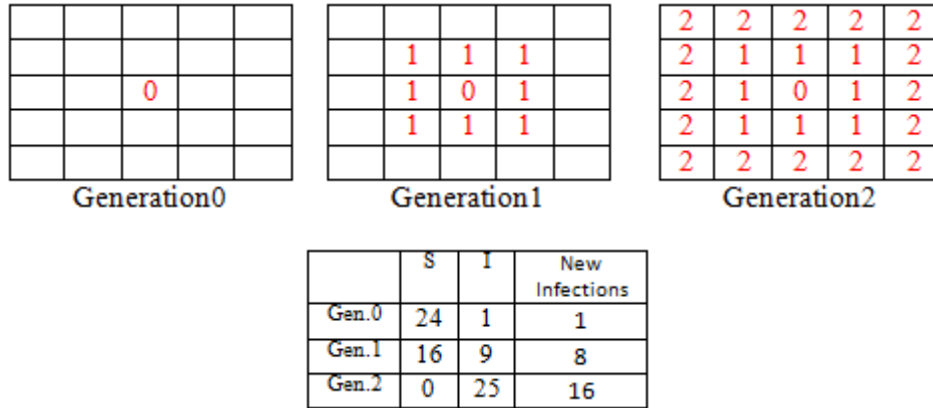
Figure 6. The spreading of a disease in a 5x5 lattice with $\rho = 8$ (8-connected cells)

For the 8-connected neighborhood, the evolution of the infected people can be carried out by means of the equation $(2\rho+1)^2$.

		0		

Generation0

	1		1	
1				1
		0		
1				1
	1		1	

Generation1

	1	2	1	
1	2	2	2	1
2		0		2
1	2	2	2	1
	1	2	1	

Generation2

3	1	2	1	3
1	2	2	2	1
2	3	0	3	2
1	2	2	2	1
3	1	2	1	3

Generation3

	S	I	New Infections
Gen.0	24	1	1
Gen.1	16	9	8
Gen.2	6	19	10
Gen. 3	0	25	6
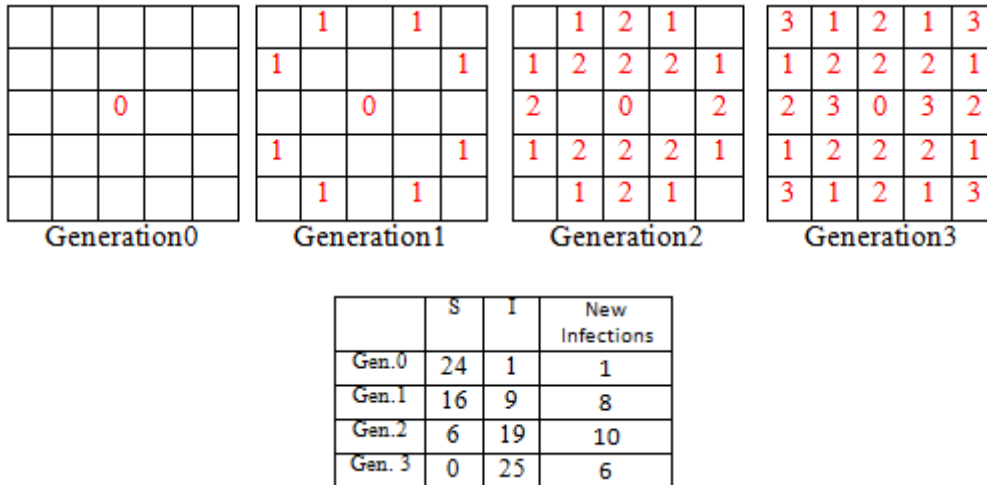
Figure 7. The spreading of a disease in a 5x5 lattice with $\rho = 8$ (horse jumping chess connected cells)

3.3. Comparison Between the Traditional SI Model and the 4-Connected Neigborhood Model

The following graph represents the 4-connected case in a 32x32 lattice, equivalent to N=1024 (in order to approximate the graphic shown in Figure 2 where N=1000).
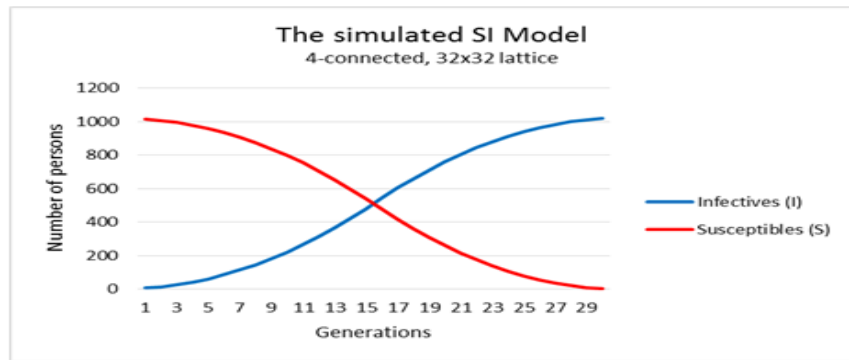
Figure 8. The spreading of a disease in a 32x32=1024 lattice with $\rho = 4$ (4 -connected cells).

The comparison between Figures 2 and 8 suggests that the 4-connected cells (Von Neumann neighbourhood) could be a suitable approximation to approach the traditional SI model. In order to better compare the models, we now compare the previous simulation based on Equation 2 with our 4-connected neighborhood model for similar populations, lattices 10x10, equivalent to N=100 and 100x100, equivalent to N=10000, See Figures 9 and 10.

Figure 9. The spreading of a disease in a 10x10 lattice with a $\rho = 4$ (4 -connected cells), compared to the simulation of the traditional model (N=100, $\beta r = 0,82$).

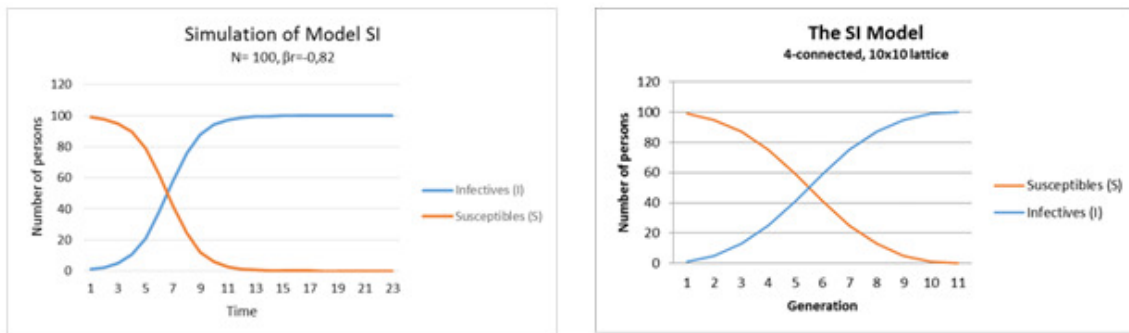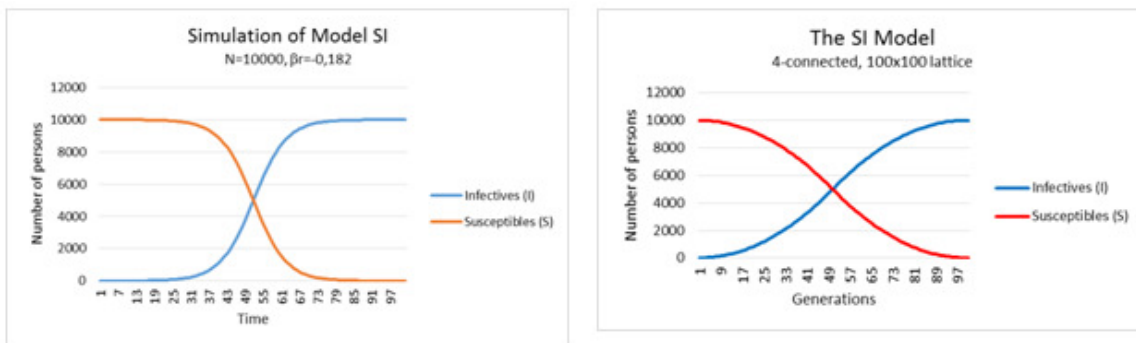

Figure 10. The spreading of a disease in a 100x100 lattice with $\rho = 4$ (4 -connected cells), compared to the simulation of the traditional model (N=10000, $\beta r = 0,182$).

Empirically, the rate βr of the traditional simulation of Model SI has been modified when N varies in order to impose the crossing point between (I) and (S) occurs at the same "time" (obviously we have assumed that "Time" in the traditional model is equivalent to "Generation" in ours, as explained in Section 4.).

4. DISCUSSION

Our model establishes a link between the traditional ODE simulation of the SI deterministic compartmental model of disease expansion and a very simple model based on both the connectivity between people and different rules that define the results of the contacts. The parameters of each model must harmonize. Our lattice size stands for N, the number of persons (N = S + I). In the ODE model, the number of contacts each host has per unit time, r, and the probability of transmission of infection per contact, β, have been englobed in a unique variable, βr, which stands for the number of actual infections that occur in a unit time, which is equivalent to the connectivity, ρ, in our model. Obviously, the time scale is different in each model, so, unit time may be days, hours, etc... In our model, "Generation" is a dimensionless unit, which only means how the sequence of infections occur. In order to allow meaningful comparisons between the different approaches we have harmonized the parameters of the models as follows, N ≡ size of the lattice and βr ≡ ρ. In this initial paper we have equalized the values of the studied population (N=100 or 10000 vs lattice size 10x10 or 100x100, respectively) as well as the crossing point of the plotted values of S and I, which is 49, by tuning the value of βr (see Figures 9 and 10). This empirical approach reveals the capability of our model to meet the desired values. For a more refined modelling some attention should be paid to the slopes, by means of a balanced choice of (lattice size, rate, generation).

5. CONCLUSION

We have presented a new approach to the SI deterministic compartmental model. Our proposal is based on both connectivity between people and elementary binary rules that quantify the contacts between people. Our model fits the results carried out by traditional simulation of ODE. This encouraging empirical result must be improved in the future by means of a deeper analysis of the connectivity including a probabilistic approach of it. The SIR, SIS, SIRS models will be also studied. Finally, experimental data coming from statistics on real cases must be directly confronted in order to validate our model. Further we envisage to adapt this model to the field of virus expansion in computers.

REFERENCES

[1]   Siettos, C.I. & Russo, L. Mathematical modeling of infectious disease dynamics Virulence (2013) May 15; 4(4): 295–306.

[2]   Choisy, M. et al. Mathematical Modeling of Infectious Diseases Dynamics. In Encyclopedia of Infectious Diseases: Modern Methodologies, by M.Tibayrenc Copyright © 2007 John Wiley & Sons,Inc.

[3]   Bernoulli, D. (Reprinted in Blower, S. 2004). An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. Reviews in medical virology, 14 (5): 275–88.

[4]   Hamer, W.H. Epidemic Disease in England: The Evidence of Variability and of Persistency of Type. Bedford Press, 1906.

[5]   Ross, R. The Prevention of Malaria. Murray, London, 1911.

[6]   Kermack, W.O & McKendrick, A.G. A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. 115 (1927), 700-721.

[7]   Isea, R & Lonngren, K.E. On the Mathematical Interpretation of Epidemics by Kermack and McKendrick. Gen. Math. Notes, Vol. 19, No. 2, December 2013, pp. 83-87.

[8]   Bailey NJT. The Mathematical Theory of Infectious Diseases and its Application. Griffin, London, 1957.

[9]   Culshaw, R.V. Mathematical Modeling of AIDS Progression: Limitations, Expectations, and Future Directions. Ph.D. Journal of American Physicians and Surgeons, Vol 11, Nº4, Winter 2006.

[10]   Marshall, Brandon D. L. et al. A Complex Systems Approach to Evaluate HIV Prevention in Metropolitan Areas: Preliminary Implications for Combination Intervention Strategies. PLOS ONE, Vol.7, (9), 2012.

[11]   Naresh, R. et al. Modelling and analysis of the spread of AIDS epidemic with immigration of HIV infectives. Mathematical and Computer Modelling, Vol. 49, Issues 5–6, March 2009, pp. 880–892.

[12]   Cassels, S. & and Goodreau, S.M. Interaction of mathematical modeling and social and behavioral HIV/AIDS research. Curr Opin HIV AIDS. 2011 Mar; 6(2): 119–123.

[13]   Van den Driessche, P. Detreministic Compartmental Models: Extension of Basic Models. Lecture Notes in Mathematics, Mathematical Epidemiology, Springer, 2008.

[14]   Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions". The Journal of Physical Chemistry, 81 (25), pp. 2340–2361.

[15]   Ricard, M.R. et al. Epidemiological model with fast dispersion. Proceedings of BIOMAT 2005, R. P. Mondaini & R. Dilao Ed.

[16]   Ahmed A & M. Hassan. A New Solution of SIR Model by using the Differential Fractional Transformation method. International Journal of Engineering and Applied Science, April, 2014. Vol. 4, No. 11

[17]   Legrand, J. et al. Understanding the dynamics of Ebola epidemics. Epid. Infect. May 2007; 135(4) 610–621.

[18]   http://www.math.washington.edu/~morrow/mcm/mcm15/38725paper.pdf

[19]   Iseal, R. & Lonngren, K.E. A Preliminary Mathematical Model for the Dynamic Transmission of Dengue, Chikungunya and Zika. American Journal of Modern Physics and Application 2016; 3(2): 11-15

[20]   Signes Pont, M.T. et al. A Computational Approach of the French Flag Model to Connect Growth and Specification in Developmental Biology, Cognitive Computation, September 2016.

[21] Signes Pont, M.T. et al. Computational primitive to model the emergence of behavioral patterns. Proceedings of the 3rd WSEAS International Conference on Automatic control, Soft Computing and Human-machine Interaction in Recent Advances in Electrical Engineering Series 49; 2015, pp.80-104.

[22] Signes Pont, M.T. et al. An approach to stigmergy issues based on the recursive applications of binary elementary operations. Proceedings of World Conference on Information Systems and Technologies. In Advances in Intelligent Systems and Computing, Vol. 275, 2014, pp.191-197.

[23] Signes Pont, M.T. et al. An approach to computations in living tissues based on logic functions. Proceedings of 20th Euromicro Conference on Parallel, Distributed and Network-Based Processing, 2012.

AUTHORS

Maria Teresa Signes Pont received the BS degree in Computer Science at the Institut National des Sciences Appliquées, Toulouse (France) and the BS degree in Physics at the Universidad Nacional de Educación a Distancia (Spain). She received the PhD in Computer Science at the University of Alicante in 2005. Since 1996, she is a member of the Computer Science and Technology department of the same university where she is currently an associate professor and researcher. Her areas of research at the Specialized Processors Architecture Laboratory include computer arithmetic and the design of floating point units, approximation algorithms related to VLSI design and natural systems modelling.

Higinio Mora Mora received the BS degree in computer science engineering and the BS degree in business studies in University of Alicante, Spain, in 1996 and 1997, respectively. He received the PhD degree in computer science from the University of Alicante in 2003. Since 2002, he is a member of the faculty of the Computer Technology and Computation Department at the same university where he is currently an associate professor and researcher of Specialized Processors Architecture Laboratory. His areas of research interest include computer modelling, computer architectures, high performance computing, embedded systems, internet of things and cloud computing paradigm.

Antonio Cortes Castillo is a computer engineer trained in the Latin American University of Science and Technology (ULACIT), Costa Rica, 1995. He obtained his bachelor's degree in computer engineering with an emphasis in Management Information Systems at the National University of Heredia, Costa Rica, 2002. Adquiere his Master's degree in Computer Science (Telematics) at the Technological Institute of Costa Rica. He is now teaching at the University of Panama and is with his PhD. at the University of Alicante.

EMOTIONAL LEARNING IN A SIMULATED MODEL OF THE MENTAL APPARATUS

Martin Fittner[1] and Christian Brandstätter[2]

[1]Institute of Computer Technology, Vienna University of Technology, Vienna, Austria fittner@ict.tuwien.ac.at
[2]Research Institute of Molecular Pathology (IMP) Vienna, Austria brandstaetter@imp.ac.at

ABSTRACT

How a human being learns is a wide field and not fully understood until now. This paper should give an alternative attempt to get closer to the answer how human beings learn something and what the relation to emotions is. Therefore, the cognitive architecture of the project "Simulation of Mental Apparatus and Applications (SiMA)" is used to fulfill two tasks. One is to give an answer to the question above and the other one is to enhance the functional model of the mental apparatus with learning. For that reason, the functions of the model are analyzed in detail for their ability to enhance them with a learning ability. The focus of the analysis lay on emotions and their impact on the ability to change memories in the model to determine a different behavior than without learning.

KEYWORDS

Learning, Emotion, Artificial General Intelligence, Simulation of Mental Apparatus and Applications (SiMA), Artificial Recognition System (ARS), Cognitive Architectures, Cognitive Automation, Psychoanalytically Inspired AI, Software Agents

1. INTRODUCTION

The target of the project Simulation of Mental Apparatus and Applications (SiMA) is the development of a holistic functional model of the human mind. It was founded by Dietmar Dietrich in 2000 [1] as a new attempt to overcome the lack of solutions for regulation of complex tasks in buildings that can't be solved with traditional algorithms. A bionic approach was used to develop a model of the human mind. As the best available holistic functional model [2] of the human mind the Metapsychology from Sigmund Freud [3] was taken as a framework for the model. This has been evaluated by [4] and [5] that stated that psychoanalysis fulfills the needed criteria of a unified holistic model of the mental apparatus. Like many other projects in cognitive science, the project moved to a multi-agent simulation to prove the functionality of the model. A methodology was developed [6] to transfer models from psychoanalysis and neurology into a holistic functional model. After 16 years, the model was enhanced and refined and remarkable results were reached [6]. But until now the functionality of learning was still missing. This paper should give the first outlook how learning can be integrated into the model with the focus on emotions.

2. BACKGROUND AND RELATED WORKS

To solve specific technical problems, a lot of learning algorithms have to be developed. They belong to the group of machine learning (reinforced learning, deep learning, neural networks, etc.) and showed remarkable results. But their strength is the solving of particular problems. The more they are used away from their application, the more they make failure in their calculations. For SiMA more general learning functions are needed to solve the problem of learning in a wider range. Therefore, machine learning can only be used in SiMA for particular data handling problems but not as a top-level learning model.

But there are more projects that deal with learning in the area of artificial general intelligence (AGI). These projects have the same target as SiMA to build a general-problem-solver (GPS). As a delegate of these projects, LIDA (Learning Intelligent Distribution Agent) should be discussed regarding their similarity to SiMA [7] and its functions for learning and emotions.

LIDA enhances its predecessor IDA (Intelligent Distribution Agent) by the ability to learn something. It combines several theories of the mind in a cognitive architecture [8]. In LIDA unconscious and conscious data processing is done due to the global workspace theory (GWT) that is used there. Learning is more conceptual but some parts are implemented like perceptual learning and episodic learning. In LIDA new information can only be stored (learned) when they first become conscious. Emotions and feelings are used as facilitators of learning [8] in the model.

3. THE SIMA PROJECT

The SiMA project uses the metapsychology as a framework for building a functional model of the human mind. This was done after an intensive search for a holistic unitary model of the human mental apparatus. To ensure a unified model different experts from other scientific disciplines are involved in the building of the model. Computer scientists are using such layer model now since the beginning of the first computers. The split of a system in its physical representation including dependencies and a pure data processing part without physics is a specialty of computer scientists. This was described together with the model in detail in [9]. So, with such a layer model, we can connect the data processing layers with the physical world. And as the physical world can be changed, as long as the interfaces to the new physics are adapted. So, it must be possible to run a human mental apparatus on a computer, as long as the hardware can provide enough power for the data processing.

The model is divided into 3 layers. Layer 1 (L1 in Figure 1) is the neural layer which handles signals from sensors and actors and represents the physical world of the model. The next layer is the neurosymbolic layer (L2 in Figure 1) where signals are transformed to neurosymbols as described by [10]. The neural and the neurosymbolic layer are not in the focus until now and realized as simple pass-thru functions. The third layer (L3 in Figure 1) was the starting point of SiMA. In this layer the human mental apparatus was located and divided into unconscious and conscious or preconscious data processing. The unconscious part was named as primary process and the conscious or preconscious part as secondary process. This separation has the advantage that a huge amount of data can be processed with fast automatic functions and only a small number of data is passed to the conscious processing part. This ensure that the system still can make decisions in nearly every situation. In Figure 1 the track view of SiMA is shown which is an abstract view of the model.
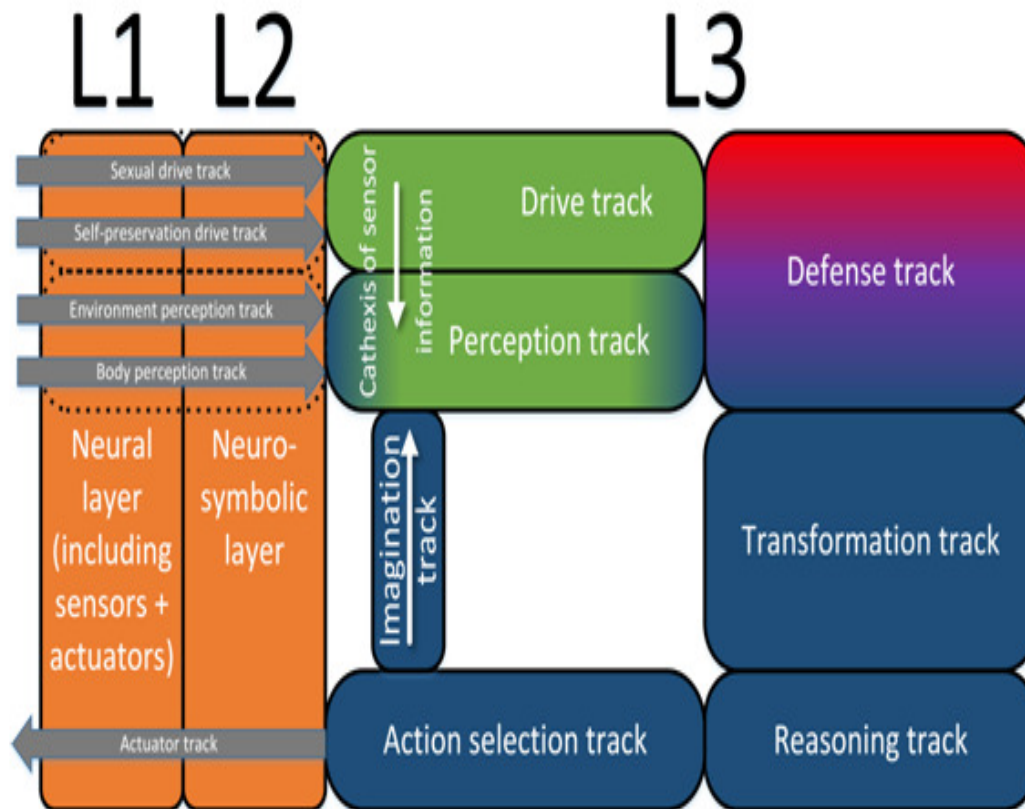
Figure 1. Functional model of SiMA in an abstract view [9]

4. AN EMOTIONAL LEARNING MODEL

To create a model for learning the following parts are needed:

- Memory

- Trigger for learning

- Change or creation of new data structure parts

How these parts are realized in SiMA and how they should be adapted to achieve a learning functionality are described in the following.

4.1. The Memory Model

The memory is located in the 1st layer of the SiMA model [9]. It corresponds to the multi storage model of Atkinson und Shiffrin [11]. A new interface is needed to access the memory from the functions. This can be seen as a separate data storage layer model to the data processing layer model of the functions as shown in Figure 2.
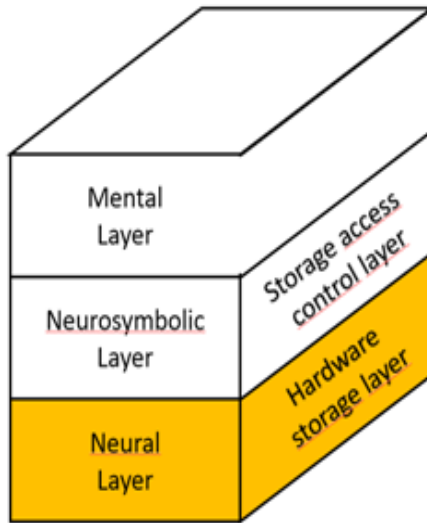
Figure 2. Memory model in SiMA   Figure 3. Psychic intensity and its components

The memory is further divided into short-term memory and long-term memory. An ultra-short-term memory can't be used due to the missing implementation of the sensor functions, but in the model the values, that are handed from the simulator to the mind, are static for one processing cycle which goes a little bit in the direction of a sensory memory. For the hardware storage layer an ontology is used as a database and without production rules. A detailed description of the memory model can be found in [9].

4.2 Emotions

Emotions are one form of valuation in the SiMA model. All valuation terms are summarized as psychic intensity. As stated in [9] "In the SiMA project, psychic intensity is used as an umbrella term for all valuation quotas…Valuation ultimately serves to prioritize actions in order to mediate between the demands of the outside world (the own body and the environment) and the inner (psychic) needs (e.g. satisfying psychic and physiological needs within the environment, or adapting (psychic) wishes to the external circumstances)". In Figure 3 all valuations are pictured. Let us start with valuations of the secondary process. There we have feelings and neutralized intensity, where feelings are built up on emotions (basic and extended emotions) and neutr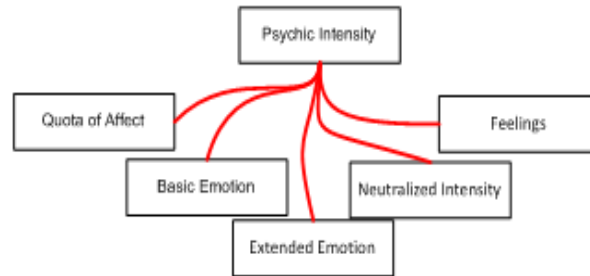alized intensity is created out of the quota of affect. So, the fundamentals of secondary process valuations reside in the primary process. There we have the quota of affect which is a valuation of the tension of sexual and self-preservation drives. The quota of affect was already equipped with a learning functionality. The outcome of the simulation shows that with the adaption of the quota of affect, that is connected with objects and actions, these objects or actions are more or less valuated and on the basis of these valuations the agents in the simulation decided how to interact with these objects. So, the same agent with the same personality in the same environmental setup chooses e. a. to eat an apple instead of a carrot just depending on the learned values of the quota of affect. As the valuation of the quota of affect is a part of the primary process it is unconscious. This matches perfectly with the human decision making when choosing something to eat. There is often no conscious reason why we sometimes choose this and sometimes that but there is an undeniable tendency that drags us to choose one of them.
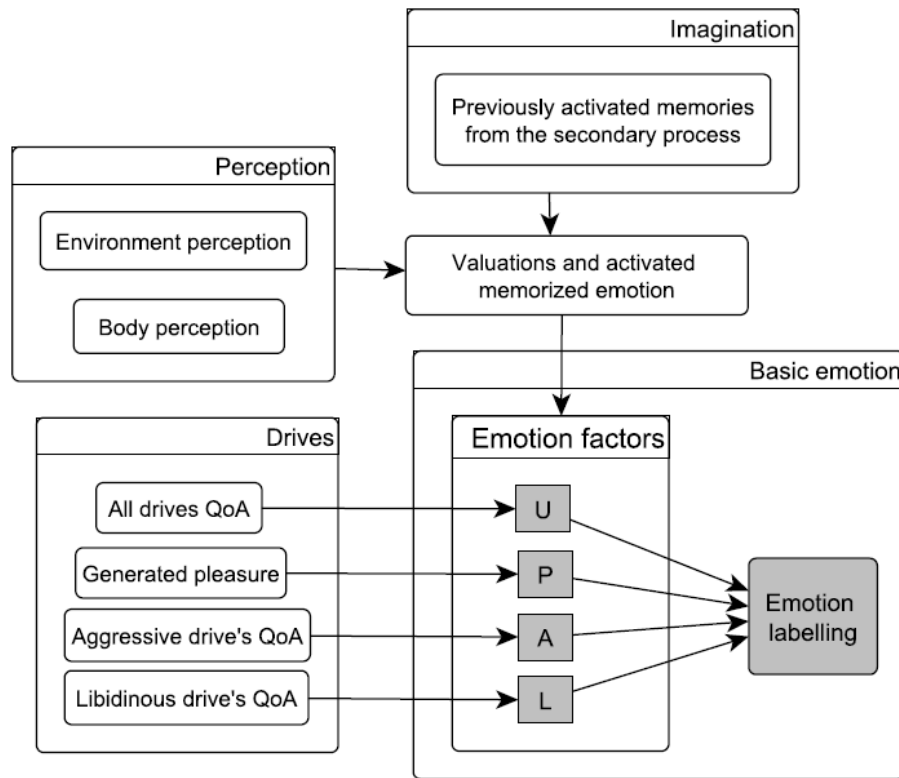
QoA ... Quota of Affect, U ... Unpleasure, P ... Pleasure,
A ... Sum of aggressive drives' QoA , L ... Sum of libidinous drives' QoA

Figure 4. Creation of basic emotion in SiMA [6]

Emotions were separated into basic and extended emotions. Basic emotions are identical in every mammal and are formed out of four factors which can be seen in Figure 4. These factors evolve out of changes of the quota of affect of all drives. Basic emotions are further extended by conflicts of the defense mechanism. In this paper, basic emotions should be in the focus and therefore extended with a learning functionality to create new or change emotionally weighted memories of images. Images are snapshots of a current moment which combine all perceived information together.

4.3 Exemplary Case

The exemplary case [12] is used in the SiMA project to validate the correct function of new implementations. It also supports discussion of certain functionality of the model in an interdisciplinary collaboration. In this paper, the exemplary case should show how the change of experiences effects the agent's behavior.

The starting point of the exemplary case are two agents (Adam and Bodo) and two food sources (an apple and a carrot). At first, Adam perceived the situation and decided according to his hunger to go to the food source that attracts him more (the apple). His valuation of Bodo's bodily expressions doesn't lead to a change in his plan to eat the apple. As he reached the food source, he starts to eat. In the meantime, Bodo watched Adam and due to his predefined anger, he decided to beat Adam. When Bodo beats Adam, Adam perceives unpleasure and flees from Bodo.

5. SIMULATION

For the simulation, the above described exemplary case is used and simulated in a multi-agent environment based on MASON [13]. There are two simulation runs with different emotional valuations. The valuation is adapted manually until now but should be done by the agent in the near future. After the first run (Figure 5) the perceived unpleasure of Adam is captured with inspectors (numerical or graphical descriptions of certain internal states and decisions of the agent during the simulation) and used to adapt the emotional valuation of the bodily expressions of Bodo.
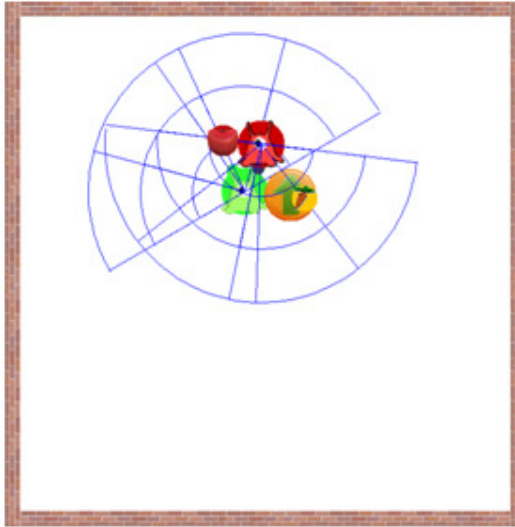
Figure 5. Learning phase   Figure 6. Behavior after Learning in 2nd run
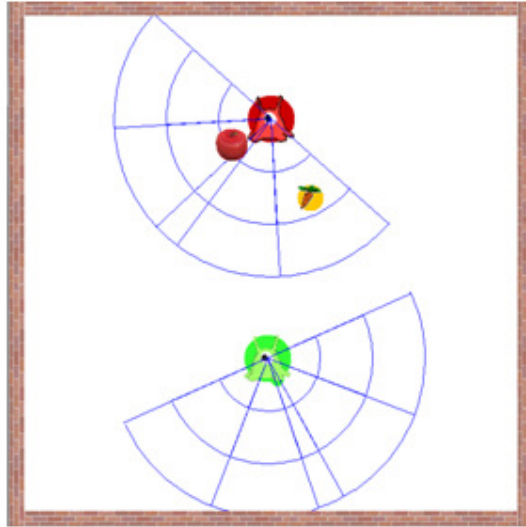
After this adaption, the simulation is started again with the same parameters except the adapted emotional valuation. Now Adams emotions don't match with the plan to eat the apple and he decides to execute the second-best plan and flees (Figure 6).

6. CONCLUSIONS AND OUTLOOK

In contrast to the results of the learning with quota of affect valuations the valuations of emotions are unconscious, but with the transformation to feelings they become conscious. So the agent is also able to reason about the learned feelings that are a result of the changed emotions.

This paper shows a first attempt for the implementation of a learning functionality in regard to emotions. Now the change of emotional valuations must be implemented to produce the expected changes in memories. Then more complex exemplary cases can be created to show that the model is applicable for detecting dangerous situations. Therefore, the system could be used to valuate certain bodily expressions and actions of human beings. Of course, then we have to include a capable perception unit and the until now missing self-categorization.

REFERENCES

[1] Dietrich, Dietmar (2000) "Evolution potentials for Feldbus systems. In: Factory Communication Systems", Proceedings, 2000 IEEE International Workshop on Bd. 1, 2000, 145-146. - Invited Talk

[2] Kandel, E. R. (1999) "Biology and the future of psychoanalysis: A new intellectual framework for psychiatry revisited" American Journal of Psychiatry, 505-524.

[3] Freud, S. (1915) "The Unconscious, volume XIV (1914-1916) of On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works. Vintag.

[4] Jakubec, M., Doblhammer, K., Fittner, M., Wendt, A.: Logical Thought Based on Word Presentations, in proceedings of EAPCogSci 2015, EuroAsianPacific Joint Conference on Cognitive Science, Torino, Italy, September 25-27, pp. 95-100, 2015.

[5] Dietmar Dietrich, D., Fodor, G., Zucker, G., Bruckner, D. (2009) "Simulating the Mind - A Technical Neuropsychoanalytical Approach" (2st ed.), Vienna: Springer.

[6] Schaat, S. (2016) "Simulation of Foundational Human Information-Processing in Social Context", Unpublished undergraduate dissertation, Vienna University of Technology.

[7] Wendt, A., Gelbard, F., Fittner, M., Schaat, S., Jakubec, M.: Decision-Making in the Cognitive Architecture SiMA. To be published at the 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI 2015), Taiwan, 2015.

[8] Franklin, S., Madl, T., D'Mello, S. Snaider, J., (2014) LIDA: A Systems-level Architecture for Cognition, Emotion, and Learning. IEEE Transactions on Autonomous Mental Development 6 (1), (pp. 19 – 41), IEEE.

[9] Dietrich, D., Brandstätter, C., Doblhammer, K., Fittner, M., Fodor, G, Gelbard, F., Huber, M., Jakubec, M., Kollmann, S., Kowarik, D., Schaat, S., Wendt, A., Widholm, R. (2015). Natural Scientific, Psychoanalytical Model of the Psyche for Simulation and Emulation; Scientific Report III, http://publik.tuwien.ac.at/files/PubDat_240983.pdf, Vienna University of Technology, Institute of Computer Technology.

[10] Velik, R. (2008). A Bionic Model for Human-like Machine Perception. Unpublished undergraduate dissertation, Vienna University of Technology, Institute of Computer Technology, Vienna.

[11] Atkinson, R. C., Shiffrin, R. M. (1968). Chapter: Human memory: A proposed system and its control processes. In Spence, K. W., & Spence, J. T. The psychology of learning and motivation (Volume 2). New York: Academic Press. pp. 89–195.

[12] Schaat, S., Kollmann, S., Zhukova, O., Dietrich, D., Doblhammer, K. (2015) Examination of foundational AGI-agents in Artificial-Life simulations. Conference on Technologies and Applications of Artificial Intelligence (TAAI), Tainan, 2015, pp. 344-351.

[13] Luke, S., Cio-Revilla, C., Panait, L., Sullivan, K. (2004). Mason: A new multi-agent simulation toolkit. In Proceedings of the 2004 Swarmfest Workshop. Retrieved August 21, 2016, from http://cs.gmu.edu/~eclab/projects/mason/publications/SwarmFest04.pdf.

AUTHORS

Martin Fittner was born 1984 in Mistelbach (Austria), studied Embedded Systems at the University of Applied Science Technikum Wien in Vienna and finished with the degree Master of Science. He works as a senior developer for software and system design for automotive electronic control units at the Robert Bosch AG. His main topics are testing software for factory testing, customer bootloader programming and security protection of the system. He is a member of the SiMA (Simulation of the Mental Apparatus and Applications) team. Currently, he works on his dissertation at the Vienna University of Technology in the field of the SiMA project.

Christian Brandstätter was born 1979 in Tulln (Austria), studied Informatics and Informatics-Management at the Vienna University of Technology and finished both studies with the degree Master of Science. He works as System Administrator for general IT in a pharmaceutical research institute. His responsibility there is the IT infrastructure (network, storage, and security). Currently, he works on his dissertation and is a member of the Institute of Computer Technology at the Vienna University of Technology. Brandstätter, M.Sc. is a member of IEEE.

An Open Shop Approach in Approximating Optimal Data Transmission Duration in WDM Networks

Timotheos Aslanidis[1] and Stavros Birmpilis[2]

[1]National Technical University of Athens, Athens, Greece
`taslan.gr@gmail.com`
[2]National Technical University of Athens, Athens, Greece
`s.birmpilis@gmail.com`

ABSTRACT

In the past decade Optical WDM Networks (Wavelength Division Multiplexing) are being used quite often and especially as far as broadband applications are concerned. Message packets transmitted through such networks can be interrupted using time slots in order to maximize network usage and minimize the time required for all messages to reach their destination. However, preempting a packet will result in time cost. The problem of scheduling message packets through such a network is referred to as PBS and is known to be NP-Hard. In this paper we have reduced PBS to Open Shop Scheduling and designed variations of polynomially solvable instances of Open Shop to approximate PBS. We have combined these variations and called the induced algorithm HSA (Hybridic Scheduling Algorithm). We ran experiments to establish the efficiency of HSA and found that in all datasets used it produces schedules very close to the optimal. To further establish HSA's efficiency we ran tests to compare it to SGA, another algorithm which when tested in the past has yielded excellent results.

KEYWORDS

WDM networks, packet scheduling, preemption, approximation

1. PROBLEM DESCRIPTION-WDM NETWORKS

In the past decade Optical WDM Networks (Wavelength Division Multiplexing) are being used quite often and especially as far as broadband applications are concerned. They provide high quality services with a large bandwidth and are therefore ideal in providing multimedia, telematic, fast internet browsing and many more communication services. In WDM network an optic fiber may be split in a number of channels depending on the frequency. Each of these channels corresponds to a specific wavelength. WDM Networks are used vastly by telecommunication companies because they increase capacity without the need to install new lines. Message packets can be interrupted using time slots in order to maximize network usage and minimize the time required for all messages to reach their destination. However, pre-empting a packet will result in time cost. Time to setup for the next packet can often be significant. The problem of scheduling message packets through such a network is referred to as PBS (Preemptive Bipartite Scheduling).

Even though numerous algorithms have been designed in an effort to produce efficient schedules there seems to still exist room for further research.

2. THE GRAPH MODEL FOR PBS

As there are 2 sets between which the multiplexing and demultiplexing process takes place the ideal representation seems to be a bipartite graph G(U,V,E,w). Source stations will be assigned to U, destination stations to V, messages to be transmitted will be the edges connecting nodes of U to nodes of V. w: E→Q_+ will be a weight function giving each edge e=(u,v) a weight equal to the duration of the transmission for u to v. Given a matching M in G we will denote by w(M) the maximum weight of any edge e∈M, that is w(M)=max{w(e), e∈M}. Following the notation used in previous research on the problem, Δ will denote the degree of G, W the maximum sum of edge weights incident to any of the nodes and d the setup cost to prepare for the next packet transmission. Thus, a feasible schedule for PBS would cost $\sum_{i=1}^{N} w(M_i) + d \cdot N$, where N is the number of times the network has to reconfigure so that all data will be transferred.

Using these notations, the value L=W+d·Δ represents a lower bound. L is not always achievable but is easy to calculate and is considered to be a good approximation of the optimal solution when designing near optimal algorithms for PBS.

3. PAST RESEARCH ON PBS

The NP-Hardness of PBS derives from the fact that it is a bicriteria minimization problem, namely the objective function to be minimized depends on two different criteria each of which affects the other. Regardless the hardness of minimizing both criteria simultaneously, minimization of each criterion separately is relatively easy. Algorithms proposed by the authors of [10] and [8] minimize the number of preemptions while the one in [15] minimizes the transmission time. In general, the problem is 4/3-ε inapproximable for all ε>0 as shown in [6]. The best guaranteed approximation ratio of any algorithm proposed for the problem is $2 - \dfrac{1}{d+1}$. Proof of that can be found in [1]. Many other algorithms have been proposed in order to provide solutions close to the optimal. Experimenting on test cases has yielded good results in [2], [3] and [7].

In this manuscript we try to exploit in the best way possible a reduction of PBS to the open shop scheduling problem (Om| |C_{max}), in order to use polynomial time algorithms proposed for some special instances of it, to minimize each criterion separately and combine the results to design a hybrid algorithm (HSA-Hybridic Scheduling Algorithm), which will tackle the bicriteria problem efficiently.

4. REDUCING PBS TO OPEN SHOP AND DESIGNING HAS

Theorem 1: Any instance of PBS can be transformed to an instance of Open Shop and vice versa.

Proof: Let G(U, V, E, w) be the graph corresponding a PBS instance. We transform this graph to an open shop instance in the following way: U={u_1, u_2,... u_n} will be the set of processors P={p_1, p_2, ..., p_n}. V={v_1, v_2, ..., v_m} will be the set of Jobs J={J_1, J_2, ..., J_m} and E={(u_i,v_k) | u_i∈ U, v_i∈ V} will be the set of operations O={O_{ik} | i=1, 2, …, n and k=1, 2, …, m}. O_{ik} is the operation of J_k to

be processed on machine i. The processing time of each operation will be calculated by the function p: O→Q^+, where $p(O_{i\,k}) = \begin{cases} w(u_i, v_j) & ,\text{if} \quad (u_i, v_j) \in E \\ 0, & \text{otherwise} \end{cases}$.

The inverse transformation is straight forward.

Unfortunately the above reduction does not imply of a way to solve PBS using open shop algorithms as a PBS schedule would preempt all transmission simultaneously, while open shop scheduling does not have such a requirement. Yet, there exist two special instances of the Open Shop problem that are known to be solvable in polynomial time and are exactly right for our purposes. Om|prpm|C_{max} in which preemption is allowed and Om|p_{ij}=0,1|C_{max} in which all processing times are either 0 or 1.

The polynomial time algorithm described in [15] minimizes a preemptive open shop makespan by preempting all processor tasks simultaneously. We will refer to this algorithm by LLA (Lawler-Labetoulle Algorithm). LLA uses linear programming techniques to define a set of tasks in order to reduce the workload of all stations that, in each step of the process are assigned with the maximum workload W. The authors of [15] call this a decrementing set. The number of preemptions is O(m^2+n^2). In order to improve the results of LLA instead of using a random decrementing set to reduce the workload of the stations we use one produced by a maximum weighted perfect matching algorithm. We will call this variation of LLA, POSA (Preemptive Open Shop Algorithm). We will use POSA to minimize HSA's makespan.

To complete HSA we also need an algorithm which will minimize the number of preemptions. A linear programming algorithm for Om|p_{ij}=0,1|C_{max} is described in [8]. Yet, in order to better fit the requirements of our WDM network transmission we used the following Open Shop algorithm:

OS01PT Algorithm (Open Shop 0, 1 Processing Times)

Step1: Add the minimum number of nodes needed to G(U, V, E) so that |U|=|V|. Call the induced graph G´

Step2: Add edges to G´ to make it degree-regular.

Step3: Assign weights to the edges of G´ in the following way: Edges of the initial graph will weigh 1 while newly added edges in step2 will weigh 0.

Step4: calculate a perfect matching M in G´.

Step5: remove all edges of M from G´.

Step6: Repeat step4 and step5 until G´=∅.

To make the graph degree regular we use the subroutine described in [10].

Theorem: OS01PT will produce a schedule for PBS with exactly Δ transmissions.

Proof: By induction on the value of Δ.

For Δ=1: Since G´ is degree regular, all nodes have exactly one adjacent edge. These edges form a perfect matching for G´ and the transmission will conclude in one step.

Let the theorem stand for any regular graph with Δ=n-1.

Suppose that Δ=n. A perfect matching in G´ will reduce the degree of all nodes by one, thus making G´´s degree n-1. From the inductive hypothesis an n-1 degree graph will need n-1 transmissions to schedule its data. Therefore, to transmit all data 1+(n-1)=n transmissions will be needed. Proof that a perfect matching can always be found in a graph with Δ=n>1 can also be found in [10].

We now have all the necessary tools to design HSA:

HSA (Hybridic Scheduling Algorithm)

Step1: Let S1 be the feasible schedule produced for PBS using POSA. Let C1 be the cost of S1.

Step2: Let S2 be the feasible schedule produced for PBS using OS01PT. Let C2 be the cost of S2.

Step3: If C2<C1 then transmit as in S2 else transmit as in S1.

5. DECIDING A CRITICAL VALUE OF D FOR HSA

Five hundred test cases following a uniform distribution have been ran for a 30 source-30 destination system for values of setup cost varying from 0 to 100 and message durations varying from 0 to 120. We have to point out that since PBS is an NP-Hard problem, calculating an optimal schedule is inefficient therefore to estimate the approximation ratio we have used the lower bound to the optimal solution namely W+Δ·d.

Figure 1 depicts the deviation from the optimal solution when using POSA to calculate a schedule for PBS. Figure 2 depicts the corresponding results when using OS01PT, while Figure3 shows the results yielded by HSA. In order for OS01PT to better suit the requirements of our problem, instead of calculating a perfect matching as described in step4 of the algorithm's description, we use a maximum weighted perfect matching algorithm just as in the case of POSA.
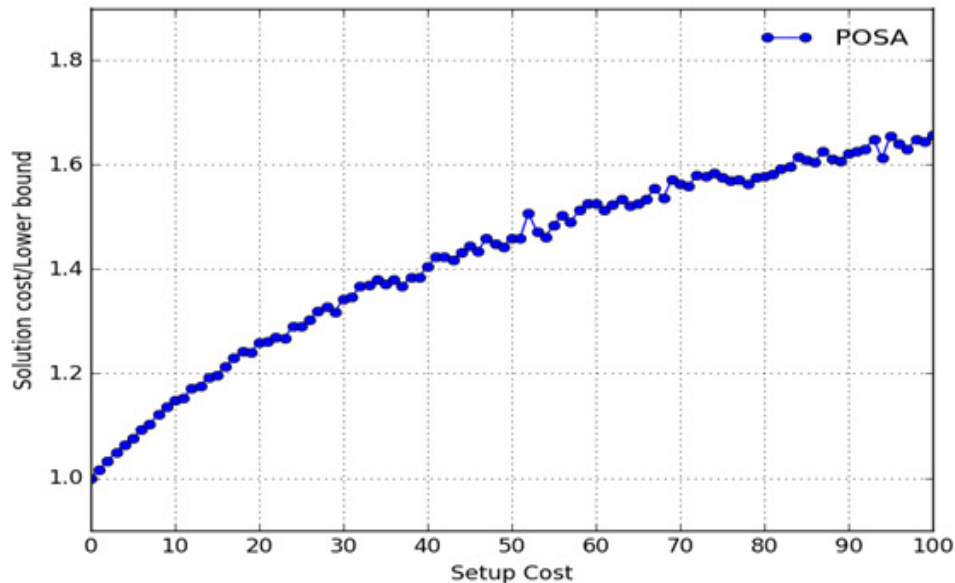
Figure 1. Average Solution cost/lower bound using POSA
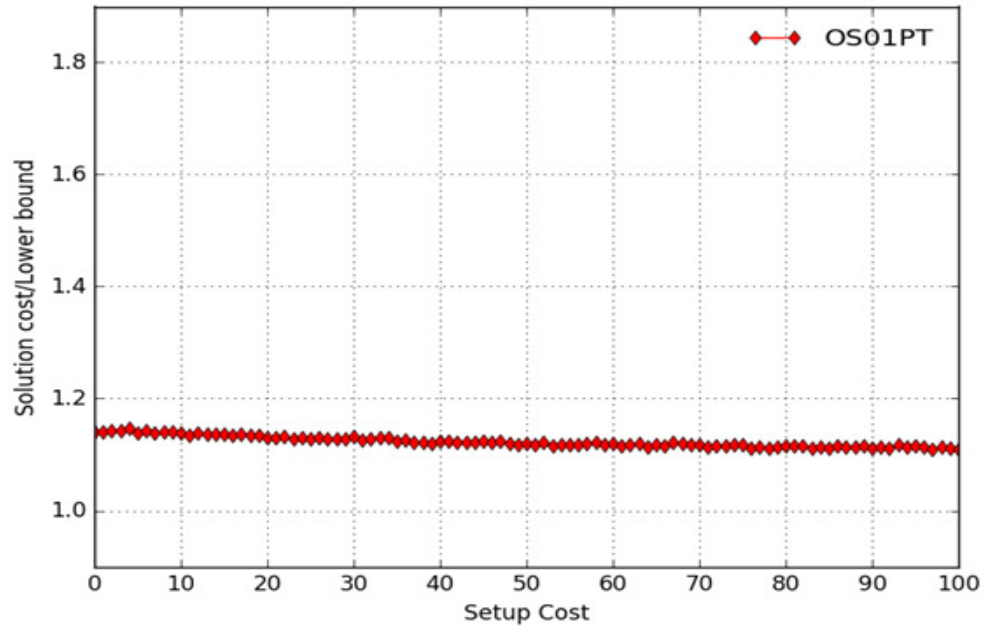
Figure 2. Average Solution cost/lower bound using OS01PT
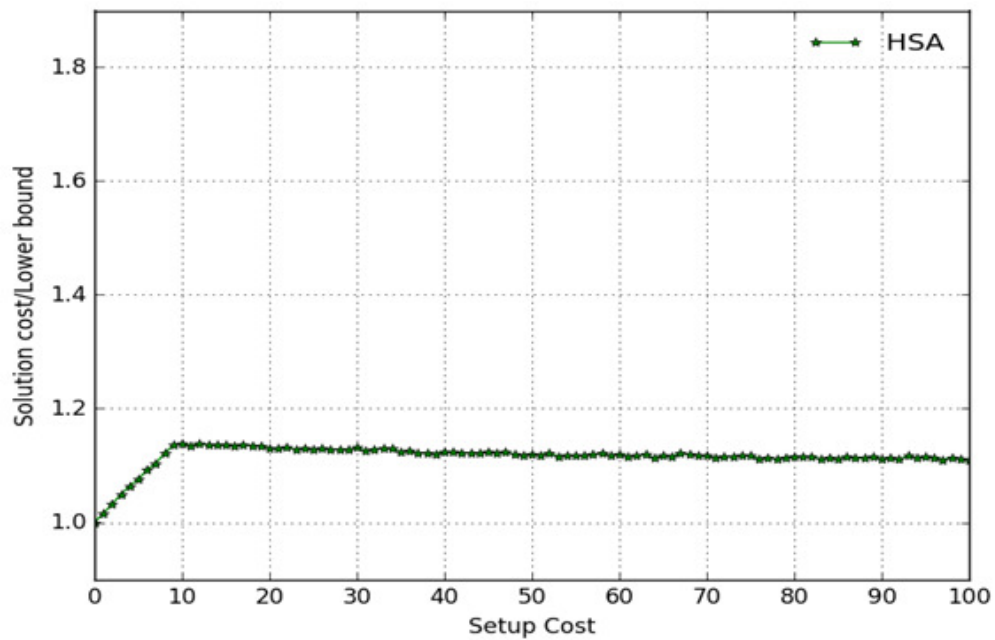
Figure 3. Average Solution cost/lower bound using HSA

According to figures1 and 2 the appropriate value of d to switch from POSA to OS01PT is d=9.

Figure 4 shows the (worst solution cost)/(lower bound) ratio of HSA for any of the instances used for each value of d. Note that it never exceeds 1.3.
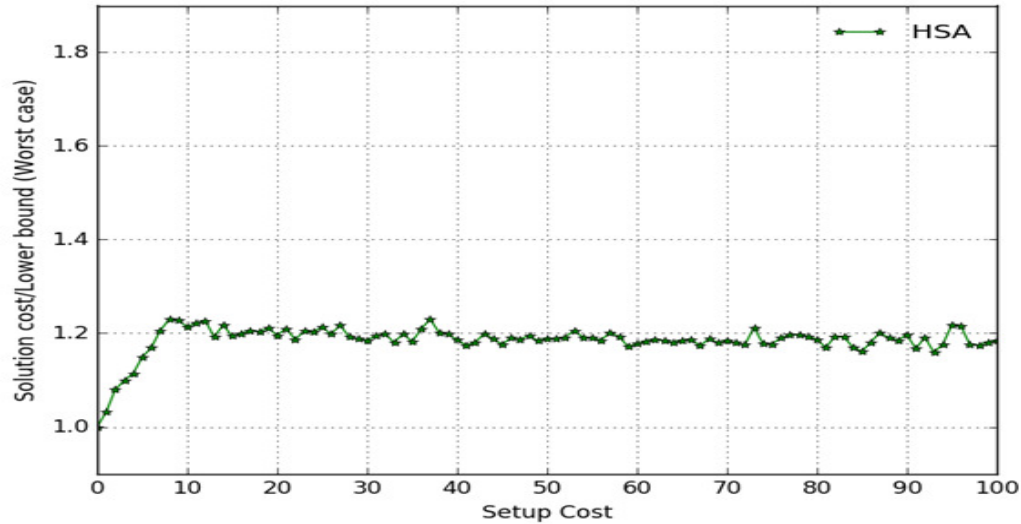
Figure 4. Worst solution cost/lower bound using HSA.

6. COMPARING HSA TO ANOTHER EFFICIENT ALGORITHM FOR PBS

One of the most efficient algorithms designed by researchers for PBS in the past is the SGA (Split Graph Algorithm). SGA splits the initial graph in two subgraphs, one with messages of duration less than d and another one with messages of duration at least d. The larger messages are scheduled first and then the small ones. It was found to be very efficient when tested in comparison to other efficient algorithms and it appears to be one of the top heuristics for PBS. We ran tests to compare HSA with SGA which show that HAS always produces a schedule at least as good as SGA. HSA's approximation ratio is, for some values of d up to 8% better than the one of SGA. As in paragraph 5, we used five hundred test cases following the uniform distribution for a 30 source-30 destination system for values of setup cost varying from 0 to 100 and message durations varying from 0 to 120.
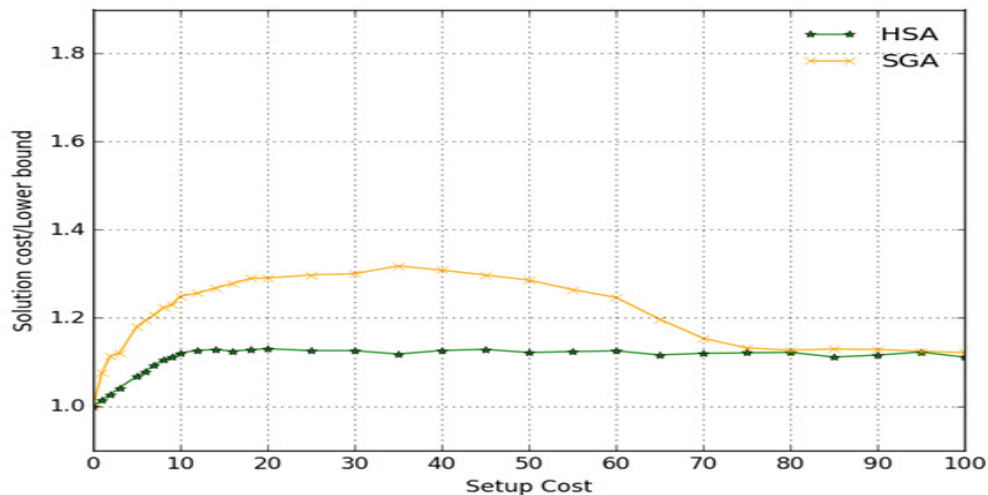
Figure 5: Comparison of HSA with SGA

7. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a reduction of a network transmission problem (PBS) to a scheduling problem (Open Shop). Based on this reduction we have designed a hybrid algorithm (HSA) using suitable variations of polynomial time algorithms for special instances of Open Shop (POSA and OS01PT) designed for the purposes of this paper, in order to design an efficient transmission strategy for WDM network transmissions. We have ran tests to establish the efficiency of our hybrid algorithm and to suggest the appropriate value of network delay to switch from POSA to OS01PT. Knowing this value of d improves the computational complexity of HSA. In these experiments we used data following a uniform distribution. Furthermore we tested HSA against SGA, one of the most efficient algorithms designed for PBS in the past to conclude that HSA's results have in most cases an approximation ratio up to 8% better than SGA's.

Future research could focus on further improvement of the time complexity of HSA. The fact that HSA's approximation ratio even for the worst data tested has always been less than 3/2, suggests that a formal mathematical proof of an approximation ratio lower than 2 might be possible. Furthermore, tests could be ran for data following non uniform distributions such as Gaussian or Exponential. To further improve performance and complexity a new hybrid algorithm could be designed using different approaches on how to minimize each criterion separately. This algorithm might also be independent of the open shop approach. It would aim in minimizing just one of the criteria under the constraint that the other one is minimum.

REFERENCES

[1] F. Afrati, T. Aslanidis, E. Bampis, I. Milis, Scheduling in Switching Networks with Set-up Delays. Journal of Combinatorial Optimization, vol. 9, issue 1, p.49-57, Feb 2005.

[2] T. Aslanidis, M.E. Kogias, Algorithms for Packet Routing in Switching Networks with Reconfiguration Overhead. In Proceedings, Second International Conference on Computer Science and Engineering (CSE-2014), April 2014.

[3] T. Aslanidis, L. Tsepenekas, Message Routing In Wireless and Mobile Networks Using TDMA Technology. International Journal of Wireless & Mobile Networks. Vol. 8, Num. 3, June 2016

[4] G. Bongiovanni, D. Coppersmith and C. K.Wong, An optimal time slot assignment for an SS/TDMA system with variable number of transponders, IEEE Trans. Commun. vol. 29, p. 721-726, 1981.

[5] J. Cohen, E. Jeannot, N. Padoy and F. Wagner, Messages Scheduling for Parallel Data Redistribution between Clusters, IEEE Transactions on Parallel and Distributed Systems, vol. 17, Number 10, p. 1163, 2006.

[6] J. Cohen, E. Jeannot, N. Padoy, Parallel Data Redistribution Over a Backbone, Technical Report RR-4725, INRIA-Lorraine, February 2003.

[7] P. Crescenzi, X. Deng, C. H. Papadimitriou, On approximating a scheduling problem, Journal of Combinatorial Optimization, vol. 5, p. 287-297, 2001.

[8] H. M. Gabow and O. Kariv, Algorithms for edge coloring bipartite graphs an multigraphs, SIAM Journal on Computation, 11, p. 117-129, 1982

[9] I. S. Gopal, G. Bongiovanni, M. A. Bonucelli, D. T. Tang, C. K. Wong, An optimal switching algorithm for multibeam satellite systems with variable bandwidth beams, IEEE Trans. Commun. vol. 30, p. 2475-2481, Nov. 1982.

[10] I. S. Gopal, C. K. Wong Minimizing the number of switchings in an SS/TDMA system IEEE Trans. Commun. vol. 33, p. 497-501, 1985.

[11] T. Inukai, An efficient SS/TDMA time slot assignment algorithm IEEE Trans. Commun. vol 27, p. 1449-1455, Oct. 1979.

[12] E. Jeannot and F. Wagner, Two fast and efficient message scheduling algorithms for data redistribution over a backbone, 18th International Parallel and Distributed Processing Symposium, 2004.

[13] A. Kesselman and K. Kogan, Nonpreemptive Scheduling of Optical Switches, IEEE Transactions in Communications, vol. 55, number 6, p. 1212, 2007.

[14] M.E. Kogias, T. Aslanidis, A comparison of Efficient Algorithms for Scheduling Parallel Data Redistribution, International Journal of Computer Networks & Communications, May 2014, vol. 6, num. 3.

[15] E.L. Lawler and J. Labetoulle, On preemptive Scheduling of Unrelated Parallel Processors by Linear Programming. Journal of the Association of Computing Machinery 25: 612-619, 1978

[16] K.S. Sivalingam, J. Wang, X. Wu and M. Mishra, An internal based scheduling algorithm for optimal WDM star networks. Journal of Photonic Network Communications, vol. 4, No 1, p. 73-87, 2002

[17] B. Towles and W. J. Dally, Guaranteed Scheduling of Switches with Configuration Overhead, in Proc. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM '02. pp. 342-351, June 2002.

AUTHORS

Timotheos Aslanidis was born in Athens, Greece in 1974. He received his Mathematics degree from the University of Athens in 1997 and a master's degree in computer science in 2001. He is currently doing research at the National and Technical University of Athens in the School of Electrical and Computer Engineering. His research interests comprise but are not limited to computer theory, number theory, network algorithms and data mining algorithms.
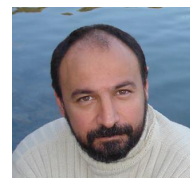
Stavros Birmpilis was born in Athens, Greece in 1994. Currently, he is a senior student in the School of Electrical and Computer Engineering at the National and Technical University of Athens, expecting to receive his diploma by July 2017. His research interests lie in the field of Algorithms and Discrete Mathematics.
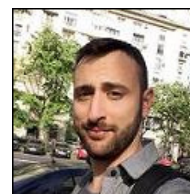
COMPUTER VISION PERFORMANCE AND IMAGE QUALITY METRICS: A RECIPROCAL RELATION

Christopher Haccius and Thorsten Herfet

Telecommunications Lab, Saarland University, Saarbrücken, Germany
{haccius/herfet}@nt.uni-saarland.de

ABSTRACT

Computer vision algorithms are essential components of many systems in operation today. Predicting the robustness of such algorithms for different visual distortions is a task which can be approached with known image quality measures. We evaluate the impact of several image distortions on object segmentation, tracking and detection, and analyze the predictability of this impact given by image statistics, error parameters and image quality metrics. We observe that existing image quality metrics have shortcomings when predicting the visual quality of virtual or augmented reality scenarios. These shortcomings can be overcome by integrating computer vision approaches into image quality metrics. We thus show that image quality metrics can be used to predict the success of computer vision approaches, and computer vision can be employed to enhance the prediction capability of image quality metrics – a reciprocal relation.

KEYWORDS

Computer Vision Performance, Image Quality Assessment, Subjective Quality

1. INTRODUCTION

In today's world computer vision systems have become a central part of modern life. Computer vision in cars reads street signs and markers, in assembly lines checks production and processes, and almost every camera uses computer vision for face detection or artistic effects. In most scenarios Computer Vision is employed to analyze visual information. However, Computer Vision is also increasingly used to generate visual information, for example in augmented reality applications.

For all the different scenarios of computer vision the robustness of the computer vision algorithms is important. As robustness we consider the impact that common types of image errors have on a given computer vision algorithm. Classical image errors stem from image acquisition, and are given by thermal noise or blur. Each computer vision system relying on cameras needs to be robust against such noise, at least to a certain degree. Compression artifacts, like JPEG blocking or JPEG2000 ringing artifacts, become a matter of concern as soon as data for computer vision algorithms is retrieved from space limited storage or after distribution over throughput-limited channels, which make data size and respectively compression critical.

Today we see an increasing amount of visual information that is synthetically generated. For such content novel types of errors occur, which are scene composition errors. Such scene composition

errors occur when synthetic objects are algorithmically merged with captured content, and the synthetic addition is positioned incorrectly, scaled wrongly or not aligned with the captured environment.

This paper analyzes the impact of classical image errors and novel scene composition errors on standard computer vision approaches. For this analysis an image database is necessary, which contains both classical image distortions and scene composition errors, and comes with additional information like ground truth segmentation data, object information or subjective evaluations. Section 2 introduces this database.

We analyze the impact of distortions for three very basic computer vision algorithms, which are object segmentation, object tracking and object detection. In Section 3 we introduce these computer vision algorithms and the experiments we have set up to analyze the distortion impact. Knowing the impact of a distortion on computer vision algorithms for a single image is interesting, yet far more interesting is the ability to predict the impact of distortions. A good prediction can enable system designers to define robustness levels for computer vision systems. In Section 4 we correlate the distortion impact to three image quality metrics, which are subjective opinion scores, scores based on image statistics and scores based on the human visual system.

Finally, in Section 5, we switch the perspective. After having evaluated how image quality metrics can be used to predict computer vision performance, we introduce an approach which employs computer vision algorithms for enhanced image quality assessment. We thus have image quality metrics to predict computer vision performance and computer vision to enhance image quality metrics - a reciprocal relation.

2. THE IMAGE DATABASE

In order to conduct experiments according to the above mentioned motivation an image database is necessary that fulfills several requirements. Most importantly, color images are required which are distorted by classical errors such as noise and compression artifacts. Second, the images need to contain synthetic objects which can be modified to model scene composition errors. Third, ground truth segmentation data needs to be available to conduct object segmentation and tracking experiments. Last, to allow comparisons to subjective assessments, mean opinion scores (MOS) need to be available for the distorted images.

Several image databases exist that fulfill one or few of these requirements. The LIVE and TID2013 database [1], [2] are databases based on real images which are distorted by classical image errors and subjectively evaluated. Both databases lack the ability to modify scene objects, and ground truth segmentation data is missing. The BSDS500 database [3] contains images and segmentation data, yet also lacks the ability to modify objects and has no subjective evaluations for the images. As - to our knowledge - no suitable database exists, we present the Synthetic Image Database SSID, which we have created from fully synthetic scenes with the goal to enable scene composition modifications, image distortions and additional data like depth maps and segmentation data. This database was evaluated by human assessors, and MOS have been calculated for all distorted images [4].

Figure 1a shows a set of scenes contained in the database and presents a depth map (Figure 1b) and a segmentation map (Figure 1c) which can be easily rendered from the synthetic data. In the database Gaussian blur, white noise, JPEG and JPEG2000 coding artifacts as well as object scaling, rotation and translation are implemented as distortions. For the distorted images roughly 20.000 opinions have been obtained from 200 assessors in subjective evaluations, and mean

opinion scores have been calculated. The database is available online for image quality associated research [5].

(a) Scenes contained in the Synthetic Image Database.

(b) Depth Map.   (c) Segmentation Map.
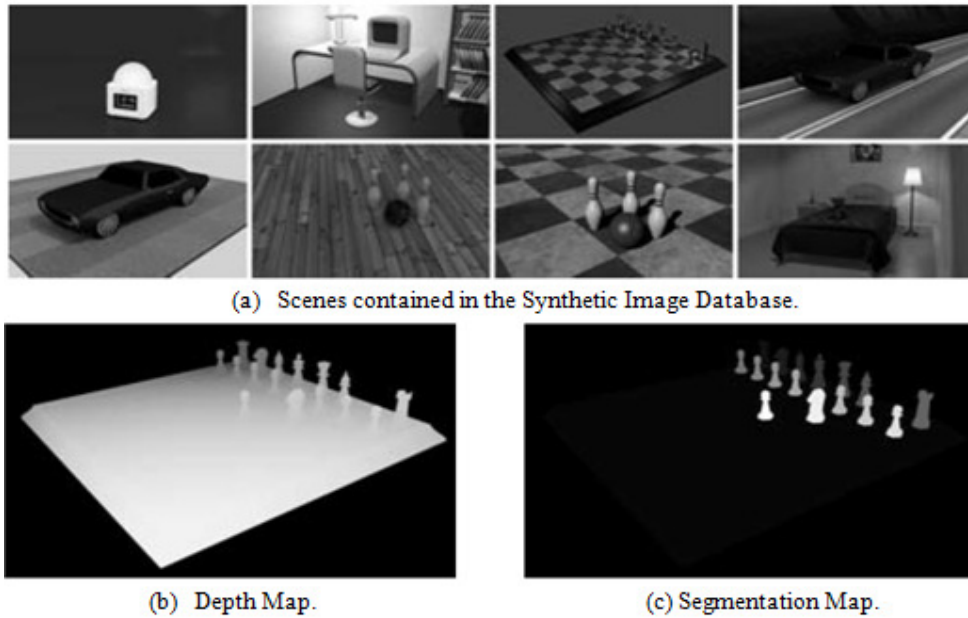
Figure 1.The Synthetic Image Database [4].

3. COMPUTER VISION – ALGORITHMS AND EXPERIMENTS

On the image database presented in Section 2 we test three computer vision algorithms, which are object segmentation, object tracking and object detection. By measuring the success of the computer vision algorithm we can derive the impact that a given image distortion has on a computer vision approach. For these tests we observe a certain object per scene (e.g. the computer, given in Figure 2 which is segmented, tracked and detected in the experiments described in the following sections.

3.1. Object Segmentation

A well established and widely used image segmentation approach is presented by Achanta et al. [6]. SLIC Superpixels are known for their high boundary recall at a low computational complexity, at the cost of oversegmentation. The low computational complexity makes SLIC Superpixel segmentation applicable even for real-time requirements. We evaluate experimentally, how much the boundary recall of SLIC Superpixel segmentation is affected by image distortions. Boundary recall br here is defined as the ratio between correctly recalled boundary pixels in the test image segmentation S and total number of boundary pixels in the ground truth segmentation T:

$$br(S,T) = \frac{bp(S,T)}{bp(T)} \qquad (1)$$

where $bp(S,T)$ are the boundary pixels of ground truth T matching the boundary pixels of test image segmentation S, and $bp(T)$ are the boundary pixels of T only. We compare the boundary recall of each test image segmentation S to the boundary recall of its reference image segmentation R, and record the impact on segmentation $I_{segment}$ as the ratio

$$I_{segment} = \frac{br(S,T)}{br(R,T)} = \frac{bp(S,T)}{bp(R,T)}. \qquad (2)$$

3.2. Object Tracking

Optical flow was introduced by Horn and Schunck in 1981 already [7] and presents a common basis for object tracking between frames. In our experiment we "track" scene objects from a reference image to the distorted test images. Using the ground truth segmentations we can evaluate how many object pixels of the distorted image are tracked correctly from the reference image. For a total number of pixels k and correctly tracked pixels t we calculate the impact on tracking I_{track} as the ratio

$$I_{track} = \frac{t}{k}. \tag{3}$$

Figure 2. Exemplary Training Object for Object Detection.

3.3. Object Detection

A common way of detecting objects in images is to compare image features. The Scale-Invariant Feature Transform (SIFT) was introduced by Lowe in 1999 [8], and in 2006 the faster Speeded-Up Robust Features (SURF) were made public by Hay et al. [9]. SURFs can be learned on a reference object, and then be used to detect the same object in a scene. For our experiment we create renderings of objects outside of their scene and train SURFs on this image. We than compare the matched SURFs between training object and reference image m to the matched SURFs between training object and test image n. The distortion impact on object detection I_{detect} is than

$$I_{detect} = \frac{m}{n}. \tag{4}$$

For all impact measures I it is I = 1 if segmentation, tracking and detection remains as good in the test image as in the reference, and I < 1 if the computer vision results are deteriorated in the test cases compared to the reference.

4. IMAGE QUALITY – METRICS AND RESULTS

Image quality is usually assessed using image statistics or methods modeling the human visual system. To predict the impact of image distortions on computer vision algorithms we calculate two algorithmic image quality metrics and compare to subjective quality scores as well. Mean Opinion Scores (MOS) are already available through subjective tests for the database described in Section 2.

For a reference image R and a test image T with dimensions $x \times y$ an average value expressing the overall statistical image error is the Mean Squared Error (MSE) calculated as

$$MSE(T,R) = \frac{1}{x \cdot y} \sum_{i=0}^{x} \sum_{j=0}^{y} \big(R(i,j) - T(i,j)\big)^2. \tag{5}$$

In the Peak Signal to Noise Ratio ($PSNR$) the MSE is related to the amplitude of the original signal:

$$PSNR(T,R) = \frac{max_{i \in [0,x]}\big(max_{j \in [0,y]}(R(i,j)^2)\big)}{MSE}. \tag{6}$$

Table 1. Table of Spearman Rank Correlations between Distortion Impact and Quality Measures.

	Noise	Compression	Transformation	All
$I_{segment}$ to MOS	0.429	0.557	0.365	0.415
$I_{segment}$ to $PSNR$	0.697	**0.574**	0.655	**0.650**
$I_{segment}$ to $SSIM$	0.649	0.371	**0.680**	0.545
$I_{segment}$ to $PARA$	**0.708**	0.015	0.023	0.171
I_{track} to MOS	0.372	0.636	0.349	0.592
I_{track} to $PSNR$	0.849	**0.693**	0.880	0.746
I_{track} to $SSIM$	**0.850**	0.626	**0.957**	**0.875**
I_{track} to $PARA$	0.649	0.095	0.039	0.056
I_{detect} to MOS	0.450	0.639	0.218	0.408
I_{detect} to $PSNR$	0.141	**0.643**	**0.571**	**0.490**
I_{detect} to $SSIM$	0.032	0.628	0.532	0.403
I_{detect} to $PARA$	**0.494**	0.112	0.085	0.165

The $PSNR$ is still a very common metric for image quality analysis. It can be easily implemented and has a very low computational complexity, which is an important criterion for real-time applications. For image quality assessment $PSNR$ has been shown to relate poorly to subjective image quality findings [10]. Therefore metrics based on the human visual system have been developed, of which the Structural Similarity ($SSIM$) is a widely established one [11]. The Structural Similarity index ($SSIM$) compares three different image components: luminance, contrast and structure. Structural similarity $SSIM$ between a test image T and a reference image R is calculated as the weighted product of luminance l, contrast c and structure s:

$$SSIM(T,R) = l(T,R)^\alpha \cdot c(T,R)^\beta \cdot s(R,T)^\gamma \tag{7}$$

with $0 < \alpha, \beta, \gamma$.

A fourth measure for image quality which we analyze in the context of this work is the parameter which was used to distort the image. As the scene composition errors are assigned with error parameters for 3 dimensions, we map the three parameters to one error parameter $PARA$ by calculating the absolute rotation angle, absolute size deviation and vector sum of transitions.

We than calculate the Spearman rank correlation [12] between the distortion impact on segmentation $I_{segment}$, tracking I_{track} and detection I_{detect} and the quality measures MOS, $PSNR$, $SSIM$ and $PARA$. Table 1 presents the correlation values for different image distortion classes. These distortion classes are noise (including white noise and Gaussian blur), compression artifacts (including JPEG and JPEG2000 compression artifacts), transformation errors (including object rotation, scaling and translation errors) and all (superclass of all previous classes). A bold font indicates the best correlation in each error class for one computer vision approach.

Table 1 indicates that *PSNR* and *SSIM* are good measures to predict the success of computer vision algorithms. Especially object tracking is very well correlated to *SSIM* and *PSNR*. Only for the noise error class the error parameter (if obtainable) is suggested as a success indicator, while *SSIM* has almost no correlation to the impact of noise on object detection. At the same time, when correlating the presented quality metrics *PSNR*, *SSIM* and *PARA* to the subjective *MOS*, it becomes clear that neither those presented metrics (nor any other metrics known to us) present suitable predictors for image distortions resulting from scene composition errors. This observation is confirmed by Caviedes at al. who note that subjective quality is more aesthetically-oriented whereas computer vision may have different quality requirements [13].
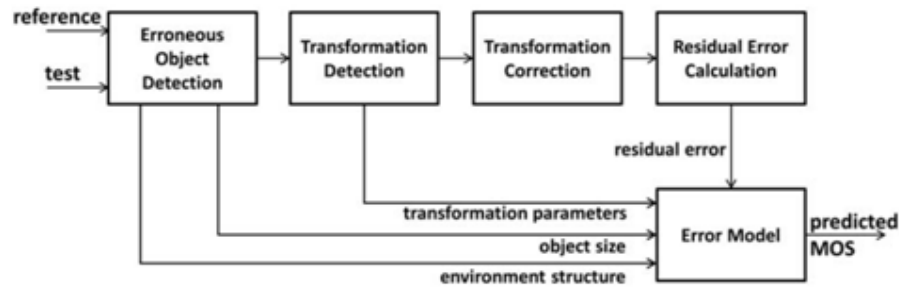
Figure 3. Structure of our proposed SC-VQM.

5. A RECIPROCAL RELATION

In [14] we approach this problem of predicting subjective image quality and develop a Visual Quality Metric for Synthetic Content (SC-VQM) that employs computer vision algorithms to better predict subjective image quality in virtual worlds or augmented reality scenarios. The approach to achieve this goal is straight forward: We detect object changes and correct those before calculating a residual error. This idea is outlined in the block diagram in Figure 3 and described by the following six steps:

1. Erroneous object detection: Distorted objects in a scene composition are detected
2. Erroneous object matching: Objects in test image are matched with objects in reference image
3. Object size calculation: The portion of the image affected by the distorted object is calculated
4. Environment structure analysis: The environment of the distorted objects is analyzed for the amount of structures contained
5. Object correction: The object in the test image is corrected according to the reference object, transformation parameters are recorded
6. Residual error calculation: The residual error between corrected object and reference image is calculated
7. Approximate *MOS* by detected parameters: All parameters from the previous analysis steps are combined in an error model to predict a *MOS*

5.1. Implementation

From a computer vision perspective the following three steps are most interesting: object detection, object matching and object correction. They closely relate to the above mentioned computer vision approaches of object segmentation, object detection and object tracking. To illustrate these three steps we employ a sample image, which is introduced in Figure 4.

a) Erroneous Object Detection: A characteristic of erroneous objects is that image errors accumulate in the areas of these objects. We use this characteristic and in a first step compute the

average image error as the Mean Square Error. If objects are misplaced the image error in these areas is above the average image error, while the error is below in other areas. By filtering the error areas with a disk-shaped stencil, object areas can be distinguished. Two things are important to note: First, the object outline is only rough, but covers the whole area in which an object is misplaced with respect to the original. Second, the averaging disk size depends on image size and viewing conditions, to differentiate between noise and relevant objects.

(a) Rereference Image.   (b) Test Image.

Figure 4. Example Image Set illustrating the implementation

The result of this detection step is a mask with outlined areas. If multiple objects in an image are moved, all of these areas are marked and noted. For the sample images shown in Figure 4 the object detection mask is given in Figure 5.

b) Erroneous Object Matching: To match objects between test and reference images there are two possible cases: a transformed object may be overlapping in reference and test image (only one erroneous region detected) or they may be spatially distinct (two erroneous regions). With the additional possibility to have several wrong objects in an image, we need to match each region with itself and with all other error regions. For region matching we employ Scale Invariant Features (SIFT) as proposed by Lowe [8]. For each area detected in the previous step we record the closest match between reference and test image. Figure 6 shows detected features between reference (top) and test image (bottom). The translation of the car between test and reference image can already clearly be seen by the feature lines (white) running slightly tilted between both images.

c) Object Correction: Reallocating the distorted object from the test image to its original position in the reference image is an important task to calculate the visual disturbance of the picture irrespective of any transformations. Initially, we remove the misplaced object from the test image and fill the created hole with an inpainting algorithm. Second, we use the SIFT feature correspondences to get a rough registration of the object in the test image [8]. As SIFT feature matching leaves inaccuracies in the order of single pixels we employ a Levenberg-Marquardt least-square optimization with a Fourier-Mellin transform module to achieve an image registration with sub-pixel precision for exact object placement [15]. The order of applying the SIFT registration before the Fourier-Mellin transform based registration is advantageous, as the SIFT registration works robustly, but with a certain inaccuracy, while the Fourier-Mellin transform becomes unstable for images that are too different from each other but works with a high precision when images are closely aligned already. Our implemented concatenation is both robust and precise. Finally, the registered object is fitted onto the filled background image. Filled background image and test image after object registration are shown in Figure 7. Next to the registered image this step retrieves the scaling, translation and rotation values between reference and test object.

(a) Mask outlining erroneous object.   (b) Environment of erroneous object
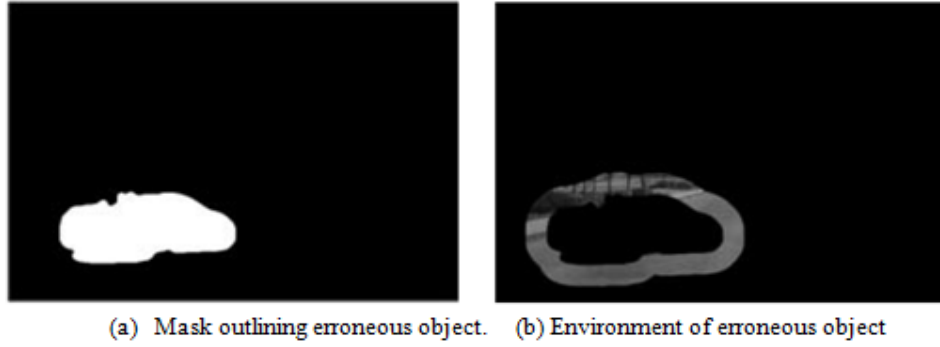
Figure 5. Mask and Environment of Erroneous Object

Figure 6. SIFT Matching between test and reference image

5.2. Metric Results and Relation to Computer Vision

The SC-VQM analyzes scene objects for transformations, and employs detected transformation parameters as well as the object size and its environment structure for visual quality prediction. We have tested this metric on the $SSID$ database, presented in Section 2. A comparison of correlations between the different metrics shows that the SC-VQM increases the correlation between MOS and predicted MOS for transformation errors by 28% compared to currently existing and established metrics. This result is visualized exemplary in Figure 8. While SSIM assigns a MOS score of "Fair" to the image ($MOS_p = 3$), our metric evaluated the test image close to "Excellent" ($MOS_p = 4.6$). The statistical error map (Figure 8.c) indicates why traditional metrics fail: a shift in image textures causes large parts of the image to be fully wrong, yet the human brain judges this error to be fairly unimportant.

We therefore observe a reciprocal relation between Computer Vision Performance and Image Quality Metrics. Image Quality Metrics can be used to predict the performance of Computer Vision algorithms; Image Quality Metrics can therefore play an essential role in the design and development of Computer Vision algorithms. At the same time, ideas from Computer Vision are employed in Image Quality Metrics to increase the correlation between predicted quality and subjective evaluations.

(a) Filled Background.   (b) Registered Object.
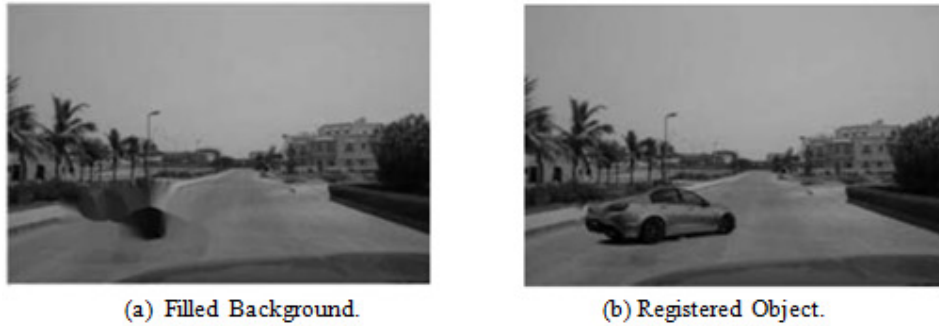
Figure 7. Filled Background and Registered Object for Example Image Set.

(a) Original.   (b) Wrong Shirt.   (c) Error Map.

Figure 8. Perceived quality does not always correspond to statistics.

6. CONCLUSION

In the previous sections we have shown that there exists a reciprocal relationship between image quality and computer vision. As a basis for research connecting both domains we have developed an image database, SSID, which contains synthetic images with classical image distortions and scene composition errors. This database is subjectively evaluated and contains additional data such as depth and segmentation maps, as well as the raw data to produce further information.

We have introduced three basic computer vision algorithms and four quality measures for visual information and have analyzed the image quality measures concerning their suitability for computer vision success prediction. Especially *PSNR* and *SSIM* were found to predict the impact of image distortions on computer vision algorithms well.

On the other hand we have observed that image quality metrics fail for visual content produced with computer vision approaches. Therefore a novel visual quality metric, SC-VQM, was developed, which is especially designed to analyze synthetic contents in virtual worlds or augmented reality scenarios. This metric can increase the quality prediction by 28% compared to current standard quality metrics.

Thus image quality metrics can be used to predict the success of computer vision approaches and computer vision can be employed to enhance the prediction capability of image quality metrics - a reciprocal relation.

REFERENCES

[1]   Hamid R Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik, "Live image quality assessment database release 2," 2005.

[2] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lukui Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Color image database tid2013: Peculiarities and preliminary results," in Visual Information Processing (EUVIP), 2013 4th European Workshop on. IEEE, 2013, pp. 106–111.

[3] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in Proc. 8th Int'l Conf. Computer Vision, July 2001, vol. 2, pp. 416–423.

[4] Christopher Haccius and Thorsten Herfet, "An image database for design and evaluation of visual quality metrics in synthetic scenarios," in International Conference on Image Analysis and Recognition, ICIAR, Póvoa de Varzim, Portugal. IEEE, July 2016.

[5] Christopher Haccius, "SSID - saarbru¨cken synthetic image database," http://ssid.nt.uni-saarland.de/, Apr. 2016, Accessed: 2016-06-14.

[6] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 11, pp. 2274–2282, 2012.

[7] Berthold K Horn and Brian G Schunck, "Determining optical flow," in 1981 Technical symposium east. International Society for Optics and Photonics, 1981, pp. 319–331.

[8] David G Lowe, "Object recognition from local scale-invariant features," in Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Ieee, 1999, vol. 2, pp. 1150–1157.

[9] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in Computer vision–ECCV 2006, pp. 404–417. Springer, 2006.

[10] Zhou Wang, Hamid R Sheikh, and Alan C Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in Image Processing. 2002. Proceedings. 2002 International Conference on. IEEE, 2002, vol. 1, pp. I–477.

[11] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," Image Processing, IEEE Transactions on, vol. 13, pp. 600–612, 2004.

[12] Charles Spearman, "The proof and measurement of association between two things," The American journal of psychology, vol. 15, no. 1, pp. 72–101, 1904.

[13] Kalpana Seshadrinathan, Jorge E Caviedes, Audrey C Younkin, and Philip J Corriveau, "Visual quality management in consumer video r&d," in International Workshop on Video Processing and Quality Metrics (VPQM), 2010.

[14] Christopher Haccius and Thorsten Herfet, "SC-VQM - a visual quality metric for synthetic contents," in Picture Coding Symposium, PCS, Nuremberg, Germany (submitted to). IEEE, Dec. 2016.

[15] George Wolberg and Siavash Zokai, "Robust image registration using log-polar transform," in Image Processing, 2000. Proceedings. 2000 International Conference on. IEEE, 2000, vol. 1, pp. 493–496.

AUTHORS

Prof. Thorsten Herfet was born in Bochum, Germany, on April 26th, 1963. Thorsten received a Diploma in Electrical Engineering in 1988 and a Ph.D. in telecommunication in 1992. Having been a PostDoc for 4 years he joined industry in 1996, finally being appointed Director of Research & Innovation of GRUNDIG. In 2004 he rejoined academia and became Full University Professor at Saarland University. His fields of research are cyber-physical networking, low latency streaming, computational videography and high mobility.

Prof. Herfet 2006-2008 served as the Dean for Informatics and Mathematics, in 2009 has been appointed Director of Research and Operations of the Intel Visual Computing Institute at Saarland University and since 2014 is Saarland University's Vice President for Research and Technology-Transfer. Thorsten published more than 100 papers, holds 15+ patents and has initiated and led several multi-million € collaborative research projects. He is a Senior Member of the IEEE, member of ACM SIGGRAPH, member of the German VDI/FKTG and serves as Steering Board member and Curator for various consortia and institutes.

Christopher Haccius was born in Lünen, Germany, on December 7, 1986. After his university-entrance diploma he received a BSc in Computer Science from the International University in Germany, Bruchsal, in 2009 and a MSc in Computer Science with focus on computer vision and telecommunications from Saarland University, Saarbrücken, Germany, in 2013. As a researcher his major fields of study are computer vision and telecommunications.

Mr. Haccius has work experience as a software developer with Fluid Operations and as a researcher at the telecommunications lab of Saarland University. Currently he is working for Continental Automotive in R&D of vehicle electronics.

INTENTIONAL BLANK

A Collage Image Creation & "Kanisei" Analysis System by Combining Multiple Images

Anju Kawamoto[1], Yasuhiro Hayashi[2], Izumi Fuse[3] and Yasushi Kiyoki[1]

[1]Faculty of Environmental Information, Keio University
5322, Endo, Fujisawa, Kanagawa, 252-0816, Japan
t13309ak@sfc.keio.ac.jp, kiyoki@sfc.keio.ac.jp
[2]Graduate School of Media and Governance, Keio University
5322, Endo, Fujisawa, Kanagawa, 252-0816, Japan
yasuhiro.hayashi@keio.jp
[3]Information Initiative Center, Hokkaido University
Nishi 5, Kita 11, Kita-ku, Sapporo, Hokkaido, 060-0811, Japan
ifuse@iic.hokudai.ac.jp

ABSTRACT

The collage is an artistic method to create a new image by combining multiple images under some rules or conditions for collage creators. To realize a mechanism to interpret "Kansei" of the collage creator by the combination process of multiple image-materials and a formed collage image, we propose a new system to analyze the collage work by using a database stored to each collage creation information. The collage works which a creator made by using this system was evaluated. And also, we clarified how to combine image-materials to make the good collage image having specific impression.

KEYWORDS

Collage Image, Image Creation, Semantic Analysis, Image Processing, Image Database

1. INTRODUCTION

The collage is an artistic method to create a new image by combining multiple images under some rules or conditions for collage creators. The outline of the collage is shown in Figure 1.
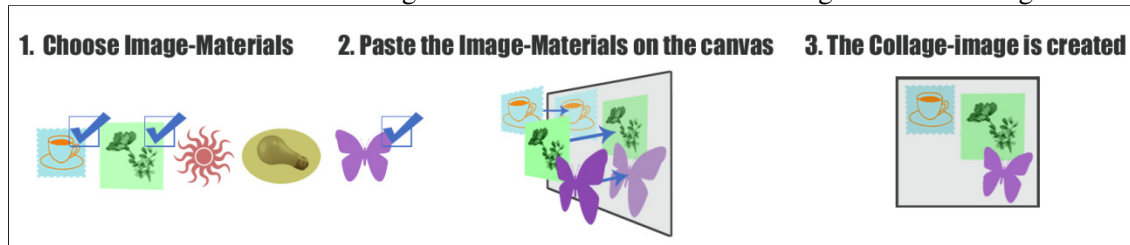

Figure 1. The Outline of the Collage

The creator chooses image martials for make a collage image. Then, the creator arranges images on the canvas while the creator considers position and size of each image. Finally, the collage image is made.

The purpose of the collage creation is to understand "Kansei" of human by combining process of the images as graphic materials to create and a combined image. The concept of "Kansei" includes several meanings on sensitive recognition,   such as human senses, feelings, sensitivity and psychological reactions.

The collage has been already used in various fields. In the real world, it is applied to the artistic expression [1] and the psychoanalysis [2, 3]. People enjoy humor of the combined image which do not exist in the real world and also they perform the self-understanding or stress emission through a series of processes for forming a collage image. The collage has been also utilized in the information retrieval research field [4]. In this field, Query-by-multiple-image is one of methods to retrieve images suited to retrievers' intention [5].

In this method, the retriever combines images his/her own have and creates an image-query for searching image data by the image. The retrieval method can create the image-query easily in comparison with the drawing method for CBIR (Content-based Image Retrieval) [6, 7].

Our goal is to realize a mechanism to understand "Kansei" by the combination process of the material images and a formed collage image on computers.

In the collage technique as previous research, psychologists have considered that facets of something of a creator are included to the collage work. For this reason, following points are important in order to interpret "Kansei" and intention of human.

- What kind of psychological meanings of a collage creator do the image-materials to be selected in the collage work have?

- How is the mental symptom expressed as combination of images?

However, the relevance between image-materials of collage and "Kansei" has not been fully researched because the collage technique is not able to solve the following points [8]:

- The format to analyze a collage work

- Characteristic expression and meanings in a collage work

From the viewpoint of this among "Kansei" analysis, we have already proposed an image-material selection method for the collage work by using images and "Kansei" databases. In this method, we had examined the relevance between the image-materials and the creator's conscious mind. By the results of this examination, we got knowledge as many collage creators choose image-materials to make the collage even if they dislike them. The collage image is not only combined images the collage creator likes but also new image including semantic values.

In this paper, we propose a new system to analyze the collage work by using a database stored to collage creation information. Each image material used for creation collage corresponds with

atoms of chemical substances. We explore meanings of the collage image that consists of atomic images. This operation is like analyzing molecular combination. Our system was developed and we evaluated collage works which a user made by using this system. And also, we clarified how to combine image-materials to make the collage having specific impression.

2. A COLLAGE IMAGE CREATION & "KANISEI" ANALYSIS SYSTEM BY COMBINING MULTIPLE IMAGES

The system enables the "Kansei" analysis of an image formed by the combination of multiple images as the collage image. The features of this system are shown as follows:

1. This system is able to analyze the relevance of each impression of image-materials and whole impression of the combined image as collage.

2. This system is also able to analyze the relevance of attributes of a collage creator and an impression of the combined image as collage.

The system consists of a database, an image editing tool and several functions to analyze the collage work. The target users of this system are a collage analyst as the administrator and a collage creator as the user. The collage analyst sets the image-materials for making collage images in the collage image database and analyze the created collage images by this system. The collage creator selects the image-materials that he/she wants to use to compose the collage image on the image editing tool.

2.1. Definitions of the Collage Database Schema

The schema of the collage database is defined as follows. A kind of data stored in the database are the image-materials, the collage-image combined the image-materials, and procedures and parameters of the image processing that correspond with the collage creators' intentions. In this system, impression of each image-material and the collage image is managed independently. These data are stored to each table having following structures.

2.1.1. The Image-Material Table:

The information about the image-material is recorded to the image-material table. The details of the image-material table are shown in Table 1.

Genre, format, impressions and data of the image-material are stored in this table. As the genre, this system handles 5 types genres ('nature', 'food', 'human', 'animal', 'symbol') based on the results of the previous research. The format of the image-material is stored into the format column as a binary value. If the format is picture, the stored value is expressed as 'true'.

This system deals with four basic impressions ('warm', 'dynamic', 'happy', 'light'). If the numerical value of each basic impression is 0, it means that an image is unrelated to the word of the basic impression. If the numerical value is positive, then it means that an image is related to the word. On the other hand, if the numerical value is negative, then it means that an image has the opposite impression ('cold', 'quiet', 'sad', 'heavy') to each impression word. These five basic

impression words were extracted from the impression word group for searching for paintings proposed by Kurita, Kato (1993) [9]. Moreover, we had added the impression words that suitable for the collage.

Table 1. The Image-Material Table

Property	Description	Data Type
m_id	Image-Material ID	INTEGER
m_url	Image-Material URL	TEXT
format	Flag for Distinguishing Photo whether or not	BOOLEAN
genre	Genre of Image-Material	TEXT
warm	Degree of Impression about 'warm'	INTEGER
dynamic	Degree of Impression about 'dynamic'	INTEGER
happy	Degree of Impression about 'happy'	INTEGER
light	Degree of Impression about 'light'	INTEGER

2.1.2. The Collage Image Table:

The information of the collage image and profile of a collage creator is stored to the collage image table. The structure of the collage image table is shown in Table 2.

Table 2. The Collage Image Table

Property	Description	Data Type
c_id	Collage Image ID	INTEGER
c_url	Collage Image URL	TEXT
name	Name of Creator	INTEGER
gender	Gender of Creator	CHAR(1)
age	Age of Creator	INTEGER
c_imp	Impression of Collage Image	TEXT
rule	Characteristic Representation	TEXT

The entire impression of the collage work is stored to 'c_imp' column as text type. The gender of the collage creator is stored into 'c_gender' column as M or F. The characteristic representation on the collage image is recorded to 'rule' column as text data.

2.1.3. The Image Processing Table:

A series of procedures and parameters of image processing for creating a collage image by combination of multiple images as many record data are saved into this table. The details of the image processing table are shown in Table 3.

Table 3. The Image Processing Table

Property	Description	Data Type
m_id	Image Material ID	INTEGER
c_id	Collage Image ID	INTEGER
manipulation	Manipulation Class of Image Processing	TEXT
parameter	Parameter for Manipulation	TEXT

Each record on the image processing table means that a collage image includes an image-material that is edited by a parameter. The system provides four functions ('zoom', 'layer', 'rotation', 'position') of the image processing. Selected image processing is marked to the 'manipulation' column as text data. The zoom means the magnification ratio and the ratio is stored as the parameter. The hierarchy order of each image on the canvas is stored as the parameter of the layer. We regard the magnification ratio and the order of the image layer as emphasized impression. The rotation angle is also stored as the parameter of the rotation. The parameter of the position corresponds to x and y coordinate of a collage image. All parameters are described as text type in the database to store different types of each parameter.

2.2. A Construction Method of the Collage Database for Creation and Analysis

The construction method of the collage database is shown as follows.

Step 1. Collecting The Image-Materials

The image-materials being used to create the collage image are prepared by the administrator of this system in advance. The administrator needs to keep impartial impression of the prepared image-materials. An image that strongly has a specific impression is not suited as the image material because there is potential for bias in the use frequency of the image-materials.

Step 2. Storing The Image-Materials to The Collage Database

The administrator distinguishes elements of each image-material and stores them into the image-material table in the collage database. Especially, the impression words of each image are defined by the administrator while referring to the results of a questionnaire that is related to relevance between an image and impression. The degree of impression is represented as an integer value from -2 to 2.

Step 3. Creating The Collage Image and Storing It into The Collage Database

The collage creator edits the image-materials to create the collage image by using the image editing tool. The created collage image and its entire impression are stored to the collage image table on the image database.

Step 4. Storing The Procedures and Parameters of the Image Processing to The collage database

The procedures and parameters of the image processing that the collage creator uses the image editing tool for are stored to the image processing table on the image database.

2.3. Implementation Method of the "Kansei" Analysis System for the Collage Images

In order to develop our system, PostgreSQL is used for the database system. Additionally, as an image editing tool, we utilized a web-based application for the image processing that is developed by Fuse, Okabe and Makino [10]. The screenshot of the image editing tool is shown in Figure 2.

Figure 2. The Screenshot of The Image Editing Tool

The collage creator can arrange the image-materials on canvas and also perform basic image processing (resize, rotation, layer management). The tool can output the created collage image and a XML file in which the procedures and the parameters of the image processing are described.

In this time, the collage creator submits the collage image and the XML file to the collage analyst. After that, the collage analyst converts the XML file to the SQL file, and stores the collage image and the image processing information to each table in the collage database. Additionally, collection of the image-material is selected from the web [11, 12, 13] by the administrator of this system.

2.4. An Analysis Method of Appropriate Combination of the Image-Materials and the Image Processing for the Collage Image

We show an analysis method of appropriate combination of the image-materials and the image processing for making a collage image.

Step 1. Choosing Specific Collage Images in The Collage Image Table

For example, when the collage creator produces a collage image representative of 'funny', the data where the 'c_imp' attribute of the collage image table is 'funny' is extracted by the system. Then, the extracted data is defined as a new table on the database by using The view function of SQL.

Step 2. Sorting The Image-Materials Obtained by Step 1 into The Frequency Order and Extracting Top 3 of The Image-Materials

There are attributes of 'm_id' and 'c_id' on the image processing table on the image database. The collage image 'c_id' and the used image-materials 'm_id' are related by this table. The image-materials with high frequency are searched by the 'm_id' attribute defined in Step 1. In this time, we extract top 3 of the image-materials based on the results of several experiments.

Step 3. Outputting Combination of Manipulations and Parameters of Three Image-Materials

After finishing 3 steps, characteristic impression is extracted on the specific collage image.

3. EXPERIMENTS AND DISCUSSION

In the experiments, while following the steps to analyze the impression of the image-material, we had carried out a questionnaire survey to 31 persons (12 males and 19 females) to set numerical values of four impression words ('warm', 'dynamic', 'happy', 'light') of each image-material. The most frequent value on the results of the questionnaire were adopted as numerical values of each impression word. The screen of the questionnaire is shown in Figure 3. And also, the image-material A is shown in figure 4. The data of the impression word for the image-material A is shown in Table 4. In the results of the questionnaire survey for the image-material, the most frequent value of 'light' was -2. In this reason, the image-material A is regarded as 'heavy'.

①

Do you feel what impression in this image (①) ? *

	2	1	0 (Neither Agree nor Disagree)	-1	-2
warm (+2) ⇔ cold (-2)	○	○	○	○	○
happy (+2) ⇔ sad (-2)	○	○	○	○	○
light (+2) ⇔ heavy (-2)	○	○	○	○	○
dynamic (+2) ⇔ static (-2)	○	○	○	○	○

Figure 3. The Screenshot of The Questionnaire

Figure 4. The Image-Material A

Table 4. The Numerical Value and Impression Word in The Image-Material A

Impression	Value
warm	1
dynamic	2
happy	1
light	-2

After finishing the questionnaire survey, we had 20 creators (11 males and 9 females) created collage images having impression of 'funny' as a fundamental experiment in order to analyze relevance between the image and the impression. We also analyzed following points and considered about representation of 'funny'.

- What kind of impression and image-materials of the genre does a collage creator?

- What kind of the image processing is operated the image-materials by a collage creator to make a collage work?

The results of the experiment are shown in table 5. Figure 5. and 6. are collage image that the creators actually created at the expreiments. The values of top five of each impression word that is frequently used were as follows: 'human' and 'animal' were used for the genre of the image-materials used for the collage images. And also, the image processing used by the collage creators most was 'rotation' in this time.

Table 5. The Results of The Experiment

Ranking of Usage Frequency	Genre of Image-Material	warm	dynamic	happy	light
1st (17 times)	'human'	2	1	1	1
2nd (13 times)	'animal'	1	-1	1	1
3rd (11 times)	'animal'	-1	-2	-1	-1
3rd	'food'	2	-1	2	0
3rd	'symbol'	-2	0	-2	2

Figure 5. The Collage Image 1

Figure 6. The Collage Image 2

Therefore, we noticed following conditions by the experiments. For making 'funny' collage images,

1. The genre of image-materials is 'human' or 'animal'.

2. The impression of an image-material is that 'warm' is positive value and 'dynamic' has negative value.

3. The same image-material is used in collage image many times.

4. A collage creator turns each image-material in the collage image appropriately.

4. CONCLUSTION AND FUTURE WORK

In this paper, we proposed a new system to support the creators of collage works to find a new expression method by providing an environment to analyze the impression of collage works.

In the collage database in our system, several kinds of collage works created by various human "Kansei", and the numerical values of impression metadata of each collage work. By our system, users and analyzers are able to discover a kind of "collective knowledge" in the collage creation. In this sense, our system leads to a new supporting environment for collage creators and analyzers by providing several new expression methods of collage from this collective knowledge database.

In this study, we performed a series of experiments under the following conditions, and examined the feasibility of our system. We set "interesting" as an input impression of collage, 'nature', 'animal', 'food', 'human' and 'symbol' as image genre, 'warm', 'dynamic', 'happy' and 'light' as impressions of the image-materials. We will continue to examine the effectiveness of our system based on the experimental results presented in this study.

Our system seems to have a wide applicability in various academic, art, and psychological field. First, our system has an applicability in the field of art and psychological therapy, because our system is able to collect and archive various collage-work data, and provide various "Kansei"-based expression of collage work. In the collage therapy, it is pointed that making a patient create collage works using unusual and unfamiliar materials and arrangement give a good effect to the patient by recognizing the self-understanding and changing the mental situation. Our system gives the new environment to the user (the creator or the patient) to challenge to find a new expression and a human creativity. Second, our system has an applicability in the field of the image retrieval by the following two points: (1) the creation of new images by the combination of multiple images, and (2) the interpretation of the creator's "Kansei" and the intention from the creation process and the newly-created image. In the field of the image retrieval, (2) is not achieved yet.

In our research, as feature work, we focus on the following three points to make our system practical: (1) automatic extraction of the most important expression way in each collage work, (2) database design to correspond to various and heterogeneous expressions of collage work, and (3) detailed checking of the genre and impression-words for collage material images. Our ultimate goal is to get to the essence of psychological situation of human and the rule of "Kansei". To achieve this goal, we have tried to collect the vast amount of real collage data and analyze them quantitatively and deeply.

REFERENCES

[1]   H.B. Landgarten, "Magazine Photo Collage: A Multicultural Assessment and Treatment Technique," London: Routledge, 1993.

[2]   J.W. Stallings,"Collage as a Therapeutic Modality for Reminiscence in Patients With Dementia," Journal of the American Art Therapy Association Volume 27, 2010 - Issue 3.

[3]   S. Forzoni, M. Perez, A. Martignetti, S.Crispino, "Art therapy with cancer patients during chemotherapy sessions: An analysis of the patients' perception of helpfulness," Palliative and Supportive Care, Volume 8, Issue 1, pp. 41-48, March 2010.

[4]   M. Flickner, H. Sawhney, W. Niblack, "Query by image and video content: the QBIC system," Computer 28, Issue: 9,23-32, 1995.

[5]   Y. Hayashi, Y Kiyoki, X. Chen, "A Combined Image-Query Creation Method for Expressing User's Intentions with Shape and Color Features in Multiple Digital Images," Information Modeling and Knowledge Bases, Vol. XXII, pp.258-277, May 2011.

[6]   Shaw, J. and Fox, E. Combination of Multiple Searches in Proceedings of TREC-3 (April 1995) 105-108.

[7]   Yoshitaka, A., and Ichikawa, T. A Survey on Content-Based Retrieval for Multimedia Databases, IEEE Transactions on Knowledge and Data Engineering, 11(1), 81-93.

[8]   N.Sonoda, A.Kondo, "Relationship between the style of collage and self," Kurume University psychological research No.5, 13-20,2006.

[9]   T. Kurita, T. Kato, "Learning of Personal Visual Impression for Image Database Systems," IEEE Xplore, Oct. 1993.

[10]   I.Fuse, S.Okabe, K.Makino, "Copyright Education using Manga materials with awareness of the creativity," JSiSE2014, 323-324, 2014.

[11]   Wafu-Sozai.com, http://www.wafusozai.com/.(accessed:2016-01-10)

[12]   Photo AC, http://www.photo-ac.com/.(accessed:2016-01-10)

[13]   Illust AC, http://www.ac-illust.com/ .(accessed:2016-01-10)

INTENTIONAL BLANK

HUMAN COMPUTER INTERACTION ALGORITHM BASED ON SCENE SITUATION AWARENESS

Cai Mengmeng[1,2], Feng Zhiquan[1,2] and Luan Min[1,2]

[1]School of Information Science and Engineering ,
University of Jinan, Jinan ,China 250022
[2]Shandong Provincial Key Laboratory of Network-based Intelligent Computing,
Jinan, 250022, P.R.
Corresponding author: Cai Mengmeng, E-mail: `1414663370@qq.com`
Corresponding author: Feng Zhiquan, E-mail: `ise_fengzq@ujn.edu.cn`

ABSTRACT

Implicit interaction based on context information is widely used and studied in the virtual scene. In context based human computer interaction, the meaning of action A is well defined. For instance, the right wave is defined turning paper or PPT in context B, And it mean volume up in context C. However, Select object in a virtual scene with multiple objects, context information is not fit. In view of this situation, this paper proposes using the least squares fitting curve beam to predict the user's trajectory, so as to determine what object the user's wants to operate .And fitting the starting position of the straight line according to the change of the discrete table. And using the bounding box size control the Z variable to move in an appropriate location. Experimental results show that the proposed in this paper based on bounding box size to control the Z variables get a good effect; by fitting the trajectory of a human hand, to predict the object that the subjects would like to operate. The correct rate is 88.6%.

KEYWORDS

Least-squares method; gesture recognition; human-computer interaction; visualization; implicit interaction; Context information

1. INTRODUCTION

With the continuous development of computer science and technology, intelligent human-computer interaction has gradually become the dominant trend in the development of computing model. And this trend becomes more obviously after Weiser Mark [1] putting forward the concept of "Ubicomp" in 1990s. In order to lighten the load of people's operation and memory, during the interaction, the traditional way of interaction need to be expanded. And integrate the implicit human-computer interaction into the explicit human-computer interaction.

At present, implicit human-computer interaction has become an important research frontier in the field of interaction. The universities and research institutes of the United States, Germany, China,

Austria and so on , has been carried out in-depth study to IHCI theory and application gradually. Schmidt in the University of Karlsruhe in Germany conducted an earlier study of the theory of implicit interaction [2].He believes that the two elements of implicit interaction are perception and reasoning, and he also put forward that contextual information is very important for interaction. Hamid Mcheick [3] presents a context aware model with ability to interact. This model adapt to dynamically environment and can interact with the user flexibility. The implicit interaction based on context is also applied in other aspects. Young-Min Jang [4] proposed a novel approach for a human's implicit intention recognition based on the eyeball movement pattern and pupil size variation. Bojan Blažica [5]introduces a new more personal perspective on photowork that aims at understanding the user and his/her subjective relationship to the photos. It does so by means of implicit human-computer interaction, this is, by observing the user's interaction with the photos.

In China, Tao Linmi [6] of Tsinghua University developed an adaptive vision system to detect and understand user behaviour, and to carry out implicit interaction. At the same time, Tian Feng in software research institute of Chinese Academy of Sciences also studied the characteristics of implicit interaction from the perspective of post WIMP user interface [7]. Wang Wei proposes that more use of user context information in the process of implicit human-computer interaction [8], Including user behaviour, emotional state (for example: The emotional design method of Irina CRISTESCU[9]), and physiological state. But there is also some use of Environmental context information, such as location-based services, etc. And it pointed out that the implicit human-computer interaction technology is one of the development directions in the future. Gao Jun pointed out in the article[10] Semantic Analysis is the importance and difficulty of high-level interpretation in image understanding, in which there are two key issues of text image semantic gap and text description polysemy. Yue Weining [11] proposed a context aware and scheduling strategy for intelligent interactive systems, which improves the system's intelligence. And Feng Zhiquan [12] uses the context information in the gesture tracking, and has achieved a good interaction effect.

2. RELATED WORK

2.1. Image segmentation

Before image segmentation, the image should be filtered to remove the noise. At present, the common methods of image segmentation [13] can be divided into: Threshold segmentation method [14], edge detection method [15], region segmentation method and the method of combining the theorem of the segmentation method. Besides, Qu Jingjing[16] proposed the segmentation method of continuous frame difference and background subtraction. This article uses the skin colour model [14] (YCbCr) to separate the human hand and the background, and the image banalization. Segmentation results are shown in Figure 1:
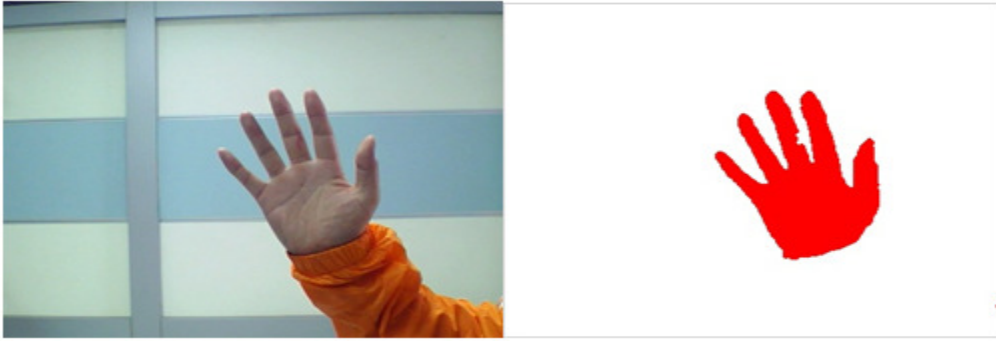
Figure 1. Original image and segmented image

2.2. Feature Extraction

The method of feature extraction is varied. Tao Sangbiao [17] proposed a static gesture contour feature extraction algorithm based on contour and skin colour. It extracts the gesture contour though skin's colour, and then extracts contour information. ZHU Jiyu [18] proposed a novel gesture segmentation algorithm can be divided global and local features. A fuzzy set is used to describe the background, colour and motion of the spatial and temporal information in the video stream. Ren Haibing [19] used a variety of information such as colour, motion and edge to extract features that can reflect the structure characteristics of the human hand, And he divides the characteristic lines into small curve segments. and track the movement of these curve segments. Feng Zhiquan [20] proposed the gesture features separation algorithm, gesture circumcircle radius is divided into different regions, and then features extraction. This method is not only simple but also has certain rotation and scaling invariance. In this paper, we use the method of document [20] to extract the feature points of the hand gesture. The specific methods as follows:

First, get the segmentation of the coordinates of the hand gesture, and the point of the greatest distance from the coordinates. Second, I using the centroid point as the centre point and the centroid of the farthest point distance concentric circle radius, divided into 7 layers as show in figure 2. Third, these 7 layers are divided into 3 categories: Fingertip layer, Finger heel layer, Joint point layer. In the end, get the fingertip and the number of layers and the number of connectivity.
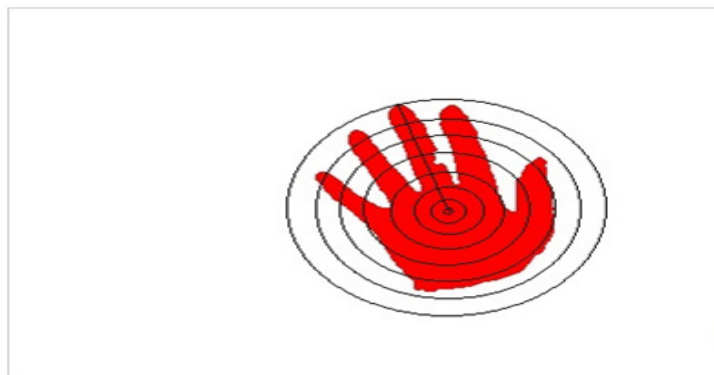
Figure 2. Feature Extraction

2.3. Gesture Recognition

Hand gesture recognition methods include: template matching method, statistical recognition method, fuzzy recognition method and artificial neural network classification, and shape matching method. Commonly used in the shape matching method [21] has the invariant moment method, the geometric parameter method, the characteristic model representation, the boundary direction histogram method, the wavelet importance coefficient method, as well as our country scholar studies the wavelet contour representation and so on. The method of gesture recognition based on Hausdorff distance [22] template matching algorithm is used in this paper. It is to obtain the characteristics of the library files and calculate the Hausdorff distance, the smaller the distance, the better the matching of the feature points. Specific algorithms are as follows: Assuming that A, B for the two sets has N and M elements respectively, then The distance Hausdorff(A, B) between A and B is H(A, B)

$$H(A, B) = \max\{h(A, B), h(B, A)\} \qquad \text{(formula 1)}$$

```
Int  Temp = Cnt =0;
For i=0: N
     For j=0: M
```

$$Temp = \min\|a_i - b_j\|;$$

$$Cnt = \max\{Temp, Cnt\};$$

$$h(A, B) = Cnt;$$

In the same way, you can calculate thus obtained H(A,B).

3. SCENE MODELLING

3.1. Brief introduction of image display

The principle of image display using OpenGL [23] in the virtual environment is exemplified here in Figure 3.
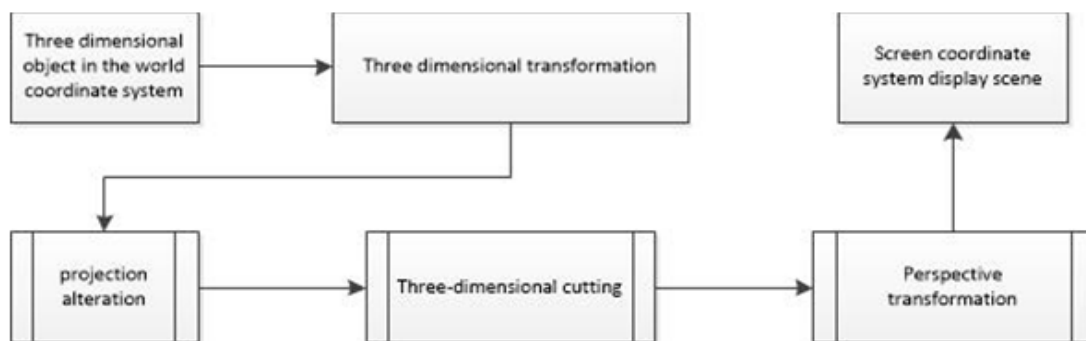
Figure 3. The principle of OpenGL image display

For different Z plane(Z=C, C is a constant), moving the same distance in one condition while the output of distance is not the same (The closer to the point of view, the greater the moving distance on the screen is). Therefore, objects at different coordinates in the virtual scene needs different functions to move them. Moreover, two-dimensional image obtained by the common camera is not good at controlling the movement of three - dimensional hand in three - dimensional spaces. So many researchers have used animation as a method to avoiding this problem. Using the principle that the bounding box size is proportional to the image display is the key to control changes in the Z-axis coordinate.
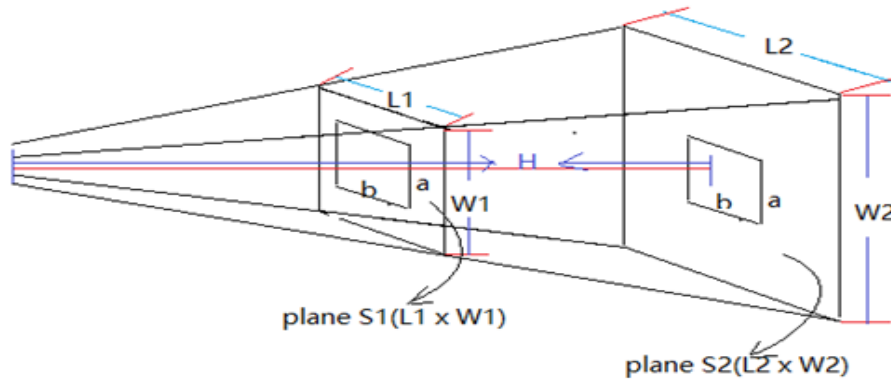
Figure 4. Camera image acquisition principle

The captured image (the size of determined acquisition: 400x300) is mapped to a window to display. So the length of a in plane S1 is $\dfrac{W_2}{W_1}$ times in plane S2 at the time of display.

3.2. Determine the Relationship of Mapping

Collect and record the size of the bounding box and get its average size when each of the experimenter is operating in the 3D scene. And Mapping shown by MATLAB is shown in Figure 5.
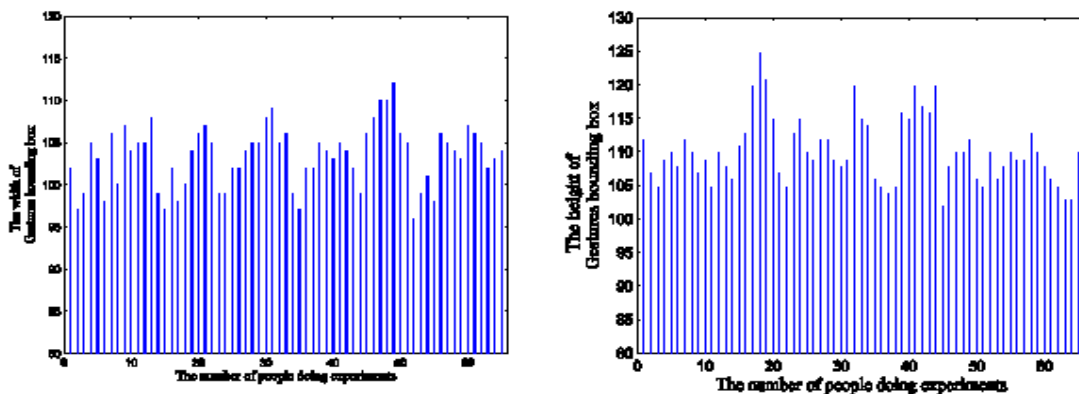
Figure 5. The Height and Width of Gesture Bounding Box

According to the probability formula in Statistics:

$$L = \frac{1}{n} \sum_{i=0}^{n-1} l_i \qquad\qquad (formula\ 2)$$

$$W = \frac{1}{n} \sum_{i=0}^{n-1} w_i \qquad\qquad (formula\ 3)$$

Calculate the initial value of L and W, $L_0 = 110$, $W_0 = 80$.

Calculate the corresponding relationship between real and virtual human hand (when the bounding box keeps the same size).

$$\frac{\Delta L}{\Delta L'} = \frac{400 - 2 * L_0}{L_0'} = 0.18125 \qquad\qquad (formula\ 4)$$

$$\frac{\Delta W}{\Delta W'} = \frac{300 - 2 * W_0}{W_0'} = 0.1833 \qquad\qquad (formula\ 5)$$

In the virtual scene, $L_0' = 1600$, $W_0'' = 1200$. Real hand moves in the horizontal direction a unit, virtual hands should move 5.51 units; and moving a unit in the vertical direction means virtual hands should move 5.45 units. For other positions virtual hands should move $5.51\dfrac{L}{L_0}$

units; virtual hands should move $5.45\dfrac{W}{W_0}$ units.

For Z coordinates, Position of each object in virtual scene is in [20 30]. The variation of bounding box's length is 80Pixel to 130Pixel. So the congruent relationship of bounding box's length and Z coordinates is:

$$\frac{Z - 20}{L - 80} = \frac{30 - 20}{130 - 80} = 0.2 \qquad\qquad (formula\ 6)$$

That is:

$$z = 20 + (l - 80) * 0.2$$

4. INTERACTION ALGORITHM BASED ON SCENE SITUATION AWARENESS

4.1. Based on Least Square Method [24] to fit the Motion Trajectory (Broken Line Segment) Algorithm

In order to better fit the motion trajectory of the hand gesture, in this paper, the least square method is used to fit the nonlinear equation.as shown in formula 7:

$$y_i = a * x_i + b * \sin(x_i) + c \qquad (i = 1,2,3, \ldots, n) \qquad (formula\ 7)$$

Formula (x_i, y_i) is the observation coordinate, a is first-order coefficients, b is sine coefficients, and c is constant. a, b and c is the parameter to be solved, assume a_0, b_0, c_0 for their approximate value. Order:

$$a = a_0 + \delta a,\ b = b_0 + \delta b,\ c = c_0 + \delta c$$

Taking y as the dependent variable and X as the independent variable, the error equation is:

$$v_{yi} = \begin{bmatrix} x_i & \sin(x_i) & 1 \end{bmatrix} \begin{bmatrix} \delta a \\ \delta b \\ \delta c \end{bmatrix} + a_0 * x_i + b_0 * \sin(x_i) + c_0 - y_i \quad (formula\ 8)$$

Error equation matrix can be expressed as:

$$A\delta X = 1 + V \qquad\qquad (formula\ 9)$$

Among them:

$$A = \begin{bmatrix} x_1 & \sin(x_1) & 1 \\ x_2 & \sin(x_2) & 1 \\ M & M & M \\ x_2 & \sin(x_n) & 1 \end{bmatrix},\ \delta X = \begin{bmatrix} \delta a \\ \delta b \end{bmatrix}$$

$$1 = \begin{bmatrix} a_0 x_1 + b_0 \sin(x_1) + c_0 - y_1 \\ a_0 x_2 + b_0 \sin(x_2) + c_0 - y_2 \\ M \\ a_0 x_n + b_0 \sin(x_n) + c_0 - y_n \end{bmatrix},\ V = \begin{bmatrix} V_{y1} \\ V_{y2} \\ M \\ V_{yn} \end{bmatrix}$$

According to the least square rule (formula 10), fitting a straight line.

$$\sum_{i-1}^{n} \left\| ax_i - b\sin(x_i) + c_0 - y_i \right\|^2 = \min \qquad (formula\ 10)$$

And Dependent variable residual is:

$$V = A\delta X - 1 \qquad\qquad (formula\ 11)$$

Because the cycle of $\sin(x_i)$ is $2 * \pi$, $b\sin(x_i)$ is periodic oscillation among in [0 400], so equation of a curve is

$$y_i = ax_i + b\sin(0.01x_i) + c \qquad (i = 1,2, \ldots, n)$$

In the end, According to the coefficient to confirm the good and bad fit.

4.2. Scene Situation Awareness and Interaction Algorithm

Calculate the size of the bounding box, and determine the corresponding relationship. According to the moving direction and distance of the 3D human hand of the two frame image, the movement of the centroid of the human hand is determined. The feature data of the multi frame images is used to synthesize nonlinear curve to predict the direction of human hand movement. And then determine the object at the direction and get the distance to human hand. Therefore, perform the corresponding operation; the specific algorithm is as follows:

First step: Capture a RGB image using a common camera .The height of the image is 400, and the width is 300. Then carry out image segmentation, and image banalization.

Second step: According to the formula (12) of the centroid of mass coordinates [25]:

$$r_c = \frac{\sum_i m_i r_i}{\sum_i m_i} \qquad\qquad (\text{formula } 12)$$

Figure out the centroid of coordinates after banalization; According to the formula-13 figure out bounding box size.

$$X_l = \min_{f(x_i, y_i) \neq 0} \{x_i\} \qquad\qquad (\text{formula } 13 - 1)$$

$$X_r = \max_{f(x_i, y_i) \neq 0} \{x_i\} \qquad\qquad (\text{formula } 13 - 2)$$

$$Y_l = \min_{f(x_i, y_i) \neq 0} \{y_i\} \qquad\qquad (\text{formula } 13 - 3)$$

$$Y_r = \max_{f(x_i, y_i) \neq 0} \{y_i\} \qquad\qquad (\text{formula } 13 - 4)$$

X_l is the left edge of the bounding box, X_r is right edge; Y_l is the upper boundary of the bounding box, Y_r is lower boundary. $f(x_i, y_i) \neq 0$ means that the pixels of the (x_i, y_i) coordinates are skin colour.

Third step: Calculate the vector (the size and direction) between two different centroid of coordinate and determine the direction and distance of the human hand movement in the 3D virtual scene according to the size and coordinates of the bounding box.

$$(\delta x, \delta y) = (x_{i+1}, y_{i+1}) - (x_i, y_i) = (x_{i+1} - x_i, y_{i+1} - y_i)$$

Fourth step: Using the glTranslatef (Dx, Dy, Dz) belonging to OpenGL to change the movement of the three-dimensional human hand in the virtual environment. If the moving amount of one direction (assumed to be X axis direction) is much greater than the other direction (Y axis) so you can only consider the direction where the moving amount is larger.

Fifth step: determine whether frames is greater than a threshold(set to 10).If less than, then return to the first step; Else, use the least square method to simulate curve.

Sixth step: Judge whether the fitting is good. If good, go to step seven; if not, adjust dynamically the number of the current frame according to the change of the discrete table, return to the fourth step.

Seventh step: Determine the number of objects that are in the prediction direction; if there is only one: move the object to the human hand. If not, adjust dynamically the number of the current frame according to the change of the discrete table, return to the fourth step.

At last, carry out the corresponding operation on the object by identifying a series of actions, for example: rotation, scaling, translation, and so on.

Algorithm flow chart is shown in figure 6.
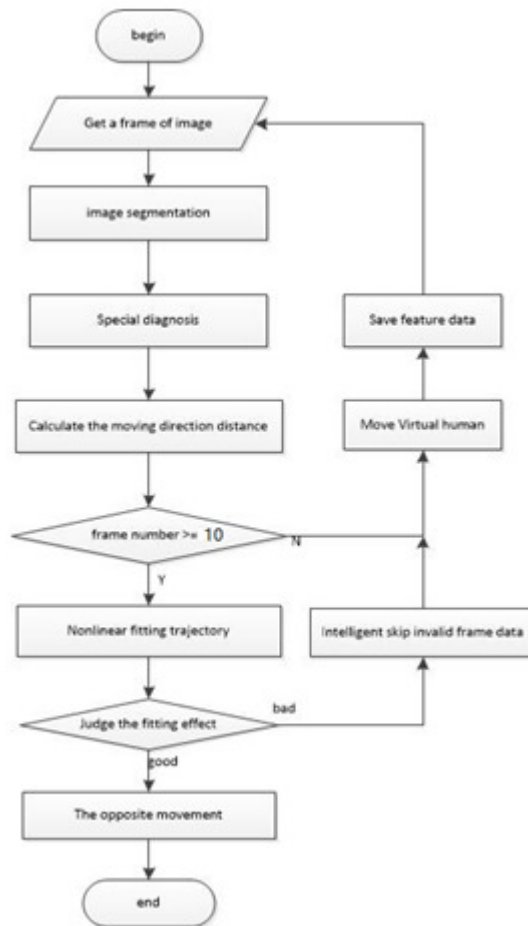
Figure 6. Algorithm flow chart

5. EXPERIMENTAL RESULT

The experiment is divided into two parts. A part is to be familiar with the experimental environment, Operation method and procedure, to determine the mapping relationship. Experimental interface is shown in Figure 7.
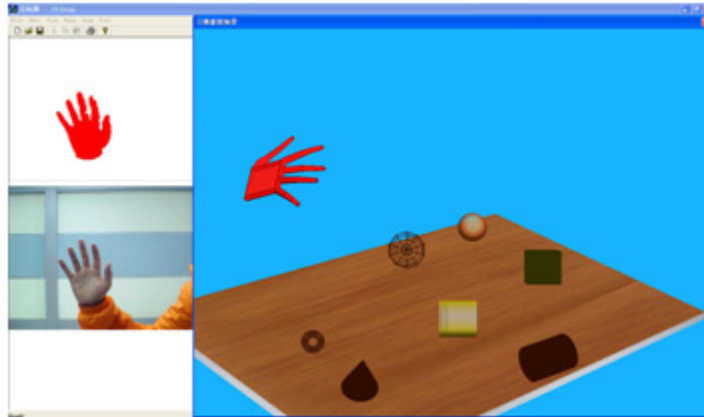
Figure 7. Virtual scene

On the right is a virtual 3D scene. Scene consists of virtual hands, and small balls, cylinder, cone and other three-dimensional objects. Each object is fixed, and is not in a z plane. On the left, there have two pictures, one picture is the original, and the other is the split hand. Real hand and virtual hand there is a certain relationship.

I find 65 students to do the experiment in the laboratory environment, under the constant light environment, the completion of the virtual scene to grab objects A, B, C, D, and other simple operation of the experiment. I recorded the size of their gestures when they were in the experiment, calculate the average and mapped with MATLAB, as shown in the figure 5. Determine the corresponding relationship and discrete table data, the content of the discrete table is related to the size of the bounding box and the speed of motion.

Another part is to select the object in the virtual scene as show in figure 7, and then do grab translation and other movements. First, I find 66 students again, divided into equal groups: A team, B team. Secondly, it is clear to tell the A team members of the experimental content: the object of the movement, the speed of movement, etc. Wait until the A team members are all familiar with the experimental environment and operating procedures, to do the experiment. Record the time it takes. And output the experimental data to a text file. I import the experimental data into MATLAB to fit the curve. According to the characteristics of the trajectory of human hand, I fit a curve. As shown in the figure 8.
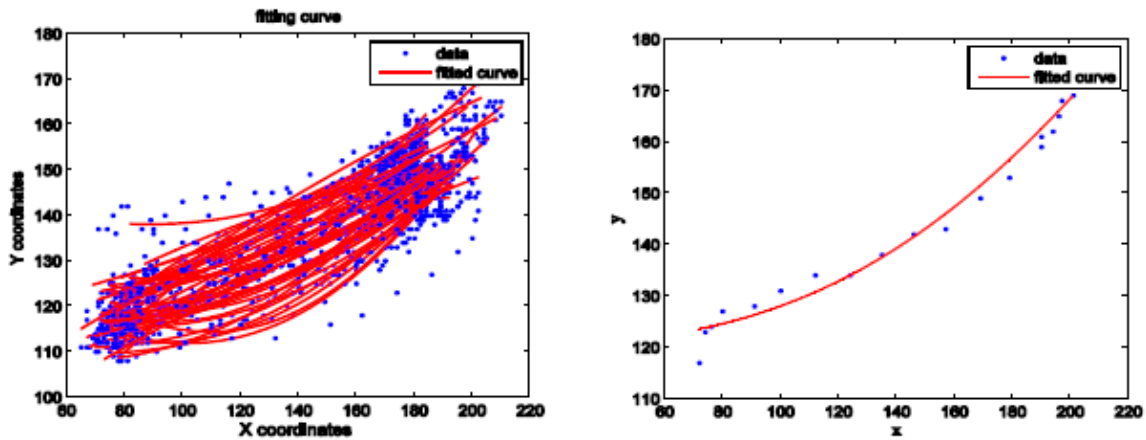
Figure 8. matlab fitting curve

The picture on the left is the fit figure of all members of the team A grab objects one. In the picture, the blue points are the centres' of Actual trajectory of hand and the red curve is the curve after fitting. The picture on the right is one of them in the left. The curve fitting coefficient is shown in figure 9. By analysing the motion trajectory of the A team, we can see that the trajectory of the hand is similar, and the movement of the hand tends to be circular. According to the trend of curve we can general position of object. We can be seen through the figure 9, the average fitting coefficient is higher than 0.95. This indicates that the selected curve is appropriate. In addition, the blue point is relatively dense in the upper right. That means it will cast lots of time to collision detection.
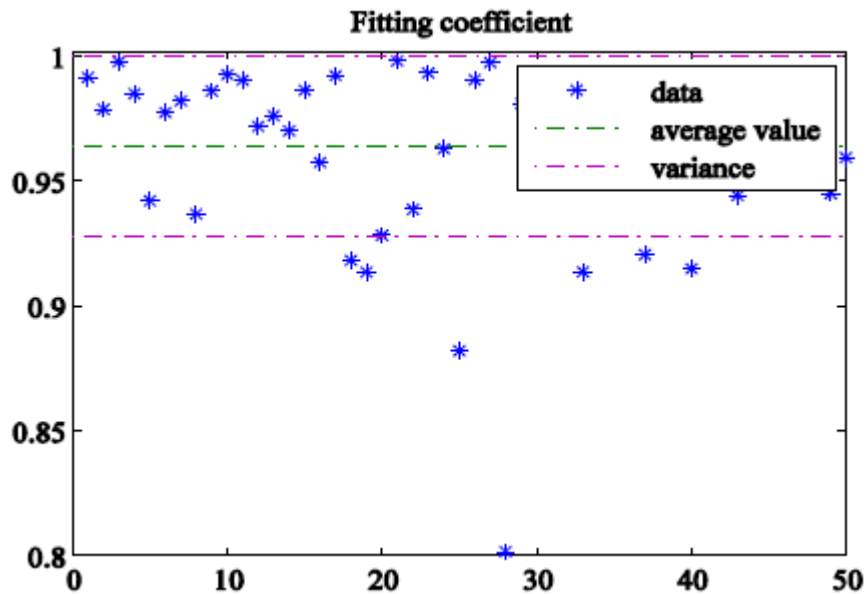
Figure 9. Correlation coefficient

Without specifying which object to select, Let the B team complete the experiment in the case of prediction and no prediction. Tell everyone select the same object twice. Record the number of

frames for each person to complete the experiment. In the experimental process with predictive function, according to the predicted result, it is judged which object in the virtual scene is to be selected. It will change the object colour and save the current number of frames. Wait the end of the experiment, record forecast result for the wrong. Repeat the experiment 5 times per person, seek its average. The final results are shown in Table 1.

Table 1. Experimental prediction results

Prediction Correct	1	2	3	4	5
The number of people	0	0	2	33	30

The 6 picture in Figure 10 is the screen in the experimental process with predictive function.
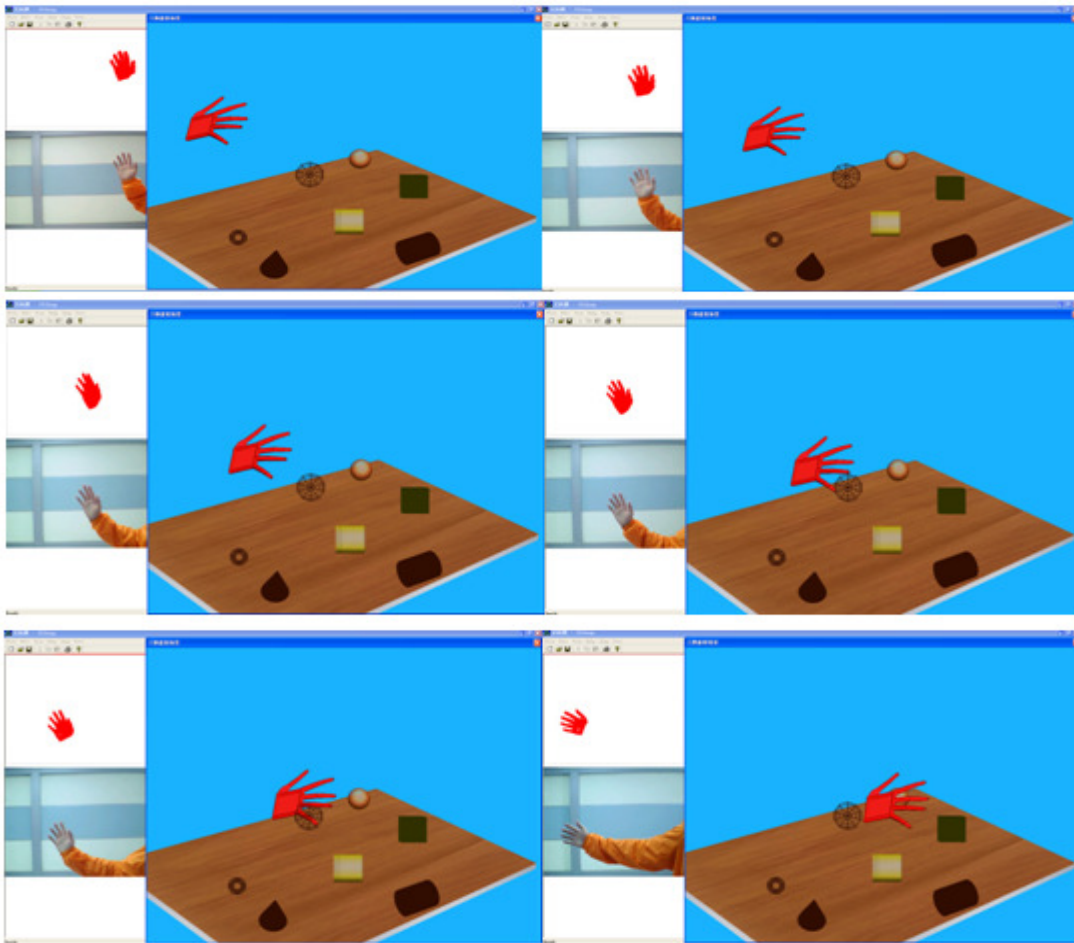
Figure 10. Grab the red ball

In the experimental process without predictive function, Use the method of Team A member to select the object to complete the experiment. Save the feature data and the number of frames used to complete the experiment.

The resulting number of frames is plotted in MATLAB as show in Figure 11.
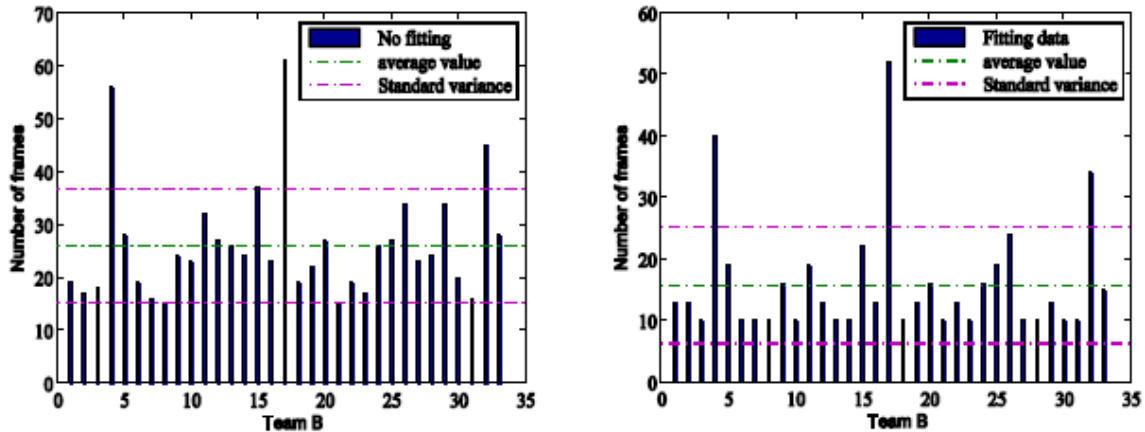
Figure 11. Time comparison chart

In Figure 11 we can see B team members to complete the experiment in the experimental process without predictive function; the average number of frames required is greater than 25. In the experimental process with predictive function, the average number of frames required is about 17. And apart from special case, the number of frames required to complete the experiment are within 20. In addition to through the table 1 we can be seen Most people can predict success 5 times, and a little of people predicted success 3 times, nobody can predict success less than 3 times. So, we can get a conclusion: in a specific virtual environment, the use of curve fitting method can be very good to predict the subjects want to operate the object.

6. CONCLUSIONS

According to the movement characteristics of the human hand and people's behaviour habits in the real scene, this article uses the least square method fitting a curve to predict direction of hand movement. This method has achieved very good results. And it can greatly reduce the time of selection. Secondly, the size of the bounding box is used to control the change of the Z axis variable in the appropriate range, and to realize the real manual control of the virtual human hand movement in the three-dimensional space. It conforms to the people in the three-dimensional environment in the operation of the habit (hand before and after the change, the virtual hand before and after the move). It also achieved good interaction effects. Finally, it is opposed to achieve human-computer interaction. And it has a certain effect. But for implementation a more intelligent human-computer interaction, there are a lot of problems to be solved. For example: the speed of the active object near the person, There is occlusion problem, as well as the computer automatically judge whether people have the purpose of the operation, etc.

REFERENCES

[1]   Lee, S.hyun. & Kim Mi Na, (2008) "This is my paper", ABC Transactions on ECE, Vol. 10, No. 5, pp120-122.

[2]   Weiser Mark . The computers for the twenty-first century [J].Scientific American, 1991, 265(3) : 94-104

[3]   Schmidt A. Implicit human computer interaction through context [J] .Personal Technologies, 200, 4 (2/3):191-199.

[4]   Hamid Mcheick. Modeling Context Aware Features for Pervasive Computing [J]. Procedia Computer Science, 2014, 37.

[5]   Young-Min Jang, Rammohan Mallipeddi , Sangil Lee, Ho-Wan Kwak, Minho Lee. Human intention recognition based on eyeball movement pattern and pupil size variation [J]. Neurocomputing, 2013.

[6]   Bojan Blažica, Daniel Vladušič, Dunja Mladenić. A personal perspective on photowork: implicit human–computer interaction for photo collection management [J]. Personal and Ubiquitous Computing, 2013, 178

[7]   Wang Guojian, Tao Linmi. Distributed Vision System for Implicit Human Computer Interaction [J]. Journal of Image and Graphics, 2010,08:1133-1138

[8]   TIAN Feng, DENG Changzhi, ZHOU Mingjun, et al. Research on the implicit interaction characteristic of Post-WIMP user interface. Journal of Frontiers of Computer Science and Technology, 2007, 1(2): 160- 169.

[9]   WANG Wei, HUANG Xiaodan, ZHAO Jijun, et al. Implicit Human_Computer Interaction [J]. Information and Control, 2014, 01:101-109.

[10]   Irina CRISTESCU. Emotions in human-computer interaction: the role of nonverbal behaviour in interactive systems [J]. Informatica Economica Journal, 2008, XII2:.

[11]   GAO Jun, XIE Zhao, ZHANG Jun, et al. Image Semantic Analysis and Understanding A Review [J]. Pattern Recognition and Artificial Intelligence, 2010, 02:191-202.

[12]   Yue Weining, WangYue, Wang Guoping et al. Architecture of Intelligent Interaction Systems Based on Context Awareness [J]. Journal of computer-Aided Design and Computer Graphics, 2005, 01:74-79.

[13]   Feng ZQ, Yang B, Zheng YW, Xu T, Tang HK. Hand tracking based on behavioural analysis for users. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2101-2116 (in Chinese). http://www.jos.org.cn/1000-9825/4368.htm

[14]   S. M. Lock, D. P. M. Wills. VoxColliDe: Voxel collision detection for virtual environments[J]. Virtual Reality, 2000, 51: .

[15]   Haokui-tang. Study of Skin Segmentation Based on Double-Models [D]. Shandong University, 2009.

[16]   Lu Kai, Li Xiaojian, Zhou Jinxing. Hand Signal Recognition Based on Skin Colour and Edge Outline Examination [J]. Journal of North China University of Technology, 2006, 03:12-15.

[17] QU Jing-jing, XIN Yun-hong. Combined Continuous Frame Difference with Background Difference Method for Moving Object Detection [J]. Acta Photonica Sinica, 2014, 07: 219-226.

[18] Tao Sangbiao, Jiao Guotai. Study on Extraction Algorithm for Static Hand Gesture Contour Feature [J]. Shanxi Electronic Technology, 2015, 02:90-91.

[19] ZHU Ji-Yu, WANG Xi-Ying, WANG Wei-Xin, et al. Hand Gesture Recognition Based on Structure Analysis [J]. Chinese Journal of Computers, 2006, 12:2130-2137.

[20] REN Hai-bing, XU Guang-you, LIN Xue-yin. Hand Gesture Recognition Based on Characteristic Curves [J]. Journal of Software, 2002, 05:987-993.

[21] FENG Zhi-quan, YANG Bo, ZHENG Yan-wei, et al. Gesture features detection based on feature points distribution analysis [J]. Computer Integrated Manufacturing Systems, 2011, 11:2333-2338+2340-2342.

[22] Li Junshan, Li Xuhui, Digital Image Processing [D]. Bei Jing, tsinghua university press, 2007：264

[23] ZHANG Liang-guo, WU Jiang-qin, GAO Wen, et al. Hand Gesture Recognition Based on Hausdorff Distance [J]. Journal of Image and Graphics, 2002, 11:43-49.

[24] kandyer. OpenGL Transform (EB/OL). http://blog.csdn.net/kandyer/article/details/12449973, 2016 - 01-18.

[25] School of Geodesy and Geomatics of Wuhan University. Error theory and measurement adjustment [M].Wuhan: wuhan university press, 2003.

[26] Zhang Mengzhong. The formula of centroid is derived by mathematical induction. [J].Journal of Jiujiang Teacher's College, 2002, 05: 46-4

AUTHORS

Cai Mengmeng

Master's degree of University of Jinan
The main research direction is Human-computer interaction and virtual reality.
E-mail: 1414663370@qq.com

Zhiquan Feng

Feng Zhiquan is a professor of School of Information Science and Engineering, Jinan University. He got the Master degree from north-western polytechnica university, china in 1995, and Ph.D degree from Computer Science and Engineering Department, shandong university in 2006. He has published more than 50 papers on international journals, national journals, and conferences in recent years. His research interests include: human hand tracking/recognition/interaction, virtual reality, human-computer interaction and image processing.
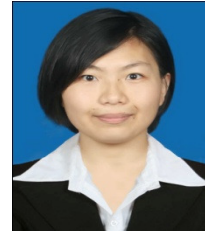
Luan Min

Master's degree of University of Jinan.
The main research direction is Human-computer interaction and virtual reality.
E-mail: 1562920346@qq.com

MULTIMODAL BIOMETRICS RECOGNITION FROM FACIAL VIDEO VIA DEEP LEARNING

Sayan Maity, Mohamed Abdel-Mottaleb, and Shihab S. As

University of Miami; 1251 Memorial Drive; Coral Gables; Florida 33146-0620
s.maity1@umail.miami.edu, mottaleb@miami.edu, sasfour@miami.edu

ABSTRACT

Biometrics identification using multiple modalities has attracted the attention of many researchers as it produces more robust and trustworthy results than single modality biometrics. In this paper, we present a novel multimodal recognition system that trains a Deep Learning Network to automatically learn features after extracting multiple biometric modalities from a single data source, i.e., facial video clips. Utilizing different modalities, i.e., left ear, left profile face, frontal face, right profile face, and right ear, present in the facial video clips, we train supervised denosing autoencoders to automatically extract robust and non-redundant features. The automatically learned features are then used to train modality specific sparse classifiers to perform the multimodal recognition. Experiments conducted on the constrained facial video dataset (WVU) and the unconstrained facial video dataset (HONDA/UCSD), resulted in a 99.17% and 97.14% rank-1 recognition rates, respectively. The multimodal recognition accuracy demonstrates the superiority and robustness of the proposed approach irrespective of the illumination, non-planar movement, and pose variations present in the video clips.

KEYWORDS

Multimodal Biometrics, Autoencoder, Deep Learning, Sparse Classification.

1. INTRODUCTION

There are several motivations for building robust multimodal biometric systems that extract multiple modalities from a single source of biometrics, i.e., facial video clips. Firstly, acquiring video clips of facial data is straight forward using conventional video cameras, which are ubiquitous. Secondly, the nature of data collection is non-intrusive and the ear, frontal, and profile face can appear in the same video. The proposed system, shown in Figure 1, consists of three distinct components to perform the task of efficient multimodal recognition from facial video clips. First, the object detection technique proposed by Viola and Jones [1], was adopted for the automatic detection of modality specific regions from the video frames. Unconstrained facial video clips contain significant head pose variations due to non-planar movements, and sudden changes in facial expressions. This results in an uneven number of detected modality specific video frames for the same subject in different video clips, and also a different number of modality

specific images for different subject. From the aspect of building a robust and accurate model, it is always preferable to use the entire available training data. However, classification through sparse representation (SRC) is vulnerable in the presence of uneven number of modality specific training samples for different subjects. Thus, to overcome the vulnerability of SRC whilst using all of the detected modality specific regions, in the model building phase we train supervised denoising sparse autoencoder to construct a mapping function. This mapping function is used to automatically extract the discriminative features preserving the robustness to the possible variances using the uneven number of detected modality specific regions. Therefore, by applying Deep Learning Network as the second component in the pipeline results in an equal number of training sample features for the different subjects. Finally, using the modality specific recognition results, score level multimodal fusion is performed to obtain the multimodal recognition result.
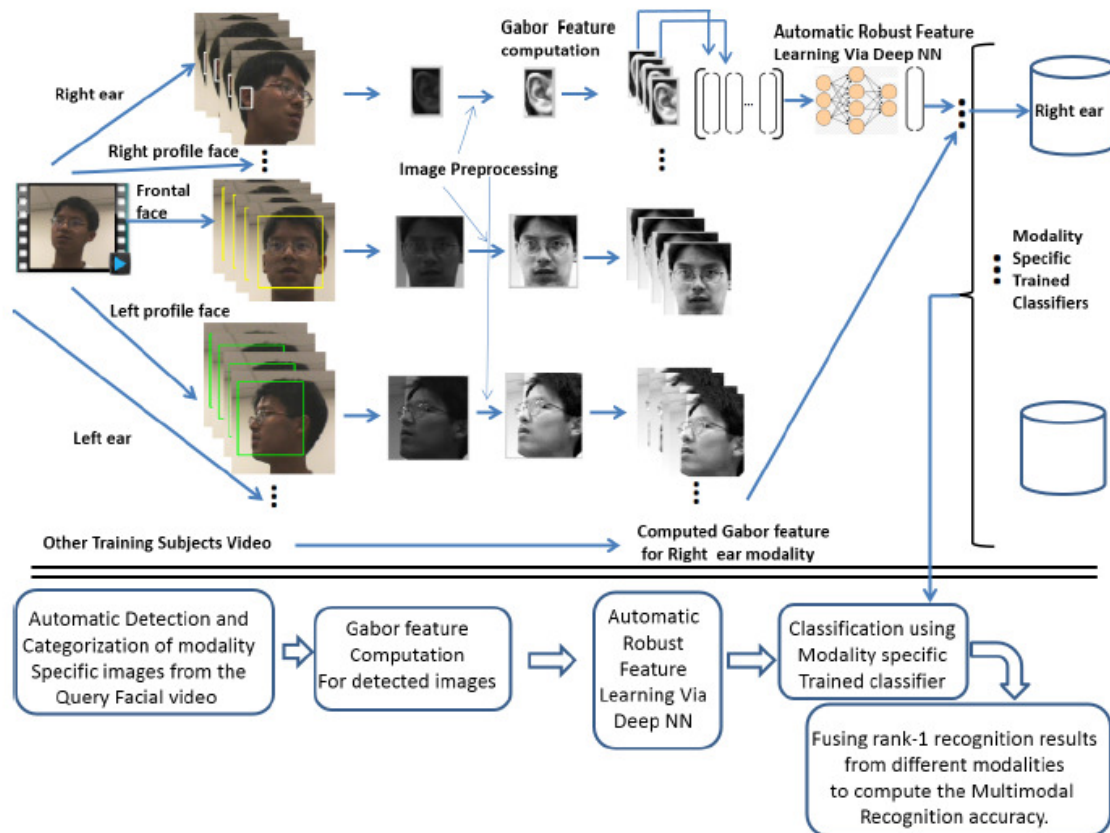
Fig. 1. System Block Diagram: Multimodal Biometrics Recognition from Facial Video

Due to the unavailability of proper datasets for multimodal recognition studies [2], often virtual multimodal databases are synthetically obtained by pairing modalities of different subjects from different databases. To the best of our knowledge, the proposed approach is the first study where multiple modalities are extracted from a single data source that belongs to the same subject. The main contributions of the proposed approach is the application of training a Deep Learning Network for automatic feature learning in multimodal biometrics recognition using a single source of biometrics i.e., facial video data, irrespective of the illumination, non-planar movement, and pose variations present in the face video clips.

The remainder of this paper is organized as follows: Section 2 details the modality specific frame detection from the facial video clips. Section 3 describes the automatic feature learning using supervised denoising sparse autoencoder (deep-learning). Section 4 presents the modality specific classification using sparse representation and multimodal fusion. Section 5 provides the experimental results on the constrained facial video dataset (WVU [3]) and the unconstrained facial video dataset (HONDA/UCSD [4]) to demonstrate the performance of the proposed framework. Finally, conclusions and future research directions are presented in Section 6.

2. MODALITY SPECIFIC IMAGE FRAME DETECTION

To perform multimodal biometric recognition, we first need to detect the images of the different modalities from the facial video. The facial video clips in the constrained dataset are collected in a controlled environment, where the camera rotates around the subject's head. The video sequences start with the left profile of each subject (0 degrees) and proceed to the right profile (180 degrees). Each of these video sequences contains image frames of different modalities, e.g., left ear, left profile face, frontal face, right profile face, and right ear, respectively. The video sequences in the unconstrained dataset contains uncontrolled and nonuniform head rotations and changing facial expressions. Thus, the appearance of a specific modality in a certain frame of the unconstrained video clip is random compared with the constrained video clips.

The algorithm was trained to detect the different modalities that appear in the facial video clips. To automate the detection process of the modality specific image frames, we adopt the Adaboost object detection technique, proposed by Viola and Jones [1]. The algorithm is trained to detect frontal and profile faces in the video frames, respectively, using manually cropped frontal face images from color FERET database, and profile face images from the University of Notre Dame Collection J2 database. Moreover, it is trained using cropped ear images from UND color ear database to detect ear images in the video frames. By using these modality specific trained detectors, we can detect faces and ears in the video frames. The modality specific trained detectors are applied to the entire video sequence to detect the face and the ear regions in the video frames.

Before using the detected modality specific regions from the video frames for extracting features, some preprocessing steps are performed. The facial video clips recorded in the unconstrained environment contain variations in illumination and low contrast. Histogram equalization is performed to enhance the contrast of the images. Finally, all detected modality specific regions from the facial video clips were resized; ear images were resized to 110 X 70 pixels and faces images (frontal and profile) were resized to 128 X 128 pixels.

3. AUTOMATIC FEATURE LEARNING USING DEEP NEURAL NETWORK

Even though the modalitiy specific sparse classifiers result in relatively high recognition accuracy on the constrained face video clips, the accuracy suffers in case of unconstrained video because the sparse classifier is vulnerable to the bias in the number of training images from different subjects. For example, subjects in the HONDA/UCSD dataset [4] randomly change their head pose. This results in a nonuniform number of detected modality specific video frames across different video clips, which is not ideal to perform classification through sparse representation.

In the subsequent sections we first describe the gabor feature extraction technique. Then, we describe the supervised denoising sparse autoencoders, which we use to automatically learn equal number of feature vectors for each subject from the uneven number of modality specific detected regions.

3.1 Feature Extraction

2D Gabor filters [5] are used in broad range of applications to extract scale and rotation invariant feature vectors. In our feature extraction step, uniform down-sampled Gabor wavelets are computed for the detected regions:

$$\psi_{\mu,\nu}(z) = \frac{||k_{\mu,\nu}||^2}{s^2} e^{\left(\frac{-||k_{\mu,\nu}||^2||z||^2}{2s^2}\right)} [e^{ik_{\mu,\nu}z} - e^{\frac{-s^2}{2}}], \tag{1}$$

where $z = (x, y)$ represents each pixel in the 2D image, $k_{\mu,\nu}$ is the wave vector, which can be defined as $k_{\mu,\nu} = k_\nu e^{i\phi_\mu}$, $k_\nu = \frac{k_{max}}{f^\nu}$, k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain, $\phi_\mu = \frac{\pi\mu}{2}$, and the value of s determines the ratio of the Gaussian window width to wavelength. Using equation 1, Gabor kernels can be generated from one filter using different scaling and rotation factors. In this paper, we used five scales, $\nu \in 0, ..., 4$ and eight orientations $\mu \in 0, ..., 7$. The other parameter values used are $s = 2\pi, k_{max} = \frac{\pi}{2}$, and $f = \sqrt{2}$.

Before computing the Gabor features, all detected ear regions are resized to the average size of all the ear images, i.e., 110×70 pixels, and all face images (frontal and profile) are resized to the average size of all the face images, i.e., 128×128 pixels. Gabor features are computed by convolving each Gabor wavelet with the detected 2D region, as follows:

$$C_{\mu,\nu}(z) = T(z) * \psi_{\mu,\nu}(z), \tag{2}$$

where $T(z)$ is the detected 2D region, and $z = (x, y)$ represents the pixel location. The feature vector is constructed out of $C_{\mu,\nu}$ by concatenating its rows.

3.2 Supervised Stacked Denoising Auto-encoder

The application of neural networks to supervised learning [6] is well proven in different applications including computer vision and speech recognition. An autoencoder neural network is an unsupervised learning algorithm, one of the commonly used building blocks in deep neural networks, that applies backpropagation to set the target values to be equal to the inputs. The reconstruction error between the input and the output of the network is used to adjust the weights of each layer. An autoencoder tries to learn a function $x_i = \hat{x}_i$, where x_i belongs to unlabeled training examples set $\{x_{(1)}, x_{(2)}, x_{(3)}, ..., x_{(n)}\}$, and $x_i \in \mathbb{R}^n$.

In other words, it is trying to learn an approximation to the identity function, to produce an output $\hat{x}$ that is similar to x, in two subsequent stages: (i) An encoder that maps the input x to the hidden nodes through some deterministic mapping function $f : h = f(x)$, then (ii) A decoder that maps the hidden nodes back to the original input space through another deterministic mapping function $g : \hat{x} = g(h)$. For real-valued input, by minimizing the reconstruction error $||x - g(f(x))||_2^2$, the parameters of encoder and decoder can be learned.

To learn features, which are robust to illumination, viewing angle, pose etc., from modality specific image regions, we adopted the supervised autoencoder [7]. The supervised autoencoder is trained using features extracted from image regions $(\hat{x}_i)$ containing variations in illumination, viewing angle and pose, whereas the features of selected image regions, (x_i), with similar illumination and without pose variations are utilized as the target. By minimizing the objective criterion given in Equation 3 (subject to, the modality-specific features of the same person are similar), the supervised autoencoders learn to capture the modality specific robust representation.

$$\min_{W,b_e,b_d} \frac{1}{N} \sum_i \left(||(x_i - g(f(\hat{x}_i))||_2^2 + \lambda||(f(x_i) - f(\hat{x}_i)||_2^2 \right) ; \qquad (3)$$

where the output of the hidden layer, h, is defined as $h = f(x) = tanh(Wx + b_e), g(h) = tanh(W^T h + b_d)$, N is the total number of training samples, and λ is the weight preservation term. The first term in Equation 3 minimize the the reconstruction error, $i.e.$, after passing through the encoder and the decoder, the variations (llumination, viewing angle and pose) of the features extracted from the unconstrained images will be repaired. The second term in Equation 3 enforces the simillarity of modality specific features corresponding to the same person.

After training a stack of encoders its highest level output representation can be used as input to a stand-alone supervised learning algorithm. A logistic regression (LR) layer was added on top of the encoders as the final output layer which enable the deep neural network to perform supervised learning. By performing gradient descent on a supervised cost function, the Supervised Stacked Denoising Auto-encoder (SDAE) automatically learned fine-tuned network weights. Thus, the parameters of the entire SDAE network are fine-tuned to minimize the error in predicting the supervised target ($e.g.$, class labels).

3.3 Training the Deep Learning Network

We adopt the two stage training of the Deep Learning Network, where we have a better initialization to begin with and a fine tuned network weights that lead

us to a more accurate high-level representation of the dataset. The steps of two stage Deep Learning Network training are as follows:

*Step*1. Stacked Denoising Autoencoders are used to train the initial network weights one layer at a time in a greedy fashion using Deep Belief Network (DBN).

*Step*2. The weights of the Deep learning network are initialized using the learned parameters from DBN.

*Step*3. Labelled training data are used as input, and their predicted classification labels obtained using the Logistic regression layer along with the initial weights of the network used as an objective function to fine tune the entire network .

*Step*4. Finally, the learned network weights are used to extract image features to train the sparse classifier.

The network is illustrated in Figure 2, which shows a two-category classification
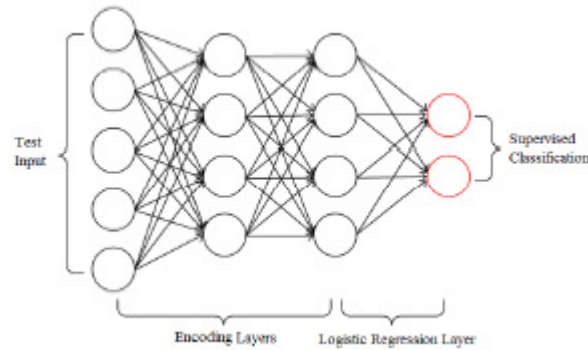
Fig. 2. Supervised Stacked Denoising Auto-encoder

problem (there are two output values), where the decoding part of SDAE is removed and the encoding part of SDAE is retained to produce the initial features. In addition, the output layer of the whole network, which is also called logistic regression layer, is added. The following sigmoid function is used as activation function of the logistic regression layer:

$$h(x) = \frac{1}{e^{-Wx-b}};$$
(4)

where x is the output of the last encoding layer y^l, in other words the features are pretrained by the SDAE network. The output of the sigmoid function is between 0 and 1, which denotes the classification results in case of two class classification problem. Therefore, we can use the errors between the predicted classification results and the true labels associated with the training data points to fine-tune the whole network weights.The cost function is defined as the following cross-entropy function:

$$Cost = -\frac{1}{m}\left[\sum_{i=1}^{m} l^{(i)} log(h(x^{(i)})) + (1 - l^{(i)})log(1 - h(x^{(i)}))\right];$$
(5)

where $l^{(i)}$ denotes the label of the sample $x^{(i)}$. By minimizing the cost function, we update the network weights.

4. MODALITY SPECIFIC AND MULITMODAL RECOGNITION

The modality specific sub-dictionaries (d_j^i) contain feature vectors generated by Deep Learning Network using the modality specific training data of each individual subject; where i represents the modality, $i \in 1, 2, ..., 5$; and j stands for the number of training video sequence.

Later, we concatenate the modality specific learned sub-dictionaries d_j^i of all the subjects in the dataset to obtain the modality specific (*i.e.*, left ear, left profile face, frontal face, right profile face, and right ear) dictionary D_i, as follows.

$$D_i = [d_1^i; d_2^i; ...; d_j^i]; \forall i \in 1, 2, ..., 5 \tag{6}$$

4.1 Multimodal Recognition

The recognition results from the five modalities — left ear, left profile face, frontal face, right profile face, and right ear are combined using score level fusion. Score level fusion has the flexibility of fusing various modalities upon their availability. To prepare for fusion, the matching scores obtained from the different matchers are transformed into a common domain using a score normalization technique. Later, the weighted sum technique is used to fuse the results at the score level. We have adopted the *Tanh* score normalization technique [8], which is both robust and efficient. The normalized match scores are then fused using the weighted sum technique:

$$S_p = \sum_{i=1}^{M} w_i * s_i^n; \tag{7}$$

where w_i and s_i^n are the weight and normalized match score of the ith modality specific classifier, respectively, such that $\sum_{i=1}^{M} w_i = 1$. In this study, the weights $w_i, i = 1, 2, 3, 4, 5$; correspond for the left ear, left profile face, frontal face, right profile face, and right ear modalities, respectively. These weights can be obtained by exhaustive search or based on the individual performance of the classifiers [8]. Later, the weights for the modality specific classifiers in the score level fusion were determined by using a separate training set with the goal of maximizing the fused multimodal recognition accuracy.

5. EXPERIMENTAL RESULTS

In this section we describe the results of the modality specific and multi-modal recognition experiments on both datasets. The feature vectors automatically learned using the trained Deep Learning network resulted in length of 9600 for frontal and profile face; 4160 for ear. In order to decrease the computational complexity and to find out most effective feature vector length to maximize the recognition accuracy, the dimensionality of the feature vector is reduced to a lower dimension using Principal Component Analysis (PCA) [9]. Using PCA, the number of features is reduced to 500 and 1000. In Table- 1 the modality specific recognition accuracy obtained for the reduced feature vector of 500, 1000 is shown. Feature vectors of length 1000 resulted in best recognition accuracy for both modality specific and multimodal recognition.

Table 1. Modality Specific and Multimodal Rank-1 Recognition Accuracy

Gabor Feature Length	Frontal face	Left profile face	Right profile face	Left ear	Right ear	Multimodal
No feature reduction	91.43%	71.43%	71.43%	85.71%	85.71%	88.57%
1000	**91.43%**	**71.43%**	**74.29%**	**88.57%**	**88.57%**	**97.14%**
500	88.57%	68.57%	68.57%	85.71%	82.86%	91.42%

The best rank-1 recognition rates, using ear, frontal and profile face modalities for multimodal recognition, compared with the results reported in [10{12] is shown in Table 2.

Table 2. Comparison of 2D multimodal (frontal face, profile face and ear) rank-1 recognition accuracy with the state-of-the-art techniques

Approaches	Modalities	Fusion Performed In	Best Reported Rank-1 accuracy
Kisku et al.[11]	Ear and Frontal Face	Decision Level	Ear: 93.53%; Frontal Face: 91.96%; Profile Face: NA; Fusion: 95.53%
Pan et al. [12]	Ear and Profile Face	Feature Level	Ear: 91.77%; Frontal Face: NA; Profile Face: 93.46%; Fusion: 96.84%
Boodoo et al. [10]	Ear and Frontal Face	Decision Level	Ear: 90.7%; Frontal Face: 94.7%; Profile Face: NA; Fusion: 96%
This Work	Ear , Frontal and Profile Face	Score Level	Ear: **95.04%**; Frontal Face: **97.52%**; Profile Face: **93.39%**; Fusion: **99.17%**

6. CONCLUSION

We proposed a system for multimodal recognition using a single biometrics data source, i.e., facial video clips. Using the Adaboost detector, we automatically detect modality specific regions. We use Gabor features extracted from the detected regions to automatically learn robust and non-redundant features by training a Supervised Stacked Denoising Auto-encoder (Deep Learning) network. Classification through sparse representation is used for each modality. Then, the multimodal recognition is obtained through the fusion of the results from the modality specific recognition.

REFERENCES

[1]   Viola, P. and Jones, M.: Grid Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, pp. 511{518.(2001).

[2]   Zengxi Huang and Yiguang Liu and Chunguang Li and Menglong Yang and Liping Chen: A robust face and ear based multimodal biometric system using sparse representation. In: Pattern Recognition, pp.2156{2168.(2013).

[3]   Gamal Fahmy and Ahmed El-sherbeeny and Susmita M and Mohamed Abdel-mottaleb and Hany Ammar: The effect of lighting direction/condition on the performance of face recognition algorithms. In:SPIE Conference on Biometrics for Human Identification, pp.188{200.(2006).

[4]   K.C. Lee and J. Ho and M.H. Yang and D. Kriegman: Visual Tracking and Recognition Using Probabilistic Appearance Manifolds. In:Computer Vision and Image Understanding.(2005).

[5]   Chengjun Liu and Wechsler, H.:Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. In: IEEE Transactions on Image Processing, pp.467{476.(2002).

[6]   Rumelhart, David E. and McClelland, James L.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. In:MIT Press. Cambridge, MA, USA.(1986).

[7]   Shenghua Gao and Yuting Zhang and Kui Jia and Jiwen Lu and Yingying Zhang: Single Sample Face Recognition via Learning Deep Supervised Autoencoders. In:IEEE Transactions on Information Forensics and Security, pp.2108{2118.(2015)

[8]   Ross, A. A. and Nandakumar, K. and Jain, A. K.: Handbook of multibiometrics. In:Springer.(2006)

[9]   Turk Matthew and Pentland Alex: Eigenfaces for recognition.In: J. Cognitive Neuroscience. MIT Press, pp.71{86.(1991)

[10]   Nazmeen Bibi Boodoo and R. K. Subramanian: Robust Multi biometric Recognition Using Face and Ear Images. In:J. CoRR.(2009)

[11]   Dakshina Ranjan Kisku and Jamuna Kanta Sing and Phalguni Gupta: Multibiometrics Belief Fusion. In:J. CoRR.(2010)

[12]   Xiuqin Pan and Yongcun Cao and Xiaona Xu and Yong Lu and Yue Zhao: Ear and face based multimodal recognition based on KFDA. In:International Conference on Audio, Language and Image Processing. pp.965{969.(2008)

INTENTIONAL BLANK

CLASSIFICATION OF UPPER AIRWAYS IMAGES FOR ENDOTRACHEAL INTUBATION VERIFICATION

Dror Lederman

Holon Institute of Technology, Holon, Israel
drorl@hit.ac.il

ABSTRACT

This paper addresses the problem of classification of upper airways images for endotracheal intubation verification in order to improve the safety of patients undergoing general anaesthesia. The proposed method is based on textural features utilized in a continuous probabilistic framework using parallel Gaussian mixture models (GMMs). The classification decision is made based on a maximum likelihood approach, which is insensitive to the angle at which the image was taken. Evaluation of the proposed approach is done using a dataset of 200 images that includes three classes of anatomical structures of the upper airways. The results show that the approach can be used to efficiently and reliably represent and classify medical images acquired during various procedures.

KEYWORDS

Artificial intelligence, intubation verification, Gaussian mixture models, textural features.

1. INTRODUCTION

Intubation is performed widely in hospitals and emergency medical units. During intubation, a flexible tube is used to secure passage of air to and from the lungs. The procedure is performed by manually opening the mouth, lifting the tongue using a device called laryngoscope in order to reveal the vocal cords, and inserting an endotracheal tube (ETT) through the vocal cords. The ETT should be positioned between 2 and 5 cm above the bifurcation of the trachea into the two primary bronchi ("carina").

The anatomy of the patient does not always allow easy insertion of the ETT and consequently it might be incorrectly positioned, usually either in the esophagus or in the right main bronchus. Both of these conditions can produce catastrophic results, as the patient might be deprived of oxygen. Unintentional esophageal intubation has been associated with high mortality rate [1, 2]. In cases of right lung intubation (also termed one-lung intubation (OLI)), only one lung is ventilated. Prolonged one lung ventilation might cause serious pulmonary complications such as collapse of the contralateral lung and hyperinflation of the ventilated lung, which might eventually result in hypoxia and pneumothorax, respectively, and has been associated with a significant increase in morbidity [3, 4] and Pneumonia [5]. Both esophageal and OLI may occur after the ETT was positioned correctly ("dislodgement") from many reasons, for example, due to neck flexion during general anesthesia [6, 7].

Confirmation of correct tube positioning is a challenging task that requires high skills and the use of secondary objective devices. Numerous studies, which investigated endotracheal misplacement rates in hospital and pre-hospital settings, reported rates between 0% and 25%, depending among others, on study design [1, 2, 5, 8-13].

There are various methods and techniques for endotracheal intubation confirmation. The most common technique is auscultation to lung sounds using a stethoscope. This technique requires high attention, and its reliability has been questioned in many studies [14-20]. The use of exhaled carbon dioxide detection (CO_2) measurements (termed end-tidal CO_2 ($ETCO_2$)), has become the gold standard-de-facto for confirming correct tube positioning. However, the method has been found to be unreliable in many emergencies [21-25]. In addition, the method can not be used to detect OLI incidents as in such cases the capnogram is generally typical in shape and shows normal $ETCO_2$ values [22, 25]. Other techniques have been proposed (e.g., [26-30]), but none of them has been proven effective. Therefore, attempts to find the ultimate technique for correct tube position confirmation have been continued.

In this paper, we further develop a previously-proposed system [31, 32] for endotracheal intubation confirmation. The system is based on identification of specific anatomical landmarks as indicators of correct or incorrect tube positioning, based on a continuous probabilistic approach using textural features.

2. MATERIAL AND METHODS

2.1 The System

Intubation is usually performed using an intubating stylet, used to control and guide the ETT. We designed and assembled a designated video-stylet. The tip of the stylet comprises a miniature complementary metal oxide silicon (CMOS) sensor. The inner part of the stylet contains wires to transfer the image and a narrow lumen to spray water or air in order to clear blood and secretions away from the camera sensor. The image sensor is connected to a processor with an integrated image acquisition component. During intubation, this rigid stylet is inserted into a standard ETT with its camera at the tip. Video signals are continuously acquired and processed by the confirmation algorithm implemented on the processor.

2.2 Pre-Processing and Textural Features

The classification system is a supervised one and comprises three main steps. The first step is conversion of the acquired image from a color image (RGB) to a grey scale image, which is followed by feature extraction. Various features have been utilized in medical image applications, including squared grey-level difference [33], cross correlation [33] and localized intensity features [34]. In this work, textural features [35, 36] were used. Textural features contain important information about the structural arrangement of surfaces and their relationship to the surrounding environment. In particular, features based on grey level co-occurrence matrices (GLCM) were utilized. These features are based on the assumption that textural information on an image is contained in the overall or "average" spatial relationship, which the grey tones in the image have to one another [37]. More specifically, it is assumed that this textural information is adequately specified by a set of grey tone spatial dependence matrices which are computed for various angular relationships and distances between neighbouring resolution cell pairs on the image. The advantages of these features is that they are robust to small differences in the imaging directions and scaling. This property is of great importance in many medical imaging applications.

2.3 The Gaussian Mixture Model (GMM) Probabilistic Framework

In order to classify the images, a probabilistic framework is utilized, in which the images are represented in the feature space using Gaussian mixture models (GMMs). GMM based classification methods have been largely applied to speech recognition [38], and recently to some medical image categorization applications [31, 34]. Mixture models, in particular GMM, form a common technique for probability density estimation. This is justified by the fact that any density can be estimated, in a required degree of approximation, using finite Gaussian mixture [39]. Their mathematical properties, as well as their flexibility and the availability of efficient estimation algorithms, make them attractive for classification problems. The most popular algorithm for GMM parameters estimation is the expectation-maximization (EM) [39]. This algorithm allows iterative optimization of the mixture parameters, under monotonic likelihood requirements, and has a relatively simple implementation. In this study, the probability density functions (pdf) of the images were represented, in the feature space, by three classes, where each class was modeled by a combination of four GMMs. Each GMM, representing a random process, x, is defined as a weighted sum of K Gaussian components, as follows:

$$f_K(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \phi_{\theta_k}(\mathbf{x}), \qquad (1)$$

where π_k represents the mixing weight such that $\sum_{k=1}^{K} \pi_k = 1, \pi_j \geq 0 \forall j$ and $\phi(\mathbf{x}; \theta_j)$ represents the jth d-dimensional Gaussian mixture component given by:

$$\phi(\mathbf{x}; \theta_j) = (2\pi)^{-d/2} \left| \mathbf{S}_j \right|^{-1/2}$$
$$\exp\left[-0.5(\mathbf{x} - \mathbf{m}_j)^T \mathbf{S}_j^{-1} (\mathbf{x} - \mathbf{m}_j) \right], \qquad (2)$$

which is parameterized on the mean $\mathbf{m}_j$ and the covariance matrix $\mathbf{S}_j$, collectively denoted by the parameter vector θ_j.
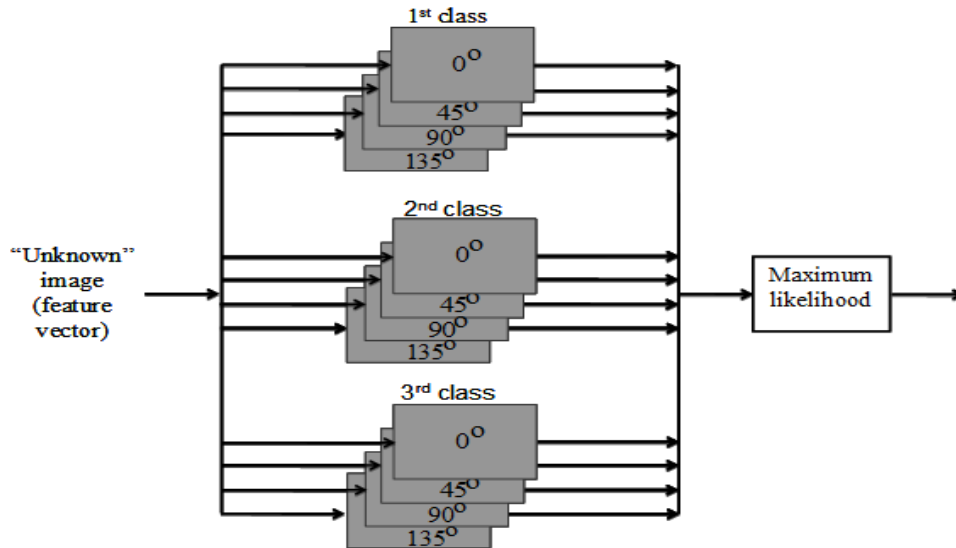

Figure 1. The parallel model approach

2.4 The Parallel GMMs Scheme

During endotracheal intubation, a common procedure in operating rooms, the tube is inserted into the patient's mouth in an horizontal direction, i.e., an angle of 0°, where the head of the tube turns away from the patient. While inserting the tube into the trachea, the medical professional rotates the tube counter clockwise until it reaches a vertical direction (90°). Therefore, the acquired images might be rotated with respect to the images that were used to train the models and therefore pre-alignment of the images prior to their classification is required. Traditionally, two approaches have been used to overcome this common problem: (a) employ an image registration technique prior to classification [40], (b) utilize features which are insensitive to the imaging angle [37]. Common to these approaches is their computational complexity, which is a major drawback when real-time classification is required. An alternative model based approach is proposed, according to which several models are used to represent each class of images, where each "sub-model" represents images taken at a specific range of angles, i.e., a GMM for angles around 0°, a GMM for angles around 45°, etc. During the training phase, a model is estimated using the images of each and every class, resulting in a total of 12 models (4 models for each one of the classes). During the testing phase, an "unknown" image is classified according to a generalized maximum likelihood rule, i.e., the selected model is the one which one of its sub-models provides the maximum likelihood (ML). The classification decision rule is therefore given by:

$$m = \arg_l \left\{ \max_{l=1,2; \ s=1,2,\ldots,S} f(O \mid \lambda_{ls}) \right\}, \tag{3}$$

where λ_{ls} represents the s th sub-model of the l th group. Figure 1 presents the parallel models scheme.

3. RESULTS

In order to evaluate system performance, we used a database of 200 images recorded from the upper airways of 10 patients where the images were taken at different angles and directions. The details of this database appear elsewhere [41]. The dataset used in this work includes 200 randomly-chosen images from the available database, that were visually inspected by a medical expert and classified into one of the following categories: upper-trachea, carina and esophagus. Each of these classes was further divided into four sub-groups representing 4 quantized imaging angles- 0°, 45°, 90° and 135°. Evaluation of the proposed approach was performed using a leave-one-out validation method: in each iteration, 199 of the images were used to train the models and the remaining image was used to test system performance. This process was repeated 200 times, such that each image participated once in the testing phase.

The classification results are summarized in Table 1, where the rows represent the predicted (identified) classes and the columns represent the actual classes. In this case, the order of each one of the GMMs was set to 8. The system achieved an overall classification rate of 92.0%.

Table 1. Confusion matrix of the three classes

Identified	True		
	Upper tracheal	Carina	Esophagus
Upper tracheal	92%	3%	5%
Carina	5%	94%	6%
Esophagus	3%	3%	89%

4. CONCLUSIONS

A method for classification of upper airways images based on a probabilistic framework using textural features was presented. A parallel scheme utilizing GMMs is used to represent images taken at different angles. While the training phase, which is usually a one-time phase, requires a tedious manual segmentation of the dataset according to the different sub-groups (for each range of imaging angles), the testing phase is based on a simple ML decision rule among the models of each class. Therefore, the proposed scheme is simple and computationally efficient. It should be noted however that further work is needed in order to evaluate the performance of the proposed approach and to compare it with other methods proposed in the literature.

All of the algorithms used in this work were implemented in Matlab R2016a 64bit. Using a conventional PC equipped with Dual Intel Xeon 3.4 GHz with a 4 GBytes of RAM, feature extraction requires about 1.5 second for each image and the parallel ML based decision rule requires approximately 1 second for each image. Efforts are currently being invested to collect a larger database in order to allow a more thorough investigation of the proposed approach.

REFERENCES

[1]   S. Silvestri, G. A. Ralls, and B. Krauss, "The effectiveness of out-of-hospital use of continuous end-tidal carbon dioxide monitoring on the rate of unrecognized misplaced intubation within a regional emergency medical services system," Ann. Emerg. Med., vol. 45, pp. 497-503, 2005.

[2]   A. Timmermann, S. G. Russo, C. Eich, M. Roessler, U. Braun, W. H. Rosenblatt, and M. Quintel, "The out-of-hospital esophageal and endobronchial intubations performed by emergency physicians," Crit. Care and Trauma, vol. 104, pp. 619-623, 2007.

[3]   R. L. Owen and F. W. Cheney, "Endobronchial intubation: a preventable complication," Anesthesiology, vol. 67, pp. 255-257, 1987.

[4]   C. W. Zwillich, D. J. Pierson, and C. E. Creagh, "Complications of assisted ventilation, a prospective study of 354 consecutive episodes," Am. J. Med., vol. 57, pp. 161-170, 1974.

[5]   H. E. Wang, L. J. Cook, C. H. Chang, D. M. Yealy, and J. R. Lave, "Outcomes after out-of-hospital endotracheal intubation errors," Resuscitation, vol. 80, pp. 50-55, 2009.

[6]   S. T. Vergese, R. S. Hannallah, M. C. Slack, R. R. Cross, and K. M. Patel, "Auscultation of bilateral breath sounds does not rule out endobronchial intubation in children," Anesth. Analg., vol. 56-58, 2004.

[7]   S. J. Yap, R. W. Morris, and D. A. Pybus, "Alterations in endotracheal tube position during general anesthesia," Anaesth. Crit. Care, vol. 586-588, 1994.

[8]   L. M. Jacobs, L. D. Berrizbeitia, B. Bernnett, and C. Madigan, "Endotracheal intubation in the prehospital phase of emergency medical care," JAMA, vol. 250, 1983.

[9]   M. E. Jemmet, K. M. Kendal, and M. W. Fourre, "Unrecognized misplacement of endotracheal tubes in a mixed urban to rural emergency medical services setting," Acad. Emerg. Med., vol. 10, pp. 961-965, 2003.

[10]   J. H. Jones, M. P. Murphy, and R. L. Dickson, "Emergency physician-verified out-of-hospital intubation: miss rates by paramedics," Acad. Emerg. Med., vol. 11, pp. 707-709, 2004.

[11]   S. H. Katz and J. L. Falk, "Misplaced endotracheal tubes by paramedics in an urban emergency medical services system," Ann. Emerg. Med., vol. 37, pp. 32-37, 2001.

[12] J. E. Pointer, "Clinical characteristics of paramedics' performance of endotracheal intubation," J. Emerg. Med., vol. 6, 1988.

[13] R. D. Steward, P. M. Paris, and P. M. Winter, "Field endotracheal intubation by paramedical personnel: success rates and complications," Chest, vol. 85, 1984.

[14] W. Brunel, D. L. Coleman, and D. E. Schwartz, "Assessment of routine chest reoentgenograms and the physical examination to confirm endotracheal tube position," Chest, vol. 96, pp. 1043-1045, 1989.

[15] T. H. Howells, "Oesophageal misplacement of a tracheal tube," Anaesthesia, vol. 40, pp. 398-389, 1985.

[16] I. D. Klepper, R. K. Webb, J. V. D. Walt, G. L. Ludbrooks, and J. Cockings, "The stethoscope: Application and limintation- an analysis of 2000 incidents reports," Anaesth. Intens. Care, vol. 21, pp. 575-578, 1993.

[17] K. Linko, M. Paloheimo, and T. Tammisto, "Capnograhphy for detection of accidental oesophageal intubation," Acta Anaesthesiol. Scand., vol. 27, pp. 199-202, 1983.

[18] A. W. Peterson and L. M. Jacker, "Death following inadvertent esophageal intubation: a case report," Anesth. Analg., vol. 52, pp. 398-401, 1973.

[19] H. E. Wang, J. R. Lave, C. A. Sirion, and M. Yealy, "Paramedic intubation errors: isolated events or symptoms of larger problems?," Health Affairs, vol. 25, pp. 501-509, 2006.

[20] G. R. Wodicka, P. D. DeFrain, and S. S. Kraman, "Bilateral asymmetry of respiratory acoustic transmission," Med. Biol. Eng. Comp., vol. 32, pp. 489-494, 1994.

[21] S. M. Bhende and A. E. Thompson, "Evaluation of an end-tidal CO2 detector during pediatric cardiopulmonary resuscitation," Pediat., vol. 95, pp. 395-399, 1995.

[22] J. S. Gravenstein, M. B. Jaffe, and D. A. Paulus, Capnograhpy clinical aspects: Cambridge University Press, 2004.

[23] J. Li, "Capnography alone is imperfect for endotracheal tube placement confirmation during emergency intubation," J. Emerg. Med., vol. 20, pp. 223-229, 2001.

[24] J. P. Nolan, C. D. Deakin, and J. Soar, "European resuscitation council guidelines for resuscitation," Resuscitation, vol. 67, pp. S39-S86, 2005.

[25] R. K. Webb, J. H. V. D. Walt, W. B. Runciman, J. A. Williamson, J. Cockings, W. J. Russel, and S. Helps, "Which monitor? an analysis of 2000 indicent reports," Anaesth. Intens. Care, vol. 21, pp. 529-542, 1993.

[26] D. Lederman, "An energy ratio test for one lung intubation detection," in 18th Biennial International EURASIP conference, Brno, Czech Republic, 2006.

[27] C. J. O'connor, H. Mansy, R. A. Balk, K. J. Tuman, and R. H. Sandler, "Identification of endotracheal tube malpopsitions using computerized analysis of breath sounds via electronic stethoscopes," Anesth. Analg., vol. 101, pp. 735-739, 2005.

[28] S. Tejman-Yarden, D. Lederman, N. Weksler, and G. Gurman, "Acoustic monitoring of double lumen ventilated lungs for the detection of selective unilateral lung ventilation," Anesth. Analg., vol. 103, pp. 1489-1492, 2006.

[29] S. Tejman-Yarden, A. Zlotnik, L. Weizman, J. Tabrikian, A. Cohen, N. Weksler, and G. M. Gurman, "Acoustic monitoring of lung sounds for the detection of one-lung intubation," Anesth. Analg., vol. 105, pp. 397-404, 2007.

[30] L. Weizman, J. Tabrikian, and A. Cohen, "Detection of one-lung intubation incidents," Annals of Biomed. Eng., vol. 36, pp. 1844-1855, 2008.

[31] D. Lederman, "An endotracheal intubation confirmation system based on carina image detection- a preliminary assessment," Med. & Biol. Eng. & Comp., vol. 49(1), pp. 75-83, 2011.

[32] D. Lederman, "Endotracheal intubation confirmation based on video image classification using a parallel GMMs framework- a preliminary evaluation," Annals of Biomed. Eng., vol. 39(1), pp. 508-513, 2011.

[33] D. Deguchi, K. Mori, M. Feuerstein, T. Kitasaka, C. R. Maurer, Y. Suenaga, H. Takabatake, M. Mori, and H. Natori, "Selective image similarity measure for bronchoscope tracking based on image registration," Medical Image Analysis, vol. 13, pp. 621-633, 2009.

[34] H. Greenspan and A. T. Pinhas, "Medical image categorization and retrieval for PACS using the GMM-KL framework," IEEE Trans. on Inform. Tech. in Biomed., vol. 11, pp. 190-202, 2007.

[35] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images," Clinical Radiology, vol. 59, pp. 1061-1069, 2004.

[36] J. Tan, B. Zheng, W. Wang, D. Lederman, J. Pu, F. C. Sciurba, D. Gur, and J. K. Leader, "Texture-based segmentation and analysis of emphysema depicted on CT images," presented at the Proc SPIE, 2011.

[37] R. M. Haralick, M. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. on Systems, Man, and Cybernatics, vol. SMC-3, pp. 610-621, 1973.

[38] R. C. Rose and D. A. Reynolds, "Text-independent speaker identification using automatic acoustic segmentation," presented at the IEEE Int. Conf. Acoust., Speech and Signal Proc., 1990.

[39] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Royal Stat. soc., vol. 39, pp. 1-38, 1977.

[40] Y. Keller, Y. Shkolnisky, and A. Averbuch, "The Angular Difference Function and its application to image registration," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, pp. 969-976, 2005.

[41] D. Lederman, S. Lampotang, and M. Shamir, "Automatic endotracheal tube position confirmation system based on image classification- a preliminary assessment," Med. Eng. & Phys., vol. 33(8), pp. 1017-1026, 2011.

AUTHORS

Dror Lederman, PhD, BEMS, LLB
Faculty of Electrical Engineering
Holon Institute of Technology, Holon, Israel

Short Biography

Dr. Dror Lederman received the B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering from Ben Gurion University of the Negev, Beer Sheva, Israel in 1998, 2003 and 2009, respectively, the B.EMS. in Emergency Medicine from Ben Gurion University of the Negev, Beer Sheva, Israel in 2005, and the LL.B. degree in Law from Tel-Aviv University, Tel-Aviv, Israel. He is currently a senior lecturer at the Holon Institute of Technology. His research interests include mainly machine learning, artificial intelligence and computer-aided diagnosis. Dr. Lederman also serves as a paramedic in the Red Cross Chapter in Israel ("Magen David Adom").

INTENTIONAL BLANK

BIG DATA TECHNOLOGY ACCELERATE GENOMICS PRECISION MEDICINE

HAO LI

Software Architect, Datacenter Health and Life Science,
Intel Corporation, Shanghai, China
`hao.h.li@intel.com`

ABSTRACT

During genomics life science research, the data volume of whole genomics and life science algorithm is going bigger and bigger, which is calculated as TB, PB or EB etc. The key problem will be how to store and analyze the data with optimized way. This paper demonstrates how Intel Big Data Technology and Architecture help to facilitate and accelerate the genomics life science research in data store and utilization. Intel defines high performance GenomicsDB for variant call data query and Lustre filesystem with Hierarchal Storage Management for genomics data store. Based on these great technology, Intel defines genomics knowledge share and exchange architecture, which is landed and validated in BGI China and Shanghai Children Hospital with very positive feedback. And these big data technology can definitely be scaled to much more genomics life science partners in the world.

KEYWORDS

GenomicsDB; Lustre; Big Data; Cloud; Life Science

1. INTRODUCTION

In genomics life science research, the data volume is going bigger and bigger, which is calculated as TB, PB or EB etc. For example, Genomics sequencing generates more than 1TB data per patient. During 2015, 1.65 million new patients in US generates more than 4EB data. In CNGB (China National Gene Bank), there is 500PB data volume deployed for now and it is estimated that volume will be increased by 5-10PB per year. In SCH (Shanghai Children Hospital) and Sjtu (Shanghai Jiao Tong University) Super Computing Center, there have hundreds of nodes for totally 30PB storage deployment. The key priority will be how to store and analyze data with optimized way and how to exchange and share the data with each other.

Intel defines and deploys scalable genomics knowledge share and exchange architecture, which is landed and validated in BGI China and Shanghai Children Hospital with very positive feedback. The solution stores data by using high density big data Lustre file system and leverages the cloud storage for hot and cold data efficient management. And solution architecture provides customized GenomicsDB engine for genomics variant call data search by position with very fast speed. Because genomics position is discrete, GenomicsDB also optimizes the sparse array storage by saving only the "useful" data. Intel also defines the genomics knowledge data sharing process and architecture for making the genomics knowledge be consolidated and utilized more efficiently.

2. INTEL BIG DATA ARCHITECTURE FOR LIFE SCIENCE

In real scenario, the research on genomics must work on big data mode. In this solution, it provides big data architecture (Figure 1). It is separated into 2 layer. One is application framework, which is likely interface to end user. It supports Genomics Knowledge App, Genomics DB UI as well as Genomics Work Flow etc. The other is core framework and Linux kernel, which provides services to support request from application framework. The architecture supports big data level core framework. Such as TileDB and GenomicsDB Engine for big genomics variant data, Lustre file system with HSM (Hierarchal Storage Management) to leverage local and cloud storage for scaling to bigger data.
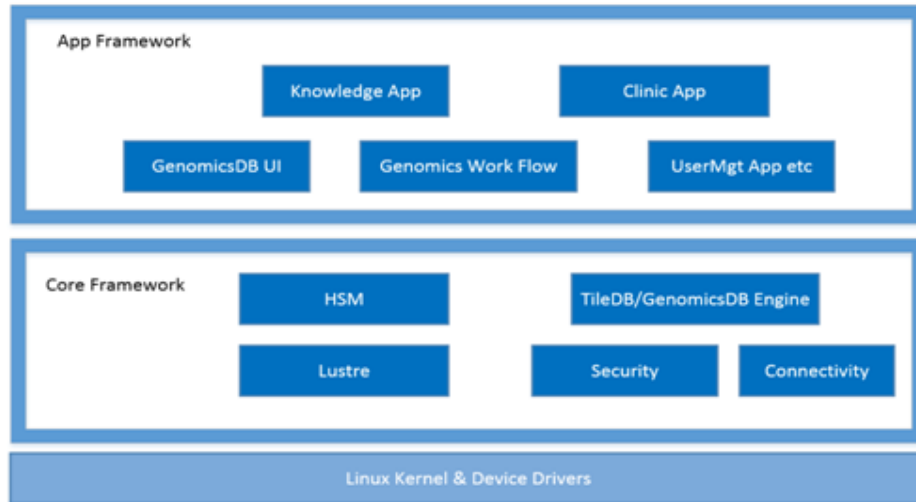

Figure 1. Big Data Architecture for Life Science

2.1. GENOMICS KNOWLEDGE SHARE MODEL

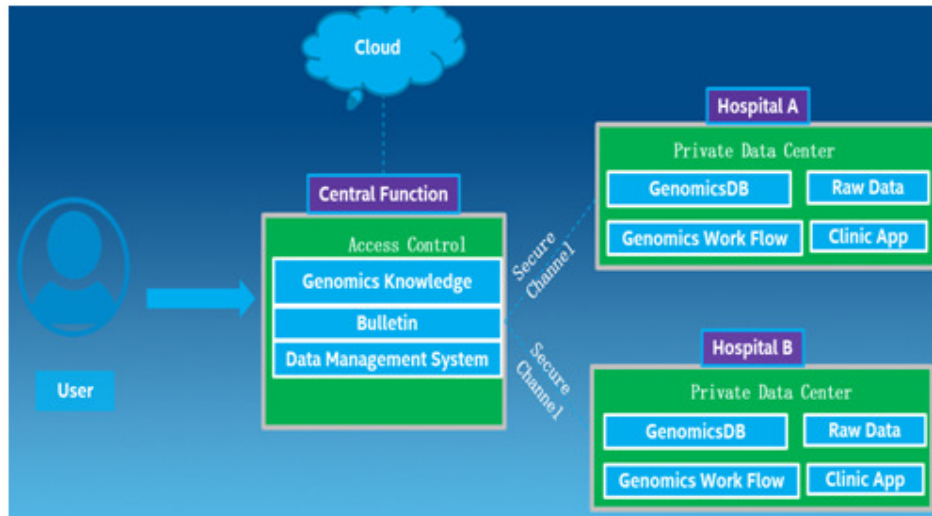Figure 2. Genomics Knowledge Share Model

Figure 3. Genomics Knowledge Share Architecture

In Figure 2 and Figure 3, it shows key usage model and architecture of Genomics knowledge share. There is Central Function defined as data share agent. The purpose is for secure and consolidated data share and exchange. In real scenario, the genomics data is relate to privacy, hospitals can't share too much publicly. In this solution, hospitals keeps the raw data in local private data center, Central Function provides statistical and summarized genomics knowledge database share for query. For example, hospital A and B store raw genomics data with genomics work flow and create private GenomicsDB in their private data center. At the same time, hospital A and B can contribute statistical and representative variant call data to Central Function, data management system can create consolidated GenomicsDB knowledge center for share. Then end user can query consolidated GenomicsDB knowledge from both hospital A and B. And the data share is not for raw genomics data but statistical and GenomicsDB knowledge. Other than hospitals' private data center, Central Function can be deployed on public cloud with secured access control for data share.

2.2. Visualized Genomics Work Flow
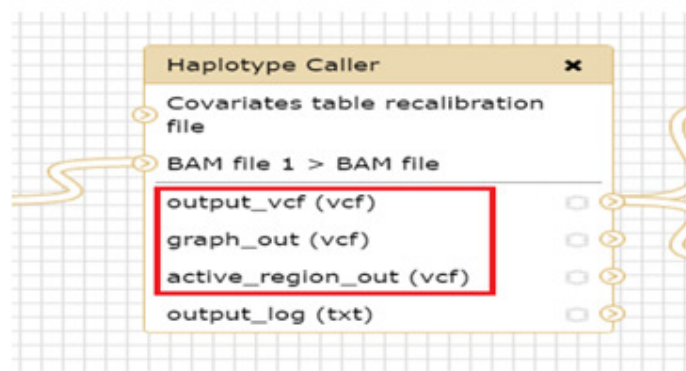
Figure 4. Visualized Genomics Work Flow

In Figure 4, it shows example of visualized genomics work flow, like data convert from fastq to BAM and then create final vcf (variant call format) data. In life science, some bio researchers don't have too much IT background. The visualized UI can help bio researcher much during

customization of genomics data analysis and conversion. These raw data should be put in secure environment like hospital private data center and will be used to create GenomicsDB knowledge.

2.3. GenomicsDB and TileDB
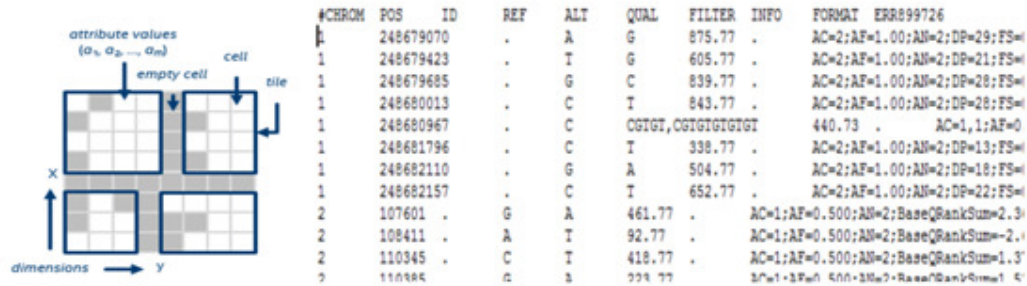
Figure 5. TileDB and GenomicsDB

In Figure 5, it shows TileDB work model. TileDB is a system for efficiently storing, querying and accessing sparse array data. It is optimized for sparse data and supports high performance linear algebra. For example, when storing data and querying cell, TileDB skips the empty cell to save much storage and query time. The GenomicsDB is instance of TileDB, which stores variant data in a 2D TileDB array. Each row corresponds to sample in a vcf and each column corresponds to a genomic position. Figure 5 also shows example of discrete genomics position data.
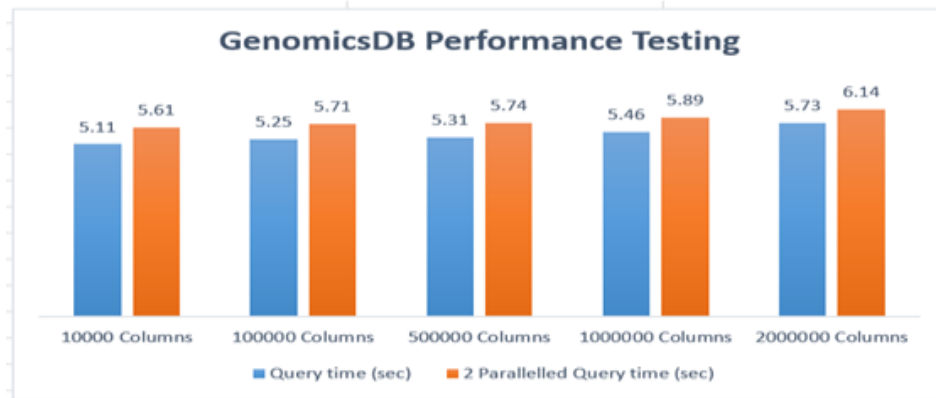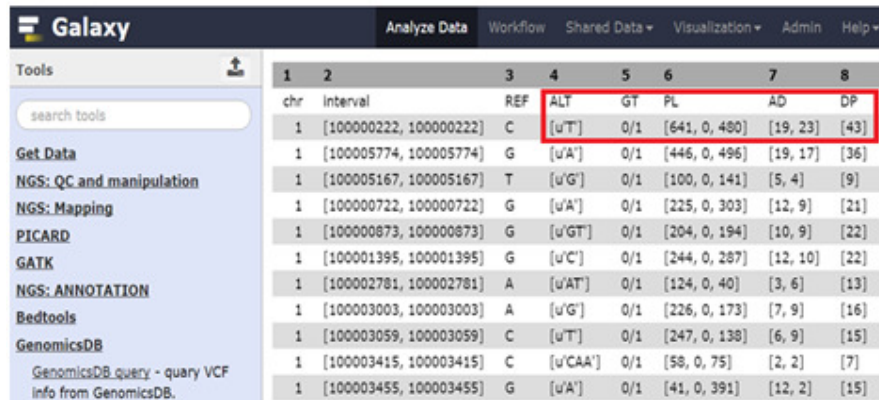
Figure 6. GenomicsDB Performance Report

Figure 6 shows real GenomicsDB testing report from shanghai children hospital with 11G sample vcf. It takes seconds time to response user and shows better performance when doing paralleled query and scaling to millions of variant data column range.

Figure 7. GenomicsDB UI

Figure 8 GenomicsDB UI Testing Report

Since some of bio researchers don't have much IT background, this solution provides very friendly web UI to support the query by end user. In Figure 7, it shows the convenient UI that has been landed in shanghai children hospital. In Figure 8, it shows the GenomicsDB UI testing report. Although it takes longer time to UI parse, render as well as annotation process etc, it still takes only seconds time to response user and shows great performance when scaling to bigger data with bigger column rage.

2.4. Lustre Big Data File System

Lustre file system is an open-source, parallel file system that supports many requirements of leadership class HPC simulation environments. Lustre is designed as high density file system, which utilizes physical storage as few as possible. It is easily scaled to bigger data in life science and supports standard file system behavior without additional protocol support. So it can be seamlessly integrated with legacy ecosystem. In Figure 9, it shows architecture of Lustre file system. The MGS (Management Servers) controls and manages entire Lustre file system. The MDS (Metadata Servers) manages the Meta data of multiple storages such as storage distribution, coordinators for data movers etc. The OSS (Object Storage Servers) manages the detailed storage devices. The Lustre Client is user interface, which acts as "normal" file system with POSIX interface. Once the client machine mounts Lustre file system, it can be treated as Lustre Client and Lustre details is transparent to end user.
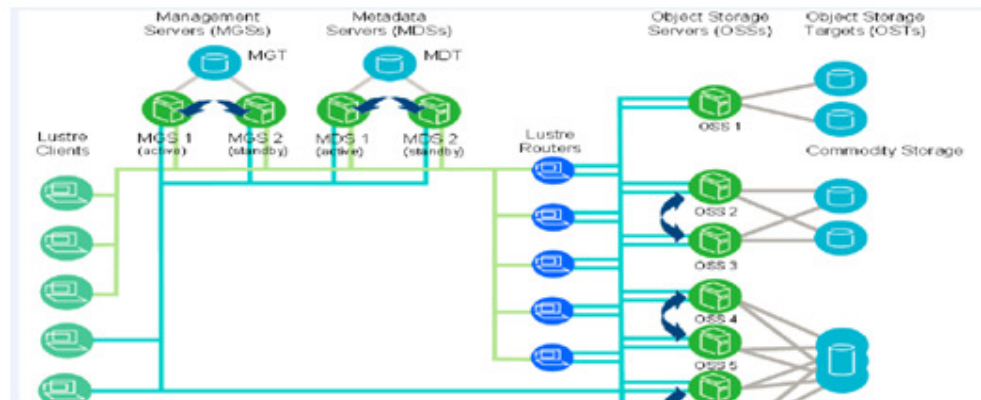
Figure 9. Lustre Architecture

2.5. Lustre HSM (Hierarchal Storage Management) Architecture
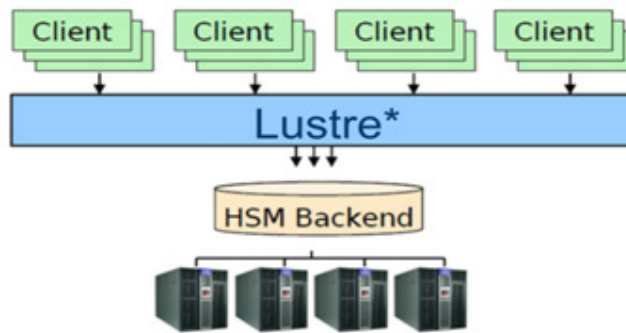
Figure 10. Lustre HSM Architecture
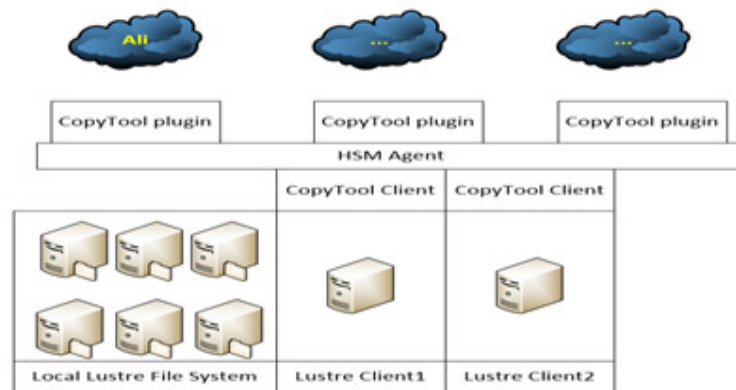
Figure 11. Scale by Plugin

Figure 10 shows Lustre HSM (Hierarchal Storage Management) Architecture, which supports bigger and bigger data. In this solution, hot data will be put on local Lustre and cold data will be put on cloud. HSM Agent controls the interface between different cloud services by plugin mode. And HSM Agent manages the file status system automatically, which makes correct action and status update be feasible and avoids incorrect file action that might causes to file missing.

Furthermore, it supports parallelized upload and download with many Lustre clients. So as long as there is enough network bandwidth, the bandwidth of file data exchange between local Lustre and remote cloud is unlimited.

It can be scaled easily to new cloud services by creating plugin mode. For bio researchers, they don't need to get deep knowledge about how Lustre works. They just need to know the interface provided by HSM Agent and provide copy tool plugin then the tool can be scaled to more cloud services. See Figure 11.
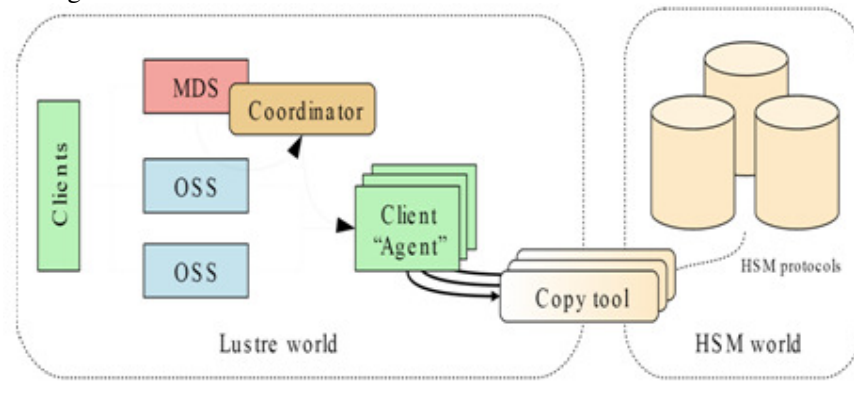
Figure 12. HSM Work Model

Figure 12 shows how Lustre HSM works. It utilizes Copytool as interface for data exchanging between local Lustre system and remote. The end users send request from copy tool client, then backend agent will work with Lustre internal coordinator to do the data exchange from/to remote cloud services by different HSM protocols.
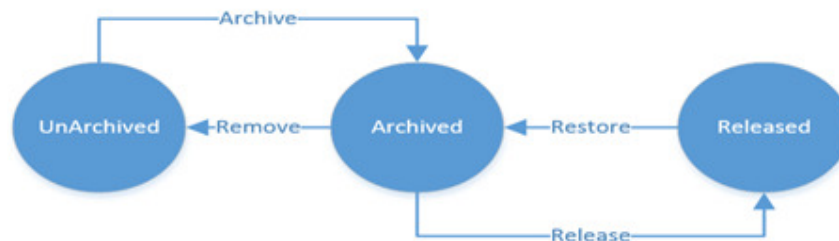
Figure 13. Standard Lustre Status Machine

Figure 13 shows standard Lustre file status machine

- Archive: Copy data from a Lustre file system file into the HSM solution.

- Release: Remove file data from the Lustre file system.

- Restore: Copy back data from the HSM solution into the corresponding Lustre file system file.

- Remove: Delete the copy of the data from the HSM solution.

Figure 14. Nature Language Interface

The solution supports nature language interface, which can make end users who don't have IT background be more comfortable. As it has been released as open source project, it can be easily to scale to more language support such as Chinese, Japanese, Korean, Spanish, Portuguese, and Russian etc. Figure 14 shows the example of nature language interface. For example, background part is used for checking runtime environment like runtime user, Lustre file system mount, HSM agent start etc. The scenario part is used for data exchange between local Lustre file system and remote cloud services.

3. CONCLUSIONS

As big data era of genomics life science industry is coming, previous traditional way can't fulfil the data growing request. By using Intel Big Data Architecture with GenomicsDB and Lustre, genomics data can be stored and utilized with optimized way. Central Function with genomics knowledge share will be scaled as future commercial standard. It can definitely facilitate life science research and accelerate the genomics precision medicine.

REFERENCES

[1]   Robert Read, Lustre file system for cloud and Hadoop, pp.1-25.
       https://www.openfabrics.org/images/eventpresos/workshops2015/UGWorkshop/Thursday/thursday_1
       3.pdf

[2]   Karthik Gururaj & Kushal Datta, GenomicsDB. https://github.com/Intel-HLS/GenomicsDB

[3]   Brian Cremeans & Kushal Datta & Karthik Gururaj & Samuel Madden & Timothy Mattson &
       Mishali Naik & Paolo Narvaez & Stavros Papadopoulos & Jagannath Premkumar, TileDB. http://istc-
       bigdata.org/tiledb/tutorials/index.html

[4]   Intel Stands at Center of Huge Health Data Exchange Project, https://www.premisehealth.com/intel-data-exchange-project/

[5]   Tim Mattson Intel labs, Polystore, Julia, and produc3vity in a Big Data world, pp. 1-38. http://www.clsac.org/uploads/5/0/6/3/50633811/mattson.pdf

[6]   Lustre Software Release 2.x Operations Manual, Part III. Administering Lustre, pp. 172-178. http://doc.lustre.org/lustre_manual.pdf

[7]   Jennie Duggan Northwestern U. & Aaron Elmore U of Chicago & Tim Kraska Brown U. & Sam Madden M.I.T. & Tim Mattson Intel Corp. & Michael Stonebraker M.I.T., The BigDawg Architecture and Reference Implementation, pp. 1-2. http://users.eecs.northwestern.edu/~jennie/research/BigDawgShort.pdf

[8]   A. Elmore Univ. of Chicago & J. Duggan Northwestern & M. Stonebraker MIT & M. Balazinska Univ. of Wash. & U. Cetintemel Brown & V. Gadepally MIT-LL & J. Heer Univ. of Wash. & B. Howe Univ. of Wash. & J. Kepner MIT-LL & T. Kraska Brown & S. Madden MIT & D. Maier Portland St U. & T. Mattson Intel & S.Papadopoulos Intel / MIT & J. Parkhurst Intel & N. Tatbul Intel / MIT & M. Vartak MIT & S. Zdonik Brown, A Demonstration of the BigDAWG Polystore System, pp. 1-4. http://livinglab.mit.edu/wp-content/uploads/2016/01/bigdawg-polystore-system.pdf

[9]   Usegalaxy, https://wiki.galaxyproject.org/Admin/GetGalaxy

AUTHORS

HAO LI

Hao is software architect in Intel Datacenter Health and Life Science Group. Hao joined Intel in 2005 and worked in many Intel internal and external open source product design and development such as Lustre HSM, GenomicsDB, Tizen, MeeGo etc. Hao received Master Degree of Computer Science from Fudan University in 2003.

INTENTIONAL BLANK

OPENSKIMR A JOB- AND LEARNING-PLATFORM

Andreas Kofler[1] and Marianne Prast[2]

[1,2]Management Center Innsbruck (MCI), Innsbruck, Austria
`andreas.kofler@mci.edu, marianne.prast@mci.edu`

ABSTRACT

This paper is concerned with the mathematical aspects of the development of the job- and learningplatform OPENSKIMR. The platform should enable users to be matched with jobs based on their individual skill profile and the job skill requirements. Further, once users have chosen jobs they like, possible learnings in order to better fit the job's requirements are recommended. We give a short introduction to the data model we use and show the mathematical framework we act within. Further, we present our Route Planner Algorithm and discuss its functionality.

KEYWORDS

OpenSkiMr, knowledge engineering, data mining, digitalization, e-learning, learning roadmap, matching, recommender systems

1. INTRODUCTION

OPENSKIMR (Open European Skill Match Maker) is a project funded by the European Union. The goal is to develop a matching system between users, jobs and education. In doing so, the platform should help stabilize the European labor market (especially for young people), foster lifelong learning and vocational education and enhance geographic mobility within Europe. The philosophy standing behind OPENSKIMR is the picture of a route. Users should be able to find new ways, new goals as well as get information on how to reach them.

The platform should perform the following tasks:

- *Submitting skills or dream job:* OPENSKIMR offers two different ways to submit personal information that builds the basis for the subsequent matching. The user can either sketch a personal skill profile or submit a dream job.

- *Matching the users according to their skill profile with occupations and jobs:* Skill mismatch is a complex problem which does affect not only those who are looking for a job but most of the workforce. The phenomenon of skill mismatch involves underskilling as well as overskilling. Both can undermine the long-term potential of the workforce [1].

- *Suggesting an education route*: The user receives recommendations for learnings which it might be interested in in order to better fit a job's requirements

2. DATA MODEL

The functionality of OPENSKIMR is based on ESCO (European Skills, Competences and Occupations) catalogue. The outcome is a standardized data set with a common understanding of occupations and their related skills and knowledge, as well as the relationships between these concepts. It is part of the Europe 2020 strategy [2].

2.1. NOTATION

Let $S = \{s_1, \ldots, s_n\}$ denote the set of skills and tools and $I = \{1, 2, \ldots, n\}$ the set of skills and tools indices. From a semantic point of view, it is necessary to distinguish between skills and tools. Skills such as *programming* or *data mining* refer to the ability to apply knowledge and use know-how in a certain area, while so-called skill applications or tools denote, for example, concrete programming languages such as C++ or Java. Therefore, the set S is given as $S = S_{skills} \cup S_{tools}$ with $n_{skills} + n_{tools} = n$ and $I = I_{skills} \cup I_{tools}$.

By the italic letter O we denote the set of occupations with cardinality N_O. The *ESCO catalogue* delivers information about occupations and their related skills. In particular, it is stated which skills are required by the occupations. Therefore, we can identify each occupation $o \in O$ with its binary vector representation $\mathbf{o}$, where $o_i = 1$ denotes that the skill or tool s_i is required and $o_i = 0$ denotes that the skill s_i is not necessary for the occupation o. Note that we define all occupations only by skills and not by tools, i.e. $o_i \in \{0, 1\}$ for $i \in I_{skills}$ and $o_i = 0$ for $i \in I_{tools}$.

The italic letters U and V denote the set of users and jobs on the platform. Again, we denote by N_U and N_V the cardinality of the respective sets. Note that, while the sets U and V are sets whose cardinality and elements vary in time when users and jobs are added to or deleted from the platform or user profiles change due to qualification acquirements, the cardinality and the elements of the set O are constant over time.

Finally, let $u \in U$ denote a user and let $\mathbf{u}$ denote its respective vector representation where each entry u_i of the vector corresponds to the level of knowledge of the user u with respect to the skill or tool s_i. Analogously, a job is denoted by $v \in V$ with vector representation $\mathbf{v}$ where each entry v_i corresponds to the level of knowledge with respect to the skill or tool s_i required for the job. A user u builds up its profile by rating a skill or a tool s_i with a score from 1 (low) to 5 (expert). If a user does not deliver any information about its knowledge with respect to the skill s_i than we set $u_i = 0$. By $R = \{0, 1, 2, 3, 4, 5\}$ we denote the set of possible skill or tool ratings for jobs and users. The choice of the set R relies both on usability reasons as well as on the outlook of a possbile implementation of recommender systems where we might benefit from the exhaustive research that has been conducted in the last years on the basis of the Netflix data set [3]. Recommender systems might be applied to recommend users skills they might be interested to rate.

For a user $u \in U$ we denote by $I_u \subset I$ the set of skills and tools indices it has rated. For a job $v \in V$ and an occupation $o \in O$ the sets $I_v, I_o \subset I$ are analogously defined and determine the job's and occupation's skills requirements, respectively.

As the total number of skills and tools is expected to be quite high ($n > 12000$) we can consider users as well as jobs and occupations to be represented by sparse vectors in a n-dimensional space with $\mathbf{u}, \mathbf{v} \in R^n$ and $\mathbf{o} \in \{0,1\}^n$.

3. SKILLS AND TOOLS ASSESSMENT

Since we base the matching of users and job profiles on the skills and tools required, the first problem we are faced with is to offer the user an accessible way to build its profile by rating some skills and tools.

On the one hand, a user may have a quite clear idea of the area it wants to work in and therefore might rate skills which are related to certain occupations. On the other hand, a user may not have a specific choice on the area but would rather like to get an overview of jobs it would fit to. In this case the user might be interested in rating its skills by choosing them from some skills clusters. When a user rates its skills by browsing occupations, it finds tools when they are related to skills. For example, when a user states that it has knowledges about programming, the user has the possibility to rate the programmning languages C++ or Java. Recruiters are treated the same as users when they create a job advertisement.

3.2. CLUSTERING OF SKILLS AND OCCUPATIONS

We build the occupations clusters as well as the skills clusters from which users may choose and rate their skills using an agglomerative hierarchical clustering algorithm.

An Agglomerative hierarchical clustering algorithm requires an initial dataset $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_O}\}$, where each data point $\mathbf{x}_k$ is considered to be a single cluster. Then, the algorithm repeats merging the two closest clusters according to some similarity measure between clusters. The procedure is repeated until all data points are contained in one single cluster. The output of the algorithm is a sequence of nested partitions of the dataset.

Let $\mathbf{O}$ denote the binary $N_O \times n$ matrix whose rows $\mathbf{o}_1, \ldots, \mathbf{o}_{N_O}$ correspond to the binary occupations vectors. Analogously, the rows of the transposed matrix $\mathbf{O}^T$ contain the binary skills vectors whose entries define by which occupations the skills are required.

Let C_i and C_j denote two non-empty, non-overlapping clusters of data points which might represent two clusters of occupations or skills. We merge occupations clusters as well as skill clusters according to the complete link method, i.e. the distance measure $D(\cdot, \cdot)$ between the clusters C_i and C_j is given by

$$D(C_i, C_j) = \max_{\mathbf{x} \in C_i, \ \mathbf{y} \in C_j} d(\mathbf{x}, \mathbf{y}) \tag{1}$$

where $d(\cdot, \cdot)$ is given by the cosine distance measure

$$d(\mathbf{x}, \mathbf{y}) = 1 - \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \tag{2}$$

Note that, though $d(\cdot,\cdot)$ is not a proper distance measure since it does not fulfill the triangle inequality property, the similarity measure given by $c(\mathbf{x},\mathbf{y}) = 1 - d(\mathbf{x},\mathbf{y})$ is a commonly used measure to compute the similarity between points, especially when the data one handles can be represented as binary data, e.g. text documents [4]. Furthermore, the computation of the similarity measure between all the occupations vectors and the skills vectors is extremely efficient as it reduces to the multiplication $\mathbf{O}_{\text{norm}}\mathbf{O}_{\text{norm}}^T$ and $\mathbf{O}_{\text{norm}}^T\mathbf{O}_{\text{norm}}$, where $\mathbf{O}$ is sparse.

Since we are going to use the cosine similarity later within the paper, we denote the cosine similarity measure by $c(\mathbf{x},\mathbf{y}) = 1 - d(\mathbf{x},\mathbf{y})$. We apply the clustering algorithm to the set of occupations with initial set $\mathcal{D}_{\text{occupations}} = \{\mathbf{o}_1,\ldots,\mathbf{o}_{N_O}\} \in \{0,1\}^n$ and to the set of skills $\mathcal{D}_{\text{skills}} = \{\mathbf{o}_1^T,\ldots,\mathbf{o}_{n_{\text{skills}}}^T\} \in \{0,1\}^{N_O}$. Be aware that tools are not related to occupations and therefore are not clustered. With the help of the constructed set of clusters we can decide how we group the data and approximately how large the clusters shall be in order to guarantee best usability. We use the output of the clustering algorithm as an initial suggestion to group occupations and skills. Still, the correctness of the clustering from a semantic point of view is checked by hand and the clusters are given an appropriate name.

4. FUNCTIONALITY OF OPENSKIMR

Within the mathematical framework just presented, the process of matching a user u and a job v corresponds to the calculation of a similarity measure between their vectors $\mathbf{u}$ and $\mathbf{v}$ which expresses how the user's skills profile satisfies the job's requirements.

4. 1 MATCHING FUNCTION

In order to measure how good the user u fits the job v, we are looking for a function $r_{\mathbf{v}} : R^n \rightarrow [0,1]$ for each job $v \in \mathcal{V}$ which is dependent on the job's requirements given by the vector $\mathbf{v}$. Thereby, $r_{\mathbf{v}}(\mathbf{u}) = 0$ should denote a complete mismatch of the user u with respect to the job v and $r_{\mathbf{v}}(\mathbf{u}) = 1$ should indicate that the user u would be the perfect candidate.

As often used for job advertisements, a job v might require some skills and tools to be essential in order to successfully perform the job related tasks and others to be optional. Therefore, let us assume that for every job $v \in \mathcal{V}$ we are given information about a skill or a tool to be essential or optional and let the sets $I_{v,\text{essential}} \subset I$ and $I_{v,\text{optional}} \subset I$ denote the indices of the essential and the optional skills and tools, respectively. When a job profile is created and a skill or a tool is declared to be only optional for the job v, recruiters are not able to rate the skill s_i, i.e. $v_i \in \{0,1\}$ for $i \in I_{v,\text{optional}}$ which means that the type of information is binary as for occupations.

Let $s : R^2 \rightarrow [0,1]$ be a function with $s(x,y) = 1$ whenever $x = y$ and $s(x,y) < 1$ otherwise. Then, for a fix index set $\varnothing \neq J \subset I$ we define the function $q_{s,J} : R^n \times R^n \rightarrow [0,1]$

$$q_{s,J}(\mathbf{x},\mathbf{y}) = \frac{1}{|J|}\sum_{i \in J}\text{sign}(x_i)\, s(x_i,y_i). \tag{3}$$

Finally, for a given job v with vector representation $\mathbf{v}$ and related information about which skills and tools are essential and which are optional, we define the matching function $r_{\mathbf{v}}$ by

$$r_{\mathbf{v},\lambda_1,\lambda_2}(\mathbf{u}) = \lambda_1 \left(q_{s,I_{v,\text{essential}}}(\mathbf{u},\mathbf{v}) \right) + \lambda_2 \left(\frac{|I_u \cap I_{v,\text{optional}}|}{|I_{v,\text{optional}}|} \right) \tag{4}$$

with parameters $\lambda_1,\lambda_2 > 0$ with $\lambda_1 + \lambda_2 = 1$. Here, we assumed that $I_{v,\text{optional}} \neq \varnothing$. Otherwise, the second term is set to zero and $\lambda_1 = 1$. We measure the matching with respect to the essential and optional requirements separately and use the parameters λ_1 and λ_2 to weight them differently. Typically, a parameter setting with $\lambda_1 > \lambda_2$ will be chosen in order to give the matching score regarding the essential skills and tools higher importance.

4.2. MATCHING USERS WITH JOBS AND OCCUPATIONS

Once a user u has completed the skills rating process, it can search for jobs or occupations it would fit to. Since an occupation vector $\mathbf{o}$ is a binary vector, we define the vector $\bar{\mathbf{o}}$ as a representative average of the jobs related to the occupation $o \in O$. As we do not define an occupation by concrete tools as they are rather jobs related than occupation related, we restrict the calculation of the mean to the indices which refer to the skills. For an index set $J \subset I$ we define the restriction of a vector $\mathbf{x} \in R^n$ to the entries with indices in the set J by

$$\mathbf{x}_J = (x_{j_1},\ldots,x_{j_{n_J}})^T \in R^{n_J} \tag{5}$$

where $n_J = |J|$ denotes the cardinality of the index set J. Then, we calculate the restriction of the vector $\bar{\mathbf{o}}$ to the entries indexed by the elements of I_o by

$$\bar{\mathbf{o}}_{I_o} = \left[\frac{1}{|\mathcal{N}(\mathbf{o})|} \sum_{\mathbf{v} \in \mathcal{N}(\mathbf{o})} \mathbf{v}_{I_o} \right] \tag{6}$$

where $[\,\cdot\,]$ denotes the function which rounds each entry of the vector to its nearest integer. The set $\mathcal{N}(\mathbf{o})$ denotes the set of job vectors whose binary vector representations are most similar to $\mathbf{o}$ according to the cosine similarity measure $c(\cdot,\cdot)$, i.e

$$\mathcal{N}(\mathbf{o}) = \{\mathbf{v} \mid \mathbf{o} = \underset{\tilde{\mathbf{o}} \in \mathcal{D}_{\text{occupations}}}{\arg\max}\ c(\tilde{\mathbf{o}}_{I_o}, \text{Sign}(\mathbf{v}_{I_{\tilde{\mathbf{o}}}}))\}. \tag{7}$$

where, in this context, the function Sign shall denote the componentwise application of the signum function. For all $i \notin I_o$ it holds $\bar{o}_i = 0$. Let $\bar{O}$ denote the set of the average of the jobs whose vector representations are calculated as in 6. Then, for each $\bar{o} \in \bar{O}$ we can match users with occupations by calculating $r_{\bar{\mathbf{o}}}(\mathbf{u})$. Note that in the calculation no distinction is made between the set $I_{\bar{o},\text{essential}}$ and $I_{\bar{o},\text{optional}}$ since occupations do not distinguish between essential and optional skills. Therefore, the matching of a user u with an occupation described by $\bar{\mathbf{o}}$ reduces to calculating

$$r_{\bar{\mathbf{o}}}(\mathbf{u}) = q_{s,I_{\bar{o}}}(\mathbf{u},\bar{\mathbf{o}}). \tag{8}$$

We can calculate $r_{\bar{o}}(\mathbf{u})$ for all occupations $\bar{o} \in \bar{O}$ and recommend the user u the K_1 best occupations where the matching scores are the highest. This recommendation only serves as an initial suggestion in order to show users which occupations they fit to. Then, the user is able to select one of the shown occupations. For the chosen occupation we compute the matching scores $r_{\mathbf{v},\lambda_1,\lambda_2}(\mathbf{u})$ for all jobs $v \in \mathcal{N}(\bar{o})$ and show the user u the K_2 jobs with the highest matching score.

4.3. ROUTE PLANNER ALGORITHM

Once a user u has chosen a concrete job v it is able to see its matching score $r_{\mathbf{v}}(\mathbf{u})$. As a next step, we would like to recommend the user content and learnings which increase its level of knowledge regarding several skills and tools required by the job.

4.3.1. DEFINITION OF A LEARNING

Formally speaking, a learning or a training transforms a user u into a user $\tilde{u}$ with different skills profile. Due to the type of learning or training the user u attends, different type of knowledge is gained. This is expressed by the fact that the user's vector representation $\mathbf{u}$ turns into $\tilde{\mathbf{u}}$ where the entries which indeed change their value are the ones who denote the skills indices of the skills which the learning or training is focused on.

Note that, due to the requirements we set for a matching function $r_{\mathbf{v}}$ for a job $v \in \mathcal{V}$, there is no guarantee that a learning actually increases the matching score of a user u with respect to the job v. Consider the case, for example, that a job v denotes a junior software developer job role which only requires moderate qualifications for certain skills. If a learning or a training turns the user into an expert with respect to those skills it will be overqualified and get a lower contribute for the rating regarding those skills or tools. We assume that in general a learning might cover several topics and therefore might increase your knowledge in more than one single skill or tool.

We denote a learning or a training by the capital letter L and the set of learnings and trainings by $\mathcal{L}$.

Let $\varnothing \neq S_L \subset S$ be a subset of skills or tools with indices in an index set I_L and $R_L := R_{L,1} \times R_{L,2} \subset R^2$ be a subset of rating pairs with

$$\forall i \in I_L : \forall \left(r_1(i), r_2(i)\right) \in R_L : r_1(i) < r_2(i). \tag{9}$$

The set S_L is the set of skills which are related to the learning L and the set R_L can be seen as a set of rating pairs where the first rating denotes the minimal knowledge a user should have with respect to a certain skill in order to be able to participate in the course and the second entry denotes the level of knowledge a user reaches when it succesfully completes the learning. Given the two sets I_L and R_L we define a learning L by the set

$$L = I_L \times R_L. \tag{10}$$

Therefore, the learning L can intuitevely be seen as a set of triples $(i, r_1(i), r_2(i))$ where i is the index of the skill or tool in which a user increases its knowledge from level $r_1(i)$ to level $r_2(i)$ by attending the learning.

Further, for each learning L we define a learning function

$$l_L : R^n \longrightarrow R^n,$$

$$\mathbf{u} \longmapsto \tilde{\mathbf{u}} = \begin{cases} \mathbf{u}, & \exists\, i \in I_L : u_i < r_1(i), \\ t_L(\mathbf{u}), & \text{else} \end{cases} \tag{11}$$

where $t_L(\mathbf{u}) = \left(t_L^{(1)}(u_1), \ldots, t_L^{(n)}(u_n) \right)$ and

$$t_L^{(i)}(u_i) = \begin{cases} r_2(i), & i \in I_L,\ u_i \le r_2(i) \text{ for } r_2(i) \in R_{L,2}, \\ u_i, & \text{else .} \end{cases} \tag{12}$$

The function l_L acts on the vector representation $\mathbf{u}$ of the user and maps it to a vector $\tilde{\mathbf{u}}$. Its entries which correspond to the skills and tools which are related to the learning L are updated if the user is qualified to attend the learning. Entries which correspond to skills and tools not related to the learning L are left unchanged.

4.3.2. CONSTRUCTION OF A LEARNING ROUTE

The function l_L describes how the vector representation of a user profile is updated when a learning or a training L is successfully completed by the user. Of course, a user might be interested in attending more than one learning in order to increase its matching score with respect to a job v.

Therefore, we define a learning route as a finite sequence of learnings $(L_q)_{q=1}^N$ with $L_q \in L$ and L^N as the set of all sequences of learnings of length N. Let N_L denote the number of learnings and trainings. The best updated profile a user u with respect to a job v is given by

$$\mathbf{u}_v^* = \underset{N \in \{1, \ldots N_L\}\, (L_q)_{q=1}^N \in L^N}{\arg\max}\ \arg\max\ r_v\left(\left(l_{L_1} \circ \ldots \circ l_{L_N} \right)(\mathbf{u}) \right). \tag{13}$$

Of course, the calculation of $\mathbf{u}_v^*$, as just defined, is a computationally too expensive is not applicable. We now discuss our approach in order to still be able to recommend a the user a proper learning route.

We formulate the problem of finding an optimal learning route for a user u with respect to a job v by finding the shortest path of a graph $G = (V, E)$ with initial node u and destination node $\tilde{u}$ for some node $\tilde{u}$ with $r_v(\tilde{u})$ as close as possible to one.

We start with a graph $G = (V, E)$ with $V = \{u\}$ and $E = \varnothing$ where u represents the initial user. Let l_L be a learning function for a given learning $L \subset L$ and let u_L denote the user after it has successfully attended the learning L, i.e. $\mathbf{u}_L = l_L(\mathbf{u})$. Then, if $r_v(l_L(\mathbf{u})) > (1+\beta) r_v(\mathbf{u})$ for a given matching function r_v and $\beta \ge 0$, we add the edge $e_{u,u_L} = (u, u_L)$ to the set of edges E and we add the node u_L to the set V. This step can now be repated by applying further learnings to the user profile u. After having added additional nodes and edges to the respective sets E and V we repeat the procedure by applying learnings to the nodes which represent already updated user profiles.

Algorithm (2) describes the construction of the graph. Note that we might weight an edge (u, u_L) using any numerical attributes related to the learning L where there exists a order relation such as

---

Algorithm 1 Route Planner Algorithm

Require: $u, v, r_{v, \lambda_1 \lambda_2}, L, l_L, d_{max}, \alpha_1, \alpha_2, \beta$

    $m = 0$

    $V_{seen} = \emptyset$

    $V^{(m)} = \emptyset$

    $V^{(m+1)} = \{u\}$,

    $E = \emptyset$,

    Pre-process the learnings with $\alpha_1, \alpha_2 > 0$

    $\mathcal{L}_{\alpha_1} = \{L \in \mathcal{L} \mid c((x_L, H_0(v - u)) \geq \alpha_1\}$

    $\mathcal{L}_{\alpha_2} = \{L \in \mathcal{L} \mid 1 - p_{x_{R_{1,2}}}(v) \geq \alpha_2\}$

    $\mathcal{L}_{\alpha_1, \alpha_2} \leftarrow \mathcal{L}_{\alpha_1} \cap \mathcal{L}_{\alpha_2}$

    while $V^{(m+1)} \setminus V^{(m)} \neq \emptyset \land m < d_{max}$ **do**

        $V^{(m)} \leftarrow V^{(m+1)}$

        $V_{temporary} \leftarrow V^{(m+1)}$

        for $\tilde{u} \in V^{(m+1)} \setminus V_{seen}$ **do**

            for $L \in \mathcal{L}_{\alpha_1, \alpha_2}$ **do**

                $\tilde{u}_L \leftarrow l_L(\tilde{u})$

                if $r_{v, \lambda_1 \lambda_2}(\tilde{u}_L) > (1 + \beta) r_{v \lambda_1 \lambda_2}(\tilde{u})$ **then**

                    $V^{(m+1)} \leftarrow V^{(m+1)} \cup \{\tilde{u}_L\}$

                    $V_{temporary} \leftarrow V_{temporary} \cup \{\tilde{u}_L\}$

                    $E \leftarrow E \cup \{(u, \tilde{u}_L)\}$

                end if

            end for

            $V_{seen} \leftarrow V_{seen} \cup \{\tilde{u}_L\}$

        end for

        $V^{(m+1)} \leftarrow V_{temporary}$

        $m \leftarrow m + 1$

    end while

    $V \leftarrow V^{(m)}$

    $G \leftarrow (V, E)$

    $u^*_{v, d_{max}} \leftarrow \arg\max_{\tilde{u} \in V} r_v(\tilde{u})$

    Apply Dijkstra Algorithm with initial node u and destination node $u^*_{v, d_{max}}$

Ensure: $G = (V, E)$, shortest path and $r_v(u^*_{v, d_{max}})$

---

price of the learning or the duration of the learning. Then, applying Dijkstra's algorithm [5] we find the shortest path from the initial node u to the node which has the best matching score with respect to the job v.

Also note that we need to pre-process the learnings we take into consideration for the calculation as an initial step. This step is crucial in order to save computational time.

Let the vector $\mathbf{x}_L \in \{0,1\}^n$ be the binary learning vector with $x_{L,i} = 1$ if $i \in I_L$ and $x_{L,i} = 0$ otherwise. Then, let $H_0(\mathbf{v} - \mathbf{u})$ denote the binary vector whose i-th component is equal to one if the user has a knowledge gap with respect to the i-th skill, where H_0 denotes the componentwise application of the Heaviside function with $H_0(\mathbf{x}) = (h_0(x_1), \ldots, h_0(x_n))$ and $h_0(x) = 1$ for $x > 0$ and $h_0(x) = 0$ for $x \leq 0$. By applying only learnings where the cosine of the angle between a learning vector $\mathbf{x}_L$ and the user's skill gaps vector $H_0(\mathbf{v} - \mathbf{u})$ exceeds a chosen threshold $\alpha_1 > 0$, we avoid considering learnings which are not related to the skills or tools where the user has a gap. By doing so, we mainly focus on the learnings which cover topics needed for the job v. Further, the computation of the angle between the learnings and the skill gap vector can performed again extremely efficiently since it corresponds to matrix-vector multiplication. After the computation has been performed, we define the set of learnings $\mathcal{L}_{\alpha_1}$ by

$$\mathcal{L}_{\alpha_1} = \left\{ L \in \mathcal{L} \mid c\big((\mathbf{x}_L, H_0(\mathbf{v} - \mathbf{u})\big) \geq \alpha_1 \right\} \tag{14}$$

The set $\mathcal{L}_{\alpha_1}$ mainly contains learnings which are related to skills where the user has indeed deficits. The second pre-processing step ensures that the skill goals a user u reaches by attending a learning L are as close as possible to the skills the job v requires. Thus, analogously as for the matching function $r_\mathbf{v}$, for a learning $L \in \mathcal{L}$ and respective *rating-to* vector $\mathbf{x}_{R_{L,2}}$ given by

$$x_{R_{L,2},i} = \begin{cases} r_2(i), & \text{if } i \in I_L \\ 0, & \text{else}, \end{cases}$$

we define a penalty function $p_{\mathbf{x}_{R_{L,2}}}$ by

$$p_{\mathbf{x}_{R_{L,2}}}(\mathbf{v}) = \frac{1}{5 n_{v,L}} \sum_{i \in I_v \cap I_L} \left| v_i - x_{R_{L,2},i} \right|,$$

where $n_{v,L} = |I_v \cap I_L|$. For a learning L with vector $\mathbf{x}_{R_{L,2}}$ and a job vector $\mathbf{v}$, it holds $p_\mathbf{v}(\mathbf{x}_{R_{L,2}}) \in [0,1]$. The smaller the value, the nearer the vector $\mathbf{x}_{R_{L,2}}$ is to the job's requirements given by $\mathbf{v}$. Given a job and the penalty function $p_{\mathbf{x}_{R_{L,2}}}$ we define the subset of learnings $\mathcal{L}_{\alpha_2} \subset \mathcal{L}$ as

$$\mathcal{L}_{\alpha_2} = \left\{ L \in \mathcal{L} \mid 1 - p_{\mathbf{x}_{R_{L,2}}}(\mathbf{v}) \geq \alpha_2 \right\} \tag{15}$$

Further, we allow the graph only to reach a maximal length $d_{\max}$ for a learning route.

All in all, for $\mathcal{L}_{\alpha_1}$ and $\mathcal{L}_{\alpha_2}$ as described in (14) and (15) for $\alpha_1, \alpha_2 \in (0,1)$, we set the the pre-processed learning set $\mathcal{L}_{\alpha_1,\alpha_2}$ as

$$\mathcal{L}_{\alpha_1,\alpha_2} = \mathcal{L}_{\alpha_1} \cap \mathcal{L}_{\alpha_2}. \tag{16}$$

For the case that $\mathcal{L}_{\alpha_1} \cap \mathcal{L}_{\alpha_2} = \emptyset$, we might choose less restrictive parameters α_1 and α_2.

5. TESTING

To give an example, we create a sample user u with vector representation $\mathbf{u} = (1,\ldots,1)^T \in R^{15}$ which chooses to receive a recommendation for a learning path with respect to a job v with

$$\mathbf{v} = (3,0,5,0,5,4,0,5,3,5,1,1,0,1,0)^T \in R^{15},$$
$$I_{v,\text{essential}} = \{1,3,5,6,8,9,10\}, \tag{17}$$
$$I_{v,\text{optional}} = \{11,12,14\}$$

Note that $v_{11} = v_{12} = v_{14} = 1$ does not denote a low requirement of the skills but just that the skills are optionally required and therefore no weight is considered for the computation of the matching score.

The matching function $r_{\mathbf{v},\lambda_1,\lambda_2}$ is chosen according to (4) with

$$s(x,y) = s_\sigma(x,y) = \exp\left\{ -|x-y|^2/\sigma \right\} \tag{18}$$

with parameters $\sigma = 3.6$, $\lambda_1 = 0.8$ and $\lambda_2 = 0.2$ which gives an initial matching score $r_{\mathbf{v},\lambda_1,\lambda_2}(\mathbf{u}) = 0.2900$.

We randomly generate 300 learnings of the form (10) by choosing the skills indices $i \in \{1,\ldots,15\}$ and the values $r_1(i), r_2(i) \in R$ uniformly at random and apply algorithm 2 with $d_{\max} = 5$.

The nearer α is to 1, the more restrictive is the pre-processing and the less learnings we may be able to apply to the user's vector representation $\mathbf{u}$ in order to update its skills profile. With $\alpha = 0$ no pre-prcoessing step is performed and therefore all possible combinations allowed by the parameters setting are considered. The larger the parameter β the higher must be the increase of the matching score of the updated profile in order to add the new node to the graph. A smaller β allows learnings with less high impact to be part of the learning route.

The parameter β also has influence on the number of nodes the graph G might contain. By choosing β very small, every learning which increases the matching score when the profile is updated, is considered to bo a valid learning. By setting a higher β we only add a learning to a path if its impact is strong enough. Therefore, it is also possible to estimate the length of a learning path dependent on the current parameter β and the user vector $\mathbf{u}$. For example, if we had a user vector $\mathbf{u}$ with $r_{\mathbf{v},\lambda_1,\lambda_2}(\mathbf{u}) = 0.5$ for a job vector $\mathbf{v}$ and set $\beta = 0.2$, then, the maximal number of learnings of a recommended learning path cannot exceed 5, regardless of what the parameter $d_{\max}$ was actually chosen to be.

Also, we might consider, to build the learnings route in such a way that the first learnings are the ones which have a higher impact on the matching score and to adapt the parameter β depending on the current learning step matchings score.

Table 2 also shows different results when the route planner algorithm is run with different pre-processing parameters α_1 and fixed $\alpha_2 = 0$ and $\beta = 0.2$, i.e. we only apply the cosine-pre-processing step. The computational times refer to an average over 5 runs with the same initial data. The first column of the table contains the pre-processing parameter α_1. In the second column we see the cardinality of the set $\mathcal{L}_{\alpha_i}$ which corresponds to the number of learnings which are indeed applied to a user profile. The variable r_{max} denotes the highest matching score $r_v(\tilde{u})$ for some updated user profile $\tilde{u}$ which we achieve with the given parameters.

The results motivate our believe that the algorithm should be applicable even with a higher number of learnings

TABLE 2 Variation of α_1 and α_2

α_1	$\mathcal{L}_{\alpha_1}$	r_{max}	time [s]
0	300	0.9723	57.20
0.35	124	0.9723	6.7220
0.375	107	0.9723	2.46
0.4	106	0.9723	0.2.34
0.425	96	0.9723	1.64
0.45	86	0.9723	1.18
0.5	76	0.9445	0.66
0.525	69	0.7912	0.08

α_2	$\mathcal{L}_{\alpha_2}$	r_{max}	time [s]
0	300	0.9723	57.20
0.5	187	0.9723	18.36
0.6	181	0.9723	16.35
0.7	138	0.9445	11.05
0.8	104	0.8039	0.5370
0.9	200	0.6697	0.01

Note that the random learnings are not distributed according to any particular distribution which we surely can expect from real data.

From this simple example we can state that the pre-processing parameters α_1 and α_2 have a crucial effect on the computational complexity of the graph construction algorithm.

We repeat the experiment with the parameter setting $\alpha_1 = 0, \beta = 0.2$ and vary α_2. Table 2 shows the results for the variation of α_2.

6. CONCLUSIONS AND FUTURE WORK

In the last pages we have presented the mathematical framework of the OpenSkiMr project. We have exposed our approach of formulating the problem of matching users and jobs as well as users and occupations. By introducing our route planner algorithm we have shown an intuitive way of building different learning routes which may lead a user to update its skills. The next steps of our project concern a further theoretical study of the algorithm's behaviour with respect to the different parameters as well as experiments envolving different functions s. In particular, the effects of the parameters α_1, α_2 and β should be carefully studied. Eventually, as already mentioned, we also might not choose a global β but let it be dependent on the current learning step.

Also, assuming a very high number of learnings, we may think of clustering the learnings and building the initial set L by adding only some of the learnings per cluster. This avoids the problem of having solely learnings which are similar in terms of skills they are related to after the pre-processing with the parameters α_1 and α_2.

In order to be able to recommend a learning route, we need the content of learnings to be analyzed and adapted to the form our algorithm can work with. Our developing team is engaged with the implementation of the platform whose goal is to collect real data about users as well as job advertisements as soon as possible. We expect our route planner algorithm to work on real data as well as it exploits the natural underlying structure of it.

FUNDING

ACKNOWLEDGEMENTS

REFERENCES

[1]   CEDEFOP (2015). Matching Skills and Jobs in Europe. European Centre for the Development of Vocational Training, Europe 123, 570 01 Thessaloniki (Pylea), Greece, 2nd edition.

[2]   European Commission (2013). European Classification of Skills/Competences, Qualifications and Occupations. European Commission, Publications Office of the European Union: Luxembourg, 1st edition.

[3]   Ricci, F., Rokach, L., and Shapira, B. (2015). Recommender Systems Handbook -Springer, Berlin, Heidelberg, 2nd edition.

[4]   Liu, B. (2011). Web Data Mining - Exploring Hyperlinks, Contents, and Usage Data.Springer Science and Business Media, Berlin Heidelberg.

[5]   Aric A. Hagberg, Daniel A. Schult, P. J. S. (2008). Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference (SciPy2008), pages 1115, Pasadena, CA USA.

[6]   Gan, G., Ma, C., and Wu, J. (2007). Data Clustering - Theory, Algorithms, and Applications.SIAM, Philadelphia.

ENHANCING THE PERFORMANCE OF SENTIMENT ANALYSIS SUPERVISED LEARNING USING SENTIMENTS KEYWORDS BASED TECHNIQUE

Amira Abdelwahab[1], Fahd Alqasemi[2] and Hatem Abdelkader[3]

[1,2,3]Information Systems Department, Menoufia University, Menoufia, Egypt
`1amira.ahmed@ci.menofia.edu.eg`
`2fhdahmd16@yahoo.com`
`3hatem6803@yahoo.com`

ABSTRACT

Sentiment Analysis (SA) and machine learning techniques are collaborating to understand the attitude of text writer, implied in particular text. Although, SA is an important challenging itself, it is very important challenging in Arabic language. In this paper, we are enhancing sentiment analysis in Arabic language. Our approach had begun with special pre-processing steps. Then, we had adopted sentiment keywords co-occurrence measure (SKCM), as an algorithm extracted sentiment-based feature selection method. This feature selection method had utilized on three sentiment corpora using SVM classifier. We compared our approach with some traditional methods, followed by most SA works. The experimental results were very promising for enhancing SA accuracy.

KEYWORDS

sentiment analysis; opinion mining; supervised learning; feature selection; Arabic language

1. INTRODUCTION

Social networks (SN) have changed the nature of communication between people around the world, due to the big and fast tangible ability of sharing information. It also resulted in different SN applications which gain very important popularity. One of SN precious content is opinions that people share about everything: products, sports, people, politics, science ... etc. SN contents are varying, concerning stored data types. However, the most type of information that had shared is text, specially, in the context of opinion expression, since people are expressing their opinions in written words regardless used language. This made text mining importance, so as natural language processing (NLP), and sentiment analysis (SA), regarding Opinion mining (OM), which is another term for sentiment analysis.

Sentiment analysis, where is text mining, machine learning and NLP techniques are collaborating to discover the direction of text segment. This depends on the implication of an attitude behind written text. The text may be a document, review, article or SN post/comment. The discovered direction is called polarity. Text polarity is determined by agree or disagree of text writer with any subject, which text talks about. Agree and disagree is expressed in SA studies as 'positive' and

'negative', respectively. Some papers had added the case of 'neutral' to the above two polarities, which make them three directions [1]. In this work, we had applied our proposed technique only on the two first mentioned polarities, i.e. positive and negative.

SA is achieved using two popular approaches, machine learning sentiment analysis (MLSA), and lexicon-based sentiment analysis (LBSA). MLSA is intended for supervised learning approaches that need pre-annotated text data set, to train and test some selected features. On the other hand, LBSA is considered as unsupervised learning approach, which is based on a pre-made sentiment lexicon. That lexicon is used to count sentiment terms appearance on target text [1] [2].

Sentiment mining in Arabic language is faced numerous difficulties. These difficulties related to some factors; firstly, the nature of Arabic as rich morphological language, and the varying of Arabic dialects which are overlapped with the main form of Modern Arabic Standard, i.e. MSA, especially in SN web pages. Secondly, the rare of researches in these new fields in Arabic. That had led to the lack of resources and tools that may help to create more and big techniques. Although, there were good efforts of Arabic researchers in the NLP, text mining, and sentiment analysis, they need more works.

In this paper, we present an SA approach, that have implemented MLSA on Arabic language, used resources had annotated into the two polarities, 'negative' and 'positive', however. We had exploited LBSA in the purpose of improving MLSA. This had been done by using sentiment lexicon as sentiment keywords list, which implemented as a feature selection method, which helped for increasing the performance of supervised learning task on three corpora. We had compared our approach with other traditional and recent approaches, and the results were very promising.

Literature included some methods adopted Arabic text pre-processing. Our approach took an advantage step, thus we begun with the much recommended pre-processing process, which helped in removing corpus noise, and improved the computation complexity by reducing the size of feature set variables, simultaneously. Further explanation of our methodology is in section 3, which based on sentiment keywords co-occurrence measure (SKCM), SKCM is an algorithm we had used as a measure of each corpus term polarity. It was not considered as term polarity, but, sentiment weight which give better results when replaced with term appearance, and outperformed term frequency that had been used in different figures in many literatures [3] [4].

Our selected feature method had tested in support vector machine (SVM) classifier. Thus, we had classified each corpus of our three corpora three times. Firstly, via traditional feature selection method. Then with our SKCM results. After that, we tested them with the combination of these two FS methods. The results in all three corpora had showed the preference of SKCM results versus both two mentioned methods results.

This paper has related works brief in section 2. Then methodology explanation is presented in section 3, whereas section 4 shows experiments details and results illustration. Finally, we concluded this paper at section 5.

2. RELATED WORKS

The work of [5] had built domain-specific sentiment dictionary, automatically. This domain was movie reviews in Korean language. They used huge set of online movie reviews for finding out the joint probability of dictionary words This joint probability is between words and their position wither in negative annotated reviews, or positive annotated reviews. Then they give each word a polarity which was the difference between those two joint probability values. At last, they had

normalized these values and obtained 135,082 words that constitute their outcomes of sentiment dictionary in Korean language. However, the method of [5], was gave each term a polarity value, likewise our approach, but we didn't need annotated data set for finding that polarity, due to our approach do not find that polarity by advanced knowledge of document polarity, where this term appeared in.

The work of [6], had collected and prepared the biggest Arabic language opinion mining data set to-date. It is large Arabic book reviews corpus LABR, which included over 63,000 text segments, distributed between five rating values scaled from 1 and 2 for negative opinionated reviews, 3 as neutral reviews, whereas 4 and 5 as positive sentiments. Then, they study the data set characteristics, and testing the corpus for sentiment analysis. Here, we couldn't take the whole LABR corpus. So, we just take randomly two parts of LABR, one is balanced corpus, and another as unbalanced corpus.

[4] had exhibited one of the popular Arabic sentiment corpora. For the purpose of research, they had called it OCA as opinion corpus for Arabic. OCA is consisted of just 500 movie reviews, gathered from different Arabic blogs and pages in the web. OCA didn't include short text segments. Thus it had articles that led to big number of unique words. They carried on OCA some SA experiments using SVM and other classifiers. The features selection used was traditional feature selection which based on terms frequency (TF), where they concluded that TF had the same results of TFIDF, whether in the form of unigram, bigram, or trigram.

The work of [7], had utilized a mathematical tool based on rough set (RS) theory. They focused on the differences between RS reducts in Arabic SA. They used Term Weighting using TFIDF as a feature selection method, which based on counting terms per text segments, over all corpora. Rosetta toolkit which is designated for RS was used, with cross validation evaluation. They achieved an accuracy of 57% as the best result comparing other RS reducts. They concluded the applicability of using RS for Arabic SA, which need more enhancements, especially, in regard of reduct methods.

[3] had presented Arabic SA for Arabic tweets, they had collected Arabic dataset from twitter. Their results showed that SVM classifier was outperformed NB classifier, where they had used traditional feature selection that based on term frequency in the form of unigram and bigram. They had presented also that eliminating stop-words enhanced the results of sentiment classification accuracy. In our work, we re-implemented feature selection which based on traditional methods, and comparing resulted accuracy with ours.

The work of [8] had exhibited the concept of semantic orientate on (SO) for text segments. This concept was calculated based on counting the number of adjective or adverbs terms appeared in text. Also, SO was resulted using Mutual Information relation between text and polarity advanced known terms, such as "excellent" for positive polarity, or "poor" for negative polarity.

A text is considered as positive or negative based on SO calculated value. They achieved different accuracies in different domain in English language. The concept of semantic orientation had been used in literature in various applications such as in [9] [10] [8]. In our work, we didn't use SO as it is in [8]. We satisfied with the value of terms co-occurrence which is easier to be calculated than mutual information. And replaced the adjective (or adverb) usage, with sentiment keywords list that we had generated in [2] from pre-existed sentiment lexicon [6]. Also, we calculated it for terms not for segments of text. Then distributed resulted terms values over the binary matrix that record the appearance of each term in the corpus. Finally, we compared these results with the results of multiplying them by term frequency values related to each term in each text segment in the corpus.

In [11], we had been implemented number of classifiers with number of feature selection methods, comparing these accuracies resulted via number of corpus states. Different corpus states had generated from the same original corpus, after implementing NLP varied operations. We had performed the top accuracy results with the fifth corpus state and the feature selection method that based on LBSA approach, where they give each text segment a weight. This weight is based on the resulted weights from LBSA calculating, which count the number of sentiment lexicon terms appeared in each text segment, and produce the difference between positive words count and negative words count.

The feature selection in this work is different from the one used in [11]. It is actually improved from it, where in [11] we just satisfied with LBSA weights. Nevertheless, we here had take care of the occurrence of all corpus terms, utilizing the list of sentiment terms for give polarity of each term instead of text segments. Then, for each text segment, we had employed these polarities as FS vector.

In [12], seven feature selection methods had been compared, using SVM classifier via each one of these seven methods, those seven methods are: (Principal Components Analysis, Chi-squared, Relief-F, Gini Index, Support Vector Machines (SVMs), Uncertainty, and Information Gain). OCA corpus had been tested. In all SA classification experiments, they found out that SA performance is based on the method and the number of feature selection applied on it. Also, they found out that SVM is better than other method, whether as feature selection method or in classification accuracy.

They had deal with Arabic twitter for opinion mining in [13], by presenting a framework to analyze Arabic tweets into three sentimental classes: negative, positive, and/or neutral. Handling Arabizi, emoticons, and Arabic dialects was one of those sentiment framework aspects, which had done after collecting a dataset from twitter. They had used SVM with other classifiers, since they compare the results of stop-word eliminating and stemming of the corpus, using traditional methods of feature selection that based on counting terms appearing in each tweet on the corpus.
In [14], they had tried to discover useful knowledge about the opinions of Tunisian users from their posts on Facebook, and constructing an emoticons lexicon to be used in sentiment analysis tasks. SVM and NB had been utilized in SA tasks, via feature selection method based on traditional terms counting, where they had used popular terms appearance matrix, with the aid of Part-Of Speech POS, which is defined the type of each corpus term, based on grammar context.

3. METHODOLOGY

We have a hypothesis supposed that, in a number of the same domain text segments, each term has its own polarity, this polarity may be discovered by utilizing a list of sentiment keywords, and find the bigram relation between text terms and this list of keywords. We don't consider it as term polarity. We just used this value for feature selection method, which had led to enhance SA task.
In this work, we had increased the SA accuracy, using feature selection based on above hypothesis, which achieved results outperformed traditional methods in two opinion mining domains. The first domain is movie reviews, using OCA corpus collected by [4], since we used the same corpus and achieved more accuracy measure than traditional methods that applied in [4], we re-implemented it here for comparison purpose. The second domain is book reader reviews, using part of the huge corpus, i.e. LABR [6], since we re-implemented traditional method used in [3], and compared it with our approach which achieved better results.

Different literature had used different text pre-processing schemes; however, we determined the best pre-processing scheme, as a rule of thumb. This scheme is applied on each used corpus before any sentiment analysis had done, as illustrated in fig. 1. And we had noticed the efficient

benefit of this scheme, whether in feature set size reduction, which decrease computation time, or in the results of SA enhancement.

Our approach, generally, consisted of number of phases, illustrated in fig. 1, these phases are included pre-processing steps which had further details in the following subsection.
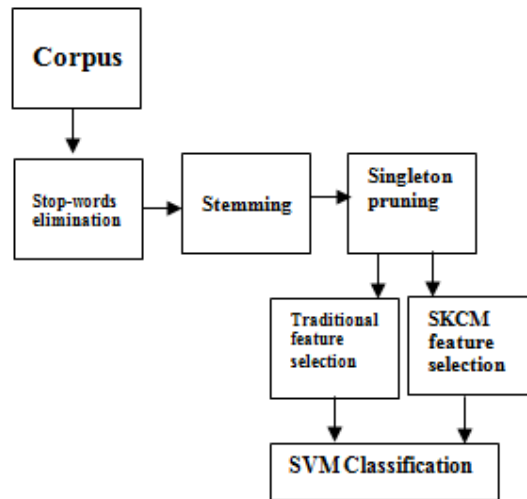
Figure 1. proposed approach general phases

3.1. Stop-words elimination:

Stop-words elimination is done without eliminating negation terms. Where stop-words are terms that known to do not have semantic affect on the written texts [3], specially, in the direction of opinion expressions. However, we preserved some words, with special exception; those are called negation terms, which are considered as sort of stop-words too. But they have an obvious impact on the polarity of terms that come after any one of them. Negation terms examples are terms like: 'no', 'but', 'more', 'less', etc, in English or opposite terms in any language, some negation words inverse the polarity of terms come after, whereas some of them affect the degree of the polarity.

3.2. Stemming:

We have done root stemming, where all unique terms, after stop-words elimination, are processed in premade tool. That converts each term into its origin: stem or root, stemming operation helps to reduce the number of terms. More precisely, those terms which have the same root. Stemming also helped to enhance feature selection, as in [4], because terms of the same root would have the same statistics, which improve semantic content on the feature selected, whereas most feature selection techniques on text mining depended on terms occur one the text.

We had utilized stemming tool that adopted by [15], the tool they had name: Alkhalil morpheme sys., which took the list of corpus unique terms as an input. Then, generated a list of terms, they were the roots of each input term. Then, we had made some refining to these root-stems before input them to SKCM algorithm.

3.3. Singleton terms pruning

Third pre-processing step was pruning corpus content. This was done by removing less impact terms in whole corpus. We choose to prune terms that occurred only once inside all corpus portions, pruning phase reduces the least significant terms in the corpus.

3.4. Traditional feature selection

The most popular method for text mining is using terms frequency TF as feature selection method, although, there were various ways to use TF, whether as one term frequency, i.e. unigram, or two terms as bigrams...etc. Whereas better TF results instance is TF-IDF, which had replaced TF mostly.

Both TF and TF-IDF are tested in [4], with unigram, bigram, and trigram variations. The results of them has supposed that no big differences between them via SA operations. This was compatible with our empirical experiments. Although, our objective herein is to proof the advantage of SKCM weights, comparing with TF weights traditional feature selection method.

3.5. SKCM algorithm

As illustrated in fig. 1, SKCM algorithm is implemented after the third pre-processing phase has finished, in parallel with the implementation of traditional FS method. SKCM algorithm is listed in fig. 2.

The outcome of SKCM algorithm had given each term a weight. This weight is different from the polarity values that had found out in semantic orientation (SO) as adopted in [8] or [9], since OS is used for constructing sentiment lexicons form seeds predefined polarity terms [9], or for determining text polarity as in [8], whereas we used this weight as feature selection method for SA. Moreover, we had used in SKCM distinct way to find terms weight, than mathematical way in [8]. Our approach is based on corpus terms discrimination using sentiment keywords co-occurrence measure (SKCM). This measure is used to find selected feature that depends on the number of corpus terms co-occurrence with sentiment keywords used for this objective.

```
SKCM algorithm
Input:
        crps; //target corpus reviews.
        utrms; // corpus unique terms.
        plst; // positive keywords list.
        nlst; // negative keywords list.
        tfm;  //terms frequency matrix.
        tdm;  // terms appearance matrix.
Output:
        skcm1;  // skcm with tdm matrix.
        skcm2;  // skcm with tfm matrix.
        tplrty; //terms polarity vector.
Begin
 nt = count corpus unique terms.
 nr = count corpus reviews.
 for i is 1 to nt
    neg=0;
    //negative co-occurrence with ith term.
    pos=0;
    //positive co-occurrence with ith term.
    for j is 1 to nr
     if uterms(i) is existed in crps(j)
       n = count nlst terms existed in crps(j).
```

```
         neg = neg + n.
         p = count plst terms existed in crps(j).
         pos = pos + p.
       end if;
     end for;
    tplrty(i) = pos – neg;
end for;
skcm1 = {replace each term's 1 over tdm, with tplrty value of the
         same term}.
skcm2 = {multiply each terms frequency over tfm, by tplrty value of
         the same term}.
END.
```

Figure 2. SKCM algorithm

Sentiment keywords are brought from premade sentiment lexicon, which we had improved in [2]. This lexicon is stemmed and used not to count lexicon words as in LBSA [9]. These keywords had extracted from a sentiment lexicon and stemmed to look for each keyword in each corpus portion. And then it gives each term in that corpus a counter. This counter is increased in both two polarities, i.e. positive and negative. Then for each corpus term, it have its own weight, which resulted by subtracting the summation of all target term co-occurrences with negative keywords, from the summation of all target term co-occurrences with positive keywords. Finally, as illustrated in fig. 2, SKCM algorithm output two selected features. The first resulted in pure terms weights, the second is the production of multiplying this results by term frequency that had been calculated in the phase of traditional FS method, as illustrated in our approach in fig. 1.

4. EXPERIMENTS AND RESULTS

We had used support vector machine (SVM) classifier, which had achieved good results in SA experiments. Also, it is a binary classifier that used to split each group of data into two categories, which is very suitable for SA tasks.

Three data sets are used. The first one is opinion corpus in Arabic (OCA) that collected by [4]. The second, two corpora are generated from the very big corpus large Arabic book reviews (LABR) [6], LABR consisted of 64000 reviews. Therefore, we, randomly, extracted two small corpora from it, a balanced and unbalanced one. UNBSLABR is unbalanced corpus that consisted of 2500 reviews, 1500 positive and 1000 negative. Whereas BLNSLABR is a balanced one which consisted of 1000 positive reviews and 1000 negative reviews, corpora statistics are available in table I.+

SA keywords list is based on SA lexicon that includes of positive bearing terms, and negative bearing terms, it was improved by [2]. This list terms are stemmed and used directly with each one of the three corpora.

Our experiments are begun with pre-processing steps, which applied to the three corpora. Then, we implemented SVM classifier using traditional selected features, which based on terms frequency (TF).

Table 1. Corpora statistics

Corpus	Reviews	Positive	Negative	Unique terms	Domain
unbSlabr	2500	1500	1000	14955	Book reviews
blnSlabr	2000	1000	1000	17801	Book reviews
OCA	500	250	250	42586	Movie reviews

Then, we applied SKCM algorithm, to assign each terms with a weight value that output from SKCM. After that, we had distributed each resulted value over text segment, based on the appearance of a term inside text segment, i.e. reviews text. These values will illustrate the relation between each term and text segment that belongs to. Then we carried on classification process using this selected feature.

Finally, for all corpora, we multiplied the selected feature of SKCM by TF values that used previously. This step was to measure and compare the effect of SKCM. If their results are better alone, or with traditional methods, feature selection based on terms frequency.

In all classification operations, we used SVM with cross validation k fold. We chose to use k=5 folds instead of 10 folds, due to computation complexity considerations.

Table II illustrates the results of our experiments. Where SKCM results are better than traditional method, i.e. term frequency depending selected features, even if we combined between SKCM results and TF, as illustrated in third column of table II, this achieved good results too. However, SKCM method results still better, when used solely.

Table 2. SVM accuracy results comparison.

	TF	SKCM	SKCM * TF
unbSlabr	74.20 %	**79.30 %**	78.00 %
blnSlabr	70.30 %	**73.25 %**	73.00 %
OCA	89.00 %	**93.00 %**	90.80 %

About the accuracy percentages, it is noticeable that movie reviews corpus made high number than book reviews two corpora. This may interpreted by that OCA has less reviews as illustrated in table I. And in the same time it is the bigger in the number of unique terms. Both two reasons helped in supervised learning, according to our practical experience. Also, another factor related to the nature of book readers reviews; this is noticed from [11] and [6], since LABR didn't achieved high accuracy results, which happened due to the language that used in that corpus, it was more complicated than another domain. And this is relevant to the richness of Arabic language morphology, which becomes as hard as movie reviews domain, perhaps due to used language, far away from dialectic or colloquial language of Arabic.

5. CONCLUSION

We had adopted new approach for enhancing sentiment analysis in Arabic language. Our approach based firstly on pre-processing steps, which reduces the size of dataset. Then we had implemented SKCM algorithm to apply our new feature selection method. After that we had used this new FS method on three corpora via SVM classifier.

Our approach results outperformed traditional methods, followed by most related works, which proof the significance impact of using the three pre-processing steps, which we had adopted in this work, and then illustrated the advantage of using sentiment keywords co-occurrence measure SKCM for discovering term weights that helped in enhancing the SA accuracy.

REFERENCES

[1]   B. Liu, Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies, Morgan & Claypool, 2012.

[2]   F. Alqasemi, A. Abdelwahab, and H. Abdelkader, "Adapting Domain-specific Sentiment Lexicon using New NLP-based Method in Arabic Language," International Journal of Computer Systems (IJCS), 2016, pp. 188-193.

[3]   A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," Collaboration Technologies and Systems (CTS), 2012 International Conference on. IEEE, 2012.

[4]   M. Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and José M. Perea-Ortega, "OCA: Opinion Corpus for Arabic," Journal of the American Society for Infomration Science and Technology, 2011, pp. 2045–2054.

[5]   H. Cho and S. Choi, "Automatic construction of movie domain Korean sentiment dictionary using online movie reviews," International Journal of Software Engineering and Its Applications 9.2, 2015, pp. 251-260.

[6]   M. Nabil, M. Aly, and A. F. Atiya, "LABR: a large scale arabic book reviews dataset," CoRR, abs, 2014 .

[7]   Q. A. Al-Radaideh and L. M. Twaiq,Rough, "Set theory for Arabic sentiment classification," Future Internet of Things and Cloud (FiCloud), 2014 International Conference on. IEEE, 2014.

[8]   P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.

[9]   A Bai, H. Hammer, A. Yazidi, and P. Engelstad, "Constructing sentiment lexicons in Norwegian from a large text corpus," IEEE 17th International Conference on Computational Science and Engineering, 2014.

[10]   M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational linguistics, 2011, pp. 37(2), 267-307.

[11]   F. Alqasemi, A. Abdelwahab, and H. Abdelkader, "An enhanced feature extraction technique for Improving se ntiment analysis in Arabic language," IEEE conference CSIST2016 Morocco, in press.

[12]   N. Omar, M. Albared, T. Al-Moslmi, and A. Al-Shabi, "A Comparative Study of Feature Selection and Machine Learning Algorithms for Arabic Sentiment Classification," Asia Information Retrieval Symposium. Springer International Publishing, 2014.

[13]   R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis in arabic tweets," Information and communication systems (icics), 2014 5th international conference on. IEEE, 2014.

[14]   J. Akaichi, Z. Dhouioui, and M. J. Pérez, "Text mining facebook status updates for sentiment classification," System Theory, Control and Computing (ICSTCC), 2013 17th International Conference. IEEE, 2013.

[15] A. L. Boudlal, and et al, "Alkhalil Morpho Sys1: A Morphosyntactic analysis system for Arabic texts," In International Arab conference on information technology, 2010.

AUTHORS

Amira Abdelwahab, received BSc degree in computer science and information systems from Faculty of Computers and Information, Helwan University, Egypt in 2000. and Ph.D. in information systems from Chiba University, Japan in 2012. In 2013, she was a postdoctoral fellow in Chiba University, Japan. Since 2012, she has been an assistant professor in information systems department, Faculty of Computers and Information, Menofia University, Egypt. Her research interests include Software Engineering, Decision Support System, database Systems, Data Mining, Machine Learning, Recommendation Systems, Intelligent Web applications, e-commerce, and knowledge discovery in Big Data.

Fahd A. Alqasemi, received his Bsc. in Mathematical and Computer from Ibb university,Yemen, then received his Master of Computer Information Systems from Arabic Academy in Sana'a, Yemen. He worked as an instructor in UST, Sana'a, Yemen. Currently, he is pursuing PhD In Menofia University, Menofia, Egypt. His major fields of interest is Image Retrieval, Information Retrieval, Data and Text Mining.

Prof. Hatem Abdelkader, obtained his BS and M.SC., both in electrical engineering from the Alexandria University, faculty of Engineering, 1990 and 1995, respectively. He obtained his Ph.D in electrical engineering also from faulty of engineering, Alexandria University, Egypt 2001. His area of interest are data security, web applications, artificial intelligence, and he is specialized neural networks. He is currently a professor in the information system department, faculty of computers and information, Menofia University, Egypt.

USING NLP APPROACH FOR ANALYZING CUSTOMER REVIEWS

Saleem Abuleil and Khalid Alsamara

MMMIS Department, Chicago State University, Chicago USA
sabuleil@csu.edu
kalsamar@csu.edu

ABSTRACT

The Web considers one of the main sources of customer opinions and reviews which they are represented in two formats; structured data (numeric ratings) and unstructured data (textual comments). Millions of textual comments about goods and services are posted on the web by customers and every day thousands are added, make it a big challenge to read and understand them to make them a useful structured data for customers and decision makers. Sentiment analysis or Opinion mining is a popular technique for summarizing and analyzing those opinions and reviews. In this paper, we use natural language processing techniques to generate some rules to help us understand customer opinions and reviews (textual comments) written in the Arabic language for the purpose of understanding each one of them and then convert them to a structured data. We use adjectives as a key point to highlight important information in the text then we work around them to tag attributes that describe the subject of the reviews, and we associate them with their values (adjectives).

KEYWORDS

Sentiment Analysis, NLP Arabic Language

1. INTRODUCTION

The web has made it possible for companies to discover what people are saying about their brands online, either in mainstream media like online newspapers and magazines, or on social media. Consumers now search for opinions online before, during, and after a purchase. The next step for brands is finding out whether people are talking positively or negatively about their brand, and why. Some online ratings provide a number but not the reasoning behind it, and may only present half of the story. The process of analyzing user's opinion or sentiment about particular services or products and their features is called opinion mining or sentiment analysis. There are two major approaches for performing sentiment analysis; statistical model based approaches and Natural Language Processing (NLP) based approaches to creating rules.

With the quick growing of e-commerce, the number of products sold on the web, the number of services offered on the web and number of online buyers and shoppers increased dramatically. It became an important task for online merchants to enable their customers to review or to express opinions on the products that they have purchased and use this feature to enhance customer

satisfaction about their products and services and also use it as a marketing tool. Almost all firms who sell goods and products on the web make it part of their business to collect and gather information about their services and goods they provide to their customers, in addition to many independent companies who collect customer opinions. Analyzing this information, summarize it, and make it available for decision makers to observe how consumers think about products and services, make it available as well for customers to assist them to make a comparison to improve their decisions before they make any order or request any service.

2. BACKGROUND AND RELATED WORK

The idea of opinion mining or sentiment analysis is to process a set of search results for a given entity, generating a list of attributes which are termed as opinion features of that entity. As a result of increasing number of people who are writing reviews on the Web, the number of reviews for products and receives grows rapidly. Some popular products can get hundreds of reviews at some large merchant sites, some reviews are short and easy to read and decision about them , but some reviews are long and have only a few sentences containing opinions on the product, that makes it hard for a potential customer to read them to make decision on them. A large number of reviews also makes it hard for product manufacturers to keep track of customer opinions of their products.

The extraction of a sentiment can be made either on a whole document (document-level SA), on each paragraph (paragraph-level SA), or on each sentence (sentence-level SA) [11]. Zen Hai and C Yang [17] proposed a method to identify opinion features from online reviews by exploiting the difference in opinion feature statistics across two corpora, one domain-specific corpus and one domain-independent corpus, this is captured by a measure called Domain relevance. They first extracted a list of candidate opinion features from the domain review corpus by defining a set of syntactic dependence rules. For each extracted candidate feature, they then estimated its intrinsic-domain relevance (IDR) and extrinsic-domain relevance (EDR) scores on the domain-dependent and domain-independent corpora, respectively. These values are compared with a threshold and are identified as best candidate features. Vasileios Hatzivassiloglou and Jance Wiebe [14] study the effects of dynamic adjectives, semantically oriented adjectives, and gradable adjectives on a simple subjectivity classifier, and establish that they are strong predictors of subjectivity. They have proposed a method for predicting subjectivity of opinions at sentence level by a supervised classification method. A trainable method that statistically combines two indicators of gradability is presented and evaluated, complementing existing automatic techniques for assigning orientation labels. Pang and Lee [3] proposed a machine-learning method that applies text-categorization techniques to just the subjective portions of the document to determine sentiment polarity. They examined the relation between subjectivity detection and polarity classification, showing that subjectivity detection can compress reviews into much shorter extracts that still retain polarity information at a level comparable to that of the full review. they have also shown that employing the minimum-cut framework results in the development of efficient algorithms for sentiment analysis.

Ryan McDonald and Kerry Hannan [13] have investigated the use of a global structured model that learns to predict sentiment on different levels of granularity for a text. The proposed model has the advantage of building the single model for all granularity levels. Labeling is done by MIRA algorithm which works at document and sentence level by applying a weight vector to each label. They showed that this model obtains higher accuracy than classifiers trained in

isolation as well as cascaded systems that pass information from one level to another at test time. Lizhen Qu and Georgiana Ifrim [10] have proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods by performing three steps: mining product features that have been commented on by customers; identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; summarizing the results.

Yessenalina and Cardie [1] Have presented a matrix-space model for ordinal scale sentiment prediction and an algorithm for learning such a model. The proposed 180 model learns a matrix for each word; the composition of words is modeled as iterated matrix multiplication. In the context of the phrase-level sentiment analysis task, their experimental results show statistically significant improvements in performance over a bag-of-words mode. Wei Jin and Hung Hay Ho [15] proposed a model that provides solutions for server problems that have been not provided by previous approaches. This system can self-learn new vocabularies based on the pattern it has learned, which is used in text and web mining. A novel approach is used to handle situations in which collecting a large training set could be expensive and difficult to accomplices. Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen [8] have emphasized on two important tasks in opinion mining, namely, opinion lexicon expansion and target extraction. they proposed a propagation approach to extract opinion words and targets iteratively given only a seed opinion lexicon of small size. The extraction is performed using identified relations between opinion words and targets, and also opinion words/targets themselves. Bo Pang and Lillian Lee [6] examine the relation between subjectivity detection and polarity classification. The subjectivity detection can compress reviews in shorter extracts that still retains polarity information at a level comparable to that of the full review. By using Naive Bayes polarity classifier the subjectivity extract are shown to be more effective input than the originating document. They show that the minimum-cut framework results in the development of an efficient algorithm for sentiment analysis. Via this framework, contextual information can lead to statistically significant improvement in polarity classification accuracy. Niklas Jacob and Iryna Gurevych [12] have shown how a CRF-based approach for opinion target extraction performs in a single- and cross-domain setting. They have presented a comparative evaluation of our approach on datasets from four different domains.

3. OUR CONTRIBUTION

Adjectives play a key role in this paper, they represent values of attributes and features of products and services. *In linguistics, an adjective is a describing word, the main syntactic role of which is to qualify a noun or noun phrase, giving more information about the object signified*[1]. Adjectives are one of the Arabic parts of speech. Arabic Adjectives are words that describe or modify another person or thing in the sentence. In Arabic adjectives are of the form فعيل Fa3iil, like كبير kabiir big, صغير saghriir small. One rule is that if a noun is definite the adjective has to be definite, like in البيت الكبير Al-bait Al-kabir The House The Big (Al is the Arabic indefinite article). Just like Spanish & German, Arabic has masculine and feminine adjective forms, in Arabic to form a feminine adjective from the masculine, you simply add "taa' marbuta" which looks like (ـة ,ة) to the end of the adjective for example (he) Beautiful Jameel جميل (masculine) and (she) Beautiful jameela جميلة (feminine). *In Classical Arabic, adjectives must agree with the nouns they modify in terms of gender (masculine or feminine),*

---

[1] Wikipedia website, the free enclobidia, "https://en.wikipedia.org/wiki/Adjective"

number (singular, dual or plural), grammatical case (subject, direct object or prepositional) and state of definiteness (whether the noun is definite or indefinite)[2].

In this paper we study customer opinions (reviews) written in the Arabic language for the purpose of understanding each one of them and then convert unstructured text to a structured data, very little work has been done in this area in the Arabic language and there is big need to contribute to it. We emphasize in this paper on two main elements: attribute and attribute value. For example TV product described by some attributes such as a screen, sound, price, size, where each attribute has a certain value such as good, bad, high, low, beautiful. Attributes are two types either simple or compound, simple attribute consists of one word such as sound, price, and size, compound attributes consists of two words to emphasise a specific feature such as sound *quality*, picture *quality*, resolution *accuracy*, and sound *clearness*, compound attribute comes in three main formats: الوان الشا *colors of screen*, الشاش الوانه *screen colors*, الشاشه ذات الوا *screen with colors*. In some cases adjective is attached to a special word to neglect it; change the status from positive to negative; such as not i.e. *not* good.

Unlike English, Arabic adjectives follow the noun they modify, which is somehow easier, because when you start with the noun first you will easily modify the adjective that comes afterward accordingly either to its masculine, feminine, dual or plural form. The noun in this context is the attribute described by the adjective. After studying hundreds of reviews we came up with novel approach consists of three steps as it shown in figure 1 to understand customer reviews written in the Arabic language.

To support our approach and to achieve our goal we collect attributes and adjectives and classify new adjectives while we are running our approach and save them in two main tables: attributes table and adjectives table, attributes tables include both simple attributes and compound attributes, each entry in this table has a pair of two roots represent a certain attribute, for simple attributes the second root is null. Adjectives table includes root of each adjectives and its classification either good or bad, we also have collected neglect tools (words) and saved them in a list.
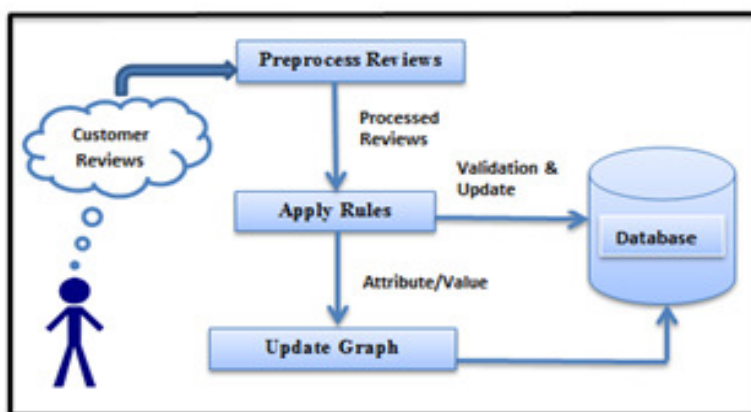
Fig 1: Approach Structure

Our approach consists of three steps as following:

---

[2] Learn Languages with Speak7 website, "http://arabic.speak7.com"

1- Preprocess Reviews: read reviews, use a morphology and part-of-speech tagging systems to:

    a. Find part-of-speech and root for each word in the text

    b. Identify adjectives in the text

    c. Check if neglected tool (word) is attached to the adjectives

2- Apply Rules: Extract attributes and associate them with their values (adjectives) that are labeled in step #1.

    a. Tag up to two words headed by an adjective, stop when encountering a verb, particle or punctuation mark.

    b. Use the following rules to form adjective phrases:

        Adjective Phrase → <Attribute> <Adjective>
                 | <Attribute> <Neglect-Tool> <Adjective>

        Attribute → Simple Attribute | Compound Attribute

    c. Check if <adjective> is already in adjectives table, find its classification, either positive or negative, otherwise classify it and update the adjectives table

    d. Check if <attribute> either if it is a simple or compound is in attributes table, if not validate it and update attributes table

3- Update Graph: use the output from step #2 (attributes/values) to update graph by updating frequency of each node and each edge. Each node in the graph contains either an attribute or a value, attribute nodes connected to values nodes through edges as shown in figure 2.

Figure 2. Graph Nodes

4. ANALYSIS

In the following example, we demonstrate how we use our approach to convert customer reviews from unstructured text to a structured data. The reviews we use in this example are about Samsung LED 4009MS-U7D 40 inch TV posted on egypt.souq.com website. First, we run a morphol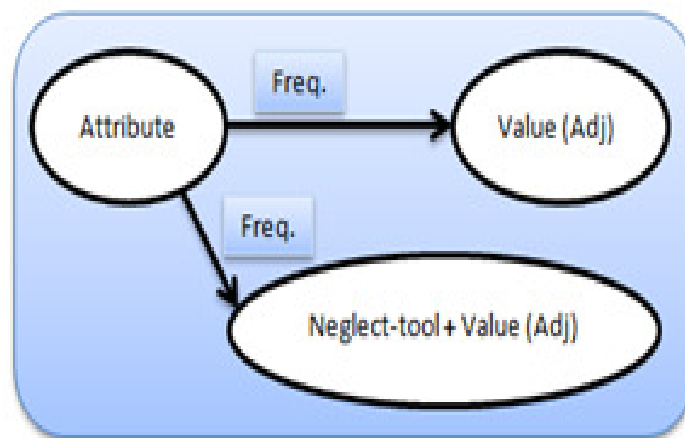ogy and part-of-speech tagging systems to identify adjectives and to find part-of-speech and root for each word in the text. The following is a sample review shows just adjectives, root and part-of-speech of each word are not shown.

احببت تلفزيون سامسونج كثيرا , استطيع ان استخدمه لليوتيوب وتطبيقات اخرى حاولت الطباعة باللغة العربية لبعض التطبيقات ولم استطع ذلك استخدمه للاكس بوكس و العب عليه طوال الوقت . صورة **نقية** وصوت **واضح** لكن الظهر الشاشه **ليس قوى** لا يوجد كابل ستريو للسماعات الخارجيه وكنت مندهشا عندما وجدت ان الجهاز ياتي معه ايضا محول اضافي

I like Samsung TV a lot; I can use YouTube and other applications I use for Xbox as well as other games most of the time. **Pure** *picture,* **clear** *sound but screen back is* **not durable***, I tried to type in Arabic while I am using different applications but I could not. There is no cable for external stereo headphones. Also, I was surprised when I have seen that the TV comes with external adapter*

Second, we tag up to two words headed by adjective, stop when encounter a verb, particle or punctuation mark, we apply some rules to form adjective phrases, check category of each adjective either positive or negative, identify attributes and associate them with their values, validate and update adjectives and attributes tables, the output of the second step is three adjective phrases as follows:

<**Simple Attribute**: صورة picture> <**Value** (*positive*): نقية pure>

<**Simple Attribute**: صوت sound> <**Value** (*positive*): واضح clear>

<**Compund Attribute**: ظهر شاشه screen back> <**Neglect-Tool**: ليس not> <**Value** (*positive*): قوى durable>

Third, we update the graph. Table 1 shows the result of 100 customer reviews.

In the above example, we found 70% of customers gave positive review for TV sound, 30% gave bad review, while 80% of customers gave positive review for TV screen and 20% gave negative review and just 8% gave positive review for the screen back and 92% gave negative review about it. Attribute frequency: TV sound repeated 50 times, TV picture 75 times and TV screen back 60 times, this gives the indication of the importance of each attribute in the reviews. In this paper we handled one side of customer reviews that when an adjectives present to describe attributes, another side needs to be studied is when customer reviews mention certain features or attributes without using adjectives to describe them such as:

I can use it to access YouTube and other applications   استطيع ان استخدمه لليوتيوب وتطبيقات اخرى
There is no cable for external stereo headphones   لا يوجد كابل ستريو للسماعات الخارجيه

In the above examples customers providing important facts about the TV, but because of the absent of any adjectives in the text we cannot catch them. One way is to look for some special phrases and work around them such as "لا يوجد" "استطيع ان استخدمه لل" *"I can use it to"* **and** "*There is no*".

Table 1: Customer Reviews for Samsung LED 4009MS-U7D 40 inch TV

Attribute	Value	Frequency	Positive	Negative
صوت Sound	رائع nice	5	5	0
صوت Sound	عالي high	5	5	0
صوت Sound	نقي pure	5	5	0
صوت Sound	جيد good	10	10	0
صوت Sound	ممتاز excellent	5	5	0
صوت Sound	واضح clear	5	5	0
صوت Sound	رديئ bad	10	0	10
صوت Sound	غير واضح not clear	5	0	5
Total		**50**	**35 / 70%**	**15 / 30%**
صورة Picture	رائع nice	15	15	0
صورة Picture	نقي pure	5	5	0
صورة Picture	جيد good	5	5	0
صورة Picture	ممتاز excellent	10	10	0
صورة Picture	واضح clear	25	25	0
صورة Picture	تغمشه noise	15	0	15
Total		**75**	**60 / 80%**	**15 / 20%**
ظهر الشاشه Screen Back	ليس قوى not durable	5	0	5
ظهر الشاشه Screen Back	غير مقوي not durable	5	0	5
ظهر الشاشه Screen Back	خفيفة light	5	0	5
ظهر الشاشه Screen Back	تعبان bad	10	0	10
ظهر الشاشه Screen Back	ليس جيد not good	15	0	15
ظهر الشاشه Screen Back	ممتاز excellent	5	5	0
ظهر الشاشه Screen Back	ضعيف weak	15	0	15
Total		**60**	**5 / 8%**	**55 / 92%**

5. CONCLUSION

In this paper, we have introduced our approach for using NLP to generate some rules to help us understand customer opinions and reviews (textual comments) written in the Arabic language for the purpose of understanding each one of them and convert them to a structured data. In future research we are going to study more reviews from different resources to test our approach on more data and generate more detailed analysis, we are going also to study cases when the adjective is absent, how to analyze text and understand it by looking for certain keywords in the reviews and work around them.

REFERENCES

[1]   A. Yessenalina and C. Cardie, "Compositional Matrix-Space Models for Sentiment Analysis", Proc. Conf. Empirical Methods in Natural Language Processing, pp. 172-182, 2011.

[2]   B. Liu, "Sentiment Analysis and Opinion Mining", Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167,May 2012 .

[3]   B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, 2004.

[4]   B Liu,"Sentiment Analysis and Opinion Mining", Synthesis Lectures on Human Language Technologies, vol.5,no.1, pp.1-167,May 2012.

[5]   E. Cambria, D. Osher and K.Kwok, "Sentic Activation : A two Level Affective Common Sense Reasoning Framework", Proc.26th AAAI Conf. Artificial Intelligence, pp.186-192, 2012.

[6]   Forman, B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics.

[7]   G.Qiu , C.Wang, J.Bu , K.Liu and C.Chen, "Incorporate the Syntactic Knowledge in Opinion Mining in User Generated Content", Proc. WWW 2008 Workshop NLP Challenges in the information Explosion Era, 2008.

[8]   G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation", Computational Linguistics, vol. 37, pp. 9-27, 2011..

[9]   L. Qu, G. Ifrim, and G. Weikum, "The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns", Proc. 23rd Int'l Conf. Computational Linguistics, pp. 913-921, 2010.

[10]   M. Hu and B.Liu, "Mining and Summarizing Customer Reviews", Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 168-177,2004.

[11]   M. Korayem, D. Crandall, and M. Abdul-Mageed. Subjectivity and sentiment analysis of arabic: A survey. In AboulElla Hassanien, Abdel-BadeehM. Salem, Rabie Ramadan, and Tai-hoon Kim, editors, Advanced Machine Learning Technologies and Applica-tions, volume 322 of Communications in Computer and Information Science, pages 128–139. Springer Berlin Heidelberg, 2012.

[12]   N. Jakob and I. Gurevych, "Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields", Proc. Conf. Empirical Methods in Natural Language Processing, pp. 1035-1045, 2010.

[13]   R. Mcdonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, "Structured Models for Fine-to-Coarse Sentiment Analysis", Proc. 45th Ann. Meeting of the Assoc. of Computational Linguistics, pp. 432-439, 2007.

[14]   V. Hatzivassiloglou and J.M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity", Proc. 18th Conf. Computational Linguistics, pp. 299-305, 2000.

[15]   W. Jin and H.H. Ho, "A Novel Lexicalized HMM-Based Learning Framework for Web Opinion Mining", Proc. 26th Ann. Int'l Conf. Machine Learning, pp. 465-472, 2009.

[16]   Y. Jo and A.H. Oh, "Aspect and Sentiment Unification Model for Online Review Analysis", Proc. Fourth ACM Int'l Conf. Web Search and Data Mining, pp. 815-824, 2011.

[17]   Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance", IEEE transactions on knowledge and data engineering, Vol. 26, NO. 3, MARCH 2014.

A NOVEL BACKGROUND SUBTRACTION ALGORITHM FOR PERSON TRACKING BASED ON K-NN

Asmaa AIT MOULAY [1] and Aouatif AMINE [2]

BOSS Team, Systems engineering laboratory, National school of applied
sciences, University Compus, Kenitra, MOROCCO
[1]`aitmoulay.asmaa@univ-ibntofail.ac.ma`
[2]`amine_aouatif@univ-ibntofail.ac.ma`

ABSTRACT

Object tracking can be defined as the process of detecting an object of interest from a video scene and keeping track of its motion, orientation, occlusion etc. in order to extract useful information. It is indeed a challenging problem and it's an important task. Many researchers are getting attracted in the field of computer vision, specifically the field of object tracking in video surveillance. The main purpose of this paper is to give to the reader information of the present state of the art object tracking, together with presenting steps involved in Background Subtraction and their techniques. In related literature we found three main methods of object tracking: the first method is the optical flow; the second is related to the background subtraction, which is divided into two types presented in this paper, and the last one is temporal differencing. We present a novel approach to background subtraction that compare a current frame with the background model that we have set before, so we can classified each pixel of the image as a foreground or a background element, then comes the tracking step to present our object of interest, which is a person, by his centroid. The tracking step is divided into two different methods, the surface method and the K-NN method, both are explained in the paper. Our proposed method is implemented and evaluated using CAVIAR database.

KEYWORDS

Video Surveillance, Object Tracking, Feature Extraction, Background Subtraction, Big data.

1. INTRODUCTION

Video tracking can be defined as an action that can describe and highlights an object so that we can follow it in a sequence of frames. It's very useful for higher level applications that rely on visual input to have the ability to consistently detect and track object motion such as humans, vehicles...etc. Therefore many developments have been affected to the methods, algorithms and techniques that are used as well as being implemented in other different areas such as security, surveillance systems, robotics...etc. In various computer vision applications, background subtraction is an easy and quick way to locate moving objects in video shot with a static camera. Even if this method will appear easy, some videos generate some false positives. Like illumination changes, an animated background...etc. Also false negatives can happen too when a moving object have the same colours as the background ones. What makes the simple inter frame difference a weak solution, so in order to deal with those challenges, different background models

have been proposed. The goal in this paper is to review several state-of-the-art background subtraction models proposed in the past years of researches and to present our proposition using this technique.

Background subtraction is detecting moving objects in videos using static cameras. It's a basic task in many video analysis applications, such as intelligence video surveillance, people detection and tracking...etc. The idea in the approach of the background subtraction is to detect the moving object of interest from the difference between the current frame and a reference frame which is called the background model. This background model must have a good quality so it can represent a scene as well as to be adapted to the luminance changes so that when we subtract it with the current image we will obtain a good background subtraction results.

1.1. Motivation

Recent events, such as terrorist attacks or the large-scale acts of banditry have led to an increased demand for security in societies. Video surveillance systems called smart are dealing with the real tile monitoring of mobile or immobile objects. The main objectives of these systems is to provide automatic interpretation of filmed scenes and to understand as well as to predict the actions and interactions of the objects observed. Video tracking helps us to map target objects in consecutive video frames, and since the demand of security in societies is growing it shall lead us to a need in surveillance application in many environments, particularly in the following domains:

1. Public safety and commercial sites: like supervising banks, shopping centres, airports, train stations, museums...etc.
2. automated video surveillance: designed to monitor the movement in an area to identify moving objects.
3. traffic monitoring: in case of breaks of traffic rules is monitored using cameras it will be tracked down easily.
4. information extraction: like to count people in public spaces, establishment of consumer profiles in malls...etc.
5. Military Application: border surveillance, monitoring secure web sites...etc.

2. MOVING OBJECT TRACKING

Object detection is an initial step of object tracking, it's a computer vision technology that deals with identifying instances of objects such as humans, vehicles, animals and other moving objects. Video sequence consist of several frame sequences which have certain surplus continuity. Frame sequences are extracted from the video sequence so we can detect moving objects.

Information about motions can be obtained through analysis and processing of frames at different periods of time. Therefore, two or more frames are acquired at different times and that contain information about relative motion between an imaging system and a scene. Video sequence analysis can be classified into three methods: optical flow method, background subtraction method, and temporal differencing method.

2.1. Optical flow

Optical flow is a vector-based method that matches points on objects over multiple frames in order to estimates motion in video. The motion field of frames is estimated to incorporate similar motion vectors into moving object. Solving transcendental equations is required in optical flow method. The calculation of this method is so complex and very sensitive to noise, which means that the amount of calculation is large. So this method is difficult to use in real time video because its performance and practicability is quite weak.[1]

To write the optical flow equation, we are going to suppose two assumptions: Assume that the apparent brightness of moving objects remains constant between frames.
And the image brightness is continuous and differentiable.

For the first assomption:

$$I(x(t+\Delta(t)),y(t+\Delta(t)),t+\Delta(t))=I(x(t),y(t),t) \qquad (1)$$

The first order Taylor expansion of the left term is

$$I(x(t+\Delta(t)),y(t+\Delta(t)),t+\Delta(t))=I(x(t),y(t),t) + \partial I/\partial x\ \partial x/\partial t\ \Delta(t) + \partial I/\partial y\ \partial y/t\ \Delta(t) + \partial I/\partial t\ \Delta(t)$$

$$= I(x(t),y(t),t) + I_x\, u + I_y v + I_t\, \Delta(t) \qquad (2)$$

From (1) and (2), we derive the optical flow equation

$$I_x\, u + I_y v + I_t = 0 \qquad (3)$$

Where I_t is the image difference between the two images

It provides a dense (point to point) pixel correspondence and Finding optic flow using edges has the advantage (over using two dimensional features) but it has also some disadvantages and problems and one problem common to almost all optical flow methods is the aperture problem, as one searches after displacement only in a small region near the previous location, one dimensional features can only be tracked in the normal direction of the object. Sources of other problems are illumination, numerical instabilities and occlusion. Illumination is a problem as motion is a geometric quantity while we are dealing with image intensity which depends on illumination. Numerical instability can be an important issue as many methods depend on numerical differentiation which is very noise sensitive. Occlusion is when a moving object occludes the background or another moving object, there are problems as patterns in the image appear and disappear.

2.3. Temporal differencing

Temporal differencing methodology is based on frame differences, it is the most simple and direct method. Moving objects are extracted according to the differences among two or three continuous frames. This method detects the moving regions and only objects marking relative motions. Moreover, given the fact that the time interval between two images is quite short, illumination changes have little influence on different images, which means the detection is effective and stable. Using frame differences can better adapt to environment in intensive fluctuation, and can easily detect those pixels causing images to change distinctly when the target moves. However, it is inadequate for dots with insignificantly changed pixels. Accordingly, the method is largely used in situations with comparatively simple background [2].

We can illustrate its equation as so :

$$|I(x,y,t) - I(x,y,t-1)| > Th \qquad (4)$$

This algorithm present an image with white pixels in the spot where the current image I at the time t is different from the previous image $I(x, y, t-1)$.

2.3. Background subtraction

Background subtraction is a widely used approach for detecting moving objects from a static camera. Objects can be detected by finding the difference between the current frame and

background frame. Background modelling can be categorised into two categories which are non-recursive and recursive techniques. For background estimation recursive technique includes frame differencing, median filter, linear predictive filter, and nonparametric model. Recursive technique is based on median filter, Kalman filter and mixture of Gaussian [1]. An image is divided into foreground and background in this method. The background is modelled, and the current frame and the background model are compared pixel by pixel. Those pixels accordance with the background model are labelled as the background, while others are labelled as the foreground. Background subtraction is a common method in moving object tracking algorithm, which is used more often in situations with relation to a background that is still. This method has low complexity. However, acquired background frames become sensitive to scene changes caused by illumination and external conditions as times goes on [2]. As the name suggests, background subtraction is the process of separating out foreground objects from the background in a sequence of a video frames, the equation that can explain this method would be like so:

$$|I(x,y,t) - B(x,y,t)| > Th \qquad (5)$$

While I is the image at the time t and B is the background at the time t. We subtract the estimated background from the input frame and apply a threshold, Th, to the absolute difference to get the foreground mask. One of the reasons that led us to choose the background subtraction method is its speed, which is illustrated in the execution time.

These background methods are used to identify moving objects in a video, which is often the first step in complex systems such as activity recognition, object tracking, and motion capture. Needless to deny the advantages of extracting the moving objects which can improve the reliability of the system by reducing the search space, reducing processing needs, and allowing the use of simpler technics for the rest of the data extraction, objects also are allowed to become part of the background without destroying the existing background model and even a different threshold is selected for each pixel and they are adapting by time what makes this method very robust against movement that are part of background (moving branches of a tree). Every method has its own disadvantages, for the background subtraction methods, there are relatively many parameters and they should be selected intelligently which make the method more complicated and also cannot deal with sudden, drastic lightning changes, shadows and gives also false positives sometimes.

3. BACKGROUND SUBTRACTION

Various methods have been proposed in the literature to perform the task of detection object of interest. In applications using fixed cameras with static background invariant, the most common approach is the detection background subtraction. But background subtraction algorithms have to stand up with several challenges arising from the nature of video surveillance.

3.1. Challenges of background subtraction

There are so many different background subtraction challenges that exist in literature [16]. Here we list some of the remarkable challenges for this algorithm [7]:

Gradual illumination changes. It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day.

Dynamic background. Some parts of the scenery may contain movement, but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees).

Occlusion. Intentionally or not, some objects may poorly differ from the appearance of background, making correct classification difficult. This is especially important in surveillance applications.

Shadows. Shadows cast by foreground objects often complicate further processing steps subsequent to background subtraction. Overlapping shadows of foreground regions for example hinder their separation and classification. Hence, it is preferable to ignore these irrelevant regions.

Video noise. Video signal is generally superimposed by noise. Background subtraction approaches for video surveillance have to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artefacts.

3.2. Background subtraction schema

In the literature, we can find so many methods in which the background subtraction algorithms are used, but most of them follow a simple flow diagram, which is shown in the figure 1. The four steps in a background subtraction algorithm are:

1. Pre-processing
2. Background modelling
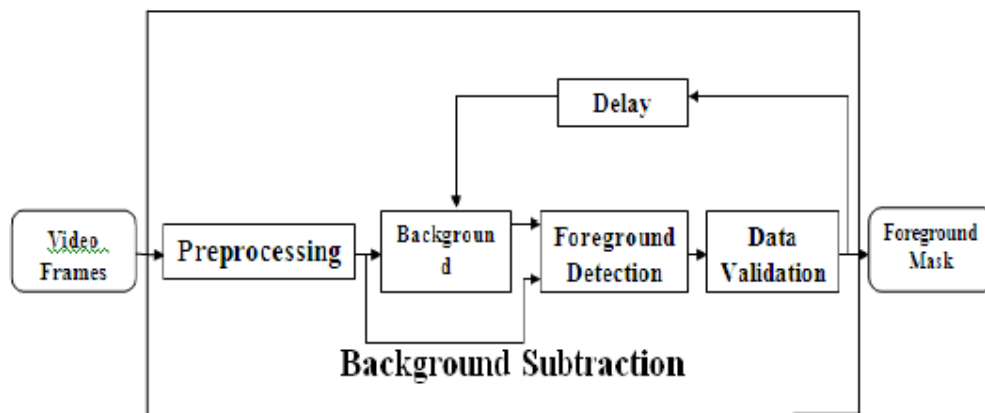*3. Foreground detection*
4. Data validation

Figure 1. Flow diagram of a generic background subtraction algorithm

Pre-processing: This step consist of a collection of simple image processing tasks that change the raw input video into a format that can be processed by the next steps. In the early stage of processing, simple temporal or spatial smoothing are used to reduce noise such as rain and snow. In pre-processing, the data format used by the background subtraction algorithm is a very important key. Most of the algorithms handle luminance intensity, which is one scalar value per each pixel. However, colour image, in either RGB or HSV colour space, is becoming more popular in the background subtraction literature [8, 15]. These papers argue that colour is better than luminance at identifying objects in low-contrast areas and suppressing shadow cast by moving objects. In addition to colour, pixel-based image features such as spatial and temporal derivatives are sometimes used to incorporate edges and motion information. For example, intensity values and spatial derivatives can be combined to form a single state space for background tracking with the Kalman filter. Pless et al. combine both spatial and temporal derivatives to form a constant velocity background model for detecting speeding vehicles. The main drawback of adding colour or derived features in background modelling is the extra

complexity for model parameter estimation. The increase in complexity is often significant as most background modelling techniques maintain an independent model for each pixel.

Background Modelling: Background modelling is at the heart of any background subtraction algorithm. Much research has been devoted to developing a background model that is robust against environmental changes in the background, but sensitive enough to identify all moving objects of interest. We classify background modelling techniques into two broad categories, non-recursive and recursive. They are described in the following subsections.

- Non-recursive techniques: A non-recursive technique uses a sliding-window approach for background estimation. It stores a buffer of the previous L video frames, and estimates the background image based on the temporal variation of each pixel within the buffer [6].

- Recursive Techniques: Recursive techniques do not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model [6]. Recursive techniques require less storage not like non-recursive techniques, but any mistake or error in the background model can lead us to a much longer period of time.

Foreground detection: This step makes us able to compare the input video frame with the background model, and identifies foreground pixels from the input frame. There is some techniques that doesn't use the same image as a background model like the Mixture of Gaussian model (MoG), but the most of techniques use a single image as their background models, so the approach for foreground detection is to check whether the input pixel is different from the corresponding background estimate [17]:

$$|I_t(x,y) - B_t(x,y)| > T \qquad (6)$$

Another popular foreground detection scheme is to threshold based on the normalized statistics:

$$(|I_t(x,y) - B_t(x,y) - \mu_d| / \partial_d) > T_s \quad (7)$$

Where μ_d and ∂_d are the mean and the standard deviation of $I_t(x, y) - B_t(x, y)$ for all special locations $(x; y)$. Most schemes determine the foreground threshold T or T_s experimentally [17].

Data Validation. To improve the candidate foreground mask based on information obtained from outside the background model, we define data validation process. All the background models have three main limitations: First, they ignore any correlation between neighbouring pixels; second, the rate of adaptation may not match the moving speed of the foreground objects; and third, non-stationary pixels from moving leaves or shadow cast by moving objects are easily mistaken as true foreground objects. When the background model adapts at a slower rate than the foreground scene, large areas of false foreground, commonly known as ghosts, often occur. Sophisticated vision techniques can also be used to validate foreground detection. Computing optical flow for candidate foreground regions can eliminate ghost objects as they have no motion. Colour segmentation can be used to grow foreground regions by assuming similar colour composition throughout the entire object [6].

4. EXPERIMENTAL RESULTS

In this section, we will track objects on a video of two seconds, this video has been taken from the CAVIAR DATA web site that contains a huge data of images and video clips with different scenarios of interest for the CAVIAR project using a wide angle lens along and across the hallway in a shopping centre in Lisbon. The resolution is half-resolution PAL standard (384x288 pixels) and compressed using MPEG2. The technique of background subtraction is used to better identify the objects in the surveillance videos. With this algorithm, we can get items to regions in

each image. We can continue to acquire the minimum bounding boxes of the objects and obtain the coordinates of the centroid of each object, which are then used to monitor the positions of objects through video frames.

4.1. Background generation

An image sequence of the background scene is first recorded. This sequence contains 68 images corresponding to two seconds with an acquisition rate of 34 frames per second.

4.2. Image segmentation

The detection image is the image resulting from comparing the current frame with the background model. Therefore it contains the elements of difference between these two images. Also, each pixel of the image is classified as part of the background or foreground element. The pixels classified in the first category will have a value of 0 (black) in a binary image of detection, and the other pixels will have a value of 1 (white).

4.3. Tracking

One of the most important issues in our subject is how an object is presented in a monitoring scenario. So an object can be defined as an object of interest for further analysis. In our case for example, we are tracking a person, so we have chosen our object which will be represented by his centroid.

- The surface method: in this method we tried to calculate the surface of each region in the binary image, then applied the code of centroid so it will associate the centroid to the largest region that represents for us the person to track.

- The K-NN method: the K-NN algorithm calculate the distance between a specific point with the others, and then find out the K nearest points to it, by then we calculate the sum of the distances related to a point so we can choose the smallest one.

The optical flow method is generally widely used for the detection and tracking of targets, this approach is an approximation of the movement of object tracked by estimating origins vectors of pixels in image sequences. Thus, the technique of calculating the optical flow is a differential method where we need it to estimate, at any point, the movement vector as a function of intensity changes of the pixel and its neighbours. These methods can accurately detect movement in the direction of the intensity gradient, but the movement that is tangent to the intensity of the gradient cannot be well represented, i.e. the optical flow methods are sensitive to illumination changes. Figure 4 shows a caption of a video tracking with the optical flow method.

To track a person on a video sequence, means that we need to trait every single image of the sequence, for example here we have 68 frames for a video of 2 seconds, so we need the whole process to be as fast as possible, because it will be applied to all the video sequence.
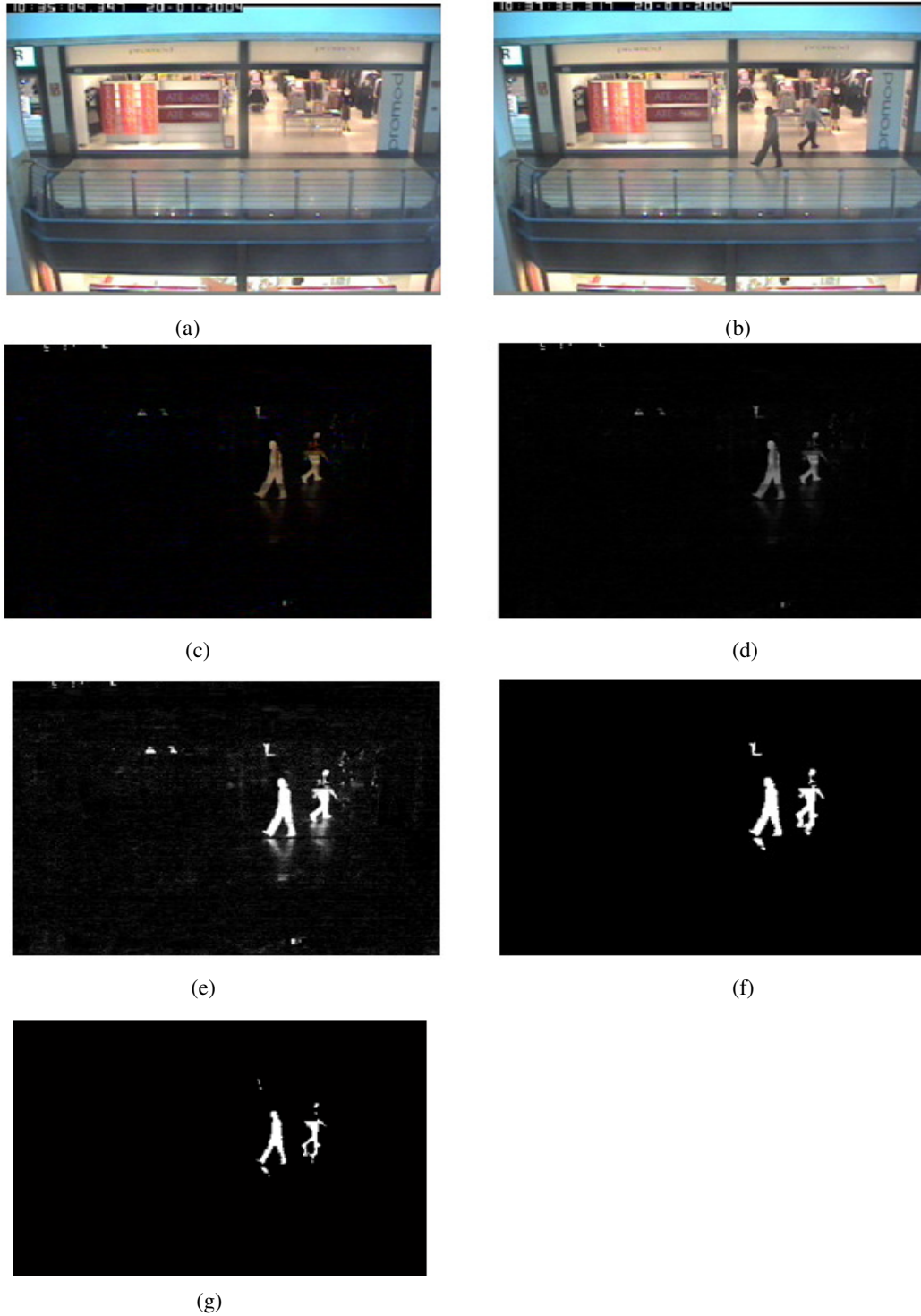
Figure 2.   (a) The background model, (b) input image, (c) difference image between (a) and (b), (d) converted (c) to grayscale, (e) adjust intensity values of (d), (f) small Object removed from (e), (g) eroded image.
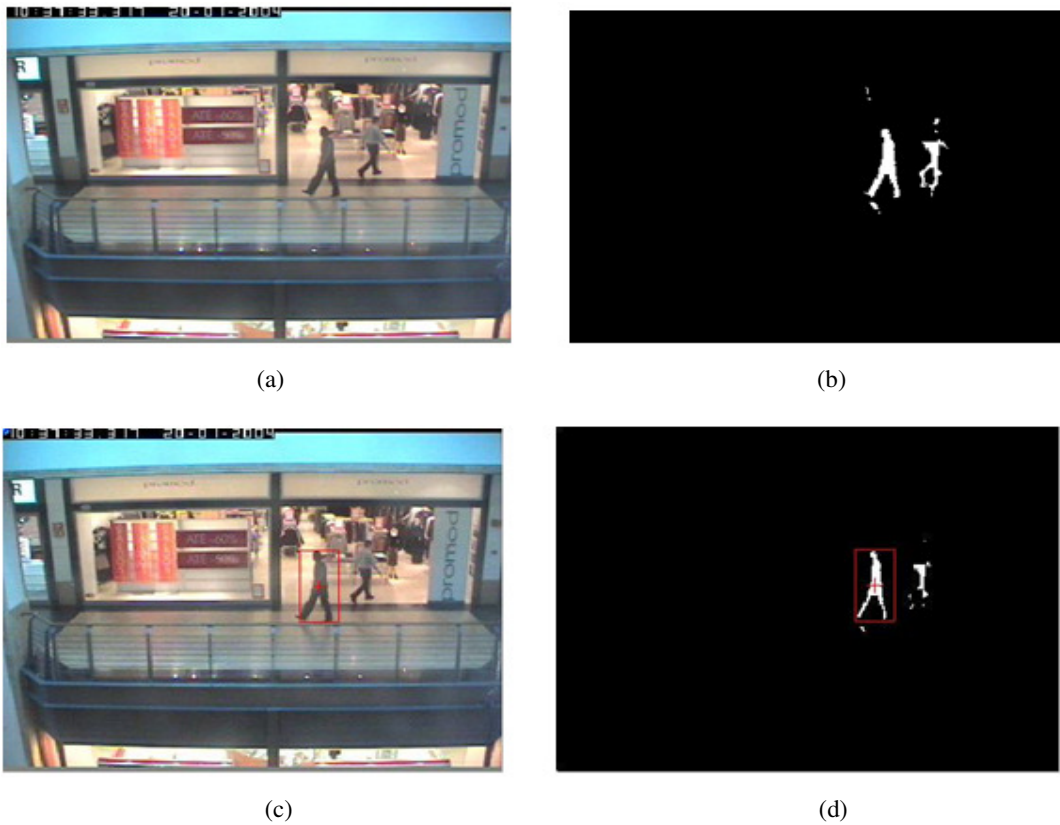
(a)

(b)

(c)

(d)

Figure 3. (a) And (c) tracking in a colour image, (b) and (d) tracking in a binary image

After executing the background subtraction method that we used, the total execution time calculated between converting the image to a grey scale and finding the centroid to track the person, was 3.797212 seconds per frame, which means 3.797212 x 68 = 258.210416 seconds, about 4.30 minutes for the whole sequence. While using the optical flow method (figure 4), the execution time was 425.59 seconds, about 7.1 minutes for the whole sequence. Therefore, the execution time is such an important task in video tracking, because here we're talking about a video of 2 seconds, but in real life, video surveillance stock huge quantities of videos and frames, which mean we need a very fast method to deal with this such a huge quantity of information.
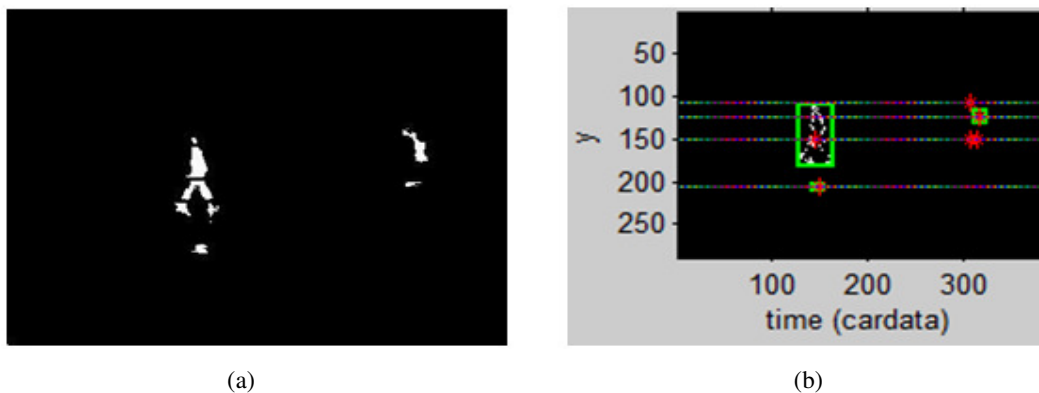

(a)

(b)

Figure 4. (a) A frame from the video, (b) the video tracking with the optical flow

3. CONCLUSION

In this paper, we introduced the object tracking and its importance now days, motivation and applications of this field and the classified methods of Video sequence analysis. We survey a number of background subtraction algorithms in the literature. We presented the challenges of this method and analyse it in pre-processing, background modelling, foreground detection, and data validation. Also, we have proposed our novel method using the background subtraction algorithm to extract object based on his surface information represented by centroid of this binary region, after that we used K-NN to track the target from frame to frame using the smallest distance between two detected regions represented by their centroids, then compare it with some other techniques.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Shipra Ojha and Sachin Sakhare, Image Processing Techniques for Object Tracking in Video Surveillance- A Survey, International Conference on Pervasive Computing, Department of Computer Engineering Vishwakarma Institute of Information Technology, india, 2015.

[2]   R.Venkatesan and A.Balaji Ganesh, Real Time Implementation On Moving Object Tracking And Recognisation Using Matlab, Dept of TIFAC CORE Velammal engineering college Chennai, india, 2012.

[3]   Junzo Watada, Zalili Musa, Lakhmi C. Jain, and John Fulcher, Human Tracking: A State-of-Art Survey, Waseda University: Japan,University Malaysia Pahang: Malaysia, University of South Australia: Australia, University of Wollongong: Australia, 2010.

[4]   Patrick Dickinson and Andrew Hunter, Scene Modelling Using An Adaptive Mixture of Gaussians in Colour and Space, Department of Computing and Informatics, University of Lincoln, UK, 2005.

[5]   Alper Yilmaz, Omar Javed and Mubarak Shah, Object Tracking: A Survey, Ohio State University, University of Central Florida, USA, 2006.

[6]   Harsha Varwani, Heena Choithwani,Kajal Sahatiya, Shruti Gangan, Tina Gyanchandani and Dashrath Mane. Understanding various Techniques for Background Subtraction and Implementation of Shadow Detection., VES Institute of Technology, Chembur, 2013.

[7]   Sebastian Brutzer, Benjamin Hoferlin and Gunther Heidemann, Evaluation of Background Subtraction Techniques for Video Surveillance, Intelligent Systems Group, Universitat Stuttgart, germany, 2014.

[8]   C.Wren, A. Azabayejani, T. Darrel, and A. Pentland, P finder: Real-time tracking of the human body, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, pp.780-785. July 1997.

[9]   R. Cutler and L. Davis, view-based detection and analysis of periodic motion. Fourteenth International Conference on Pattern Recognition, vol.1, pp. 495-500, Brisbane, Australia, Aug 1998.

[10]   A. Elgammal, D. Harwood, and L. Davis, Non-parametric model for background subtraction in Proceedings of IEEE ICCV'99 Frame-rate workshop, Sept 1999.

[11] C. Stauffer and W. Grimson, Learning patterns of activity using real-time tracking, in IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 22, pp. 747-57, Aug 2000.

[12] P. KaewTraKulPong and R. Bowden, An improved adaptive background mixture model for real-time tracking with shadow detection in Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems, Sept. 2001.

[13] M. Harville,A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models, , in Proceedings of the Seventh European Conference on Computer Vision, Part III, pp. 543-60:Copenhagen, Denmark, May 2002.

[14] D. R. Magee, Tracking multiple vehicles using foreground, background, and motion models, , in Proceedings of the Statistical Methods in Video Processing Workshop, pp. 7-12: Copenhagen, Denmark, june 2002.

[15] R. Cucchiara, M. Piccardi, and A. Prati Detecting moving objects, ghosts, and shadows in video streams, IEEE Transactions on Pattern Analysis and Machine Intelligence. vol.25, pp. 1337-1342, oct 2003.

[16] T. Bouwmans, F. El Baf, and B. Vachon,Background modeling using mixture of gaussians for foreground detection-a survey, Recent Patents on Computer Science.Vol.1(Num. 3), pp. 219-237,2008.

[17] Sen-Ching S. Cheung and Chandrika Kamath, Robust techniques for background subtraction in urban trafic video, Center for Applied Scientific Computing: Lawrence Livermore National Laboratory, 2004.

[18] C. Qin, H. R. Ren, C. C. Chang, Q. K. Chen, Novel Occlusion Object Removal with Inter-frame Editing and Texture Synthesis, Journal of Information Hiding and Multimedia Signal Processing, vol. 7, no. 2, pp. 386-398, 2016.

[19] Jean-Philippe Jodoin, Guillaume-Alexandre Bilodeau, Nicolas Saunier, Background subtraction based on Local Shape, Ecole Polytechnique de Montreal P.O. Box 6079, Station Centre-ville, Montreal, (Quebec), Canada, H3C 3A7, 2012

[20] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

AUTHORS

Ait Moulay Asmaa:

is a Phd student of computer engineering at ENSA of Kenitra, University of Ibn Tofail. She received her Master's degree in information systems security from the national school of applied sciences, and she is currently completing her researches on field of detecting and tracking individuals from multiple camera with the BOSS team in the LGS Laboratory, ENSA, Kenitra, Morocco.

Amine Aouatif:

Received her Master's degree (DESA) in Computer Sciences and Telecommunications engineering from the faculty of sciences, Mohammed V-Agdal University, Rabat, Morocco, in 2004. She remained at Computer Sciences and earned a PhD degree in 2009 for a dissertation titled "Feature Extraction and Selection for Dimensionality Reduction in Pattern Recognition and their Application in Face Recognition". Her research interests include but are not limited to dimensionality reduction, feature selection applied to face detection and recognition and driver hypo-vigilance. She joined the ENSA, Kenitra, Morocco, in 2010, as an assistant professor. Since November 2010, Aouatif Amine has been a Vice-Chair of IEEE Signal Processing Society Morocco Chapter. Since April 2015, Aouatif AMINE is the BOSS team header in the LGS Laboratory, ENSA, Kenitra, Morocco.

NEURAL NETWORKS FOR HIGH PERFORMANCE TIME-DELAY ESTIMATION AND ACOUSTIC SOURCE LOCALIZATION

Ludwig Houégnigan[1], Pooyan Safari[2], Climent Nadeu[2], Mike van der Schaar[1], Marta Solé[1], Michel André[1]

[1]Laboratory of Applied Bioacoustics (LAB), Polytechnic University of Catalonia, UPC Barcelona Tech, Spain
[2]TALP Research Center - Dept. TSC, Polytechnic University of Catalonia, UPC Barcelona Tech, Spain

ABSTRACT

Time-delay estimation is an essential building block of many signal processing applications. This paper follows up on earlier work for acoustic source localization and time delay estimation using pattern recognition techniques in the adverse environment such as reverberant rooms or underwater; it presents unprecedented high performance results obtained with supervised training of neural networks which challenge the state of the art and compares its performance to that of well-known methods such as the Generalized Cross-Correlation or Adaptive Eigenvalue Decomposition.

KEYWORDS

Time-delay estimation, neural networks, source localization, underwater acoustics, room acoustics

1. INTRODUCTION

Time-delay estimation (TDE) is a task as fundamental as spectral estimation and a key step for many popular applications such as sonar and radar direction finding, seismology, biomedicine, satellite navigation or acoustic source localization.

Recent advances in machine learning invite us to revisit classical signal processing problems and theory to renew our understanding of these problems so as to challenge well-established techniques. In that frame, over the past years, the authors of this paper, through different publications [1-3] have proposed original approaches using machine learning and data-specific modelling in order to improve TDE in the context of both air and underwater acoustic source localization (biological sources such as cetaceans, or artificial ones such as pingers, ships, navy sonar, etc).An approach using neural networks is justified by at least 3 fundamental assumptions:

Assumption (1): the most widely used methods for TDE have demonstrated their optimality [4-6] in the case of random signals and have been statistically analyzed according to that particular context, e.g. cross-correlation was demonstrated to be optimal for random data at high Signal-to-noise ratio (SNR). However, in many situations, SNR may be relatively low and the signals at stake can be far from random. Rather, they display statistical structure which can be exploited in order to achieve greatly improved results for all kinds of estimation tasks. Machine learning tools- and in particular supervised learning algorithms such as artificial neural networks- offer us a great opportunity to develop robust time-delay estimators that improve largelyon the classical estimators.

Assumption (2): Classical methods for TDE suppose models (see section 2)which typically fail to fully render the complexity of propagation in underwater contexts or in reverberant rooms in air. Using supervised learning is expected to permit to match more closely the available data to its environment.

Assumption (3): In tasks such as localization or angle estimation, much effort is done on tracking solutions such as Extended and Unscented Kalman Filters, or particle filters. Yet, putting more emphasis on reducing a priori the mean error and variance of TDE, by yielding more accurate and consistent estimates, will clearly facilitate the task of any subsequent tracking algorithm.

On the one hand, comprehensive studies on TDE [7-8]were published but none of them, to our knowledge, has ever included supervised learning. On the other hand, little has yet been published on time-delay estimation using supervised learning besides benchmark papers by Shaltaf et al. [9, 10]. With respect to the previously mentioned works, this article intends to progress by addressing the following points:

(1) Including large time delays: hence avoiding to restrict estimation to a narrow range of time-delays which fall within the Nyquist range and could in fact already be addressed by beamforming techniques, a shortcoming in [9,10].

(2) Instead of estimating only a nominal time-delay value (time position of a peak), the neural network was tasked to provide a multidimensional output representing an ideal time-delay response, (cf. section 2 and fig. 3 and 4).

(3) The number of data samples at stake is large, i.e. 8 datasets containing each 400 000 samples. Each of them was evaluated by multiple methods, in order to provide a robust and comparative statistical analysis of the various time-delay estimators at hand under different levels of noise.

2. MODELS AND METHODS FOR TIME-DELAY ESTIMATION

2.1. Ideal free-field model

Methods such as standard cross-correlation and generalized cross-correlation [6, 8, 11] or minimum entropy [12] are based on this model. It proposes to view two signals x_1 and x_2, acquired at two spatially separated sensors, as attenuated and delayed versions of a source signal plagued with additive noise. This model is well-described in [7] and can be described with the following equations:

$$x_1(n) = \alpha_1 s(n - \tau_1) + b_1(n) \ (eq.1)$$

$$x_2(n) = \alpha_2 s(n - \tau_2) + b_2(n) \ (eq.2)$$

where s is the source signal, α_1 and α_2 are attenuation factors due to propagation and b_1 and b_2 represent uncorrelated additive noise .

In this model the sample time-delay τ_{12} between signals x_1 *and* x_2 can be set as:

$$\tau_{12} = \tau_1 - \tau_2 \ (eq.3)$$

2.2 Real reverberant model

The reverberant model entails a higher level of complexity as it assumes that the original signal is deteriorated by multipaths and reverberation (walls, floors, tables in room acoustics, surface, seafloor, or scattering in underwater acoustics). Hence, signals x_1 and x_2 are modelled as convolutive mixtures of the source signal:

$$x_1(n) = h_1 * s(n) + w_1(n) \ (eq.4)$$

$$x_2(n) = h_2 * s(n) + w_2(n) \ (eq.5)$$

where h_1 and h_2 aim to model, with FIR or IIR filters, the channel impulse responses from the source to the positions of sensors 1 and 2, and where "*" indicates convolution and the noises w_1 and w_2 can be correlated. Such a model can provide a better description of the propagation environment and is typically based on adaptive algorithms such as the Adaptive Eigenvalue Decomposition (AED) [13, 14, 15].

2.3. Supervised estimation and training with neural networks

In this approach, it is assumed that the time-delay can be approximated by the output of a previously trained neural network which receives as input a combination (or a transformation) of signals x_1 and x_2. No particular assumption is made with regard to modelling. The capabilities of neural networks for system identification and interpolation permit to construct a system that minimizes the error between its output and an ideal response represented by a peak at the localization of the correct time-delay.

In this study, the input vector was set to $x = [x_1 x_2]$, a concatenation of vectors x_1 and x_2, and no particular effort was made to construct a transformed or more compact input. In [9, 10] it is also proposed to use the sum of the received signals or a transformation of those signals such as the Discrete Cosine Transform (DCT) to provide a more compact representation and hence to optimize the dimension of the neural architecture. It is proposed here to provide as target the dirac delta function $\delta(n - \tau_{12})$. The neural network then aims, through supervised training, at minimizing the distance between its output and an ideal response composed of a value 1 at the correct time-delay and zeros elsewhere.

3. IMPLEMENTATION AND TEST

3.1 Signal and noise models

Eight artificial datasets containing chirp signals were constructed. Each signal featured random duration (with 10 to hundreds of samples at a sampling rate of 16 kHz, their number being randomly selected from uniform distributions) and varying noise. The variance σ_S^2 of the signal of interest is related to the noise variance of each dataset as described in table 1.1 and 1.2. The noisy dataset aims at mimicking adverse conditions which typically cause failures in time-delay estimators as will be shown in section 4.
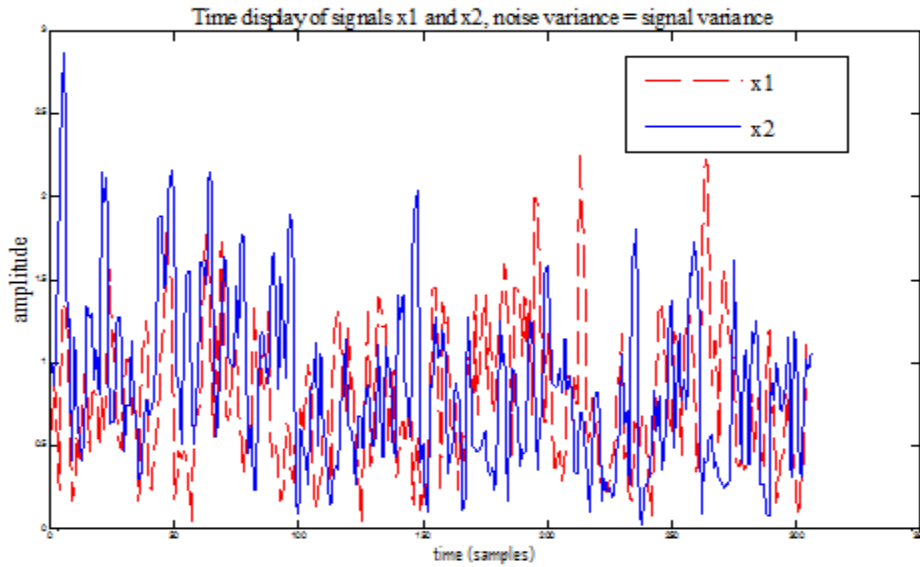

Fig.1 signals x1 and x2 are composed of noisy chirps signals of variable duration

Dataset#	1	2	3	4	
Noise var. σ_N^2	0	$0.2\sigma_S^2$	$0.4\sigma_S^2$	$0.5\sigma_S^2$	Table 1.1

Dataset#	5	6	7	8	
Noise var. σ_N^2	$0.6\sigma_S^2$	$0.8\sigma_S^2$	σ_S^2	varying	Table 1.2

Dataset 8 is made of 400000 signals corrupted with additive noise which variance is uniformly taken between $0.2\sigma^2$ and σ^2.

3.2. Neural network parameters

Multilayer perceptrons ("mlps") architectures including a single hidden layer and 30 hidden units were used. Sigmoid and linear activation functions were respectively used for the hidden and output units. The training procedure was conducted using a standard backpropagation algorithm with a fixed mini-batch size of 100 and 100 epochs. The fixed momentum and weight decay for all the systems were respectively set to 0.9 and 10^{-7}. A sparsity target of 0.05 and a sparsity penalty of 10^{-4} were used for all the networks. L2 regularization norm was set to 10^{-3}.

For each dataset, 19 neural nets were trained with varying learning rates. Among those 19 nets, for the sake of concision, only 4 were selected and are displayed here, namely the nets providing respectively the best (lowest) mean error, the worst (highest) mean error, the best (lowest) variance and the worst (highest) variance, referred to respectively as MLPA, MLPB, MLPC, MLPD. Those nets are then compared to 4 other estimators: the generalized cross-correlation with SCOT and PHAT filters (GCC-SCOT and GCC-PHAT), standard unbiased cross-correlation (XCOR) and the Adaptive Eigenvalue Decomposition (AED). In that sense, the different models already presented in section 2. are all tested here.

For each method, this work is not limited solely to the nominal time-delay, as is the case in most publications, but it also analyzes the delay distribution as indicated by the following two points:

(1) For a nominal time-delay$\hat{\tau}$: the error $\hat{\tau} - \tau_{12}$ between the estimated nominal time-delay and the ideal time delay should be minimized.

(2) Similarity of the output of the neural net with the ideal target $\delta(n - \tau_{12})$ must be evaluated. To that purpose, a similarity measure named $Q_{KL}$was derived from the Kullback-Leibler divergence to assess the resemblance of the output $NN_{12}$of the neural network to the ideal response, both seen as data distributions:

$$Q_{KL} = abs(KL(NN_{12}, \delta(n - \tau_{12})), \text{ (eq.6)}$$

where $KL(P, R)$ is a modified expression of the Kullback-Leibler divergence between two distributions P and R:

$$KL(P, R)= \quad -\sum_x p(x) \log(r(x) + 1) + \sum_x p(x) \log(p(x) + 1) \text{ , (eq.7)}$$

The metric Q_{KL} resembles a distance, inasmuch as it is positive and approaches 0 when the output of the neural net resembles its target. Yet, Q_{KL} *is* not a true metric since it is not symmetric and does not verify the triangle inequality. It was however found to be a convenient, consistent and compact measurement compared to others such as χ^2,Euclidian or Bhattacharryya distance which were also evaluated.

4. RESULTS

4.1 First overview of results

Figure 3. represents the output of cross-correlation estimator (XCOR), the output of an "mlp" and the ideal response (target). In this plot, where the variance of the noise was set to 0, it can be observed that both the cross-correlation and the neural network match closely the peak value of the target.
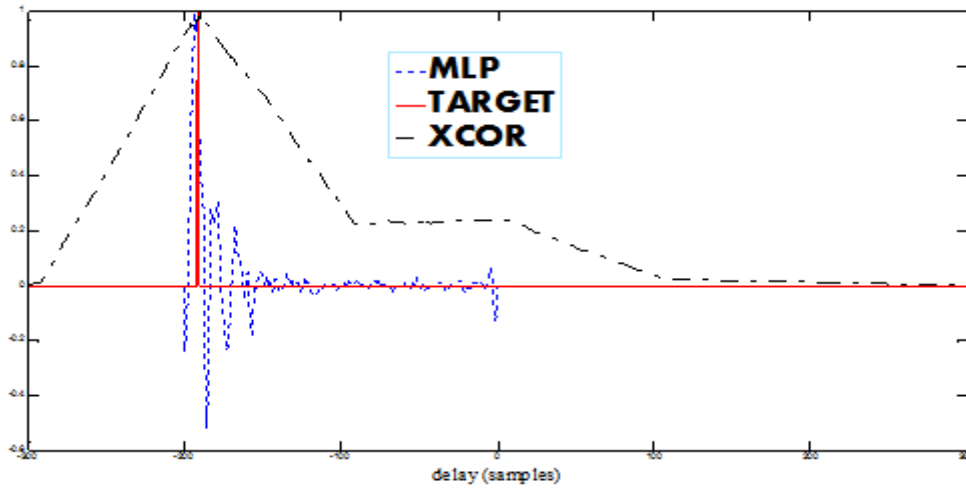
Figure 3. Output of mlp and XCOR estimators with respect to Target. Noise variance= 0

However, the shape of the target distribution is much closer to the shape of the neural network than to that of the cross-correlation. This is adequately described by the following Q_{KL} measures: $Q_{KL}(Target)$=0, $Q_{KL}(MLP) = 0.0022$, $Q_{KL}(XCOR) = 13.82$. Q_{KL} consistently produces low values when the distribution at hand is close to the target and higher values when that distributions is less close.

Figure 4. represents the output of the cross-correlation estimator (XCOR), the output of an "mlp" and the ideal response (target), this time in the presence of variable noise. It can be observed that cross-correlation is performing poorly at estimating the nominal delay whereas the neural network still closely matches the target. Beyond the nominal delay value, the overall shape of the distribution is also slightly affected by noise: the neural network, although it performs much closer to the target than cross-correlation does, is noisier than previously and has more leakage and ripples. This is adequately reflected by the Q_{KL} measures: $Q_{KL}(Target)$=0, $Q_{KL}(MLP) = 0.1895$, $Q_{KL}(XCOR) = 14.30$.
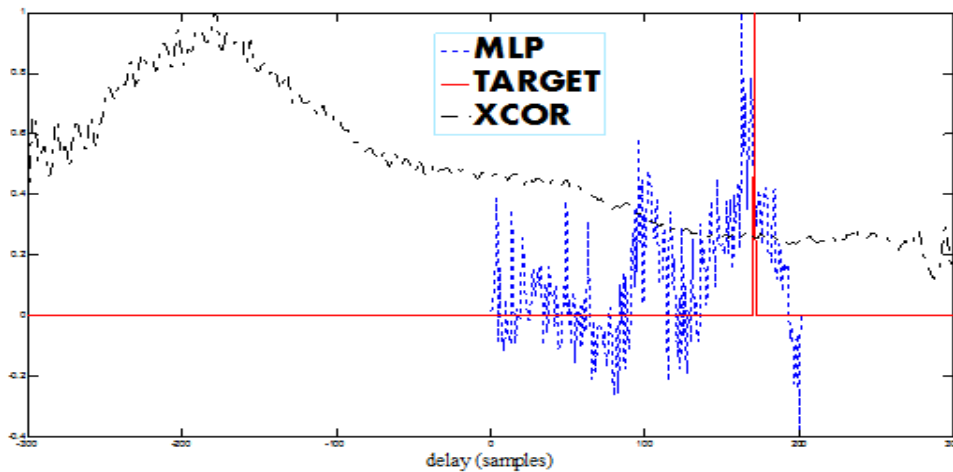

Figure 4. Output of mlp and xcor estimators with respect to target. Variance is variable (dataset 8).

4.2 Statistical significance

Tables 2 and 3 summarize the evolution of the mean of the error and its standard deviation for the tested estimators as variance changes. Among the non-supervised methods, the results obtained for GCC-SCOT, GCC-PHAT, the standard cross-correlation and AED are presented. At a noise variance of 0, SNR is infinite so that cross-correlation performs better than any of the methods under scrutiny, it is indeed an optimal estimator in such rare conditions. All "mlps" perform well: they have low variance and a small bias (one sample). On the contrary, Adaptive Eigenvalue Decomposition performs poorly, probably due to the absence of convolutive mixture.

Table 2. Standard deviation of various estimators for current datasets								
	MLPA	MLPB	MLPC	MLPD	SCOT	PHAT	XCOR	AED
$\sigma_N^2 = 0$	1.32	12.30	1.32	12.30	8.16	3.86	0	32.86
$\sigma_N^2 = 0.2\sigma_S^2$	8.89	11.29	8.78	11.60	115.97	115.70	119.86	79.63
$\sigma_N^2 = 0.4\sigma_S^2$	16.55	17.74	16.55	18.78	113.59	113.27	114.09	78.34
$\sigma_N^2 = 0.5\sigma_S^2$	19.43	21.67	18.22	22.33	110.61	110.21	106.84	78.47
$\sigma_N^2 = 0.6\sigma_S^2$	21.82	25.62	20.00	25.88	107.11	106.70	102.86	78.82
$\sigma_N^2 = 0.8\sigma_S^2$	26.29	27.92	25.30	30.40	100.01	99.47	102.37	79.08
$\sigma_N^2 = \sigma_S^2$	30.23	31.81	29.33	32.39	94.95	94.24	106.29	78.87
$variable\ \sigma_N^2$	23.35	26.95	23.31	26.95	110.85	110.48	117.78	79.61

Table 3. Mean of the error of various estimators for current datasets								
	MLPA	MLPB	MLPC	MLPD	SCOT	PHAT	XCOR	AED
$\sigma_N^2 = 0$	0.98	5.56	0.98	5.56	0.40	0.16	0	43.29
$\sigma_N^2 = 0.2\sigma_S^2$	2.13	3.46	2.19	3.19	196.74	196.38	222.51	207.95
$\sigma_N^2 = 0.4\sigma_S^2$	4.86	6.88	4.86	6.69	181.46	181.09	271.72	200.99
$\sigma_N^2 = 0.5\sigma_S^2$	6.60	9.23	8.55	9.05	171.46	171.02	292.26	197.93
$\sigma_N^2 = 0.6\sigma_S^2$	8.55	12.12	11.15	11.43	161.99	161.57	304.85	195.04
$\sigma_N^2 = 0.8\sigma_S^2$	13.09	17.66	16.49	16.57	147.86	147.38	313.32	191.80
$\sigma_N^2 = \sigma_S^2$	18.74	23.27	18.95	19.49	138.76	138.22	312.02	191.41
$variable\ \sigma_N^2$	8.84	11.62	8.84	11.62	171.77	171.35	274.01	199.54

As variance increases, the neural solutions prove to perform consistently better than any of the other methods at stake. The latter display non-monotonic and inconsistent evolutions both on their means and variances. As noise increases, all non-supervised methods face large variance and strongly biased estimates (systematically above 100 samples whereas neural solutions remain below 25). The inconsistency of these responses indicates poor estimators. Failure due to noise has been frequently demonstrated inliterature [13, 15], yet even in high noise it is found that the neural solution remains satisfactory.

Boxplots *(figure 5 to 8)* provide us additionally with a compact understanding of the performance of the various estimators and some additional statistics. For each box, the central mark is the median of the error, edges of the box are set to the 25th and 75th percentiles, the whiskers extend to the most extreme data points and outliers are displayed as triangles beyond the whiskers.

It can be observed that all trained "mlps" systematically outperform all non-supervised methods when noise is present. They also consistently perform with small and controlled bias and variance. It is also remarkable that with no noise the correlation methods perform well. In particular, the standard cross-correlation is unbiased, has no variance and thus no outliers.
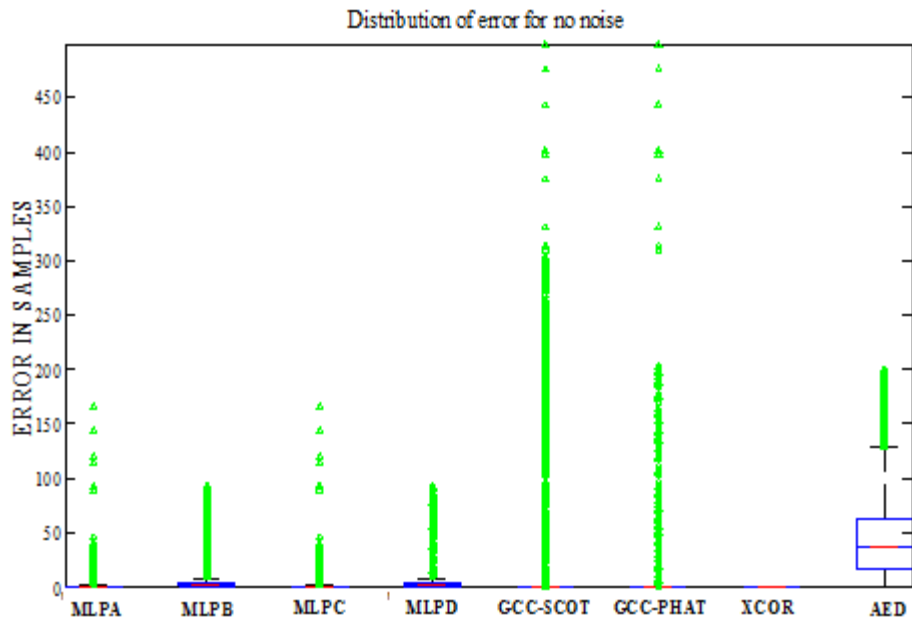
Distribution of error for no noise

Figure 5. boxplot representation of the error distribution of various estimators when noise is absent.

Distribution of error for noise with standard variation 0.5
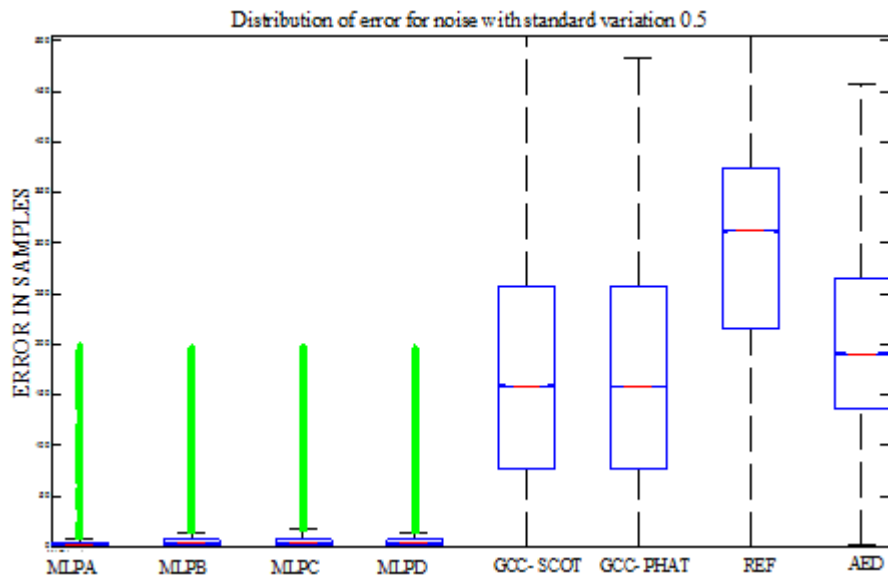

Figure 6. boxplot representation of the error distribution of various estimators when noise variance equals 0.5 signal variance.
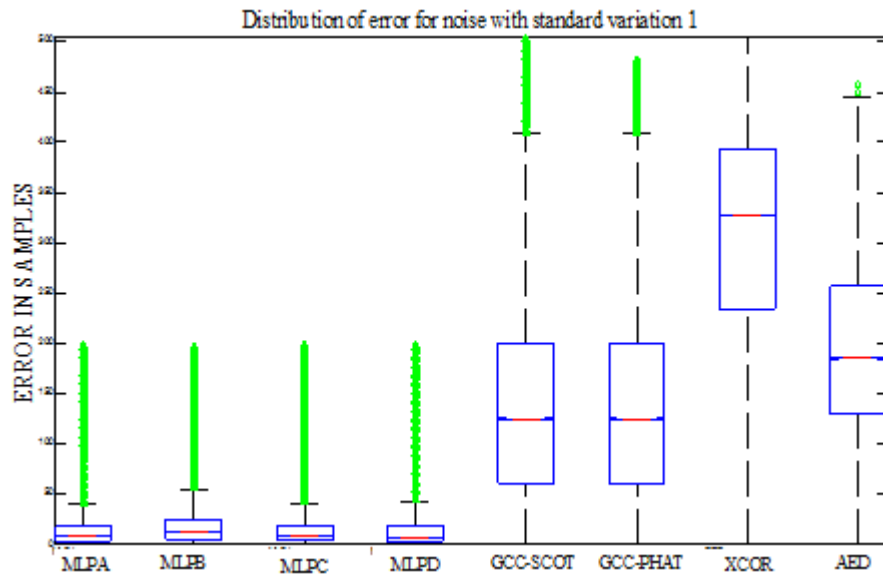
Figure 7. boxplot representation of the error distribution of various estimators when noise variance equals signal variance.
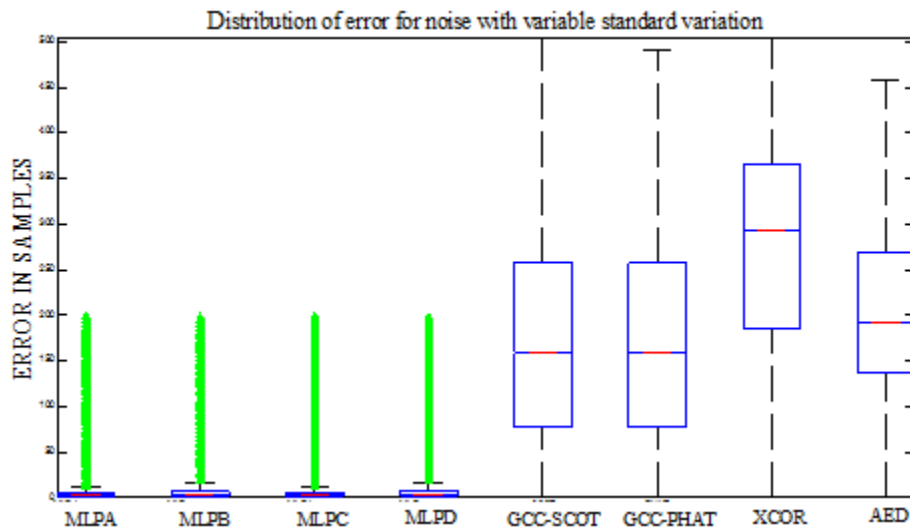

Figure 8. boxplot representation of the error distribution of various estimators with changing noise variance.

5. CONCLUSIONS

In this paper supervised neural networks were used for a successful time-delay estimation and proved to outperform benchmark methods both for the nominal estimation of time-delay and in approximating an ideal time-delay response. As an entry for localization this robust time-delay estimates would produce drastically more consistent location estimates. The integration of these improved time-delay estimators both in underwater and in room acoustics is the object of ongoing research projects.

REFERENCES

[1] Houégnigan, Ludwig, et al. "Space–time and hybrid algorithms for the passive acoustic localisation of sperm whales and vessels." Applied acoustics 71.11 (2010): 1000-1010.

[2] André, M., et al. "Localising Cetacean Sounds for the Real-Time Mitigation and Long-Term Acoustic Monitoring of Noise, Advances in Sound Localization." InTech (2011).

[3] Houegnigan, Ludwig, et al. "Neural networks for the localization of biological and anthropogenic source at neutrino deep sea telescope." OCEANS 2015-Genova. IEEE, 2015.

[4] Carter, G. Clifford. "Coherence and time delay estimation." Proceedings of the IEEE 75.2 (1987): 236-255.

[5] Fowler, Mark L., and Xi Hu. "Signal models for TDOA/FDOA estimation." IEEE Transactions on Aerospace and Electronic Systems 44.4 (2008): 1543-1550.

[6] Carter, G. C. "Time delay estimation for passive sonar signal processing." IEEE Transactions on Acoustics, Speech, and Signal Processing 29.3 (1981): 463-470.

[7] Chen, Jingdong, Jacob Benesty, and Yiteng Huang. "Time delay estimation in room acoustic environments: an overview." EURASIP Journal on applied signal processing 2006 (2006): 170

[8] Huang, Yiteng Arden, Jacob Benesty, and Jingdong Chen. "Time delay estimation and source localization." Springer Handbook of Speech Processing. Springer Berlin Heidelberg, 2008. 1043-1063.

[9] Shaltaf, Samir. "Neural-network-based time-delay estimation." EURASIP Journal on Applied Signal Processing 2004 (2004): 378-385.

[10] Shaltaf, Samir J., and Ahmad A. Mohammad. "Neural networks based time-delay estimation using DCT coefficients." American Journal of Applied Sciences 6.4 (2009): 703.

[11] Knapp, Charles, and Glifford Carter. "The generalized correlation method for estimation of time delay." IEEE Transactions on Acoustics, Speech, and Signal Processing 24.4 (1976): 320-327.

[12] Benesty, Jacob, Yiteng Huang, and Jingdong Chen. "Time delay estimation via minimum entropy." IEEE Signal Processing Letters 14.3 (2007): 157-160.

[13] Benesty, Jacob. "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization." The Journal of the Acoustical Society of America 107.1 (2000): 384-391.

[14] Reed, Feintuch, P. Feintuch, and N. Bershad. "Time delay estimation using the LMS adaptive filterstaticbehavior." IEEE Transactions on Acoustics, Speech, and Signal Processing 29.3 (1981): 561-571.

[15] Carter, G. Clifford, and E. Richard Robinson. "Ocean effects on time delay estimation requiring adaptation." IEEE Journal of oceanic engineering 18.4 (1993): 367-378.

AN EXPERT GAMIFICATION SYSTEM FOR VIRTUAL AND CROSS-CULTURAL SOFTWARE TEAMS

Isaac Chow[1,2] and LiGuo Huang[2]

[1]XO Communications, USA
[2]Department of Computer Science, Southern Methodist University,
Dallas, TX 75275-0122, USA
`isaac.chow@xo.com; lghuang@lyle.smu.edu`

ABSTRACT

Gamification is the concept of applying game elements in non-game context platforms to motivate people to participate in planned activities to achieve goals.

Gamification has been applied to academic fields including software engineering (SE) in recent years. Many gamification implementations in SE have been ad hoc and lacked standardized guidelines. This paper introduces a new concept of building an expert gamification system (EGS) to provide guidelines for the implementation of gamification for virtual and cross cultural software teams (VCCST). The system will extend the core of a regular expert system to include gamification tools, a supplementary database, and an expert knowledge source. The cross-cultural data for the EGS contains the Hofstede's cultural dimensions (HCD).

As more and more VCCST are formed in recent years, many issues have been raised in those teams stemming from miscommunication and cultural conflicts. This paper uses the EGS to help resolve the issues in VCCST.

KEYWORDS

Gamification, Software Engineering, Expert System, Hofstede's Cultural Dimensions, & Theory of Flow

1. INTRODUCTION

Gamification is the use of elements of game design in non-game contexts [1]. It is the application of game elements used to encourage engagement with a product or service. The concept of gamification has been around for more than a decade. There has been a great deal of commercial and academic interest on the use of gamification in recent years.

Some studies show that gamification can motivate engineers in SE if applied appropriately. However, many gamification implementations for SE are lacking a well-defined framework or guidelines. This paper develops an EGS that will provide guidelines for implementing gamification for VCCST.

An expert system is a computer system that emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning about knowledge

[2]. This paper extends the core of an expert system to include gamification tools, a supplementary database with the cross-cultural data, and the expert knowledge source from theory of flow to build the EGS. The core of an expert system consists of the inference engine and the knowledge base. Figure 1 is a high-level view of the EGS.
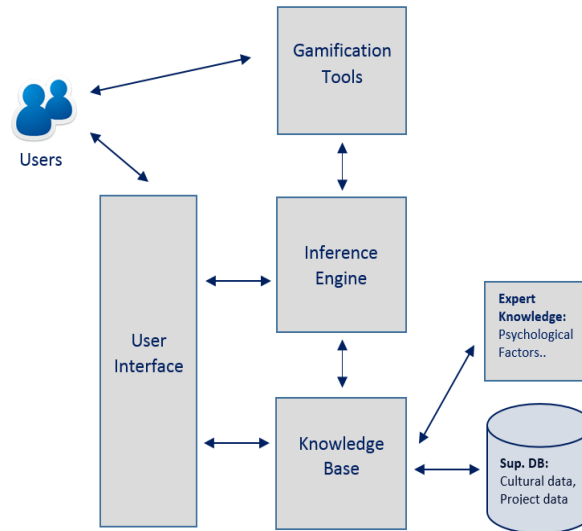

Figure 1. High Level View of the EGS

The EGS consists of the following modules:

1. Gamification Tools
2. User Interface
3. Inference Engine
4. Knowledge Base
5. Supplementary Database
6. Expert Knowledge Source

2. RELATED WORK

Gamification has been extensively researched by different types of researchers in recent years. However, there are very few studies in building an EGS as a tool for implementing gamification. Some gamification studies are intended for learning and discussed in the context of education. Other work tried to verify whether or not gamification works [3]. Some analysis has been done to investigate the gamification mechanisms and see how gamification influences behaviours [4]. Some research has indicated that gamification provides positive effects in motivation. Some studies are descriptive in nature and they have no actual listed experiment results of gamification reported [5], [6], [7], [8]. Several studies explore conceptual frameworks for classroom learning only [9].

For VCCST, some studies describe their conflicts and issues in detail but offer little help or guidance for them [10], [11]. Other research identifies the specific attributes relevant to cross-cultural concerns but does not show how to resolve the issues [12]. Some studies describe how to do cross-cultural surveys [13]. There are very few studies offering solutions for using an EGS.

3. VIRTUAL AND CROSS CULTRUAL SOFTWARE TEAM

Globalization of markets has been a major theme in recent years and more software teams are working in cross cultural environments from outsourcing or insourcing work strategies [10]. It is common for a large software project to have teams in more than one location and in different countries. Many VCCST are formed across the globe with engineers from different parts of the world [14]

It has been reported that the software projects with VCCST have quite a few issues and challenges.Most of the problems are about conflicts in culture and its adverse impact to the dynamics of the team [10], [15]. In a few instances, the software engineers view their co-workers from different cultures as rivals rather than teammates. This paper designs an EGS to promote team work and collaboration for VCCST to smoothen their conflicts and problems.

4. EXPERT GAMIFCATION SYSTEM

The EGS extends the core of an expert system and consists of six components as shown in Figure 1. The core elements of an expert system are the inference engine and the knowledge base. The inference engine is an automated reasoning system that evaluates the current state of the knowledge base, applies relevant rules, and then asserts new knowledge into the knowledge base. The knowledge base represents facts and rules. Knowledge in an expert system may originate from many sources such as reports, databases, case studies, empirical data, and domain experts [2].

Other components in the EGS are the gamification tools, a user interface, a supplementary database, and an expert knowledge source. The gamification tools include the game mechanics and the platform. For example, Microsoft SharePoint can be used as a gamification tool. These tools will contain the software team profile and personal profiles. The profiles will store and show the team scores and individual scores respectively. The user interface facilitates the communications between the project admins, users, and the EGS.

The supplementary database contains the cultural data and project data. The expert knowledge source contains the theory of flow information. The following two sections describe the cultural data and the theory of flow for this EGS.

4.1. Hofstede's cultural dimensions

Hofstede's cultural dimensions (HCD) theory is a framework for cross-cultural communication. It describes the effects of a society's culture on the values of its members, and how these values relate to behaviour using a structure derived from factor analysis [16]. According to Hofstede the most important differences between cultures can be captured by finding to what extent members of these cultures differ with regard to six values [17]:

1. Power distance index (PDI) - The power distance index is defined as the extent to which less powerful people in an organization will accept and expect power to be distributed.

2. Individualism/Collectivism (IDV) – Differences between the degrees within the Individualism vs. Collectivism index.

3. Masculinity/Femininity (MAS) - Differences between the degrees within the Masculinity vs. Femininity index.

4. Uncertainty avoidance index (UAI) - Differences between the degrees within the Uncertainty Avoidance Index.

5. Long-term/Short-term Orientation (LTO) - Differences between the degrees within the Long-Term vs. Short-Term Orientation index.

6. Indulgence vs. Restraint (IND) - Differences between the degrees within the Indulgent vs. Restraint index.

Table 1 shows the Hofstede's values for these cultural dimensions for Australia, China, India, Japan, Netherlands, UK, and USA [17]. The EGS further categorizes the countries into regions such as Australia (AUS), Asia, Europe (EU), and North America (NAM) for further region processing as shown in the second columns of Table 1.

Table 1. Hofstede's Cultural Dimensions

Country	Region	PDI	INV	MAS	UAI	LTO	IND
Australia	AUS	36	90	61	51	31	71
China	Asia	80	20	66	30	87	24
India	Asia	77	48	56	40	51	26
Japan	Asia	50	42	90	88	75	42
Netherlands	EU	38	80	14	53	67	68
UK	EU	30	89	62	30	20	69
USA	NAM	40	91	62	46	26	68

Among the HCD, three dimensions have more effect on gamification. They are Individualism, Masculinity, and Uncertainty:

1. Individualism (INV)

This is the index to indicate how people think with the mentality of "I" or "We." Some countries are very "individual" focused. They focus more on themselves and careers versus their group. The game design should take this into consideration. If the culture is more individual focused, game mechanics such as leaderboards should be used.

2. Masculinity (MAS)

Masculinity is defined as "a preference in society for achievement, heroism, assertiveness and material rewards for success." Its counterpart represents "a preference for cooperation, modesty, caring for the weak and quality of life." [17] This dimension gives important information in setting up the right game mechanics for certain cultures. For instance, many Scandinavian countries have a very low score in MAS. In those cultures, there is an important concept known as Janteloven [8]. Janteloven is essentially a set of rules for encouraging social equality, social stability, and uniformity in which one should never try to stick out from the crowd. The Netherlands is an example with a score of 14 in MAS. The EGS should take this into consideration by not focusing the games on competitiveness and leaderboards. Rather, it should promote notions of equality and creativity for them [18], [19].

3. Uncertainty Avoidance (UAI)

The uncertainty avoidance index is defined as "a society's tolerance for ambiguity" in which people embrace something unexpected, or go away from the status quo. Societies that score a high degree in this index tend to prefer strict codes of behaviour, guidelines, laws, and absolute

Truth. A lower degree in this index shows more acceptance of differing thoughts. Society tends to impose fewer regulations, ambiguity is more accustomed to [17]. The game design should take this into consideration.

4.2. Psychological factors and flow detection

Gamification uses the psychology of engagement to motivate people. Psychology can be applied to the non-technical aspects of SE like personality, teamwork, leadership, decision-making, culture, motivation and human tendencies [15].

Abraham H. Maslow was a renowned psychologist in the field of motivation. He published the famous Hierarchy of Needs in 1943 [18]. It is a pyramid depicting the five levels of needs, namely

1. Biological and Physiological: air, food, shelter, water, sex, & rest

2. Safety: health, protection, stability, freedom from fear & security

3. Love and Belonging: love, intimacy, friendship, family, & affiliating

4. Esteem: self-esteem, independence, status, confidence, achievement & respect

5. Self-actualization: realizing personal potential, self-fulfilment & peak experiences

Self-actualization is the peak level in which a person's full potential is realized. Maslow describes this level as the desire to accomplish everything that one can: to become the most that one can be [18].

4.2.1. Theory of flow and group flow

Self-actualized people are those who come to find meanings to life that are important to them and that fulfil them. Many of them experience a common phenomenon called flow, proposed by Csikszentmihalyi [20].When people are in a state of flow, they will feel focused and concentrated. They may feel a sense of ecstasy, great inner clarity, a sense of serenity, timelessness, and intrinsic motivation. The flow is the balance between the skill level and the challenge of a task. When a task is too difficult, it causes people to worry. When a task is too simple, people will be bored. When a task is balanced between skill and difficulty, the individual will reach a state of heightened focus and immersion [20]. Some research further extends the flow theory for groups. Keith Sawyer developed specific triggers needed for creating group flow [21]. Figure 2 shows the theory of flow. The EGS should try to keep the team in the flow and should be able to detect the flow zone.
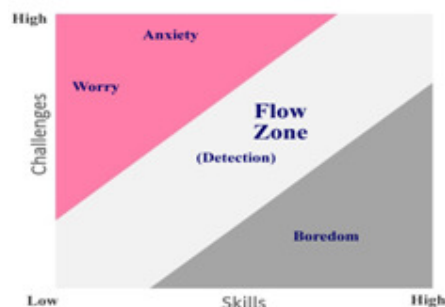

Figure 2. Flow Theory

4.2.2. Flow Zone Detection

The EGS should find ways to measure and detect the flow zone. There are some methods in measuring flow with questionnaires: Nakamura & Csikszentmihalyi enlist the various ways of measuring flow for Concept of Flow; Intrinsic Motivation Inventory with 10 Questionnaires; John Marshall Reeve: Agentic Engagement Scale Questionnaires; Davin Pavlas Play Experience Questionnaire [21]. Other than using the questionnaires, the EGS can use the game statistics to detect the flow zone. Figure 3 shows a simple game flow in which the 'flow' loop is indicated with bigger arrows.
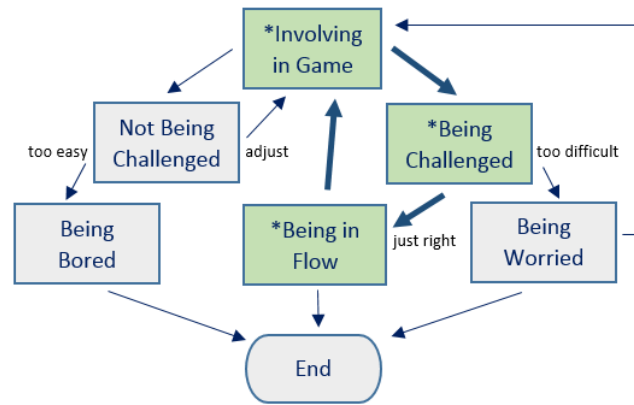

Figure 3. A Simple Game Flow

The EGS should monitor the game and collect the game statistics. The EGS should set up rules to determine if a team is in a flow state. Table 5 shows a sample of the game statistics. The average percentage of hits in that table can be a strong indicator for detecting a flow state. For example, if that percentage is at least at a certain value, then the EGS should assume that the team is in a flow state or zone. It may be a trial and error process to finalize the right value for that percentage. The percentage can be cross checked with the questionnaire method to find the optimal value.

4.3 Determining Game Mechanics and Activities

The EGS should extract the cultural information and use that information to configure the game mechanics in the system. Table 2 contains the game mechanics, the associated emotions, and their values for INV, MAS and UAI. The values represent the recommendations from the EGS for the game mechanics. For example, the value of 'all' in INV means that game mechanic is good for all values of INV. The value of 'high' in INV means the game mechanic is good if INV is high. The same rule applies to the value of 'low'. If all three HCDs have the same 'high' value for the same game mechanics, it reinforces the recommendations. If the HCDs have different recommendations, INV will have the highest preference and MAS will have the second highest preference.

Table 2. Game mechanics and cultural factors

Game mechanics	Emotions	INV	MAS	UAI
Points	Reward	all	all	all
Badges	Achievement	all	high	all
Leaderboards	Competition	high	high	high
Levels	Status	high	high	high
Progress bars	Achievement	all	all	high
Storyline	Engagement	all	all	all
Feedback	Engagement	all	all	high

Challenge/between users	Competition	all	high	all
Virtual gifts	Altruism / Self-expression	all	low	low
System for sharing/exchanging	Collaboration/ Community	low	low	all

Table 3 shows a list of activities with their corresponding values for INV, MAS, or UAI. The rules for recommendation are similar to the ones in Table 2. For example, the value of 'low' in INV for 'helping others' means that activity is recommended when it has a low INV value. The same rule applies to other activities and HCDs. Also, the HCD values in the table can be changed when more knowledge or experience is gathered.

Table 3. Activity table with cultural factors

Activities	INV	MAS	UAI
Inviting peers to design review	all	all	all
Participating in a design review	all	all	high
Inviting peers to a code review	all	all	all
Participating in a code review	all	all	all
Inviting peers to a test plan review	all	all	all
Participating in a test plan review	all	all	all
Helping others	low	low	all
Giving people credits	low	low	all
Being a team player	low	low	all
Resolving issues	all	low	high
Collaboration activities	low	low	all
Innovative idea	all	all	all
Voting for an idea	low	low	all
Beating the deadline	high	all	high
Documenting code	all	all	high
Best practice	all	all	high
Writing unit tests	all	all	high

4.4. Sample Data

A simple prototype is developing at this time. The preliminary findings are encouraging. Table 4 contains some sample inputs to the EGS for demo purpose. These inputs drive the EPS based rules mentioned in 4.3 and the 4.2.2 example.

Table 4. Inputs to the system

Attributes	Examples
Project Name	VCCST Enhancement
Project Description	A demo to show how the EGS works
Domain of interest	Software engineering
Team Info	Cross-Cultural
Team Location	Virtual
Countries	USA, India, China
# of team members	20, 12, 18
# of analyst, developers, testers	4,12,4
SE Methodology	Agile

The ESG will recommend a list of game mechanics and activities for the software project. Each individual group is customized. The EGS collects statistics from the project for analysis. One of the important tasks is to detect the flow zone for the software project team.

Table 5 shows a sample of the game statistics. The EGS uses the values from INV, MAS, UAI, and the team cultural data to determine the initial 'target %' for the flow zone. As more project data are stored and accumulated, the EGS can use the past data as reference and knowledge for further 'target %' refinement and fine tuning. If the HCD recommend the activities, the 'target %' for those activities will be higher than others.

Table 5. Sample game statistics

Point for activities	# of players	# of hits	Hit %	Target hit %
inviting peers to design review	16	15	93.75	80.00
participating in a design review	16	9	56.25	80.00
inviting peers to a code review	12	8	66.67	80.00
participating in a code review	12	6	50.00	80.00
inviting peers to a test plan review	16	7	43.75	80.00
participating in a test plan review	16	4	25.00	80.00
helping others	20	6	30.00	70.00
giving people credits	20	15	75.00	75.00
being a team player	20	9	45.00	75.00
resolving issues	20	5	25.00	70.00
collaboration activities	20	12	60.00	70.00
innovative idea	20	3	15.00	60.00
voting for an idea	20	4	20.00	70.00
beating the deadline	20	4	20.00	80.00
documenting code	12	8	66.67	80.00
best practice	16	5	31.25	75.00
writing unit tests	12	11	91.67	80.00
Average %:			47.94	75.59

A: Number of analysts	4
B: Number of developers	12
C: Number of testers	4
D: Total players	20

In Table 5, the 'Hit %' is calculated by '# of hits' / '# of players'. In this sample data, there are 4 analysts, 12 developers, and 4 testers – a total of 20 players. The system assigns players to the activities according to their roles. The system collects the games statistics including the number of hits from each activity. The system compares the average 'Hit %' and the average 'Target hit %' to determine the status of the flow zone. In Table 5, 47.94% is the average 'Hit %' and 75.59% for the average 'Target hit %'. Since the average 'Hit %' is lower than the average 'Target hit %', the EGS will infer that the project team is not in a flow zone. The EGS will send alerts to the system admins and project managers. They can choose to adjust the game mechanics/activities and or make comments about the project. The project data will be stored and served as reference and knowledge data for future projects. A SharePoint community site will be used as a platform for collaboration, community, and data collection.

5. CONCLUSIONS

This paper introduces a new concept for building an EGS for the implementation of gamification for VCCST. The goal of this paper is to open up another way of thinking in designing gamification for SE using cross cultural data, the flow theory, and the past project data for the VCCST.   Designing and building the EGS will help to create a user-friendly system tool for gamification implementation. The concept can be further extended to other domains for resolving different problems using gamification.

For future work in this EGS, one area to investigate is how to better detect the state of flow and group flow using the project data or other means. A more formal algorithm should be developed for that. Also, incorporating more cultural data to develop a robust knowledge extraction methodology for the system could be beneficial. As for the VCCST, more case studies should be done using gamification with qualitative and quantitative approaches. Supporting and fostering the gamification research to improve the software development process in the cross-cultural environment could yield impactful results for VCCST's.

REFERENCES

[1]   Deterding, S., Sicart, M., Nacke, L., O'Hara, K. & Dixon, D. (2011). Gamification: Using game design elements in non-gaming contexts. In Proceedings of CHI 2011. New York: ACM Press, 2425-2428.

[2]   Wikipedia, "Expert system", (2016), https://en.wikipedia.org/wiki/Expert_system

[3]   J Hamari, J Koivisto, H Sarsa, "Does gamification work? -a literature review of empirical studies on gamification" System Sciences (HICSS), 2014, pp. 3024-3034.

[4]   Bunchball (2010), "Gamification 101: An introduction to the Use of Game Dynamics to Influence Behavior"

[5]   Philipp Herzig, Michael Ameling, Alexander Schill, "A Generic Platform for Enterprise Gamification", 2012 Joint Working Conference on Software Architecture & 6th European Conference on Software Architecture, 978-0-7695-4827-2/12 © 2012 IEEE, pp. 217-223.

[6]   Foong Li Law, Zarinah Mohd Kasirun, Chun Kiat Gan, "Gamification towards Sustainable Mobile Application", 2011 5th Malaysian Conference in Software Engineering (MySEC), pp. 348-353.

[7]   Bogdan Vasilescu, "Human Aspects, Gamification, and Social Media in Collaborative Software Engineering", ICSE '14, May 31 – June 7, 2014, Hyderabad, India, Copyright 14 ACM 978-1-4503-2768-8/14/05

[8]   Marius Kalinauskas, "GAMIFICATION IN FOSTERING CREATIVITY", ISSN 2029-7564 SOCIAL TECHNOLOGIES 2014, 4(1), pp. 62–75.

[9]   Jorge Simões a, Rebeca Díaz Redondo b, Ana Fernández Vilas, "Computers in Human Behavior", www.elsevier.com/locate/comphumbeh, 2012 Elsevier Ltd

[10]   Michael Barrett, "Knowledge Sharing in Cross-Cultural Software Teams" Working Paper Series, 18/2007, Judge Business School, University of Cambridge

[11]   Hannu Jaakkola, Anneli Heimbürger, "Cross-Cultural Software Engineering", Informatologia 42, 2009., pp. 256–264.

[12] A Sutharshan. S P Maj "Enhancing Agile Methods for Multi-cultural Software Project Teams" Modern Applied Science Vol. 5, No. 1; February 2011, ISSN 1913-1844 E-ISSN

[13] "Guidelines for Best Practice in Cross-Cultural Surveys" THIRD EDITION, Copyright © 2011 by the Survey Research Center, Institute for Social Research, University of Michigan, ISBN 978-0-9828418-1-5, Library of Congress Control Number: 2010937486 2011, pp.163-194.

[14] Christof Ebert, "Collaboration Tools for Global Software Engineering", IEEE SOFTWARE by the IEEE Computer Society, March/April 2010

[15] Marczak, Treude, Filho, Steffens, Singer, Redmiles, "Studying Gamification as a Collaboration Motivator for Virtual Software Teams: Social Issues, Cultural Issues, and Research Methods" CSCW '15 Companion, Mar. 14–18, 2015, Vancouver, BC, Canada. ACM

[16] Wikipedia, "Hofstede's cultural dimensions theory", (2016) https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory

[17] Geert Hofstede, "Cultural Dimensions", (2010), https://geert-hofstede.com/countries.html

[18] Erica Sahlgren, Kendra Knorst, "Culture's impact on gamification", 2016 ,Linköping University | Department of Management and Engineering, ISRN-number: LIU-IEI-FIL-G--16/01497—SE

[19] Rilla Khaled, "It's Not Just Whether You Win or Lose: Thoughts on Gamification and Culture", CHI 2011, May 7–12, 2011, Vancouver, BC, Canada. ACM 978-1-4503-0268-5/11/05.

[20] Csikszentmihalyi, M, "Flow and the Psychology of Discovery and Invention". Harper, Perennial, New York (1997)

[21] Enterprise Gamification Wiki, "Flow theory", (2015), http://www.enterprise-gamification.com/mediawiki/index.php?title=Flow-theory

AUTHORS

Isaac Chow

https://www.linkedin.com/in/isaac-chow-1aa398b

Isaac got his B.S. and M.S. in CS at the University of North Texas in mid 1980's.

Isaac went to teach in Southwestern Oklahoma State University as a computer science instructor from 1985 to 1989. Isaac worked for Texas Instruments as a programmer analyst for three years working on mainframe MVS system and received quality recognition and excellence award. He worked for AT&T for 10 years as a Senior Technical Staff Member and Computer Sciences Corporation for 3 years as a Computer Scientist. He worked on Internet based multi-platform systems using MVS, Unix, C/C++, Java, Sybase, and received the Circle of Excellence award.

Isaac currently works for XO Communications on developing intranet applications using SharePoint, ASP .net, and c#. Isaac has worked on the Microsoft SharePoint and used it as a collaboration tool for many years. Isaac is a doctoral student in software engineering at SMU. He plans to contribute to the academic community through research and with his intensive software experience.

Dr. LiGuo Huang

http://lyle.smu.edu/~lghuang/

Dr. LiGuo Huang Associate Professor, Department of Computer Science and Engineering at the Southern Methodist University (SMU), Dallas, TX, USA. She received both her Ph.D. and M.S. from the Computer Science Department and Center for Systems and Software Engineering (CSSE) at the University of Southern California (USC). Her current research focuses on mining and learning from systems and software engineering repositories, software quality assurance and information dependability, process modelling, simulation and improvement, stakeholder/value-based software engineering, and software metrics. Her research has been supported by NSF, US Department of Defense (DoD) and NSA. She had been intensively involved in initiating the research on stakeholder/value-based integration of systems and software engineering and published related papers in IEEE Computer and IEEE Software.
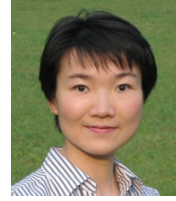
INTENTIONAL BLANK

DYNAMIC QUALITY OF SERVICE STABILITY BASED MULTICAST ROUTING PROTOCOL FOR MANETs (DQSMRP)

M.Vijayalakshmi[1], and Dr.D.SreenivasaRao[2]

[1]Department of Electronics and Communication Engineering, GNITS, India
mvlakshmi_gnits@yahoo.co.in
[2]Department of Electronics and Communication Engineering, JNTUH, India
dsraoece@gmail.com

ABSTRACT

Mobile ad hoc networking does not possess any fixed infrastructure and hence, stable routing is the major problem. The mobility nature of MANET's node facilitates rediscovery of a new path to organizing a routing. In order to intensify the Quality of Service and routing stability in MANET, we propose a Dynamic Quality of service Stability based Multicast Routing Protocol by modifying the Cuckoo Search Algorithm through a modernizing mechanism which is derived from the differential evolution algorithm. Tuned CSA is a combined feature of CSA and DE algorithms. Periodically, each node in the network creates neighbour stability and QoS database at every node by calculating the parameters like node and link stability factor, bandwidth availability, and delays. Finally, multicast path constructs route request and route reply packets, stability information and performing route maintenance.

KEYWORDS

Manets, QoS, Tuned CSA, DQSMRP

1. INTRODUCTION

Mobile Ad hoc Networks (MANETs) are foremost in class kind of wireless communication networks in which mobile nodes link on an on-demand basis. MANETs are having the ability to configure on their own and easing peer-level communications among mobile nodes independent by having no fixed infrastructure. When the route traffic occurs, constructing a MANET with each node's to continuously retain the information remains a major challenge. Normally, MANETs are used for group communications where multicast protocols are efficiently compared to unicast protocols since they improve the efficiency of the wireless links in MANETs. The multicast protocols are also used as application demands for transmitting copies from either single or multiple sources to the receivers.

Multicasting reduces the communication costs by sending the single copy of the data to multiple recipients through which it minimizes link bandwidth, processing and transmission delay. Previous researchers have developed different algorithms to improvise the ability for selecting the

routing path(s) and thereby, effectively met the desired Quality of Service (QoS) [1]. This infrastructure results in a highly dynamic topology causing a challenging task for QoS Routing .

When compared with well-established networks such as Wi-Fi, Global System for Mobile communication (GSM) and Code Division Multiple Access (CDMA), MANET was not reliable to meet QoS performances. The MANET performance is determined by the QoS through the basic parameters like bandwidth or delay or loss. QoS routing addresses both the issues of deciding a path from source to destination and fulfilling the QoS constraints like bandwidth, delay and loss. Route selection method is used to fulfill QoS requirements. The goal of QoS routing strategy is to maximize the network resources in order to reduce the constraints in MANET's mobile nodes.

In real time applications, QoS routing protocol works based on multi-hop mobile networks [3, 4]. Particularly, QoS requirement depends on constraints such as link constraints, path constraints and tree constraints. End to End delay is usually affected by path constraint whereas, delay-jitter [5] by tree constraint. In order to meet all the constraints, robust techniques are required to solve the QoS requirements. Heuristic techniques are better to solve these constraints rather than deterministic methods. Researchers have studied various improvements based on heuristic methods [5, 6] and at the same time devised various solutions to solve QoS routing problems using heuristic methods [7].

Ad hoc mobile routing (AM Routing) was studied through which a specific unicast routing protocol was identified as more feasible than other unicast routing protocols (8). In QoS routing, network topology was constructed in which bandwidth link works in destination point [9]. A new protocol was proposed for ad hoc networks which are used to establish and maintain a shared mesh for each multicast group called Policy-based Unified Multi-radio, Architecture for Agile Mesh Networking (PUMA) [10]. A new single mixed metrics have been delineated for multi-constraint QoS routing [11]. Previous literaturehas found that Agent Based Multicast Routing Scheme (ABMRS) on MANET's could be used as a set of static and mobile agents for QoS routing [12, 13]. Optimized link state routing protocol was launched and scrutinized in different research papers [14].

Besides that, several heuristic methods are examined to study the multi-constraint QoS routing. In [15], Simulated Annealing (SA) was introduced to solve the QoS routing problems. Reference [16] discussed Bees Life Algorithm (BLA) to solve the QoS multicast routing problem and constraints such as delay, allowed jitter, and requested bandwidth have been studied. Similarly, a Tabu search [17] based method concentrates on two important constraints in QoS such as bandwidth and end-to-end delay. Ant algorithm optimization was globally accepted for the effective and systematic handling of QoS routing [18, 19]. Harmony search (HS) algorithm, a unique algorithm was used to find the path of the MANET and it was discussed based on bandwidth-delay-constraints in multicast routing problem [20].

Hybrid methods have been proposed to address the delay of premature convergence [21]. A PSO-GA algorithm was evolved using a combination of both particle swarm optimization (PSO) method and genetic algorithm (GA) which provides an effective and efficient search for the solution. In [22], the new fuzzy genetic algorithm is also introduced for QoS multicast routing is also discussed.

Thus in this paper, we propose Dynamic Quality of Service Stability based multicast routing protocol (DQSMRP) for the multi-constraint QoS routing problem. Recently, Cuckoo Search (CS) is derived from reproduction strategies [23] which are widely used in heuristic based algorithms. [24, 25].It is more exhaustive and provides quick search to obtain better solutions to routing in more reliable and robust manner.

2. PROBLEM DEFINITION, RESEARCH METHODOLOGY

The optimization problem is used to solve the QoS routing problem which is the primary objective to find a multicast tree constraints based on the cost function in order to reduce the practical constraints. Four different constraints are considered for formulating the current study such as delay, packet loss, bandwidth and jitter. The following equation provides formulation data for computation [5] QoS routing function is formulated as:

$$MinC(H(x,s)) = C_c + \delta_1 C_b + \delta_2 C_d + \delta_3 C_{dj} + \delta_4 C_{pl}$$

....(1)

Reasons to merge both cost optimization and multi-constrained routing are as follows:

➢ Formulations are based on multiple constraints and the multicast structure in order to spot an attainable path for each node having source and destination.

➢ To shrink the network resource utilization by users.

In progressively,

$$C(H(x,s)) = \sum_{e \in H(x,S)}' c(y)$$

.... (2)

The Delay Constraint is defined as the allowed delay limit that exists in any branch of the network. When there is no link provided between any two nodes, a penalty occurs.

$$Delay(R(x,y)) = \sum_{e \in R(x,y)}' dl(x) + \sum_{n \in R(x,y)}' dl(y)$$

....(3)

The Bandwidth Constraint is the requirement of average bandwidth among all tree branches in the network.

$$Bandwidth(R(x,y)) = min(bw(x)), x \in R(x,y)$$ ....(4)

Specifically $\delta_1, \delta_2, \delta_3, \delta_4$ are the bandwidth, delay, delay-jitter and packet-loss for the corresponding penalty constants. The objective of the initiated solution is to solve the multicast QoS routing problem as described by numerical experiments in the coming section.

2.1 Tuned Cuckoo Search Multicast Routing Protocol

A technique that produces simple implementation and quality solutions is the best solution provider. This paper provides better results over the existing algorithms such as meta-heuristic techniques. Before analyzing the proposed method, the cuckoo algorithm is discussed as follows.

2.1.1. An Overview of Cuckoo Search Algorithm

By analogy of reproduction strategy, CS algorithm is a well known prominent heuristic search algorithms which follow the principle of cuckoos (chosen nests of other birds for laying eggs) [26, 27]. Generally, the host bird would not permit the cuckoo eggs to easily differentiate and vice-versa. The important optimization procedure is as follows:

- **Initial Solution:** Cuckoo egg symbolizes a set of solutions and its value of dimensions was extracted from various nests randomly;

- **Next Generation:** The best eggs with common solutions only able to proceed to the next generation.

- **Acceptance Rule:** New egg will succeed in the nest if eggs are unknown

Important step of CS are explained as follows:

- Step 1- Randomly set "Pop" host nests, a group of the population. (e.g., $P = \sum_{i=1}^{pop} P_i$)

- Step 2- Evaluate the fitness of all nests $F = \sum_{i=1}^{pop} f(P_i)$

- Step 3- Do Iterations < Max.Iterations

- Step 4- Produce a cuckoo egg P^i by applying a Lévy flight from a random nest

- Step 5- Compute the fitness of $f(P_i)$

- Step 6- Check for existence: choose again a random nest J

- Step 7- if $f(p_i^1) < f(P_i)$ then

- Step 8- $P_{j=}\ p_i^1$

- Step 9- end if

- Step 10- Remove a fraction of the worst nests at the rate of $R_c^{\square}$

- Step 11- Randomly, Building new nests at new locations using Levy flights replacements

- Step 12- End While

The benefit of this heuristics is its simplicity. The single parameter of CS algorithm should be considered enough when compared with PSO and Ant colony algorithms.

2.1.2 Node Stability

For efficient better packet delivery services in forwarding group, the mobile node is to be very stable .The way of providing stability is that moving node around its current position. Node stability metric have been used in previous work of authors is given in [28]. It is studied that stable path nodes have been found from packets from the source to the multicast group. Self-stability and neighbor nodes stability are two metrics to represent node stability.

Following steps to find stability of a node:

> In MANET's nodes, it establishes self-stability, i.e., node movement is relative to its previous position.

> MANET's node gives the node stability factor from self-stability and neighbor stability. Actually, stability is based on the distance between the movements of nodes positioning within transmission range. Stable node is created when movement within transmission range.

2.1.3 Neighbor node stability

It can be defined as how the node is being connected to its neighbour to obtain self-stability. Whenever nodes are in transmission range, it can exchange messages with each other. Each node consists of connectivity information, the signal stability of one-hopneighbors, and also maintains a neighbor list. The degree of a node n is denoted as number of links (or nodes) which connected to it and denoted as ND.

$$N_s(t) = \propto \times \frac{1}{ND} \sum_{i=1}^{ND} S_s^i(t) + (1-\propto) \times N_s(t-1)$$

.... (5)

Where, α is the weighting factor (lies between 0 and 1) and is distributed between 0.6 and 0.7 since they yield better results in simulation.

2.1.4 Link Stability

Link stability between the nodes defines quality and existence of the connection. It is reckoned based on two parameters:

- Received Signal Strength and

- A lifetime of the Link.

- The algorithm 1 represents a pseudocode for updating link stability status between the nodes. The different parameters used in the algorithm are as follows:

- Lifetime – Duration of continuous connectivity between the nodes

- Lifetime threshold – Indicates the maximum limit of link lifetime that decides link stability

- Link stability status – It is a Boolean variable that defines link stability between the nodes.

- Recent – Indicates most recent response received for a Hello packet from a neighbor

- P – Represents Total Number of Hello packets

- Received signal strength – Represents the strength of the signal received from a neighbor

- Signal threshold – Is acceptable signal strength to be received from neighbors

2.1.5 Algorithm: Link stability status between the nodes

- Step 1: P = No of Hello Packets;

- Step 2: Lifetime = 0;

- Step 3: Link stability status = 0;

- Step 4: Recent = 0;

- Step 5: Lifetime threshold = P × Hello Packet Interval;

- Step 6: While P > 0 do

- Step 7: If received signal strength ≥ signal threshold then

- Step 8: Lifetime = lifetime + 1;

- Step 9: Recent = 1;

- Step 10: P = P−1;

- Step 11: else

- Step 12: Recent = 0;

- Step 13: P = P−1;

- Step 14: end if

- Step 15: end while

- Step 16: lifetime sec = lifetime × Hello Packet Interval;

- Step 17: if (lifetime sec > lifetime threshold) and (Recent)

- Then

- Step 18: link stability status = 1;

- Step 19: else

- Step 20: link stability status = 0;

- Step 21: end if

2.1.6 Delay Estimation

For delay estimation, an arbitrary node that contributes to traffic forwarding using the M/M/1 queuing system is modeled. This queue represents a single queuing station with a single server [29]. The authors assume that the contributing nodes are served by a single server with first come first serve queuing policy. Packets arrive according to a Poisson distribution rate (λ), and the probability distribution of the service rate is exponential, denoted by μ. The maximum size of the queue in every node is represented by K.

To satisfy delay requirements in multimedia real-time applications, packets must be received by multicast receivers which satisfy the application delay constraints. When a packet is to be sent either by a source node or forwarding group of nodes, it experiences three types of delays: queuing, contention and transmission delay. The total delay considered over a link between two nodes is given by

$$d_{Total} = d_q + d_c + d_t$$

$$\dotfill (6)$$

The queuing delay denoted by d_q is the delay between the time the packet is assigned to a queue and the time it starts transmission. The packet waits and gives other packets in transmission queue are transmitted at this particular time. This is the amount of time a packet is spent in the interfacing queue (T).

2.1.7 Bandwidth Estimation

Bandwidth is one of the key principles for assessing Quality of service (QoS). The authors considered their previous work presented in [30], to calculate the available bandwidth based on their channel status of the radio link to calculate the idle and busy periods of the shared wireless media. Bandwidth is measured by the node activities and its surrounding neighbors using the channel utility.

In IEEE 802.11 MANETs, once it is achieved to gain the channel access, a node can be able to transmitting data packets. Hence, a node senses the channel and estimates bandwidth by using the idle and busy times in a predefined interval. This is expressed by the following equation.

$$BW = \frac{T_{idle}}{T_{interval}} \times C$$

$$\dotfill (7)$$

2.1.8 Route Discovery Process using Tuned Cuckoo Process

Multicast stable QoS path creation involves two phases: a request and a reply phase. Request phase invokes route discovery process to find routes having a group of receivers using stable and QoS intermediate nodes. It helps to reply phase for updating of RIC and also performs confirmation process. It acts as intermediate nodes that help to create a multicast mesh to the group of receivers.

RR packets are used in source node to find the route to its group of receivers. The sequences of operations that occur are as follows:

- Source node prepares an RR packet with an application of bandwidth and delay requirements.

- Selective transmission of RR packet to neighbors once satisfies stability criteria, i.e., SFBN greater than SFTH, and bandwidth requirement, i.e., estimated bandwidth greater than twice the application requirements.

- Once received, node discards RR packet (both sequence number and source address).

- If RR packet is not a duplicate, check if the availability of Routing Information Cache (RIC); if available, RP packet create and start reply propagation to the source.

- If RR packet is a duplicate, then discard it and stop transmission of RR packet.

- If not duplicate and no route available in RIC, transmit the RR packet by updating its fields (route record, Stability Factor Between Nodes (SFBN) record, bandwidth record, delay record, time to live, and nexthop address) to its neighbors as in step 2.

- Perform steps 3 to 6 (Till destination reached).

- If the receiver is not reached within certain hops, send RE packet to the source node.
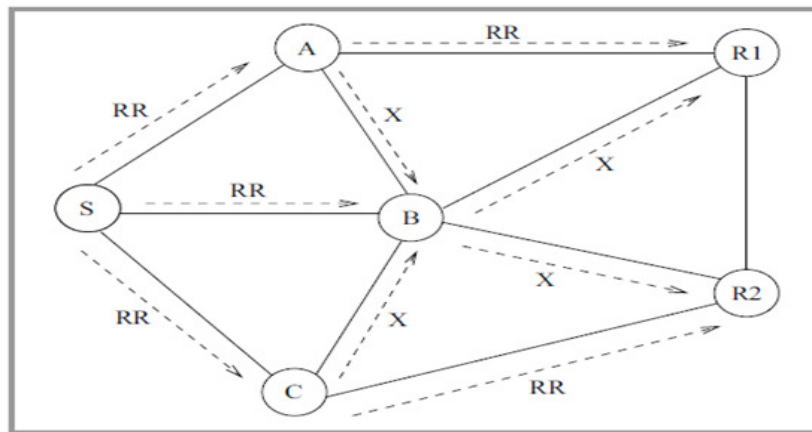
Figure 1: Route request paths from S to R1 and R2

3. RESULTS AND DISCUSSION

3.1 Simulations and Performance Evaluation

In this section, the performance of proposed protocol with DQSMRP and Fully Distributed Multicast Routing Protocol (FDMRP) [31] is compared, through a considerable set of simulations. These etiquettes have been taken for comparison because both are mesh based. These protocols are collated in terms of packet delivery ratio, control overhead and average end-to-end delay. Simulations considering the values of the production parameters are taken for several iterations, and these are computing the mean. The values lying within 95% of the confidence

interval of the mean are used for computing the mean value, which is shown in the graphs in result analysis section. The various network scenarios have been simulated. Simulation environment consists of four models: Network, Channel, Mobility and Traffic. An ad hoc network is created at an area of l × b square meters as network model; it is having N number of mobile nodes spaced randomly. Limited bandwidths in coverage area around each node have shared to its neighbors. It is assumed that the operating range of transmitted power and communication range are constant.

3.1.1 Performance Parameters

Following metrics have been used to scrutinize the performance:

- **Packet Delivery Ratio (PDR)** : The ratio of a number of average data packets received at the multicast receivers to the number of data packets sent by the source.

- **Packet Over head** : The ratio of control packets sent to the network to the total number of average data packets delivered to the receivers.

- **Average end-to-end Delay**: The average delay experienced by the successfully delivered packets in reaching their receiver.

Table 1: Simulation scenario

No. of Nodes	50,100,150 and 200.
Routing Protocol	DQSMRP
Area Size	1000 X 1000
Mac	802.11
Radio Range	250m
Simulation Time	10 sec
Traffic Source	CBR
Packet Size	512
Receiving Power	0.395
Sending power	0.660
Idle Power	0.035
Initial Energy	10.0 J
Node Speed	5,10,10 and,20 ms
No of destinations	2,4,6 and 8

3.2 Result Analysis

The simulation results of DQSMRP and FDMRP present the comparison results by varying number of multiple destinations and node speed. According to the simulation results in Fig 2 to Fig 7, the performance of TCSMRP is more reliable than FDMRP. The execution of DQSMRP packet delivery ratio is gathered more than 14.2% compare to FDMRP. Moreover, the rendering of DQSMRP startsto end delay is less than FDMRP. The performance rate is almost 18% less and the delay ratio is varied with respective of node speed and number of destinations. According to the overhead performance scenario, the overhead rate is 22% less than FDMRP.
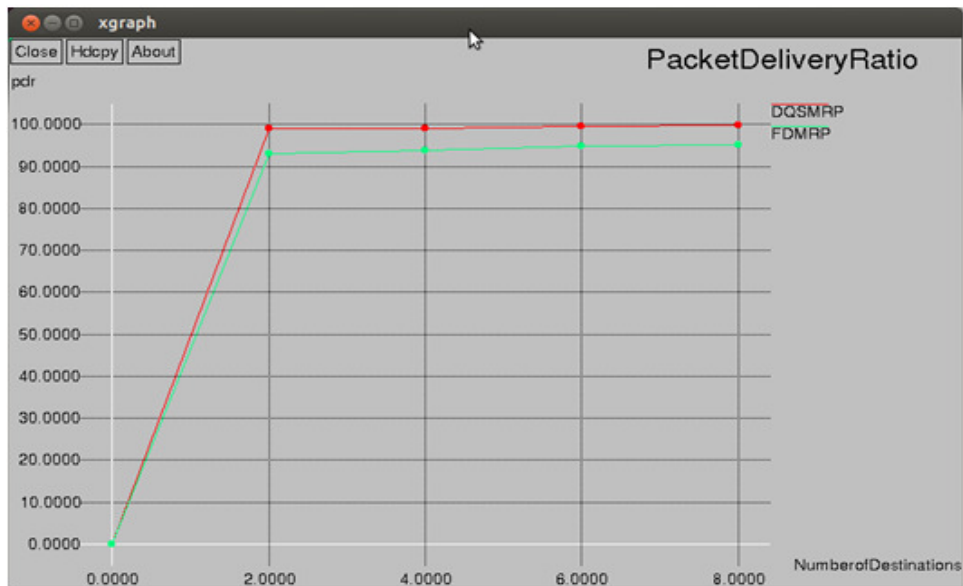
Figure 2: PDR vs Number of Destinations
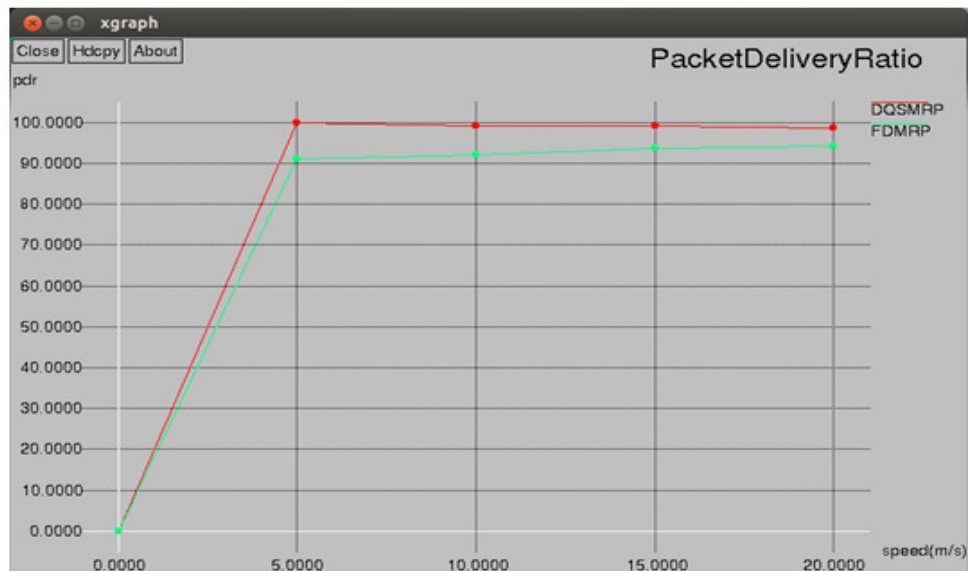
Figure 3: PDR vs Node Speed
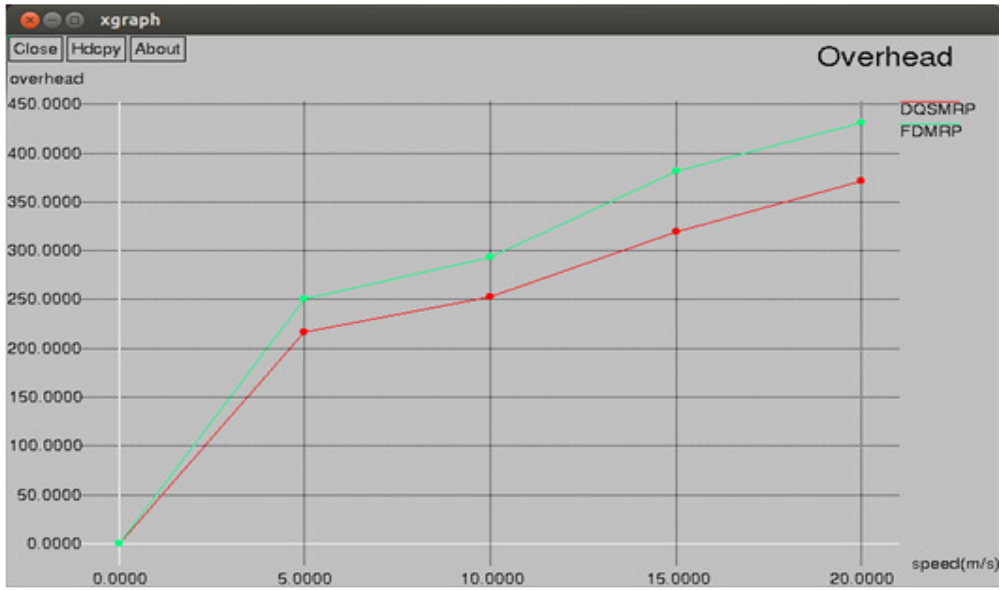
Figure 4: Overhead vs Node Speed
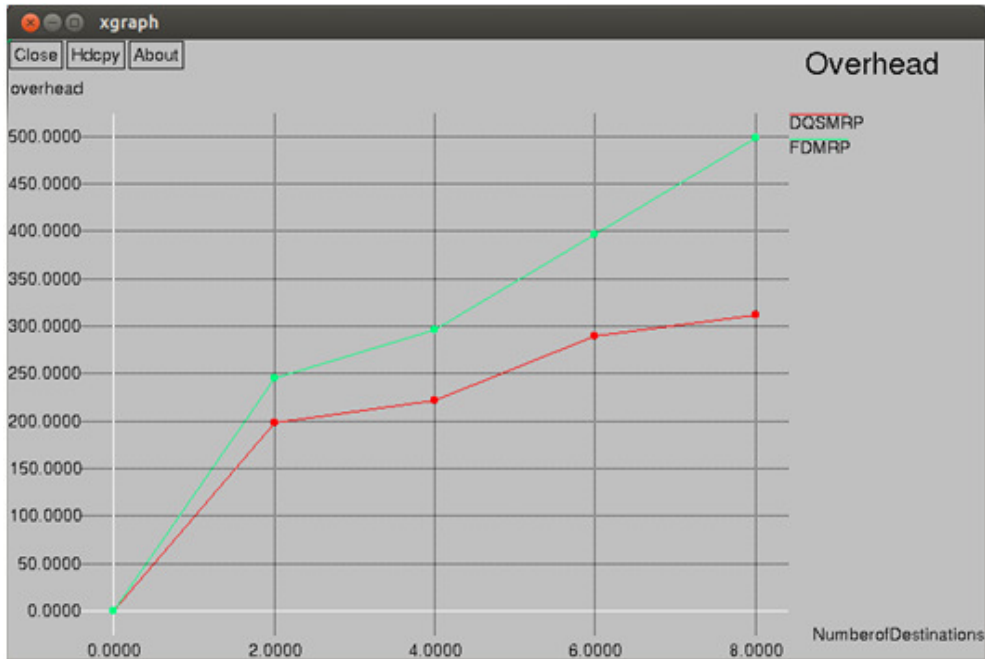
Figure 5: Overhead vs Number of destinations
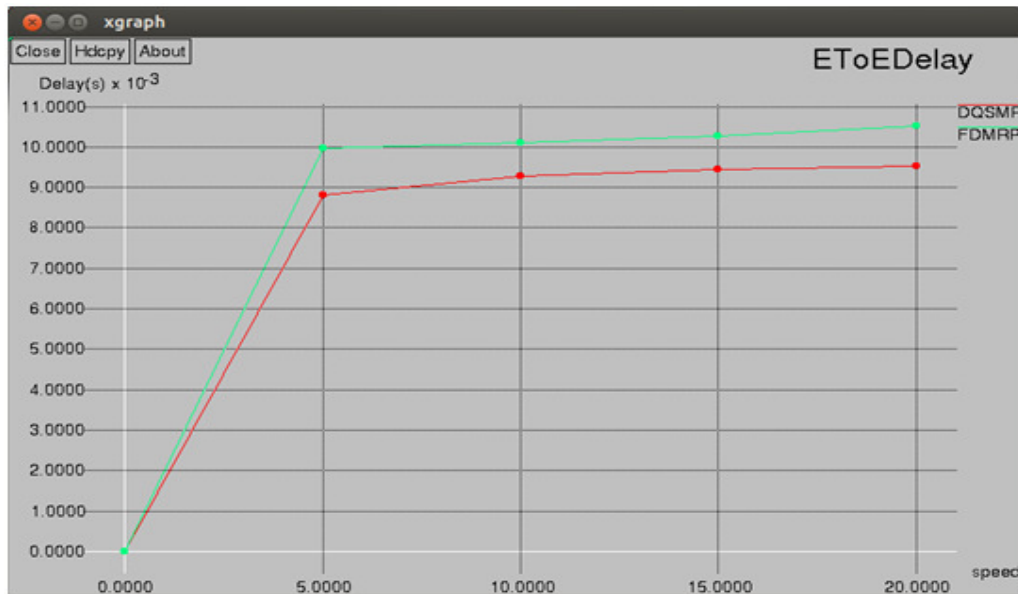
Figure 6: End - End Delays vs Speed

Figure 7: End - End Delays vs Number of Destinations

4. CONCLUSIONS

This paper clarifies about a hybrid computational intelligent algorithm i.e., Tuned Cuckoo Search Algorithm (TCSA), that includes the dominant attribute of two different heuristic techniques have been used to solve QoS multicast routing. The QoS routing in a MANET is a complicated to determine a prime routing. The general conclusion from presented simulation experiments reveals that proposed routing protocol performs better than FDMRP in terms of packet delivery ratio,

packet overhead and average delay from start to end as a function of varying number of receivers, sources and node speeds. In future works, the author's aims to study further by comparing our Tuned Cuckoo Search Multicast Routing Protocol with some more QoS based routing protocols in MANETs. Numerical experiments have shown that the propounded hybrid TCSA algorithm provides superior search multicast trees, particularly in both swift convergence and robustness, when compared to other algorithms. This proposed algorithm shall be the lead for real-time networks with suitable refinement based on when and where it utilized.

REFERENCES

[1]   R. A. Guérin and A. Orda, "QoS routing in networks with inaccurate information: theory and algorithms," IEEE/ACM Transactions on Networking, vol. 7, no. 3, pp. 350–364, 1999.

[2]   D. H. Lorenz and A. Orda, "QoS routing in networks with uncertain parameters," IEEE/ACM Transactions on Networking, vol. 6, no. 6, pp. 768–778, 1998.

[3]   L. Hanzo II and R. Tafazolli, "A survey of QoS routing solutions for mobile ad hoc networks" IEEE Communications Surveys & Tutorials, vol. 9, no. 2, pp. 50–70, 2007.

[4]   C. R. Lin and J.-S. Liu, "QoS routing in ad hoc wireless networks," IEEE Journal on Selected Areas in Communications, vol. 17, no. 8, pp. 1426–1438, 1999.

[5]   L. Zhang, L. B. Cai, M. Li, and F. H. Wang, "A method for least-cost QoS multicast routing based on genetic simulated annealing algorithm," Computer Communications, vol. 32, no. 1, pp. 105–110, 2009.

[6]   X. Yuan and X. Liu, "Heuristic algorithms for themulti-constrained quality of service routing," inProceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01), vol. 2, pp. 844–853, IEEE, 2001.

[7]   M. Abolhasan, T. Wysocki, and E. Dutkiewicz, "A review of routing protocols for mobile ad hoc networks," Ad Hoc Networks, vol. 2, no. 1, pp. 1–22, 2004.

[8]   J. Xie, R. R. Talpade, A. Mcauley, and M. Liu, "AMRoute: ad hoc multicast routing protocol," Mobile Networks and Applications, vol. 7, no. 6, pp. 429–439, 2002.

[9]   Y. S. Chen, Y. C. Tseng, J. P. Sheu, and P. H. Kuo, "An on-demand, link-state, multi-path QoS routing in a wireless mobile ad-hoc network," Computer Communications, vol. 27, no. 1, pp. 27–40, 2004.

[10]   R. Vaishampayan and J. J. Garcia-Luna-Aceves, "Efficient and robust multicast routing in mobile ad hoc networks" in Proceedings of the IEEE International Conference on "Mobile Ad-hoc and Sensor Systems", pp. 304–313, IEEE, October 2004.

[11]   P. Khadivi, S. Samavi, and T. D. Todd, "Multi-constraint QoS routing using a new single mixed metrics,"Journal of Network and Computer Applications, vol. 31, no. 4, pp. 656–676, 2008.

[12]   S. S. Manvi and M. S. Kakkasageri, "Multicast routing in mobile ad hoc networks by using a multi-agent system" Information Sciences, vol. 178, no. 6, pp. 1611–1628, 2008.

[13]   F. Kuipers, P. Van Mieghem, T. Korkmaz, and M. Krunz, "An overview of constraint-based path selection algorithms for QoS routing," IEEE Communications Magazine, vol. 40, no. 12, 2002.

[14] K. Kunavut and T. Sanguankotchakorn, "Multi-Constrained Path (MCP) QoS routing in OLSR based on multiple additive QoS metrics," in Proceedings of the International Symposium on Communications and Information Technologies (ISCIT '10), pp. 226–231, IEEE, Tokyo, Japan, October 2010.

[15] L. Liu and G. Feng, "Simulated annealing based multi-constrained QoS routing in mobile ad hoc networks," Wireless Personal Communications, vol. 41, no. 3, pp. 393–405, 2007.

[16] S. Bitam and A. Mellouk, "Bee life-based multi-constraints multicast routing optimization for vehicular ad hoc networks," Journal of Network and Computer Applications, vol. 36, no. 3, pp. 981–991, 2013.

[17] N. Ghaboosi and A. T. Haghighat, "Tabu search based algorithms for bandwidth-delay-constrained least-cost multicast routing," Telecommunication Systems, vol. 34, no. 3-4, pp. 147–166, 2007.

[18] Z. Subing and L. Zemin, "A QoS routing algorithm based on ant algorithm," in Proceedings of the IEEE International Conference on Communications (ICC '01), vol. 5, pp. 1581–1585, IEEE, 2001.

[19] C.-H. Chu, J. Gu, X. D. Hou, and Q. Gu, "A heuristic ant algorithm for solving QoS multicast routing problem," in Proceedings of the IEEE Congress on Evolutionary Computation (CEC '02), vol. 2, pp. 1630–1635, IEEE, Honolulu, Hawaii, USA, May 2002.

[20] R. Forsati, A. T. Haghighat, and M. Mahdavi, "Harmony search based algorithms for bandwidth-delay-constrained least-cost multicast routing" Computer Communications, vol. 31, no. 10, pp. 2505–2519, 2008.

[21] R. F. Abdel-Kader, "Hybrid discrete PSO with GA operators for efficient QoS-multicast routing," Ain Shams Engineering Journal, vol. 2, no. 1, pp. 21–31, 2011.

[22] P. Chen and T. L. Dong "A fuzzy genetic algorithm for QoS multicast routing" Computer Communications, vol. 26, no. 6, pp. 506–512, 2003.

[23] X. S. Yang and S. Deb, "Engineering optimization by cuckoo search," International Journal of Mathematical Modelling and Numerical Optimisation, vol. 1, no. 4, pp. 330–343, 2010.

[24] S. Walton, O. Hassan, K. Morgan, and M. R. Brown, "Modified cuckoo search: a new gradient-freeoptimization algorithm," Chaos, Solitons & Fractals, vol. 44, no. 9, pp. 710–718, 2011.

[25] X. S. Yang and S. Deb, "Cuckoo search: recent advances and applications," Neural Computing and Applications, vol. 24, no. 1, pp. 169–174, 2014.

[26] O. Baskan, "Determining optimal link capacity expansions in road networks using cuckoo search algorithm with levy flights," Journal of Applied Mathematics, vol. 2013, Article ID 718015, 11 pages, 2013.

[27] P. Civicioglu and E. Besdok, "A conceptual comparison of the Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms" Artificial Intelligence Review, vol. 39, no. 4, pp. 315–346, 2013.

[28] P. I. Basarkod, S. S. Manvi, and D. S. Albur, "Mobility-based estimation of node stability in MANETs", in Proc. Int. Conf. Emerg. Trends in Comput., Commun. and Nanotechnol. ICE-CCN 2013, Tuticorin, India, 2013, pp. 126–130.

[29] K. S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, 2 ed. Wiley Interscience, 2005

[30] P. I. Basarkod and S. S. Manvi, "Multiple parameters based approach to estimate bandwidth in mobile ad hoc networks", Int. J. Comp. Sci. Issues, Special Issue (IJCSI), vol. 1, no. 1, pp. 37–43, 2011.

[31] Woongsoo Na, Yunseong Lee, Jongha Yoon, Junho Park, and Sungrae Cho, "Fully Distributed Multicast Routing Protocol for IEEE 802.15.8 Peer-Aware Communication", International Journal of Distributed Sensor Networks Volume 2015.

AUTHORS

M. Vijayalakshmi Completed B.E and M.E in E.C.E from Osmania University. Presently pursuing Ph.D in the area of Wireless Networks at Jawaharlal Nehru Technological University, Hyderabad,(JNTUH) Under the guidance of Prof., Dr.D.Sreenivasarao, JNTUH. And working as an Assoc.Prof. in G.N.I.T.S., Hyderabad. Total teaching experience is of 16 years.

INTENTIONAL BLANK

APPLICATION BASED SMART OPTIMIZED KEYBOARD FOR MOBILE APPS

Sandeep S Machiraju, Ashok K Athukuri, Soumya Gampa, Narendra B Makela, & Venkata N Inukollu

College of Science and Engineering, University of Houston-Clear Lake, USA
Inukollu@uhcl.edu

ABSTRACT

Mobile applications are becoming an emerging property in today's world. Delivering high quality mobile applications by improving the quality of the user interface is a definitive guide in building successful mobile applications and will have a significant increase in the market for that application. This paper addresses some issues related to mobile user interface, current input system and user dictionary. After analysing these issues we have proposed a new system by making some improvements to the current input system and user dictionary. In order to improve the current system, we have taken some issues related to one of the most widely used data entry methods 'Keyboard' and proposed some improvements by introducing Smart Optimized Keyboard (STOKE). We have provided the results, which include performance, reduction in number of keystrokes, and screen space. This describes the significant change after adopting the proposed keyboard. In the end, we have proposed an optimized way of building up the user dictionary, which assists the user in inserting the input very easily, rapidly and effectively. In a combine, these two will help in providing the best user experience to the users.

KEYWORDS

Mobile App, Smart keyboard, keystroke, Predictive Text, Usability, Human Computer Interaction.

1. INTRODUCTION

Mobile applications have become an emerging property in today's world. Mobile application commonly known as a mobile app, is a software application that is designed and developed for wireless devices such as tablets and smart phones. Mobile applications are served for easy access to users compared to personal computers. Now-a-days mobile applications are vastly in use for various activities, including shopping, gaming, texting and many more. Users are capable of using the mobile applications for personal access, business needs and for day-to-day life tasks. For example, Google map has greatly simplified the navigation with GPS and is extensively applied to identify different places and also assists in knowing about traffic jams among various other services, Google notes, aids the users to save their personal information and banking apps help users to execute their banking needs including money transfers, balance inquiries, among many others.

Many research papers have been published with respect to user interface and user interface design principles. The primary factor for any mobile app is: usability: described by: the effective functioning of the app, the purpose desired by the user and served by the app, the easy access to the users   the user thoughts and feelings about the application [1]. Developing an application in accordance with the user demands and requirements along with the ease to use the app increases the lifespan of an application [2] [3]. In this research effort we discuss about the user interface, problems that a user faces while interacting with the mobiles and how a user can interact efficiently with the introduction of small improvements to the soft keyboard and Optimization of the User Dictionary. The remainder of this paper is organized as follows: Section 2 and 3 discuss about the mobile application user interface and third party smart keyboards in detail, Section 4 presents the proposed methodology in the current mobile framework and discusses the impact on usability after adopting the proposed methodology.

2. MOBILE APP USER INTERFACE AND SMART KEYBOARDS

Businesses for the mobile applications has increased rapidly and significantly all around the globe. In 2012, a study was conducted examining the growth of the market for mobile applications, according to which, by 2016, the market for mobile applications in the United States will reach $55 billion [4, 22]. Mobile user interface is one of the major aspects that needs to be considered while developing smart mobile applications and this will allow customers to decide and act immediately in their moments of need [4, 22].

The two important components that contribute significantly to mobile interfaces are Navigating and Data Entry [5] [6]. As the mobile devices are constrained in screen space, the application developers have to display the required information and need to receive the required information from the user as input by using the limited space. The system needs to provide an interface that should be clear, consistent, simple, user controlled, familiar, responsive and attractive and should be capable of efficiently taking an input from the user and providing the desired output. Such an efficient interface will enable the user to enter the input very easily, rapidly and effectively.

The widely used data entry method in mobiles is the virtual keyboard, because, it provides flexibility to the users, i.e., the Virtual Keyboard provides different modes to different text fields. In that respect are several alternative third party virtual keyboards, which are available in the marketplace. Each of the third party keyboards provides substantial benefits to the users. Despite the benefits, there are many issues that the users are currently confronting, which include: performance issues, security issues and usability issues. When the third party keyboards grant network access, there is a possibility of leakage of keystrokes and personal data of the users [7] leading to security issues. The performance issues of the third party keyboards include the unresponsive keyboard after it has been loaded [8], the unavailability and inaccessibility of the keyboard for a few seconds, etc. With respect to the usability of the third party keyboards, the users might not prefer the third party keyboards, as the users have an option to use the default keyboard that is readily available and is very good in performance, security and satisfies the basic requirements of the user. Therefore, the default keyboard provided by the system is the principal choice of the several users.

Several users prefer the default keyboard despite some drawbacks associated with the default keyboard. However, the implementation of the suggested significant improvements (discussed further in this paper) to the default keyboard from the perspective of the system and the

developer, will enable more and more users to reliably and jubilantly employ the default standard keyboard.
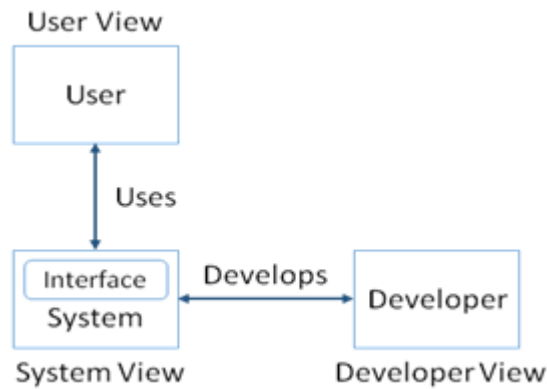

Figure1. Block Diagram that depicts different views (System, Developer & User)

- **Developer View:** This view describes about the mobile operating system guidelines and provided system functionalities that the developer needs to use in order to provide the best user interface to the users. For example, Guidelines such as offering informative feedback, design dialogs to yield closure and many more [9]. At the same time mobile manufacturers like Android and IOS enforce their own usability constraints[10] [11]

From the developer's point of view, the developer needs to follow the standards of the mobile operating system and use the provided functionalities of the system. For example, displaying the necessary type of the keyboard (the current system provides various input functions shown in below table 1) based on the input to be entered by the user. Having unnecessary keys or navigation (displaying characters instead of numbers) on the keypad will have a critical increment in number of keystrokes and errors. However, most of the software developers are not using the default input method functions provided by the system for the applications.

Table. Android Major Standard Input Functions (For more [12])

Text Field	input function (Property value)
Person Name	textPersonName
Password	textPassword
Email	textEmailAddress
Numbers	number

- **System View:** This view describes about the system and the functionalities dealt by the system, the libraries provided by the system and how the system helps the user in providing the input. For example, Libraries such as material design, material design icons, material dialogs, material menu and many more [13].

From the system's point of view, the system needs to provide an interface that will allow the users to accomplish specified goals with ease, efficiency and satisfaction in a specified context of use [9]. The interface needs to provide the necessary input method

based on the type of input, and, necessary help to the users to enter the input as efficiently and as easily as possible.

- **User View:** This view describes about how effortlessly and how effectively the users will be using the application that aids in fulfilling their requirements.

  From the user's point of view, the users desire to utilize the provided system with ease and to accomplish their objectives. Keeping in view the requirements of the users, the developer ought to build an application by following the guidelines and functionality provided by the system. The proposed system will help the developer to provide the best user interface to the user. The proposed system and its components will be explained in further sections in detail.

3. RELATED WORK

The fundamental factor in any mobile application is the user interface which depicts the effectiveness of the application, the ease with which the users are able to use the app and even measures the user thoughts and feelings about the app. The vast majority of the leading applications has been successful, essentially, by providing the best user interface and best user experience. Aside from fulfilling the business objectives, satisfying the commercial goals can be accomplished by considering the user interface part of an application.

In order to provide the best user interface to the users, the following issues associated with user interface have to be effectively addressed: Utilizing the screen space and Interaction mechanism [14]. Where utilization of screen space incorporates having different layout problems with small screens, Interaction mechanism incorporates difficulty in entering the text with more probability of entering the correct text by the user.

Among the various text entry methods for mobiles, entering the text by using soft keyboard is one of the most widely used method by users in mobiles. In 1867, Christopher L. Sholes and colleagues designed a new layout for keyboard named QWERTY and has become the de facto standard layout for both the virtual and physical keyboards [15]. The users were habituated to this kind of layout as it has been effective for several years and they tend not to see any new kind of layout as it generally consumes lot of time to learn and implement a new layout. QWERTY has a considerable effect of input speed compared to other keyboard layouts [16]. A study on various keyboard layouts in mobiles [17] demonstrates the familiarity and input speed with the QWERTY keyboard layout. This also depicts that users are able to type faster with the QWERTY keyboard as opposed to other optimized keyboards.

In 2000, in light of Fitts law, Silfverberg, et al. proposed a new mathematical model for predicting the text for the users while typing [18]. This predictive model helps users to enter the text very quickly, which was intended to diminish the number of keystrokes. The existing predictive system has a user database (Learning System/Personal dictionary) where each word entered by the users will be saved into the database at the time of typing. However, most of the users do not know how to use the personal dictionary effectively.

4. PROPOSED FRAMEWORK

The proposed framework includes four levels : Application Level, Application Framework Level, System Libraries and Kernel Level. Each level is briefly discussed below:

Application Level: All kinds of mobile applications will reside at this level and each application will have a user input that includes Input View and Input Field that allows the user to enter text.

- **Input View:** It is the user interface of all the input fields where the user inputs the text. The user selects one particular input field at a time.

- **Input Field:** Where the user can enter the data and it can vary in many ways depending on the data type. Input Field is connected to the input method editor to get the necessary keypad user interface (UI).
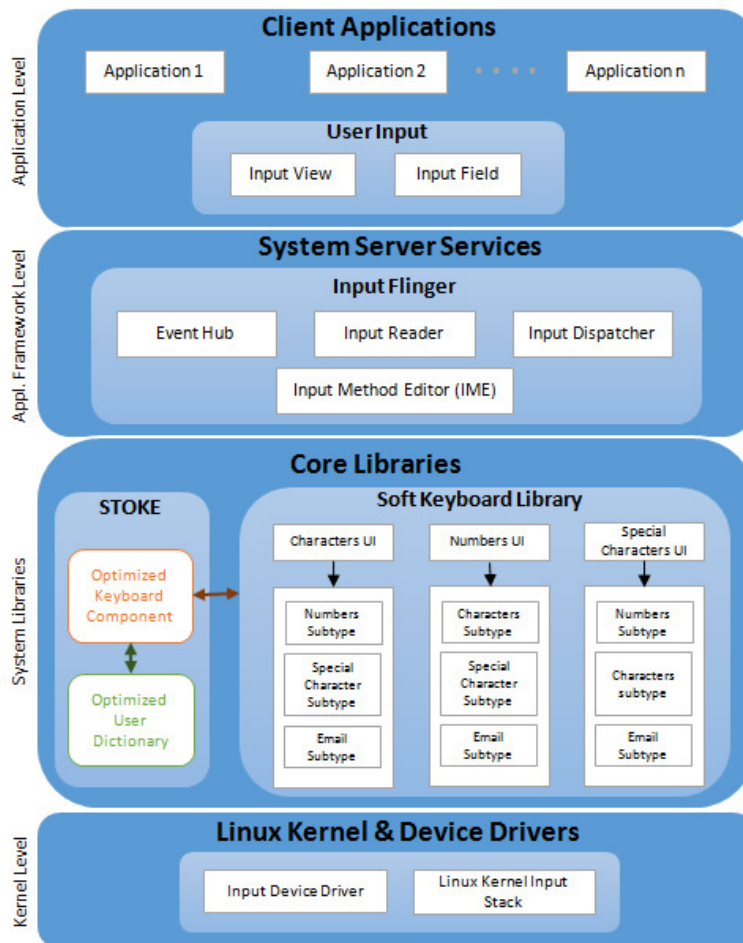
Figure2. Proposed Method in Android Framework
*STOKE = Smart Optimized Keyboard

Application Framework Level: This level has various system server services. Input Flinger, is the service which takes the input from the user. Input Flinger comprises of four components: Event Hub, Input Reader, Input Dispatcher and Input Method Editor. [20]

- **Event Hub:** The primary function of this component is to convert the raw events to operating system events. It also helps in adding and removing of devices with notifications.

- **Input Reader:** It acts like a client for the component event hub. It just reads the converted events and processes them. This processing is done with the help of input mappers. Once the processing is done, the input dispatcher is notified.

- **Input dispatcher:** This will get the processed events from input reader and identifies the target window. After identification, it dispatches to the target application. It uses the socket pair communication in order to know the status of delivery.

- **Input method editor:** This allows the user to enter the text, i.e., it provides an input method framework that helps the user to select various input methods.

Kernel Level: This level has two components: Input device driver, Kernel input stack. Where input device driver helps in registering the input_device. Input device driver also identifies the user input as events. For example, when user enters the text, device driver will create the file that has all interruptions from the user as events and these events are mapped with the default event code. This will be helpful in identifying the input from the user. [20]

System Libraries:

In this level, there are three major components: Core libraries, STOKE Keyboard component and Optimized user dictionary.

- **Core Libraries:** The following libraries are the proposed libraries which are inherited from the default libraries provided by the system. This helps STOKE keyboard to provide the respective keyboard based on the user input.

  - **Characters User Interface (UI):** This UI consists of three subtypes, namely, Numbers subtype, Special character subtype and Email subtype. These subtypes are helpful in displaying the respective keyboard on mobile screen. For example, when the input field is 'Name', then it is not required to display all the subtypes, thus increasing the screen size, decreasing the number of keystrokes and reducing human errors.

  - **Numbers UI:** This UI consists of three subtypes, namely, Character subtype, Special character subtype and Email subtype. These subtypes are helpful in displaying the respective keyboard on mobile screen. For example, when the input field is' zip code' then it is not required to display all the subtypes, thus increasing the screen size, decreasing the number of keystrokes and reducing human errors.

- **Special Characters UI:** This UI consists of three subtypes, namely, Character subtype, Numbers subtype and Email subtype. For Email the Characters UI displays first and then it navigates for selecting the appropriate emails through Email subtype.

- **STOKE Keyboard:** This is the proposed keyboard, which helps in getting the right keyboard based on the type of user input. This keyboard is an enhancement to the default keyboard and was designed by making some significant improvements from both system perspective and developer's perspective. As stated above, we have made some improvements to the existing library and likewise proposed a new library. For example, if Email is the input to be entered by the user, then STOKE keyboard will ask the Characters UI library to display the characters along with some extra features that include: '@' as a menu option. When a user selects that option the user will be navigated to a menu, where the user will be able to select the email types from the list, which includes @gmail.com, @yahoo.com, @outlook.com and many more (as shown in below Figure 3). Here the users will also be able to add their own email type that can be useful in future.

Figure3. STOKE keyboard for Email Type

This will be useful both for the users and the developers. From the users perspective, this will be useful as the number of keystrokes to enter an email, as an input and the screen size that the keyboard utilizes will be significantly decreased (Shown in below table 1). From the developer's perspective, there is no requirement to display the entire keyboard while the user is selecting the email types. Thus, developers can use the space for different purposes such as for better advertising and displaying other important aspects related to business.

Table 1. Keystrokes and Screen Size Comparison

Name of the Field	Number of keystrokes (Using existing system)	Number of keystrokes reduced (After using STOKE)	Screen Size utilization in % (Before adapting STOKE)	Screen Size reduced (After adapting STOKE)
EMAIL	1+8+1+4+1+9 = **24** keystrokes (eg: Seronika$1995@gmail.com (estimated keystroke count including characters, special characters and numbers)	8+1+4+1 = **14**	**53**	**51**

The above table clearly demonstrates that the number of keystrokes and screen size of the keyboard has been greatly reduced after STOKE keyboard was adopted. We have studied different emails of various users and calculated the average number of keystrokes required to enter an email (as shown in table 1). We have also calculated the number of keystrokes required for the user to enter an email address after adopting STOKE keyboard. There is a tremendous change in user experience as the users do not need to enter the complete email extension (ex: @gmail.com) with the STOKE keyboard. Therefore, on an average ten keystrokes will be reduced.

This research effort has also calculated the number of keystrokes and screen size for different input fields of various applications for both the default keyboard and proposed STOKE keyboard. The results demonstrate a significant difference between the keyboards, which clearly depicts an advantage in using the STOKE keyboard. The enhanced keyboard (STOKE) of our research effort does not compromise on the performance and security aspects. In addition to the above performance tests, a test was performed by calculating the time taken for both the Soft keyboard (Current) and STOKE keyboard (Proposed) to get displayed in IOS and Android. The results of both ANDROID and IOS are shown below in table 2 and figures 4&5 show the test results of both the standard keyboard and proposed keyboard in the IOS environment. Likewise, the testing process was conducted in Android environment:

Table 2. Performance Analysis

	ANDROID	IOS
Default Keyboard	0.46609 ms	0.10736 ms
STOKE	0.51201 ms	0.15782 ms

*Note that the results may vary based upon the system and the device that you use for testing. The displayed results are a resultant average of all test cases.
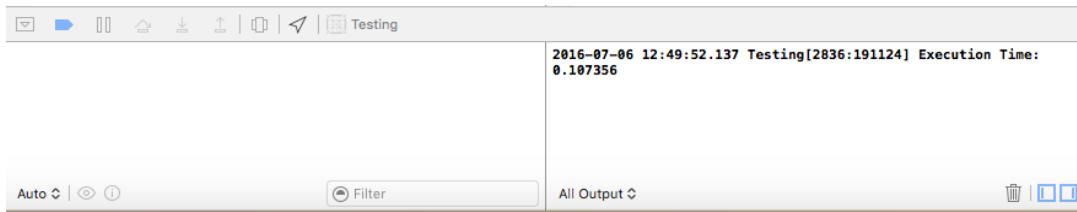
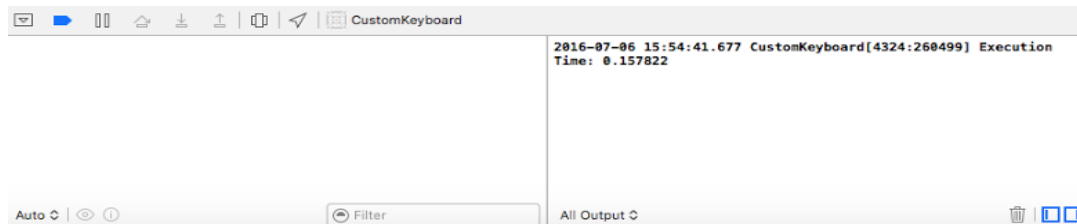Figure 4. Time taken for the default IOS standard keyboard to get displayed.

Figure 5. Time taken for the proposed(STOKE) keyboard to get displayed in IOS.

The above shown performance table clearly depicts that the STOKE keyboard has approximately consumed the same amount time in comparison to the default keyboard. Hence, this research effort claims that the users will not experience performance issues with respect to the proposed STOKE keyboard. Thus, this technique helps in providing the necessary keyboard, and also helps in reducing the unnecessary navigation by providing shortcut keys to accomplish the intended task (for example short cut keys while giving an email as input), which also significantly enhances the input speed for the user.

- **Optimized User Dictionary:** Predictive texting helps in saving a lot of time by providing some predictive words related to the entered character or word by the user. This will enable the user to enter the complete word/input in a single stroke [19]. The predictive texting, thus, reduces the number of keystrokes and also increases the accuracy, i.e., the probability of entering the wrong input is significantly diminished, thus, helping the user in finding the correct word in a shorter time. The predictive technique uses both language based dictionary and user dictionary (Personal dictionary) in order to provide predictions to the word/character entered by the user. The language based dictionary will be updated based on the language chosen by the user and the users do not have the permission to add or remove the words from the language based dictionary. In order to accomplish the addition or removal of the frequently used words by the user, the user dictionary has some into existence. This enables the users to add their own frequently used words in the dictionary which can be used in the future. Apart from this, there is a facility of text replacement that will help the users to create shortcuts for the complete phrase. This will allow user to enter the complete phrase (max length) by simply entering the corresponding shortcut word that was created.

All of the above mentioned methods have significant advantages and greatly help in improving the interaction of the users with the mobiles; iff used appropriately. But, the current system doesn't provide the users with the necessary tools to use the above facilities. With the current system, users need to enter a single word/character at a time into the dictionary or users need to build the dictionary or text replacement dictionary

before using them. These two options will consume a significant amount of time, thereby compromising the user experience. So, in order to improve the usability, we have proposed a new method named Optimized User Dictionary, which will allow the users to enter the complete set of text (of any length) or previously sent messages simply by selecting the complete content or message. The system will consider the complete text and will separate each word from the sentence using various delimiters (space and symbols). The system also provides an option for the user to enter a shortcut for the complete sentence or message, which will be added to the text replacement dictionary. Also, the system also provides an easier means to delete the set of words from the dictionary. This provides an easier, faster and effective means for managing the user dictionary.

This system will be useful for the users while interacting with the mobile applications such as texting. One of the survey states that [21], US users spend 4.8 percent of their smart phone minutes in instant messaging applications. This will be higher if considered globally. As known to all of us, for chatting and messaging, 'English keyboard type' is a standout amongst the most broadly used keyboards around the globe. Users from different parts of the world send messages in their own language by using the English keyboard and the current system provides predictions of only English words. By adopting the proposed method of keyboard, the users of different languages can build their own dictionary in English and can get the complete sentences by simply entering the shortcut created for that particular phrase/sentence.

5. FUTURE WORK

We have proposed a theoretical approach and recommendations that will significantly help to ameliorate the current system. The user has many advantages with the acceptance and implementation of the STOKE keyboard and Optimized User Dictionary. In the future, we are going to provide the usability results i.e., how user has an ease in providing the input while uisng the proposed Smart Optimized Keyboard. For this, we are developing a sample working prototype for both Android and IOS mobile users that will be helpful in getting the usability results.

REFERENCES

[1]   L. Picurelli, "The usability of mobile applications is the key element for their success," [Online]. Available: https://en.yeeply.com/blog/the-usability-of-mobile-applications-is-one-of-the-key-elements-for-their-success/. [Accessed 21 august 2013].

[2]   J. NIELSEN, "Assessing the Usability of a User Interface Standard," Nielsen norman group, [Online]. Available: https://www.nngroup.com/articles/assessing-usability-user-interface-standard/. [Accessed 28 april 1991].

[3]   J. Gould and C. Lewis, "Designing for usability: key principles and what designers think," 1985.

[4]   Schadler, et al. "Mobile is the new face of engagement."Retrieved from Forrester Research: http://cdn.blog-sap. com/innovation/files/2012/08/SAP_Mobile_Is_The_New_Face_Of_Engagement. pdf (2012).

[5]   S. Hoober, "Mobile Input Methods," [Online]. Available:
       http://www.uxmatters.com/mt/archives/2012/11/mobile-input-methods.php. [Accessed 1 november
       2012].

[6]   S. Mohamed, "Challenges facing mobile app usability testing," [Online]. Available:
       https://www.linkedin.com/pulse/challenges-facing-mobile-app-usability-testing-sally-
       mohamed?trkSplashRedir=true&forceNoSplash=true. [Accessed 21 october 2015].

[7]   Zelster, "Security of third party keyboards on mobile apps" [Online]. Available:
       https://zeltser.com/third-party-keyboards-security/

[8]   Reddit, "Performance Issues of third party keyboards", [Online]. Available:
       https://www.reddit.com/r/ios9/comments/3hn51m/does_ios_9_improve_3rd_party_keyboard_perform
       ance/?

[9]   Gong, et al. "Guidelines for handheld mobile device interface design." Proceedings of DSI 2004
       Annual Meeting. 2004.

[10]   "IOS Human Interface Guidelines - Developer.apple.com." IOS Humar Interface Guidelines. Apple,
       n.d. Web. 22 Aug. 2016. [Online]. Available: https://developer.apple.com/ios/human-interface-
       guidelines/

[11]   "The State of the Art of Mobile Application Usability" Android User Interface Guidelines. Google
       N.p., n.d. Web. 22 Aug. 2016. [Online]. Available:
       http://developer.android.com/guide/practices/ui_guidelines/index.html.

[12]   "Various Input Functions". Google n.d. Retrieved August 22, 2016. [Online]. Available:
       https://developer.android.com/guide/topics/ui/controls/text.html

[13]   Wasbeef. "Android User Interface Libraries." Google N.p., n.d. Web. 22 Aug. 2016. [Online].
       Available: https://github.com/wasabeef/awesome-android-ui

[14]   Heinecke, Andreas M., et al. "What Every Software Developer Should Know about Human-
       Computer Interaction–A Curriculum for a Basic Module in HCI in Informatics Education." ACM-
       IFIP.

[15]   Nilsson, G Erik. "Design guidelines for mobile applications." SINTEF ICT, Jun (2008).

[16]   Butts, Lee, and Andy Cockburn. "An evaluation of mobile phone text input methods." Australian
       Computer Science Communications. Vol. 24. No. 4. Australian Computer Society, Inc., 2002.

[17]   Gkoumas, Apostolos, Andreas Komninos, and John Garofalakis. "Usability of Visibly Adaptive
       Smartphone keyboard layouts." (2016).

[18]   MacKenzie, I. Scott, Shawn X. Zhang, and R. William Soukoreff. "Text entry using soft keyboards."
       Behaviour & information technology 18.4 (1999): 235-244.

[19]   Wikipedia contributors. "Predictive text." Wikipedia, The Free Encyclopedia. Wikipedia, The Free
       Encyclopedia, 30 Sep. 2016. Web. 30 Sep. 2016.

[20]   "Android Input Architecture: Android Internals"[Online]. Available:
       http://newandroidbook.com/files/AndroidInput.pdf.

[21] Perez, Sarah. "A Survey on Smart Phone Usage." A Survey on Smart Phone Usage. N.p., 22 June 2015. Web. 22 Aug. 2016. [Online]. Available: https://techcrunch.com/2015/06/22/consumers-spend-85-of-time-on-smartphones-in-apps-but-only-5-apps-see-heavy-use/.

[22] Inukollu, Venkata N., et al. "Factors influencing quality of mobile apps: Role of mobile app development life cycle." arXiv preprint arXiv:1410.4537 (2014).

SPONTANEOUS SMILE DETECTION WITH APPLICATION OF LANDMARK POINTS SUPPORTED BY VISUAL INDICATIONS

Karolina Nurzynska and Bogdan Smolka

Faculty of Automatic Control, Electronics, and Computer Science
Silesian University of Technology,
ul. Akademicka 2A, 44-100 Gliwice, Poland
Karolina.Nurzynska@polsl.pl, Bogdan.Smolka@polsl.pl

ABSTRACT

When automatic recognition of emotion became feasible, novel challenges has evolved. One of them is the recognition whether a presented emotion is genuine or not. In this work, a fully automated system for differentiation between spontaneous and posed smiles is presented. This solution exploits information derived from landmark points, which track the movement of fiducial elements of face. Additionally, the smile intensity computed with SNiP (Smile-Neutral intensity Predictor) system is exploited to deliver additional descriptive data. The performed experiments revealed that when an image sequence describes all phases of smile, the landmark points based approach achieves almost 80% accuracy, but when only onset is exploited, additional support from visual cues is necessary to obtain comparable outcomes.

KEYWORDS

Smile veracity recognition, Landmarks, Smile intensity, Classification

1. INTRODUCTION

Automatic analysis of image content finds its application in many various domains: starting from medical image analysis in order to facilitate diagnosis, going through detection of flaws in products on a production line in a factory, and finishing on human behaviour interpretation. In all these situations, similar approaches for image processing and recognition are implemented. The image content is explored in order to derive some characteristics, which later on become a feature set. These characteristics may originate from visual information, when for instance the texture is exploited as a data source, or describe abstract details, when shape and active shape features are considered.

Having an image sequence of facial gesture it is possible to recognize which emotion was presented with very high efficiency, as was proved in several research works [2, 3, 4, 5, 6]. However, the accuracy of such system performance decreases significantly, when real life scenarios are considered [7, 8, 9], because programs trained on posed emotions displayed in laboratory environment are used to very strong facial gesture presentation, which is not common

in daily life, not to mention lighting variation or occlusions. Nevertheless, in most cases recognition of the happiness emotion is on satisfactory level. Therefore, further investigations were performed, which aimed at recognition whether the smile corresponding to this emotion is a genuine one or not [10, 11, 12].

When designing a system for smile veracity recognition, two most common approaches are identified: One exploits facial landmarks for feature calculation [10], where distances and characteristic points movement relationship are considered. The other uses data sampled from the image texture [12] and combines them with smile intensity measure. The presented solution is a combination of both. The veracity of smile is described by automatically derived characteristic points, which are later normalized before the features are computed. On the other hand, a smile intensity function derived from the image textural content, following the idea of SNIP system [1], is applied.

The paper starts with description of facial landmarks determination presented in Sec. 2. Next, Sec. 3 presents the method for smile intensity function calculation. The details of image sequence corpora selected for results verification is given in Sec. 4. Then, the results are presented in Sec. 5 and conclusions are drawn in Sec. 6.

2. CHARACTERISTIC POINTS

In order to describe an emotion presented on image, the positions of facial landmarks called also characteristic points are calculated. These landmarks correspond to characteristic regions of the face, such as eyes, nose, lips, etc. and are exploited to track their changes during the presentation of emotion. Figure 1 depicts the placement of data obtained with implementation presented in [13].

2.1 Data Normalization

The subject during emotion presentation may move the head freely in any direction, therefore the landmarks are firstly normalized as was suggested in [10, 11]. It is necessary due to differences in head size, which is mostly noticeable when the head moves in and out of the camera focus as well as anatomical differences between humans should be compensated. Because the proposed automatic method for landmark annotation returns the coordinates in XY-plane only, therefore the rotation normalization was neglected and only scaling the landmarks positions was performed. It assumes a constant distance (100 pixels) between left (eye_L) and right (eye_R) eye centres, which coordinates are computed as an average of red marks presented in Fig. 1. New characteristic points coordinates p^* are computed as follows

$$p_i^* = (p_i - \frac{(eye_L + eye_R)}{2}) \cdot \frac{100}{\varrho(eye_L, eye_R)},\qquad(1)$$

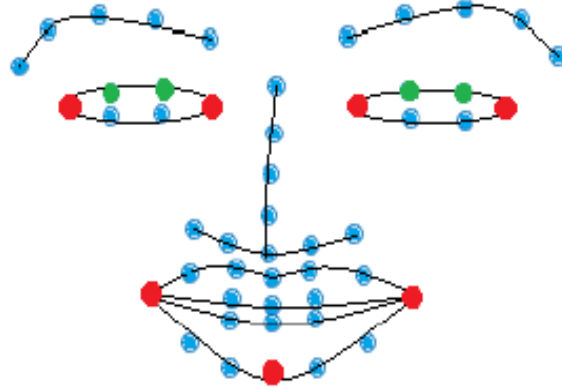where $\varrho(a, b)$ denotes the Euclidean distance between points a and b.

Figure 1: Characteristic points denoted on the facial image.

Table 1: Formulas for features calculation.

Feature	Definition
Duration	$\zeta(F_+), \zeta(F_-), \zeta(F)$
Duration Ratio	$\frac{\zeta(F_+)}{\zeta(F)}, \frac{\zeta(F_-)}{\zeta(F)},$
Maximum Amplitude	$\max(F)$
Mean Amplitude	$\frac{\sum(F)}{\zeta(F)}, \frac{\sum(F_+)}{\zeta(F_+)}, \frac{\sum(F_-)}{\zeta(F_-)}$
STD of Amplitude	$\mathrm{std}(F)$
Total Amplitude	$\sum(F_+), \sum(|F_-|)$
Net Amplitude	$\sum(F_+) - \sum(|F_-|)$
Amplitude Ratio	$\frac{\sum(F_+)}{\sum(F_+)+\sum(|F_-|)}, \frac{\sum(F_-)}{\sum(F_+)+\sum(|F_-|)}$
Maximum Speed	$\max(S_+), \max(S_-)$
Mean Speed	$\frac{\sum(S_+)}{\zeta(S_+)}, \frac{\sum(S_-)}{\zeta(S_-)}$
Maximum Acceleration	$\max(A_+), \max(A_-)$
Mean Acceleration	$\frac{\sum(A_+)}{\zeta(A_+)}, \frac{\sum(A_-)}{\zeta(A_-)}$
Net Ampl. Duration Ratio	$\frac{\sum(F_+)-\sum(|F_-|)}{\zeta(F)}$

2.2 Feature Calculation

In order to distinguish between the posed and genuine smile, its course is divided into onset, apex, and offset phases. The onset is defined as a time from the emotion start to its full presentation. The offset presents signal attenuation, while the apex depicts the full emotion facial gesture. The length ζ of each part may vary between subjects, as well as presented emotions. In order to split the data sequence into these three parts, a smile amplitude signal is considered. Figure 2 presents exemplary progress of smile amplitude change within an image sequence. Such plot is exploited for data division, with assumption that values above 1 represent smile apex. Frames preceding the apex belong to the onset phase, while those following, to the offset. In case when several smile peaks are detected, the longest one is considered in calculations.

The smile strength is given by lip movement amplitude calculated following the formula given in [10, 11]. It exploits the information of landmarks which describes the movement of lips corners and the bottom lip middle point (all pointed in red in Fig. 1). Similarly, for the eye's region, the eyelid magnitude is computed, basing on eyes corners (red colour in Fig. 1) and middle point of the upper eyelid calculated as an average of coefficients obtained for green points in Fig. 1.

Finally, the features gathered in Tab. 1 are computed separately for each smile phase as well as for each considered facial region. Moreover, information about local signal increase (denoted with '-' symbol in index) and decrease (indicated with '+' symbol in index) are exploited. For each signal F, a speed $S = dF/dt$ and acceleration $A = d^2F/dt$ are obtained. That gives 24 features for each region and phase.

Additionally, in the experiments, general approach to feature calculation was investigated. In such case, a set of features for whole smile or eyelid movement intensity was considered and as an increasing part the onset was exploited, while offset described the decreasing element in formulas in Tab. 1. Since some parameters are too strongly related to video duration and could overlap with previously calculated ones, they were removed (Duration) or redefined. For *Maximum Amplitude*, *Mean Amplitude*, *Total Amplitude*, *Maximum Speed*, *Mean Speed*, *Maximum Acceleration*, and *Mean Acceleration* only one value without distinction for increasing and decreasing parts was calculated. This generalization resulted in 14 features.
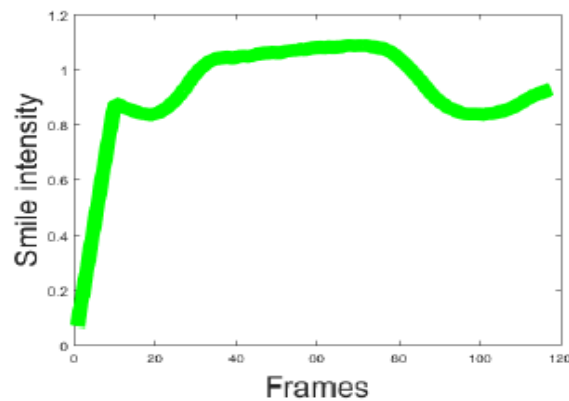
Figure 2. Smile intensity magnitude

Figure 3: Exemplary frames from image sequence in UvA-NEMO dataset.

3. SMILE INTENSITY

The smile intensity or amplitude can be also derived from the visual cues. Such a case was presented in [1], where texture operators extract feature vectors to train SVM classifier for recognition between smiling and neutral facial display. This classifier response, in other words the object distance from the division plane, is returned as a smile intensity function. In presented research, this approach for smile amplitude description is evaluated as an alternative one for these calculated exploiting lips corner movement annotated by landmarks. Similarly, the smile intensity function is divided into three phases for which parameters from Tab. 1 are computed.

4. DATABASE

The accuracy of posed and spontaneous smile recognition was verified on the UvA-NEMO database [10], which gathers images presenting only happiness emotion in scenarios allowing to record its both versions. The image sequences collected in this dataset depict happiness emotion presented by 400 subjects, whose age is in the range from 8 to 76. There are 1240 videos presenting 597 spontaneous and 643 posed facial gestures sequences. The data was recorded in controlled environment in RGB format with high resolution of 1920×1080 pixels, what results in average face resolution of 400×400 pixels. Some examples are given in Fig. 3.

5. EXPERIMENTS AND RESULTS

The goal of this experiment was to investigate not only the region influence on smile veracity detection but also the impact of the smile phase. The feature vectors built from parameters adequate for onset, apex, and offset phases of smile and eye magnitude were calculated. Moreover, a combination of all phase data was tested and the global approach for parameter calculation was exploited. The combination of all phase and global parameters was named total and tested as well. The features were fed on support vector machine (LIBSVM [14]) with linear kernel, for which the best parameter was evaluated in range $c = 10^{-5} - 10^5$. Figure 4 presents scores obtained for UvA-NEMO dataset.
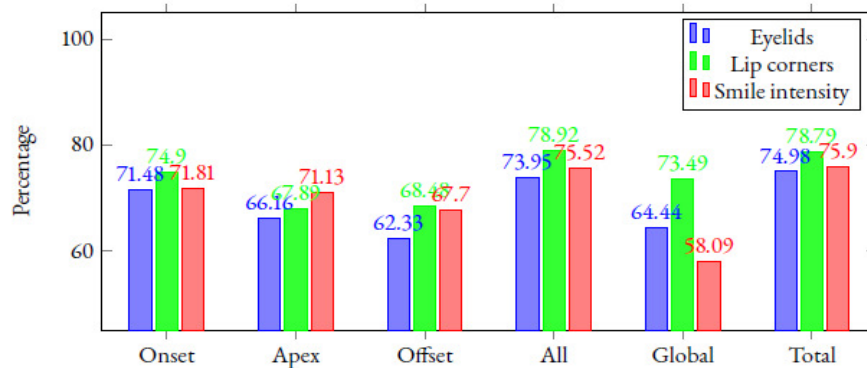
Figure 4: Influence of facial region described by features on classification rates.
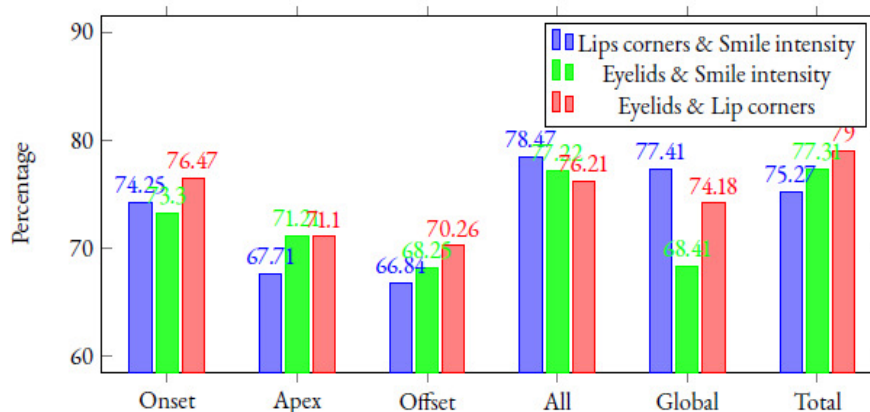
Figure 5: Classification rates when early fusion method was chosen for classification.

Results gathered in Fig. 4 revealed that it is possible to distinguish the posed from genuine smile with high accuracy. This plot shows also that the biggest influence on smile type detection has the onset phase, however for data exploiting smile intensity, similar results were obtained for the apex. Moreover, incorporating information obtained in other phases in all feature vector proved to give the best results. As presumed the global approach outcomes are slightly worse and its combination with all does not improve the accuracy (total), except for the smile intensity case. It is also worth noticing that it is easier to distinguish spontaneous emotion from posed ones, when lips magnitude is considered.

In order to create more accurate classifier, several options were considered. Figure 5 collects correct classification performances for concatenated features vectors describing all possible combinations of pairs of accessible descriptors. The best score of 79% was obtained for the total feature vector build from eyelids and lip corner amplitudes, and slightly outperformed the results when data was described by lips corners supported by smile intensity, which reached 78.47%.

Finally, Fig. 6 presents correct classification performance ratios when two different approaches fo outcome computation were investigated. The early fusion data was obtained by concatenation of feature vectors computed in first experiment and then classified with SVM. The late fusion applied three SVMs classifiers for each feature vector generated in first experiment and used voting to obtain the final score.

The conducted experiments show that it is possible to determine with high probability the veracity of smile using automatic method for landmark detection or visual approach of smile intensity description. Exploiting data derived only from landmark points with calculation of features both for local phases supported with global information enabled correct recognition with almost 80% accuracy. On the other hand, when only onset phase was accessible, combination of amplitude data derived from landmarks with those computed when visual data was exploited were not far behind.
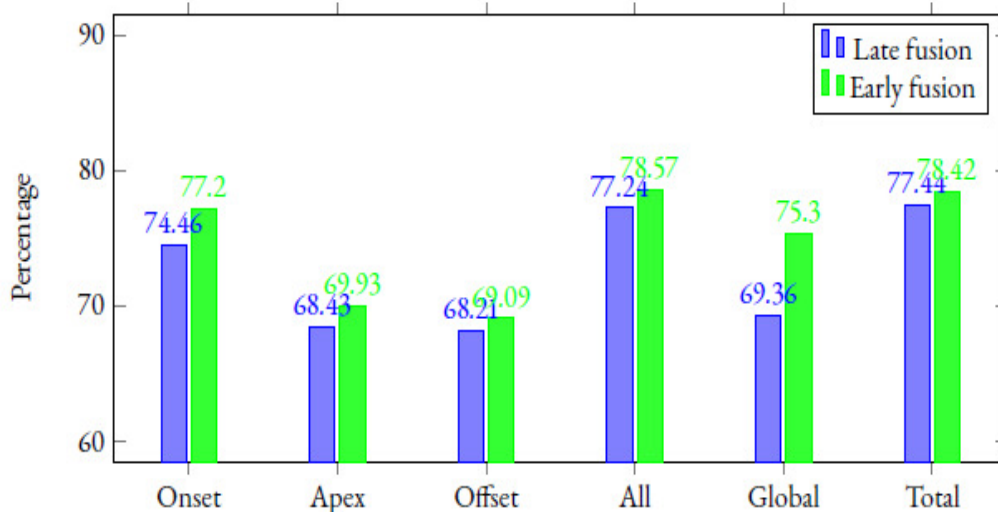

Figure 6: Classification rates when different fusion methods were chosen for classification.

6. CONCLUSIONS

This work presents fully automatic approach to smile veracity detection, which is based on landmarks movements estimation within an image sequence and is supported by information derived from visual cues. Namely, the lip corner and eyelid movements magnitudes are derived from characteristic point location and smile intensity function is achieved from the SNiP system. The feature vector was obtained as a combination of parameters computed for each input function. Using SVM for classification of parameters derived from these data, it was shown, that it is possible to achieve the accuracy of almost 80%, when all information was used. However, only slight deterioration is noticed, when parameters were reduced to describe the onset phase only. In further research, the visual content of images will be explored to improve the classification performance.

ACKNOWLEDGMENT

REFERENCES

[1]   K. Nurzynska and B. Smolka, Computational Vision and Medical Image Processing V, ch. SNIP: Smile–Neutral facial display Intensity Predictor, pp. 347–353. Taylor & Francis Group, 2016.

[2]   K. Mase, "An application of optical flow - extraction of facial expression," in IAPRWorkshop on Machine Vision Applications, pp. 195–198, 1990.

[3]   J. Cohn, A. Zlochower, J. Lien, and T. Kanade, "Automated face analysis by feature point tracking has high concurrent validity with manual facs coding," Psychophysiology, vol. 36, no. 2, pp. 35–43, 1999.

[4]   J. Cohn, A. Zlochower, J. Lien, and T. Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," in Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), pp. 396 – 401, April 1998.

[5]   B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," Pattern Recognition, vol. 36, no. 1, pp. 259 – 275, 2003.

[6]   C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image Vision Comput., vol. 27, pp. 803–816, May 2009.

[7]   A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, "Emotion recognition in the wild challenge (EmotiW) challenge and workshop summary," in Proceedings of the 15th ACM on International Conference on Multimodal Interaction, ICMI '13, (New York, NY, USA), pp. 371–372, ACM, 2013.

[8]   J. M. Girard, J. F. Cohn, L. A. Jeni, M. A. Sayette, and F. De la Torre, "Spontaneous facial expression in unscripted social interactions can be measured automatically," Behavior Research Methods, vol. 47, no. 4, pp. 1136–1147, 2015.

[9]   J. Chen, Y. Ariki, and T. Takiguchi, "Robust facial expressions recognition using 3D average face and ameliorated AdaBoost," in Proceedings of the 21st ACM International Conference on Multimedia, MM '13, (New York, NY, USA), pp. 661–664, ACM, 2013.

[10]   H. Dibeklioˇglu, A. A. Salah, and T. Gevers, Are You Really Smiling at Me? Spontaneous versus Posed Enjoyment Smiles, pp. 525–538. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[11]   H. Dibeklioˇglu, A. A. Salah, and T. Gevers, "Recognition of genuine smiles," IEEE Transactions on Multimedia, vol. 17, pp. 279–294, March 2015.

[12]   M. Kawulok, J. Nalepa, K. Nurzynska, and B. Smolka, "In search of truth: Analysis of smile intensity dynamics to detect deception," in Advances in Artificial Intelligence - IBERAMIA 2016 - 15th Ibero-American Conference on AI, San José, Costa Rica, November 23-25, 2016, Proceedings, pp. 325–337, 2016.

[13]   A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pp. 1859–1866, 2014.

[14]   C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

AUTHORS

Karolina Nurzynska received her M.E. and Ph.D. degree in the field of computer science from the Silesian University of Technology, Poland in 2005 and 2009, respectively. From 2009 to 2011 she was a Postdoc Researcher in the School of Electrical and Computer Engineering at the Kanazawa University, Japan. From 2011 to 2014 she was leading a project concerning visualisation of underground coal gasification in Central Mining Institute in Katowice, Poland, where in 2012 she was appointed the Assistant Professor position. Since 2013 she is an Assistant Profesor at the Department of Automatic Control, Electronics, and Computer Science of the Silesian University of Technology, Poland. Her research interests include image processing and understanding, data classification and 3D surface reconstruction.

Bogdan Smolka received the Diploma in Physics degree from the Silesian University, Katowice, Poland, in 1986 and the Ph.D. degree in computer science from the Department of Automatic Control, Silesian University of Technology, Gliwice, Poland, in 1998. From 1986 to 1989, he was a Teaching Assistant at the Department of Biophysics, Silesian Medical University, Katowice, Poland. From 1992 to 1994, he was a Teaching Assistant at the Technical University of Esslingen, Germany. Since 1994, he has been with the Silesian University of Technology. In 1998, he was appointed as an Associate Professor in the Department of Automatic Control. He has also been an Associate Researcher with the Multimedia Laboratory, University of Toronto, Canada since 1999. In 2007, Dr. Smolka was promoted to Professor at the Silesian University of Technology. He has published over 300 papers on digital signal and image processing in refereed journals and conference proceedings. His current research interests include low-level color image processing, human-computer interaction and visual aspects of image quality.

DISCOVERING ABNORMAL PATCHES AND TRANSFORMATIONS OF DIABETICS RETINOPATHY IN BIG FUNDUS COLLECTIONS

Yuqian ZHOU[1] and Shuhao LU[2]

[1]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
`yzhouas@ust.hk`
[2]Department of Mechanical and Aerospace Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
`slu@ust.hk`

ABSTRACT

Diabetic retinopathy (DR) is one of the retinal diseases due to long-term effect of diabetes. Early detection for diabetic retinopathy is crucial since timely treatment can prevent progressive loss of vision. The most common diagnosis technique of diabetic retinopathy is to screen abnormalities through retinal fundus images by clinicians. However, limited number of well-trained clinicians increase the possibilities of misdiagnosing. In this work, we propose a big-data-driven automatic computer-aided diagnosing (CAD) system for diabetic retinopathy severity regression based on transfer learning, which starts from a deep convolutional neural network pre-trained on generic images, and adapts it to large-scale DR datasets. From images in the training set, we also automatically segment the abnormal patches with an occlusion test, and model the transformations and deterioration process of DR. Our results can be widely used for fast diagnosis of DR, medical education and public-level healthcare propagation.

KEYWORDS

Transfer Learning, Diabetic Retinopathy, Retinal Fundus Images, Computer-aided Diagnose, Big data analytics

1. INTRODUCTION

Diabetes is one of the most common diseases in the US affecting 86 million people, which is more than 1 out of 3 adults in America [1]. Diabetic retinopathy (DR) is a retinal damage caused by long-term effect of diabetes. It has been statistically shown that around 80% of diabetes patients who have had diabetes more than 20 years suffered from DR [2]. However, timely treatment is extremely beneficial and research indicated that early diagnosis resulted in at least 90% reduction of new patients [3]. Currently, the diagnostic screening is time-consuming and laborious. Diagnosis of DR mostly depends on manual inspection of retinal fundus images, hence

several weeks are often required for the test results. The need of expensive medical equipment and expertise also lead to a considerably high cost for population screening while regular screening of diabetic patients was considered as necessary. Consequentially, the development of automated DR detection is of vital importance to assist the clinicians for better diagnosis, and reduce the time and cost in diagnostic process.

Many studies have been reported in automated diagnosis of DR using different techniques. The diagnostic scheme of different classifications of DR depends on the weighting of various patterns and their locations [11]. Oktoeberza et al. reported 94.5% accuracy using image processing based on red channel of the retinal fundus images and optic disc segmentation detection [4] while Pratt et al. reported 75% correct detection using Convolutional Neural Networks (CNN) [5]. In terms of glaucoma and DR detection based on optic disc images of healthy and unhealthy subjects, Abdel-Ghafar et al. achieved 65% of accuracy [6]. Mookiah et al. proposed a DR detection with two class of DR and non-DR using SVM and kNN algorithms and 85% of accuracy was achieved [7]. Some other studies proposed various methodologies to process the fundus images including morphological features, color histogram and deep learning with accuracy ranging from 60% to 90% for different eye-related diseases [8, 9].

However, it is shown that automated detection using algorithms was no better than manual inspection since they are all trained on hand-labeled images. The utilization of automated detection can only simplify and cut down the time cost in diagnostic process. Besides this, special situations related to abnormal patterns still require the diagnosis from professional doctors. A kind of human-involved semi-supervised learning are also recommended to improve the accuracy [10]. Therefore, in addition to improve the detection accuracy, discovering and analyzing the characteristics and mechanisms, as well as the diagnosis techniques widely utilized by clinicians is another meaningful topic.

In this paper, we treat the DR detection problem as a regression problem instead of widely used classification, by regressing the severity degree using discrete labels from a web-based large-scale eye fundus dataset. We proposed a big-data-driven automatic computer-aided diagnosing (CAD) system based on transfer learning, starting from AlexNet [15] pre-trained on generic images, and retraining the custom regressor. In addition to achieving a satisfactory diagnosing result, we are also interested in the following important questions.

First, what is the region of one specific eye image the clinicians always focus on when they diagnosed? Can we reveal the fact by big data analytics? We show those regions by an occlusion test automatically, and provide the reference and advices for young doctors, as the experience from professional clinicians.

Second, are there any characteristics of abnormal patterns in different stages of DR? Segmentation of the abnormal patches exists in large-scale database based on occlusion test. This will help the clinicians focus more on distinctive DR cases, which also improves their diagnosis skills.

Finally, we model the transformations and deterioration process of DR in a continuous way by ranking the regressive scores. In this case, the development process of DR is revealed and help doctors and the public understand DR.

2. METHODOLOGY

A large-scale dataset was used in this study. Abnormal pattern segmentations were carefully detected using transferred regressive convolutional neural network. At the end, a reasonable transformation model of DR was achieved. The details are covered as follows.

2.1. Dataset

The database we use in this paper is provided by the kaggle competition, where large-scale and high-resolution retinal images are given by an online diagnosing website called EyePACS [13]. This organization aims at matching patients potentially suffering from vision impairment from diabetic retinopathy with professional clinical doctors, by allowing the patient to upload their own eye fundus images. Some Samples in this database are shown in Figure 1. The images are labeled in five stages: Normal (0), Mild (1), Moderate (2), Severe (3) and Proliferative DR (4) according to the severity of illness. All the labels are given by the professional doctors. In total, there are 35,126 images of left and right eyes in the training set provided. However, those images are from different devices of distinct clinical organizations, and only one rater exists for one image. Therefore, noisy data requires more preprocessing and filtering before staring the training process.
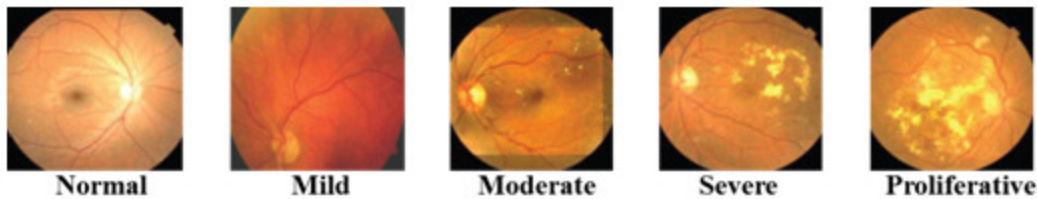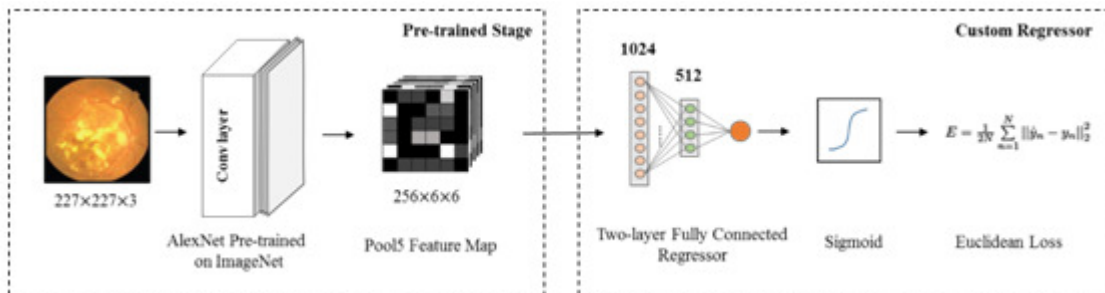

Figure 1. Some samples in the EyePACS dataset

Figure 2. System architecture of our transferred regressive CNN

2.2. Transferred Regressive Convolutional Neural Network

In the previously reported result, the whole network is designed and trained from scratch. This may yield over-fit if the training set are not large enough. However, it is shown that reusing the bottom layers of network trained on generic images will facilitate the development process of vision system, and benefit from the rich mid-level features for images in the new dataset.

The method we model the order of severity of Diabetics Retinopathy is using a regressive convolutional neural network (RegCNN). In this paper, we propose a transfer learning model, which utilizes the five bottom layers of a revised version of AlexNet [15], and connect a custom two-layer regressor with a nonlinear sigmoid function, regressing the severity according to the diagnostic labels assigned by the professional doctors. The sigmoid function is used to restrict the regression value inside a specific interval from 0 to 1. The custom regressor is fine-tuned in a fully supervised way using discrete labels, but it regresses to continuous labels. To fully utilize the high-level feature maps in pooling layer five, we do not fine-tune the bottom layers. We then utilize the Euclidean loss to train the whole network, which is defined as,

$$E = \frac{1}{2N} \sum_{n=1}^{N} \|\widehat{y_n} - y_n\|_2^2 \tag{1}$$

where N is the number of image samples, and $\widehat{y_n}$ is the estimated label, y_n is the ground truth label.

The overall structure of the network is shown in Figure 2. The regressed labels are then further used to validate the reliability of the system and model the illness severity transformations using all the training images.
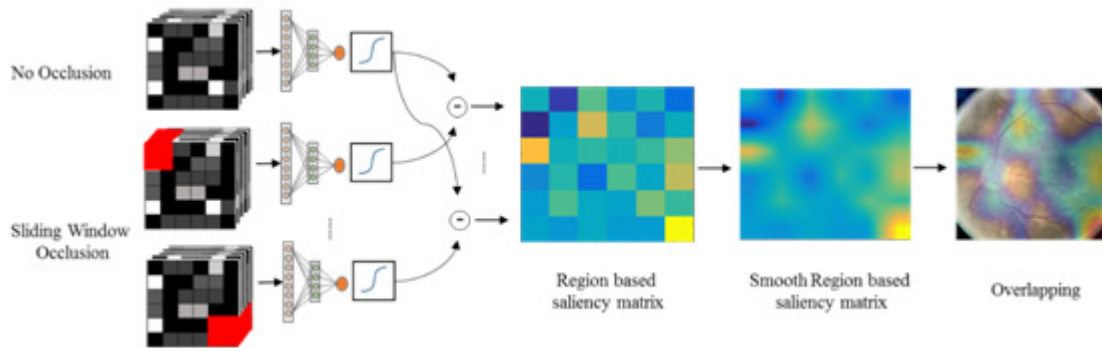
2.3. Abnormal pattern segmentation

Figure 3. The procedure of occlusion test.

Clinicians diagnose DR by checking and finding the abnormal patterns, like swelling vessels, proliferative tissues etc. We want to figure out which regions inside the eye fundus images actually help clinicians label the illness stages. Those abnormal patches contain important pattern closely related to the diagnoses. Segmenting them out automatically will highlight the critical patterns of diagnosing, assist the training process and benefit the medical education. Also, feedback given by the computer will help the clinicians prevent misdiagnosis, and mutually improve the diagnosing accuracy through human-computer interaction.

We demonstrate an occlusion test for segmentation tasks. In pooling layer five of AlexNet, the size of the feature maps is six by six, segmenting the whole pixel space into 36 small patches. We slide a window and zero out the enclosed patch across all the feature maps in pooling layer five, and monitor the regression value dropping. If the regression value of this image drops dramatically, we will regard the patch as an abnormal one, and segment it out. The process is similar in spirit to calculate the jacobian matrix with respect to neurons in each position. High

values inside the matrix mean that severity regression value is more sensitive to the patterns enclosed in this patch. Without the patterns in this patch, the patient will not be that severe since the severity degree regressed will drop. Finally, we conduct a k-means clustering on these patches and organize the typical pattern for clinical education.

We further compute the salience abnormal region for each input sample by using the above method with window size one. Patches with larger regression value dropping will be more salient and important for the model to regress the severity. The occlusion process is illustrated in Figure 3. The salient region helps the clinicians focus on the abnormal pattern localized and searched automatically by the CAD system, and verity the reliability of the diagnosis given by the computer. It will make the diagnosis process faster and more accurate.

2.4. Transformation Modelling

With the continuous regression value corresponding to the severity of DR, we want to discover the transformations and deteriorating process in the large-scale eye fundus images database. The discovery will benefit in two aspects: First, it helps the clinicians discover the intermediate states between each discrete stage label. Secondly, it helps people understand the big picture of disease transformations and will benefits the public-level healthcare propagation. In this paper, we rank and distribute the DR disease cases in the Computer-given space according to the regression value.
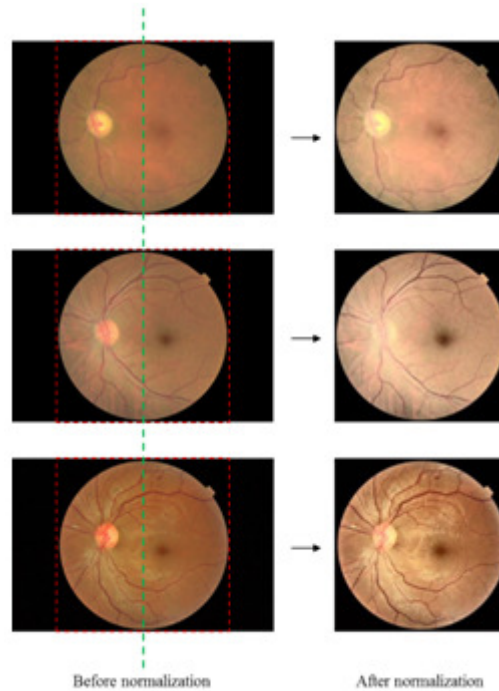
3. EXPERIMENT

Figure 4. Some examples of preprocessing result

All fundus images need to be pre-processed to eliminate the noises from both data and labels. A supervised model then was trained for regression. Besides this, occlusion test was conducted and abnormal segmentation can be found. Finally, transformations of DR were discovered. The details are illustrated as follows.

3.1. Data pre-processing

The problem of web-based database comes from plenty of noisy data and label. In the EyePACS dataset, images are collected in a crowdsourcing way, thus there is no standard for devices and illumination before collection. Normalization processing should be conducted. There are mainly three problems of the original eye fundus images, which are color variance, illumination variance, and scaling variance. Flipping and rotation do not influence the diagnosis result, thus we do not consider normalizing then in image directions.

In this case, we preprocessed the images in center cropping, color normalization, and contrast normalization. The whole normalization process is illustrated in Figure 4. For cropping, we first make the height of original images identical to each other, and crop the square regions around the center line (the blue lines). After processing, the eye fundus images seem uniform and normalized.

3.2. Regression Experiment

We want to train the model in a supervised way using the discrete labels assigned. However, there are some problems when using those labels. First, the labels are ranged from 0 to 4, but sigmoid function will only generate the values between 0 and 1. Thus, we need rescaling all the labels before training. Secondly, the samples are unbalanced across all the labels. Normal samples are far more than proliferative samples. Although the distribution somehow shows the probability of occurrence of those stages in real life, it will make the system biased to lower value. The number of samples of each label is shown in Figure 5. Finally, we redistribute the samples number by random data selection and discarding many samples. The number of samples for each stage in the whole dataset becomes normal (3000), mild (2443), moderate (2000), severe (873) and proliferative (708). Then the set is split into training set, validation set and testing set, in 80%, 10% and 10%. We stop training when the model obtains the minimum validation loss. The training process is conducted with two slaves in Amazon web platform to speed up the stochastic gradient descent. An alternative way may be data augmentation through rotation and flipping to balance the sample number.

Intuitively, regression model is more meaningful since the loss is based upon the severity, so that the loss between a proliferative sample and a normal one is larger than the loss between two closed illness stages. However, for classification model, the loss will be the same between each pair of classes. Thus, for testing result, regression model will show a clustering pattern around each value, which is more meaningful than classification, i.e. classifying a normal sample to proliferative one with only tiny probability difference. To validate the intuition, we also train a classification model using the same features and bottom layers as the regression model, where the loss function is replaced by cross entropy of softmax function. We compare the regression and classification result distribution for each class. For the classification mode, we directly got the discrete distribution through the confusion matrix. For the regression model, we use a kernel

smoothing function to estimate the distribution of regression value after scaling back for each labeled class. The result is shown in Figure 6.
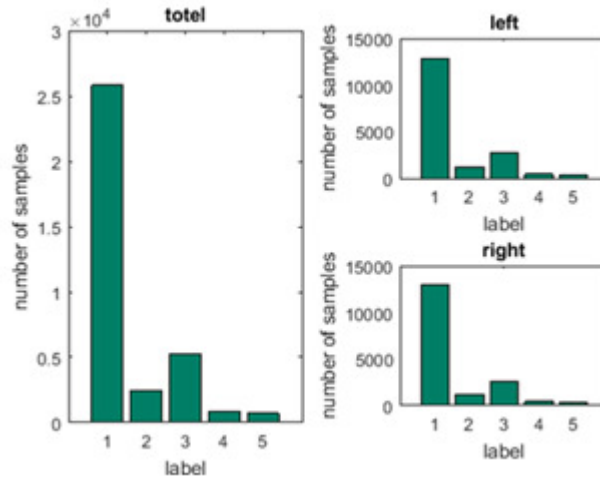

Figure 5. The label distribution in the original dataset.

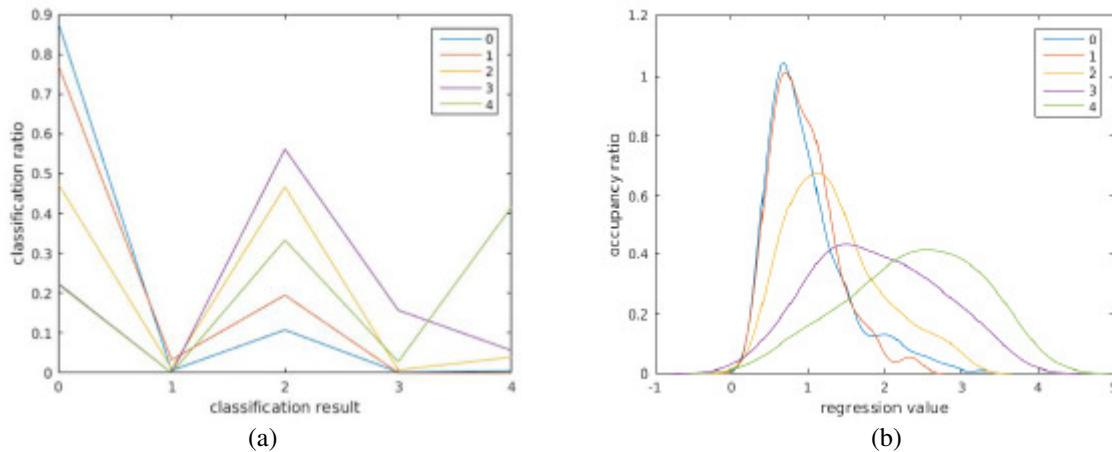(a)   (b)

Figure 6. The result distribution on the testing set of two models.

As shown, in the classification mode, it achieves a 61% accuracy and normal cases contribute a lot to the accuracy. And the detection may skip stages, i.e. classifying proliferative one to normal one, which is unacceptable. However, for the estimated distribution of regression value for the regression model, it demonstrates a more meaningful distribution. Most of the cases can be differentiated according to the severity. And the regression value for each labeled class shows a normal-distribution-like pattern both in the training and testing set.

Because of the characteristics of sigmoid function, values cannot be regressed to exact zero or one, or exceed this range, thus, the mean of the regression values of proliferative samples may not be able to fall on value 4, instead, shift to the left around 3. To compare the result with other classification models and build a detector, we consider the similar distribution in the training set, and split the range of 0 to 4 into five stages, avoiding overlapping. The method to achieve the

non-overlapped distribution is starting from the estimated peak value, and extending to both sides according to the standard deviation (SD). We then use this separation to classify the testing set. The result of range separation is illustrated in Figure 7. The testing result and one-vs-all ROC curve is shown in Figure 8.
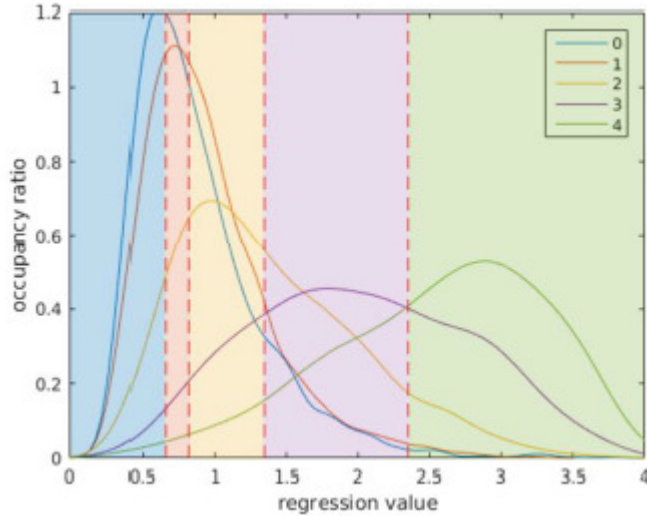

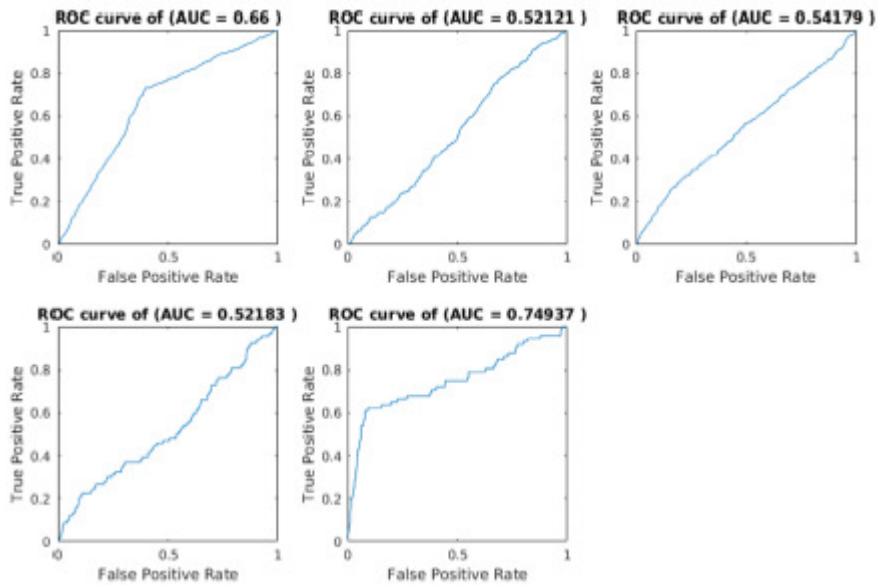Figure 7. The testing value range of regression.

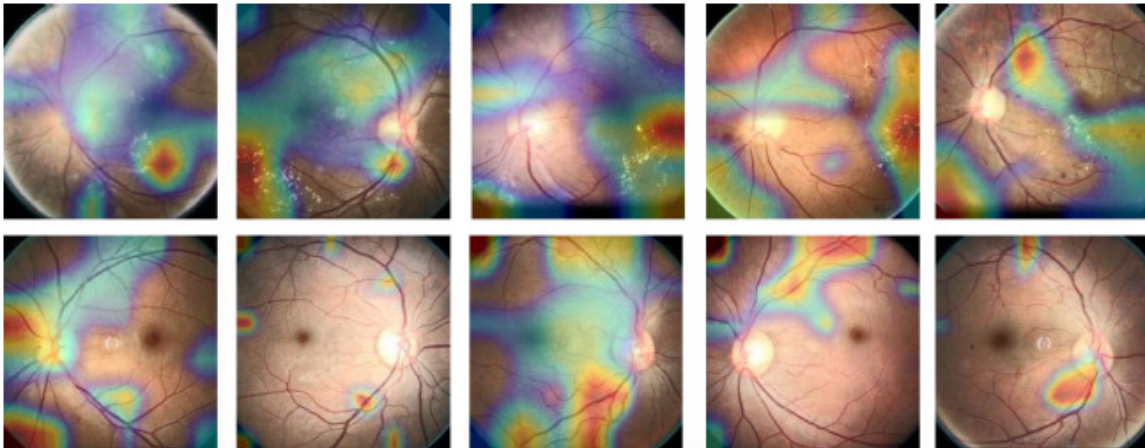Figure 8. The ROC curve of the one-vs-all classification result.

Figure 9. Some DR focus examples detected by occlusion test.

Figure 10. The clustering result and some samples of abnormal patch segmented by occlusion test.

However, the testing result is not satisfactory enough for an accurate detector. It mostly results from noises in original images and labels, and large overlapping region among classes after regression. As shown in Figure 7, the overlapping region between two classes are too large, making it hard for this model to differentiate two distinct stages. But it is still better than classification model as a binary detector (one-vs-all). Better regressive model still should be explored and developed in the future work.

3.3. Occlusion test and abnormal segmentation

We ran the occlusion test with sliding window size setting to one to generate the saliency map. Some random selected examples of the saliency map generated for each class are shown in Figure 9. As shown, images of late period of DR (the first row) will show patterns like yellow spot or proliferative tissue shape. However, images from early period of DR (the second row) only show the swelling or blocked vessels. The occlusion test is able to detect and focus on those abnormal patterns, and judge the severity. We can conclude that our regression model is using the same mechanism as the professional clinicians.

Furthermore, we segmented the abnormal patches by running the occlusion test with sliding window size setting to two. And the dropping difference threshold of segmentation is 0.2. After the test, we get totally 2772 patches. We extract the deep features using AlexNet again, and use k-means to cluster them with k=20. Finally, we visualize the segmentation using t-SNE [13] method according to the pairwise distance. The visualization is shown in Figure 10. In the first two columns, the 20 clusters are illustrated. From the graph, we found that similar pattern will be clustered together. In the third column, some segmented abnormal patches for each stage are shown. From our result, the segmented patches for one specific stage are from some fixed clusters. For the late period, the patches will be more related to proliferative tissue and yellow spot. For the early stages, the patches are related to vessels. Such segmented patches can be further used for medical education.
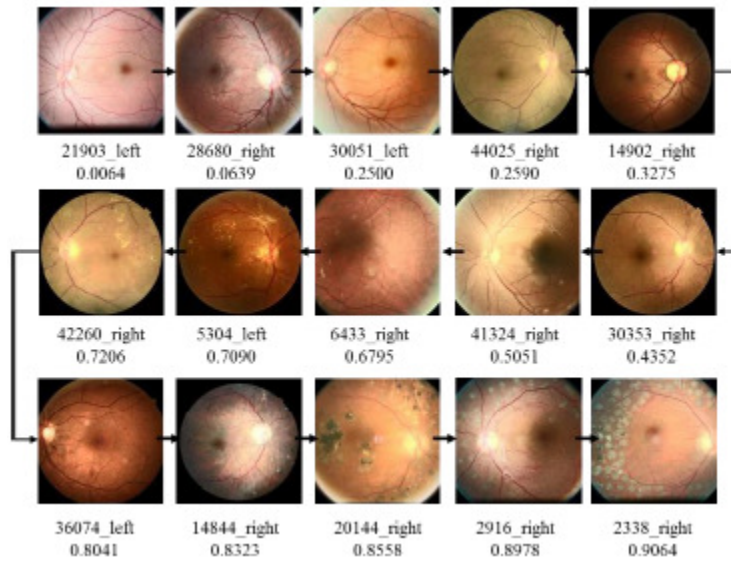

Figure 11. The transformations we model according to the regression scores.

3.4. Transformations

Finally, we discovery the transformations of DR by the ranking of the regression scores in the training set. From the ranking, we randomly select the fundus images and plot them in order of increasing regression scores. The transformation is shown in Figure 11. Through big-data-driven analysis, the system is able to regress and rank the stages of DR automatically, and we can find the deteriorating process from normal eyes in a more detailed and continuous way. The organized data will help the clinicians clearly understand the development stages of DR, and diagnose more accurately.

4. CONCLUSIONS

In this paper, we proposed a big-data-driven automatic diagnosis system based on transfer learning of deep CNN. Instead of doing classification directly, our system focuses on regressing the severity of diabetic retinopathy. Based on the trained model, we conducted an occlusion test to search for the abnormal patches in the eye fundus images, and segmented them out for medical

education. We revealed the mechanism of diagnosing DR utilized by professional clinicians and computer-aided system by showing the saliency map obtained from occlusion test. Furthermore, according to the regression value, we modeled the transformations and deteriorating process of DR. Our system will benefit efficient and accurate self-inspection for patients, and contribute to medical education and public-level healthcare propagation by assisting the learning process of young clinicians.

ACKNOWLEDGEMENTS

REFERENCES

[1]   C. f. D. Control and Prevention, "National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014," Atlanta, GA: US Department of Health and Human Services, vol. 2014, 2014.

[2]   P. J. Kertes and T. M. Johnson, Evidence-based eye care: Lippincott Williams & Wilkins, 2007.

[3]   R. J. Tapp, J. E. Shaw, C. A. Harper, M. P. De Courten, B. Balkau, D. J. McCarty, et al., "The prevalence of and factors associated with diabetic retinopathy in the Australian population," Diabetes care, vol. 26, pp. 1731-1737, 2003.

[4]   K. W. Oktoeberza, H. A. Nugroho, and T. B. Adji, "Optic Disc Segmentation Based on Red Channel Retinal Fundus Images," in International Conference on Soft Computing, Intelligence Systems, and Information Technology, 2015, pp. 348-359.

[5]   H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional Neural Networks for Diabetic Retinopathy," Procedia Computer Science, vol. 90, pp. 200-205, 2016.

[6]   R. Abdel-Ghafar and T. Morris, "Progress towards automated detection and characterization of the optic disc in glaucoma and diabetic retinopathy," Medical informatics and the Internet in medicine, vol. 32, pp. 19-25, 2007.

[7]   M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," Computers in biology and medicine, vol. 43, pp. 2136-2155, 2013.

[8]   J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, "Automated diagnosis of glaucoma using digital fundus images," Journal of medical systems, vol. 33, pp. 337-346, 2009.

[9]   U. R. Acharya, S. Dua, X. Du, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," IEEE Transactions on information technology in biomedicine, vol. 15, pp. 449-455, 2011.

[10]   M. J. Greaney, D. C. Hoffman, D. F. Garway-Heath, M. Nakla, A. L. Coleman, and J. Caprioli, "Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma," Investigative ophthalmology & visual science, vol. 43, pp. 140-145, 2002.

[11] E. T. D. R. S. R. Group, "Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10," Ophthalmology, vol. 98, pp. 786-806, 1991.

[12] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in Proceedings of the 24th international conference on Machine learning, 2007, pp. 759-766.

[13] EyePACS Dataset. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/data.

[14] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, vol. 9, pp. 2579-2605, 2008.

[15] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097-1105.

AUTHORS

Yuqian ZHOU, M.phil candidate of HKUST. His research focuses on affective computing based on deep vision, pattern recognition, machine learning and medical image processing.

Shuhao LU, Ph.D candidate of HKUST. His research focuses on Biomechanics, Ophthalmology, and medical sensor.

LARGE SCALE IMAGE PROCESSING IN REAL-TIME ENVIRONMENTS WITH KAFKA

Yoon-Ki Kim[1] and Chang-Sung Jeong[2]

[1,2]Department of Electrical Engineering, Korea University, Seoul, South Korea
[1]vardin@korea.ac.kr
[2]csjeong@korea.ac.kr

ABSTRACT

Recently, real-time image data generated is increasing not only in resolution but also in amount. This large-scale image originates from a large number of camera channels. There is a way to use GPU for high-speed processing of images, but it cannot be done efficiently by using single GPU for large-scale image processing. In this paper, we provide a new method for constructing a distributed environment using open source called Apache Kafka for real-time processing of large-scale images. This method provides an opportunity to gather related data into single node for high-speed processing using GPGPU or Xeon-Phi processing.

KEYWORDS

Real Time Image Processing, Distributed Processing, Real-Time Processing, Apache Kafka

1. INTRODUCTION

Recently, the development of video equipment such as CCTV, satellite, and drone has increased the volume of real-time image data. Large Scale of real-time image data coming from many camera channels is difficult to process at high speeds with single resources such as CPU and GPU. GPU has been developed for high-speed image processing, however it is not enough for large-scale image processing because of the limit of memory shortage. There is also an approach of HIPI [1] to large-scale image processing using the HDFS, but this is also inadequate for real-time image processing due to the limitations of the batch processing. In addition, the HDFS [2] uses a random access approach to the disk. It causes HDD I/O bottleneck. For this reason, HIPI is insufficient to process the real-time data continuously.

In order to process a large-scale image occurring in real-time, it is necessary to have a processor and memory capable of accommodating it. Large-scale image processing using multicore and GPU in a single node cannot accommodate data due to insufficient memory. To compensate for this, it is essential to distribute processing for large-scale images using cluster composed of several nodes. There are two major issues in approaching this way. The first is the memory acceptance issue at a single resource such as main memory and GPU. As the amount of data that can be accommodated in a node increases and reaches the limit of the memory capacity, the

processing speed is rapidly reduced [3]. Since the data processing speed is slower than the data occurring speed, it is important not to lose the occurring data while the data is being processed. Second issue is the communication overhead between divided tasks. When a large task is divided into small tasks and distributed processing is performed, communication occurs between the divided tasks. The more associations between divided tasks, the more communication will occur. Communication overhead is caused by dividing data between each task into more than necessary pieces [4]. This problem becomes more severe when the task is divided into highly correlated tasks. For this reason, it is important to divide tasks appropriately in order to prevent a lot of communication between divided tasks.

Apache Kafka provides features that are suitable for addressing the issues mentioned previously. Kafka is distributed messaging system for log processing. Kafka uses Zookeeper [5] to organize several distributed nodes for storing data in real time stably. It also stores data in several messaging topics. It provides a messaging queue that allows nodes which needs a message to access and process it.

In this paper, we propose a new method to process large scale image data in real-time using Kafka. This method enables large-scale image data in real time to be processed quickly. In this method, large-scale images can be processed in real time with a resilient scale-out of computing resources.

The outline of our paper is as follows: In Section 2, we describe related works for introducing Apache Kafka and some approaches to handling large-scale images in real time. Section 3, we explain a new method to process large scale image data in real-time using Kafka. Section 4 explains implementation of proposed method and shows its experimental results. Lastly Section 5 summarizes the conclusion of our research.

2. RELATED WORKS

Apache Kafka [6] can be used to collect various types of data in real time. Kafka is an advanced open source project, and its ability to handle high-volume streams such as Internet of things and log data. It performs better than existing messaging system [7-9] with specialized architecture for large-scale real-time log processing. It is suitable for both offline and online message usage. And it is based on publish-subscribe model and consists of a producer, a consumer, and a broker. Producer generates a message and publishes it to a specific topic. And the consumer takes it and processes it. A broker is a server cluster that manages messages so that producers and consumers can meet, and categorizes messages which received from producer. It can be configured as multiple broker Clusters, and each node is monitored by Zookeeper.

There are several advantages to adopting Kafka in this paper. Firstly, since messages are stored in the file system, the durability of the data is ensured without any special configuration. Kafka uses sequential access to the HDD, resulting in faster performance than memory random access methods. Performance comparison of HDD sequential access method and memory random access is compared in [10]. In the existing messaging system, the performance of the system decreases drastically as the number of messages left unprocessed increases. However, since Kafka stores the message in the file system, the performance is not greatly reduced even if a large amount of messages are accumulated. It can also be used to accumulate data for periodic batching as well as real-time processing, since many messages can be stacked. Secondly, Kafka provides fault

tolerance by storing messages in the file system and replicating them to the broker. Unlike an existing messaging system that immediately deletes an acknowledged message from a consumer, it does not delete the processed message but leaves it in the file system and deletes it after a set lifetime. Since the processed message is not deleted for a certain period of time, the consumer can rewind the message from the beginning if a problem occurs during processing of the message or if the processing logic is changed. Thirdly, Kafka provides storage capacity for large-scale data using pulling method. In traditional messaging systems, brokers push messages to consumers, while Kafka acts as a pull consumer, taking messages directly from the broker. Therefore, a consumer can get optimal performance by fetching only the messages of his processing capacity from the broker. Push-based messaging system, the broker directly calculates which messages each consumer should process, and tracks which messages are being processed. In Kafka, the consumer pulls the necessary messages directly from the broker. By pulling messages from broker, it is possible to build a batch consumer that stacks and periodically processes messages. For these reasons, we decided to adopt Kafka in the process of handling large-scale image processing on distributed environments.

3. METHODOLOGY

Real-time stream data is constantly generated so as not be accommodated in the memory continuously. Parallel processing of such a large amount of stream data in a single node is limited by the memory capacity limit. Also, communication overhead is caused by frequent data transmission between the main processor and the coprocessor. Performing only distributed processing on multiple nodes also causes communication overhead due to large-scale parameter exchange between nodes in the cluster. In order to solve this problem, we present a new architecture and several methods for large scale image processing using apache Kafka.

The model for real-time processing of large-scale images consists of three parts. The first is the part that detects the frame from the stream channel. The stream channel may be a channel directly connected to the camera or a remote channel transmitted through the RTSP protocol. This part is responsible for producing messages and consists of several nodes. One node can sense data coming from one or several channels. Depending on the resource capability of each node, it is determined how many channels can be detected in one node. In the part detecting the channel, the frame of the video is transmitted to the Kafka broker.

The second part is a Kafka broker. In this part, a large amount of stream data transmitted from the previous part is stored in a buffered queue. Generally, the processing speed of the stream is slower than generation rate. For this reason, in order to process a large amount of video data without loss, it is necessary to temporarily store the detected stream data. We used Kafka for a stream buffer. There are several reasons why we used Kafka to store large amounts of video in temporary buffers. The reasons we adopted Kafka in buffering large image processing is as follows.

- Since the video message is stored in the file system, the durability of the data is guaranteed. This advantage allows data to be permanently stored in the buffer, even if none of the nodes can process the video stream. The video stream is deleted from the file system at the time set by the user. In addition, since the video stream generated in real time has a large capacity, it can overcome the disadvantage that it is difficult to be loaded in the main memory.

- Instead of distributing messages to the distributed nodes in the master node, each node ready to process takes a message from the buffer and processes it. In this condition, even if the performance of each node is different, it is possible to perform high-speed processing without bottleneck of stream data because the node ready for processing processes the data.

The last part is that each node has an image processing application, which takes and processes stream data from the Kafka broker. The image processing application at each node is intended to perform the same operation. Since each node has one frame at a time, one frame can be seen as a set of related data. In the text processing system, data is transmitted or processed in units of tuples. In the proposed system, the image processing unit is regarded as a frame. It is a feature of image processing that many repetitions of matrix or vector operations are included. It can be processed quickly by using GPU accelerator considering frame as basic unit of processing. In this part, we use the asynchronous method of processing data regardless of the order of image generation.

We construct the model of this system with the three parts mentioned above, and this model is expressed as the following figure 1.
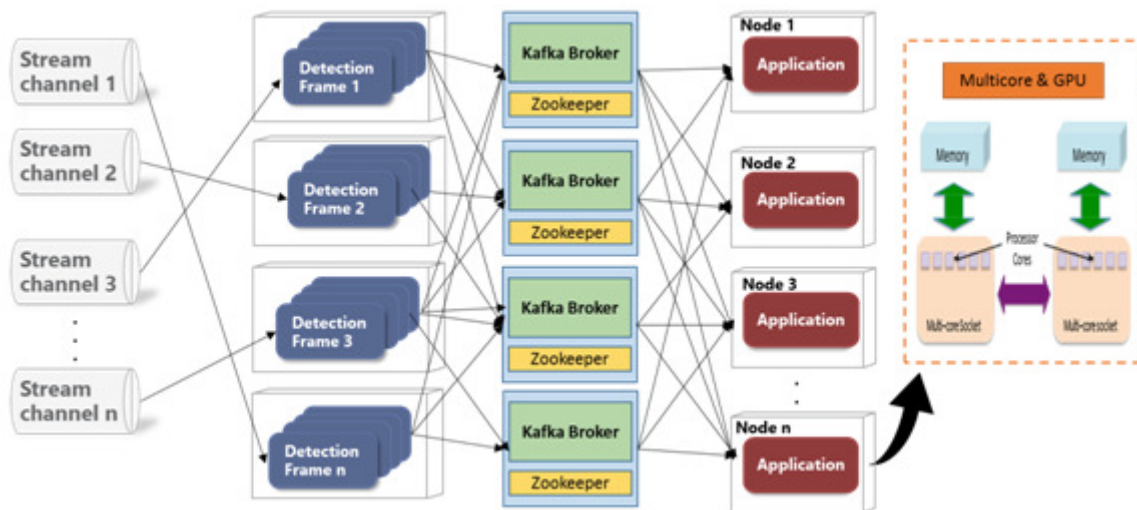

Figure 1. The overall model of large scale image processing using Kafka

3.2 Strategies

To process large-scale images in real time, we present several strategies as follows.

(1) Since the same operation is performed at the application level, the resolution of the image must be synchronized. When an image is distributed to multiple distributed resources, the image is shuffled. Therefore, when images of different resolutions are transmitted to a distributed environment, the application may require excessive workload, which may degrade performance. For this reason, the size information of the image must be transmitted to the application level in the part detecting the frame.

(2) The unit transmitted at one time is the frame of the image. In general, one frame transmitted from a camera channel has multiple channels. One frame is serialized into a byte array for transmitting to the Kafka broker. When the byte array is transmitted from the Kafka broker, the original image is obtained by re-encoding based on information such as the resolution of the image and the number of channels.

(3) Transfer image data asynchronously to Kafka Broker. This is a strategy to increase image processing speed in distributed resource environment composed of multiple nodes. This results in faster image processing, while the order of the images can be shuffled. It can be used for the purpose of identifying an object in a video generated in a large camera channel and storing the value permanently by taking only the information of the identified object.

(4) Take advantage of multiple distributed nodes to retrieve data from topics and parallelize them. A frame from one channel can be sent to multiple topics to allow distributed nodes to retrieve and process data from each topic. Alternatively, multiple nodes can access and process data with the same topic. The larger the number of channels, the greater the memory requirement for processing video frames, which increases the size of the node and maximizes memory capacity.

3.2.1 Transmission of frame

As shown in Figure 2, it is necessary to serialize the data to transfer the frame to the Kafka broker. The data coming from the channel has three channels in matrix form. In order to transmit this, a matrix composed of three channels of RBG is converted into a byte array. When this process is performed, the height and width information of the frame is lost as to how many channels are formed by one frame. This information is needed to re-encode the image at the application level, so once the channel is detected, it should be sent only once at the moment of connection with the Kafka broker.
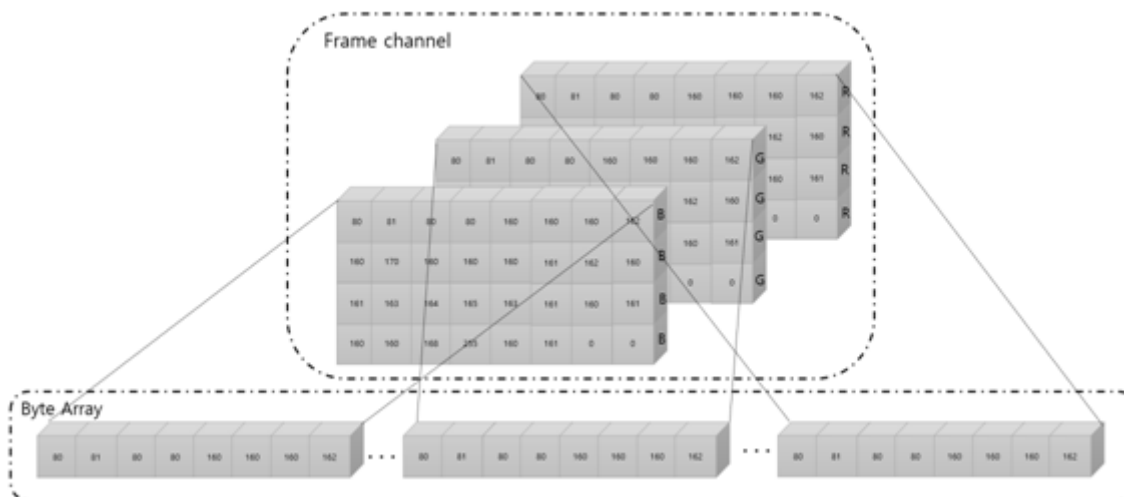

Figure 2. Conversion of frame to byte array

3.2.2 Distribution of image channel

Channels coming in on multiple channels will go into different queues through topics provided by Kafka's brokers. Figure 3 below shows how each channel is distributed across multiple topics. Multiple applications on distributed resources process data according to the resource situation and then take data by accessing the Kafka topic and bringing the message directly. As shown in the figure, the message can be retrieved by accessing multiple topics from one node, and the message stored in one topic can be distributed by allowing access by multiple nodes.
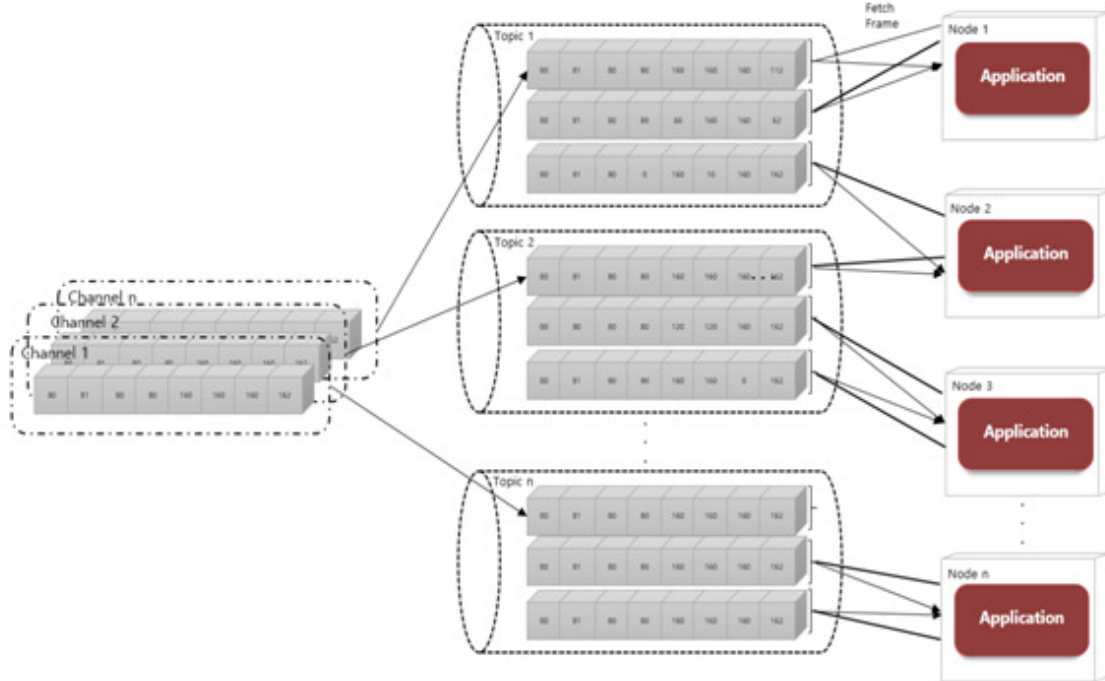

Figure 3. Distribution of each channel frame

4. EXPERIMENTAL RESULTS

We configured the Kafka broker with three physical nodes which has Intel® core™ CPU E6550 2.33 GHz processors and 4GB memory. Nodes at the application level have been run in a virtual environment to make it easier to reshape the number of nodes. Each node in the application level has a quad core virtual CPU and 8GB memory.

In the experiment, the image is processed 60 seconds after the channel is generated. The following accumulate means the total number of frames generated. And consume is the number of frames taken from the Kafka broker. Finally, lag means the number of frames remaining in the buffer without being processed. As shown in Figures 4, 5, and 6 below, the lag drops sharply as the number of processing nodes increases. As the number of nodes increases, the difference between the number of frames accumulated in the buffer and the number of processed frames decreases.
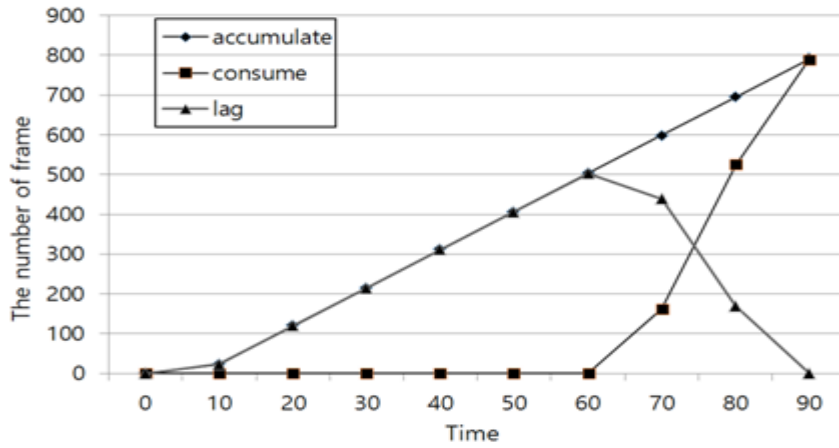
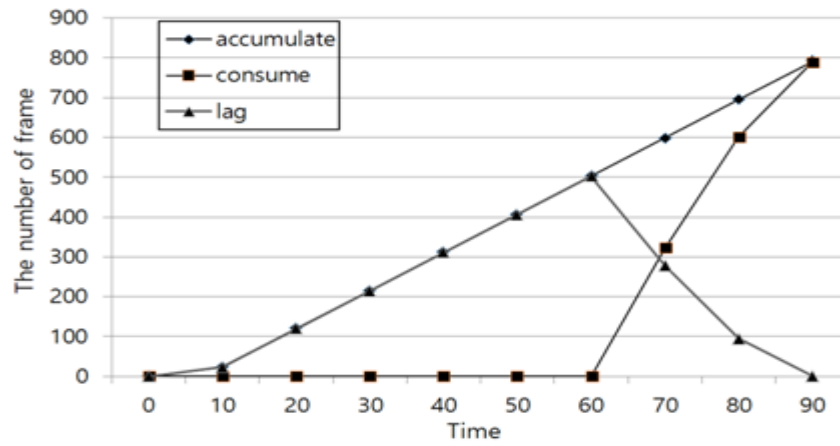Figure 4. The result of one topic and one node

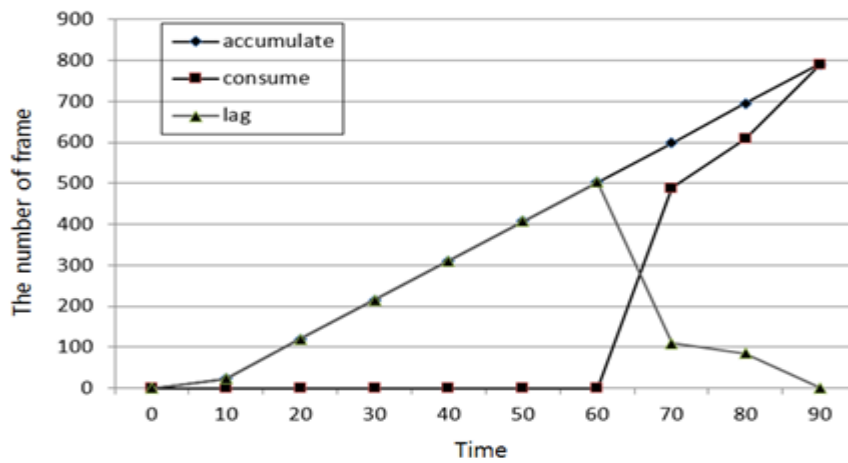Figure 5. The result of two topics and two nodes

Figure 6. The results of four topics and four nodes

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a new large-scale image processing system with Kafka. Existing systems use GPGPU on a single node to process images for high performance in real-time environments. However, this method is not suitable for large-scale image processing. For large-scale image processing, memory capacity must be available to accommodate large-scale images. Although the GPU is well suited for high-speed processing of images, it still has limited memory capacity. We suggested new large-scale image processing model that use a Kafka to create a distributed environment. It allows overcoming the memory capacity that cannot be accommodated by one node. In particular, since image data can be stored in the file system, it is advantageous to handle large-scale images without data loss. In this paper, we have not been able to perform experiments that use GPUs on a single node to achieve even greater performance. Future research will focus on how to use GPUs or Xeon-phi for each node to achieve higher performance in large-scale image processing.

ACKNOWLEDGMENTS

REFERENCES

[1]   Sweeney, Chris, et al. "HIPI: a Hadoop image processing interface for image-based mapreduce tasks." Chris. University of Virginia (2011).

[2]   Shvachko, Konstantin, et al. "The hadoop distributed file system." 2010 IEEE 26th symposium on mass storage systems and technologies (MSST). IEEE, 2010.

[3]   Gregg, Chris, and Kim Hazelwood. "Where is the data? Why you cannot debate CPU vs. GPU performance without the answer." Performance Analysis of Systems and Software (ISPASS), 2011 IEEE International Symposium on. IEEE, 2011.

[4]   Xu, Zhiwei, and Kai Hwang. "Modeling communication overhead: MPI and MPL performance on the IBM SP2." IEEE Parallel & Distributed Technology: Systems & Applications 4.1 (1996): 9-24.

[5]   Hunt, Patrick, et al. "ZooKeeper: Wait-free Coordination for Internet-scale Systems." USENIX Annual Technical Conference. Vol. 8. 2010

[6]   Kreps, Jay, Neha Narkhede, and Jun Rao. "Kafka: A distributed messaging system for log processing." Proceedings of the NetDB. 2011.

[7]   Snyder, Bruce, Dejan Bosnganac, and Rob Davies. ActiveMQ in action. Vol. 47. Manning, 2011.

[8]   Rostanski, Maciej, Krzysztof Grochla, and Aleksander Seman. "Evaluation of highly available and fault-tolerant middleware clustered architectures using RabbitMQ." Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. IEEE, 2014.

[9]   Hintjens, Pieter. ZeroMQ: Messaging for Many Applications. " O'Reilly Media, Inc.", 2013.

[10] Jacobs, Adam. "The pathologies of big data." Communications of the ACM52.8 (2009): 36-44.

AUTHORS

Yoon-Ki Kim is currently working toward the ph.D degree in Electronic and Computer Engineering at the Korea University. His research interests include real-time distributed and parallel data processing, IoT, Sensor processing and computer vision.

Chang-Sung Jeong is a professor at the department of EE/CE at Korea University. He received his MS.(1985) and Ph.D.(1987) from Northwestern University, and B.S.(1981) from Seoul National University. Before joining Korea University, he was a professor at POSTECH during 1982-1992. He also worked as an associate researcher at UCSC during 1998-1999.

INTENTIONAL BLANK

ARABIC DATASET FOR AUTOMATIC KEYPHRASE EXTRACTION

Mohammed Al Logmani[1] and Husni Al Muhtaseb[2]

[1]Information Technology, Saudi Aramco, Dhahran, Saudi Arabia
Mohammed.logmani@aramco.com
[2]Information & Computer Science Department, King Fahd University for
Petroleum & Minerals, Dhahran, Saudi Arabia
muhtaseb@kfupm.edu.sa

ABSTRACT

We propose a dataset in Arabic language for automatic keyphrase extraction algorithms. Our Arabic dataset contains 400 documents along with their keyphrases. The dataset covers eighteen different categories. An evaluation using a state-of-the-art algorithm demonstrates the accuracy of our dataset is similar to that of English datasets.

KEYWORDS

Keyphrase extraction, Arabic, dataset

1. INTRODUCTION

Automatic keyphrase extraction aims to extract a set of phrases that are highly relevant and most descriptive phrases for the input text. The process of extracting keyphrases is achieved systematically and with no or minimal human interference. Keyphrase extraction algorithms mine the data corpus to extract important phrases and label the documents with these phrases. To verify the accuracy of the extraction algorithms, datasets are used. These datasets contain training and test documents along with the keyphrases representing the content of each document.

For the English language, there are many verified datasets. To illustrate some examples, there is Reuters Dataset [1] which contains more than 20,000 of documents with focus on text classification field. Also, the work presented in [2] prepared a large dataset consists of 2000 text from scientific papers. Additionally and on the same field of scientific papers, there are datasets submitted for the Workshop on Semantic Evaluation 2010 (Sem-Eval 2010) [3]. These datasets are tailored for machine-learning automatic keyphrase extraction algorithms.

For the Arabic language, we didn't find any published datasets that target the area of Arabic keyphrase extraction. However, we did find some related research including the human annotated Arabic dataset which provides annotation on books reviews [4]. Also, the work in [5] describes a corpus that classifies newspaper text into seven domains. Whereas the work in [6] explains a dataset with more than 17,000 texts with focus on text classification problem.

In the few proposed algorithms for keyphrase extraction in Arabic, the number of documents used to experiment with the algorithm was small as mentioned in the respective publications. For example, KP-Miner [7] used a set of 100 Wikipedia documents where the algorithm proposed by El-Shishtawy [8] used a dataset consisted of 50 documents.

The main contribution of this work is the new Arabic dataset we have prepared. Below we explain the sources for the articles and the methodology used to create the dataset.

2. ARABIC DATASET

This section explains the sources for the articles and the methodology used to create the dataset. Our dataset contains 400 documents distributed on 18 different categories.

2.1. Sources

We have collected the documents from two sources: Arabic Wikipedia [9] and King Abdullah Initiative for Arabic Content [10].

- Arabic Wikipedia: is the main source of articles used in creating this dataset. Arabic Wikipedia contains more than 198,349 pages. For our goal, we obtained 365 articles from Wikipedia. Out of the 365 articles, approximately 200 articles were obtained from a previous work by Shaaban [11] where the rest were collected using BzReader [12]. BzReader is an application that allows offline browsing of the Wikipedia dump files and displays the text-only version of Wikipedia pages

- King Abdullah Initiative for Arabic Content: is an initiative aims to enrich the Arabic content on the internet after noticing the small percentage of Arabic content. According to this initiative, the percentage of Arabic digital content does not exceed 0.3% out of the world content composed of other languages. For our goal here, we obtained 35 articles with focus on medical topics.

2.2. Selection and organization

The corpus covers different knowledge areas like religion, history, geography, technology, sciences, sports…etc. Selecting the documents from different fields would help future automatic keyphrase extraction algorithm to cover general domains and not be tied to a specific domain like scientific papers. These documents vary also in size from 1 to 30 pages. The total number of words in these documents ranges approximately from 172 to 17,589 words. The number of words in the whole dataset is 1,708,168 words distributed on 288,191 lines. The documents are saved in text files with the extension (.txt) as Unicode format UTF-8.

The largest category with regard to number of documents is the people category with 59 documents where the smallest one is the food category with 3 documents. We also calculated the density percentage defined as the average number of words per file in each category. This measure shows the richness of a certain category based on the longest files they have and not based on the number of documents under that category. When calculating the density score, countries category scored the highest with 7,366. The next highest category is religion with 5,960 average words per file. This category contains 16 files. In this measure, food category scored last

with 1,233. For the health and medicine category, the density score is 1,950, which is very small comparing to the number of files (51).

The largest file in the Arabic dataset is from the history category and it is about the Ottoman Empire (الدولة العثمانية) with 17,589 words and 3,094 lines. The smallest file belongs to the environment category and it discusses radioactive pollution (التلوث الإشعاعي) with 172 words and 32 lines.

Table 1 shows the 18 categories we have chosen to use for the categorization of the files in our dataset. It also shows the sub-categories, the number of files, the total number of words, the total number of lines, and the density percentage in each category.

Table 1. Distribution of the documents in the Arabic dataset.

Category	Sub-Category	Number of Files	Number of lines	Number of words	Density
History	History	39	32,976	4,991	49.9%
Culture	Culture, Social, Cloths, Language, Buildings, palace, Festival, Flags	22	15,615	4,172	41.7%
Countries	Country, City	58	73,757	7,366	73.7%
Aviation	Airplane, Airport, Air Machine	5	1,954	2,450	24.5%
Health & Medicine	Health, Medicine, Medical	51	16,605	1,950	19.5%
Animals	Animal, Dinosaur, Zoology	29	26,606	5,459	54.6%
War	Battles, War Machines	21	8,460	2,459	24.6%
Technology	Technology, Software Engineering	12	8,631	4,237	42.4%
Sciences	Chemistry, Electricity, Energy, physics, Law	11	6,136	3,165	31.6%
Economy	Company, Economy	10	4,310	2,556	25.6%

Environment	Environmental Issues, Pollution	12	4,904	2,353	23.5%
Space	Space	20	10,621	3,258	32.6%
Entertainment	Fiction, Movie, Music	12	7,418	3,766	37.7%
Food	Fruit	3	624	1,233	12.3%
Geography	Geography, Mountain	8	2,696	1,989	19.9%
People	People	59	45,400	4,698	47.0%
Religion	Religion	16	16,400	5,960	59.6%
Sports	Sports	12	5,078	2,577	25.8%

2.3. Cleaning Up

When converting from Wikipedia pages to text format using BzReader, some clean-up for the format was needed. The clean-up process included removing some text that may confuse the readers or make the articles hard to read. Text that was generated due to the conversion from HTML/ Rich text format was eliminated. This includes place holders of graphics, sounds, and videos. To illustrate this point by an example, if the article contains several images, then the word (png) or (jpg) will be repeated several times in the text version of the article. Hence, this will increase the chance of selecting (png) or (jpg) as a keyword. This is because many of Keyphrase Extraction algorithms e.g. KEA [13] and KP-Miner [7] use term frequency as a factor when selecting candidates for keyphrases. The clean-up process included Wikipedia tables, side images captions, some references ...etc.

2.4. Manual Keyphrase Extraction

All documents were assigned to six readers to read and extract 10 keyphrases from each file. These keyphrases are stored in separate files with '.key' extension. This format is the one used by KEA and some other Automatic Keyphrase Extraction tools. In the '.key' files, each row represents a Keyphrase. They are sorted based on their importance in the article from high importance to low importance. The '.key' file name is matching exactly the '.txt' file. This is done to help the algorithm to locate the files in the training phases and help in organizing the dataset.

2.5. Keyphrase Verification

The final step in the methodology of preparing the Arabic dataset is the verification step. It included proofreading the articles and adjusting or concurring with the extracted keyphrases. This step also included reviewing and correcting the spelling mistakes, the number of keyphrases, and the '.key' files format.

As part of this work, the dataset was made available for future work related to Arabic keyphrase extraction. The dataset can be found at this link: https://github.com/logmani/ArabicDataset.

3. EVALUATION OF THE DATASET

To evaluate of the quality of our dataset, we conducted an evaluation using KEA, which is one of the most worked on algorithms in the area of keyphrase extraction. In our evaluation, we trained KEA using 300 documents, and our test set contained 100 documents. The F-measure (F-score) on our dataset was 19% which is very close to the results reported on [2] which was 19.08%. Furthermore, we ran KEA using the English dataset prepared in [3] and we found out the F-score was 15.3%. Hence, the quality of our dataset can be used reliably in future work related to keyphrase extraction algorithms. Table 2 shows a summary of our experiment on the Arabic datasets including the values of exact matching measure, precision, and recall.

Table 2. Summary of results obtained for the Arabic dataset.

Measure	Results on Arabic Dataset
Exact Match	189
Precision	18.9%
Recall	19.1%
F-Score	15.3%

4. CONCLUSIONS

In this research work, we prepared and presented an Arabic dataset for automatic keyphrase extraction. The dataset contains 400 documents along with their correspondence 400 keyphrases files. The dataset is publicly available for future automatic keyphrase extraction research on Arabic language. The evaluation showed that the quality of the dataset is reliable to be used in the future.

ACKNOWLEDGEMENTS

REFERENCES

[1] Lewis, David. "Reuters-21578." Test Collections 1 (1987).

[2] Krapivin, Mikalai, Aliaksandr Autaeu, and Maurizio Marchese. "Large dataset for keyphrases extraction." (2009).

[3] Kim, S. N., Medelyan, O., Kan, M. Y., Baldwin, T.: Semeval-2010 task 5: Automatic keyphrase extraction from scientific articles. In Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 21-26. Association for Computational Linguistics. (2010)

[4] AL-Smadi, M., Qawasmeh, O., Talafha, B., Quwaider, M.: Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis. Proceedings of 3rd International Conference on Future Internet of Things and Cloud (FiCloud 2015), Rome, Italy (2015)

[5]	Goweder, A., De Roeck, A.: Assessment of a significant Arabic corpus. Presented at the Arabic NLP Workshop at ACL/EACL 2001, Toulouse, France, (2001)

[6]	Khorsheed, M., Al-Thubaity, A.: "Comparative evaluation of text classification techniques using a large diverse Arabic dataset." Language resources and evaluation vol.47, no. 2, pp.513-538 (2013).

[7]	El-Beltagy, S., Rafea, A.: KP-Miner: A keyphrase extraction system for English and Arabic documents. Inf. Syst, vol. 34, no. 1, pp. 132–144. (2009)

[8]	El-Shishtawy, T., Al-Sammak, A.: Arabic keyphrase extraction using linguistic knowledge and machine learning techniques. arXiv preprint arXiv:1203.4605. (2012)

[9]	Wikipedia, "Arabic Wikipedia." [Online]. Available: http://ar.wikipedia.org/wiki. [Accessed: 01-Jul-2012].

[10]	King Abdullah Initiative for Arabic Content, "King Abdullah Initiative for Arabic Content." [Online]. Available: http://www.econtent.org.sa. [Accessed: 07-Oct-2012].

[11]	Shaaban, O.: Automatic Diacritics Restoration for Arabic Text. King Fahd University of Petroleum and Minerals. (2013)

[12]	Tymchenko, V.: BzReader, an application to browse Wikipedia compressed dumps offline. http://code.google.com/p/bzreader/ (2012)

[13]	Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., Nevill-Manning, C. G.: KEA: Practical automatic keyphrase extraction. In Proceedings of the fourth ACM conference on Digital libraries, pp. 254-255. ACM (1999).

AUTHORS

Mohammed Al Logmani works as Systems Analyst at Saudi Aramco. He obtained his M.S. degree in information & computer science from King zzFahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, in 2013 and the B.S. degree from King Abdul-Aziz University, Saudi Arabia in 2002. His research interests include software development and cyber security.

Husni Al-Muhtaseb Assistant Professor, Computer Science Department. He Obtained a PhD degree from the University of Bradford, UK in 2010. He received his M.S. degree in computer science and engineering from King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, in 1988 and the B.E. degree from Yarmouk University, Irbid, Jordan in 1984. His research interests include software development, Arabic Computing, computer Arabization, and Arabic OCR.

FRACTAL PARAMETERS OF TUMOUR MICROSCOPIC IMAGES AS PROGNOSTIC INDICATORS OF CLINICAL OUTCOME IN EARLY BREAST CANCER

Jelena Pribic[1], Jelena Vasiljevic[2], Ksenija Kanjer[1], Zora Neskovic Konstantinovic[1], Nebojsa T. Milosevic[3], Dragica Nikolic Vukosavljevic[1] and Marko Radulovic[1] and Natasa Zivic[4]

[1]Department of Experimental Oncology, Institute of Oncology and Radiology of Serbia,
Pasterova 14, University of Belgrade, Belgrade, Serbia
[2]Institute "Mihajlo Pupin", Volgina 15, University of Belgrade, Belgrade, Serbia
jelena.vasiljevic@pupin.rs
[3]Department of Biophysics, School of Medicine, University of Belgrade
Visegradska 26/2, Belgrade, Serbia
[4]University of Siegen, Germany

ABSTRACT

Research in the field of breast cancer outcome prognosis has been focused on molecular biomarkers, while neglecting the discovery of novel tumour histology structural clues. We thus aimed to improve breast cancer prognosis by fractal analysis of tumour histomorphology. This study included 92 breast cancer patients without systemic treatment. Fractal parameters-fractal dimension and lacunarity of the breast tumour microscopic histology possess prognostic value comparable to the major clinicopathological prognostic parameters. Fractal analysis was performed for the first time on routinely produced archived pan-tissue stained primary breast tumour sections, indicating its potential for clinical use as a simple and cost-effective prognostic indicator of distant metastasis risk to complement the molecular approaches for cancer risk prognosis.

KEYWORDS

fractal dimension, image analysis, lacunarity, breast cancer, prognosis, chemotherapy

1. INTRODUCTION

The importance of breast cancer metastasis risk prognostication derives from the fact that metastasis occurrence is exceptionally variable and that death by breast cancer is mainly caused by metastatic relapse at distant sites. Such importance of metastasis for the disease outcome means that the prognostication of metastasis risk is the central factor for decisions on the

individual therapeutic regime. For this reason, besides local surgery and radiotherapy, patients at elevated risk are increasingly treated with postoperative systemic therapy with an aim to eradicate distant micrometastatic deposits.

Breast cancer metastasis risk biomarkers have been a field of intensive research in the past decade with a focus on molecular prognosticators including transcriptional profiling [1], microRNA analysis [2], detection of circulating tumour cells in blood [3], proliferation [4], and stem cell markers [5], while research for novel histomorphological prognostic clues as a source of prognostic information has been largely neglected. The need for new prognostic approaches derives from the fact that molecular risk biomarkers often outperform the established clinicopathological prognosticators, but regrettably still exhibit insufficient prognostic accuracy, with the remaining unreliable therapeutic guidance. A therapeutic bias towards over-treatment arises as a consequence of such insufficient risk prognosis accuracy based on standard clinicopathological features such as tumour size, nodal status, metastasis (TNM stage), histological grade, steroid receptor status, age, menopausal status and the molecular prognosticators including gene expression arrays [6, 7]. The downside of over-treatment strategy is that patients which do not develop distant metastases needlessly suffer from systemic toxic side effects of chemotherapy [8]. Taken together, there is an urgent need for distant metastasis prognostic tools with improved accuracy, in order to achieve a more effective therapeutic management.

The accurate prognosis of breast cancer outcome is exceptionally challenging based on the fact that malignant transformation is a random process rendering every tumour unique. The primary tumour of patients with early breast cancer is used as the main source of information for assessment of distant metastasis risk. Apart from the above mentioned molecular traces of a tumour, its histology reveals valued prognostic clues due to the fact that the tissue structure reflects the individual patterns of malignant cell growth. A demand for improvement of breast cancer risk prognosis has triggered the idea to implement alternatives to molecular approaches, with a resulting expansion of the digital pathology as a structure examination tool for medical images [4].

2. MATHERIALS AND METHODS

Samples were obtained from surgically removed invasive primary breast tumours. Tissue was formalin-fixed, paraffin-embedded and cut to produce 4 m whole sections which were mounted on slides, deparaffinised and rehydrated. Storage time of archival embedded samples was up to 1 month.

Tissue sections were subsequently stained with haematoxylin and eosin (H&E), the most common type of staining in routine pathological examination of the tumour tissue. Haematoxylin stains the cell cytoplasm and collagen pink, while eosin stains cell nuclei blue.

Digital photographs were captured at x400 magnification under the Olympus BX-51 light microscope with a mounted Olympus digital camera. At least five representative non-overlapping colour images were captured from each section in 3638 × 2736 pixel resolution and saved as TIFF. Pixel size on image was 145 nm. Photographs were taken to comprise tumour areas with large numbers of malignant cells showing typical patterns of malignant growth for each tumour.

These images were also screened by a pathologist and those including mostly non-malignant cells, untypical growth or artefacts, were deleted.

Fractal analysis of histology images was based on the regular non-overlapping box counting method by use of the freely available ImageJ software (http://imagej.nih.gov/ij/) and its FracLac plugin [11]. The reproducibility of the analysis was based on the settings for the range of box sizes (scale window) defined as 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 pixels. With the pixel size of 145 nm the grid lengths ranged between 0.29 m and 148 m. Furthermore, scan background setting was locked to white in order to avoid the inversion of the binary image from white to black background during analysis.

Box counting delivers two main parameters: fractal dimension (FD) and lacunarity as respective measures of complexity and heterogeneity. Three control tumour histology images with known fractal dimensions and lacunarities were analysed in all runs in order to control for reproducibility. The box-counting method involves covering of the digital image with a grid of boxes with size (scale) expressed as the box size relative to image size. The count (N) refers to the space filling properties of the image measured as the minimum number of grid boxes needed to cover all parts of the image containing foreground pixels [32]. Equation (1) for FD represents the negative limit of the ratio of the logs of the number and size of boxes [33]:

$$FD = -\lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(\varepsilon)}$$

(1)

However, such limit cannot be calculated for digital images due to the lack of infinite scaling as the resolution restricts the range of available scales to one pixel as the smallest scale. Furthermore, the largest available scale is also constrained by image width and height. For these reasons, the fractal dimension of natural objects is estimated as a negative slope of the straight part of the regression line (fractal window) at the above (1) double logarithmic plot of box-size and box-count [12].

Lacunarity was also obtained by use of the regular box counting method as a measure of the heterogeneity within an object. Patterns having larger or more numerous gaps generally have higher lacunarity, while low values of lacunarity imply homogeneity based on similarly sized gaps and little rotational variance. Lacunarity is estimated on the basis of the pixel mass distribution probability determined from the number of pixels per box as a function of box size (ε). Lacunarity at a particular ε ($\lambda\varepsilon$) is calculated as:

$$\lambda_\varepsilon = \left(\frac{\sigma}{\mu}\right)^2$$

(2)

where σ is the standard deviation and μ is the mean of the foreground pixels per box at ε. To obtain a single number, the variance calculations are expressed as mean lacunarity (Λav) for the total number of box sizes used:

$$\Lambda_{av} = \frac{\sum \lambda_\varepsilon}{E}$$

(3)

Measured continuous values are converted to categorical values in order to place patients into risk strata. For the present analysis, such categorization was performed by use of "optimal" cutpoints which define the two groups with minimal P-value, calculated by the X-tile 3.6.1 software (Yale University, New Haven, CT). The prognostic accuracy for each variable was calculated based on categorisation performed by the internal split-sample cross-validation procedure. The original sample was randomly split into a training set of 44 patients and a validation set of 48 patients by use of =RAND() function in Excel, for three validation cycles [14]. The discriminating cutpoint with minimal P-value for categorisation was selected in the training partition and applied to the test partition. The training and test partitions were subsequently swapped and the process repeated.

Fractal and lacunarity features were adjusted for standard clinicopathological variables by multivariate Cox analysis in order to identify independent prognostic factors. Variables were added to a model using a training partition of 44 patients with a forward selection entry criterion of P<0.05 in univariate analysis. Features were removed by use of backward elimination according to a selection stay criterion of P<0.05. The procedure was based on internal split-sample cross-validation as described above in detail for feature categorization.

Spearman's rank correlation test was employed to calculate the correlation between variables. Analyses were completed by SPSS version 20 (IBM, Chicago, IL) and Stata/MP 13 software. Adjuvant! Online score for Breast Cancer (Version 8.0) was calculated at https://www.adjuvantonline.com/breastnew.jsp as a 10 year risk of relapse with no additional therapy based on age, tumour grade, ER status and tumour size.

3. RESULTS

Fractal analysis presents a morphometric technique useful for measurement of irregular and complex shapes [12]. Fractal dimension and lacunarity were calculated for individual histology sections as parameters which may quantify the risk diversity of primary breast tumours (FIGURE 1). Estimation of fractal dimension is a multiscale measure of space filling, based on the fact that complex objects fill more space than simpler ones. A 2-dimensional image can theoretically have a maximal FD of 2 which means that the obtained value range for FD (1.66 -1.86) indicates a geometrically rather complex tissue structure that is highly space-filling. The value ranges for lacunarity were wider at = 0.21 - 0.68. Whereas mathematical fractals are invariant in shape over an unlimited range of scales, biological fractal patterns are self-similar only statistically in terms of complexity within a fractal window, with upper and lower scaling limits spreading at least two orders of magnitude, a condition that must be experimentally established for a particular type of a sample. Therefore, we examined the statistical self-similarity based on the linearity of the double logarithmic box-number/box-size curve for the range of analysed scales (2 - 1024 pixels), as previously explained in detail [15]. The linearity was thereby established for the whole range of scales of nearly three orders of magnitude, meaning that scaling and fractal windows overlapped from 0.29 m to 148 m, based on the pixel size of 145 nm.

Several significant correlations were evident between fractal and clinicopathological variables, yet with low Spearmann coefficients. FD correlated significantly only with the pathological tumour size (Spearmann coefficient of 0.27), while lacunarity correlated with Adjuvant! (-0.22) and with the pathological tumour size (-0.25). Such low correlations suggested a possible prognostic independence of FD and lacunarities, to the advantage of their prognostic values. In agreement with this, fractal dimension was indicated as independent prognostic factor by the multivariate Cox regression.

4. DISCUSSION

We approached the task of breast cancer risk prognosis improvement by exploiting the prognostic value of tumour microscopic histology at the time of surgery. Based on the established ability of fractal geometry to quantify irregular structures of tumour microscopic histology [37, 38], we hypothesized that this approach may efficiently acquire prognostically relevant structural information and thus provide a valuable addition to existing clinicopathological and molecular prognosticators in breast cancer.

This study was performed on patients with a natural course of disease due to interference of systemic therapy with metastasis occurrence. Such patients are increasingly difficult to find due to a trend of widening systemic therapy use [16]. The patient group was assembled from a period of over 20 years ago when low metastasis risk patients have not been prescribed systemic therapy at our institution. Lower risk classification was based on a tumour grade < 3, pT < 3 and absence of lymph node metastases.

Morphological complexity is the property revealed by the retention of shapes of high structural complexity even after magnification (scaling, zooming). FD quantifies such complexity by measuring the rate of addition of structural detail with increasing magnification within the linear part of the plot designated as fractal window [17]. Our data indicate that high FD associates with high distant metastasis risk. With FD regarded as a formal means of quantifying shape complexity [11, 16], it can be considered that high structural complexity of H&E stained tumour tissue associates with high risk. This result is consistent with the previously reported prognostic value of FD obtained on pan-cytokeratin stained breast tumour tissue sections [14]. By marking epithelial cells, this kind of staining directs the fractal analysis primarily to the outline of malignant tissue growth patterns, while our analysis was performed on unspecifically stained H&E tissue sections. Furthermore, high tumour grade as an indicator of increased risk also reflected high histomorphological disorder and complexity [18].

On the other hand, lacunarity associated negatively with metastasis risk. Mandelbrot originally proposed lacunarity as an adjunct to fractal analysis because it proved useful in discerning amongst images that have similar fractal dimensions [19]. Lacunarity assesses the texture qualities of the structure based on distribution and size of the empty domains, while fractal dimension specifies how completely a fractal-like structure fills the space for decreasing scales [20]. Beyond being a measure of gappiness, lacunarity quantifies additional features such as "rotational invariance" and more generally, heterogeneity. It follows that high risk is prognosticated by low heterogeneity and high complexity of microscopic histology images, as respectively indicated by lacunarity and FD.

Results obtained in this study are the first to indicate the association between FD and lacunarity of the whole H&E tumour tissue microscopic histology images and disease outcome. Cost efficiency is an advantage of H&E tissue section analysis, as such histology samples are routinely produced following tumour extraction surgery for use in standard pathological diagnostic and prognostic assessments, without any need for imunohistochemical staining and associated expenditure for chemicals.

We investigated the prognostic potential of fractal properties in a hypothesis-free fashion as previous studies mainly analysed tumour tissue specifically stained with known prognostic markers [42]. By targeting specific features, primarily nuclei [21], even the studies using non-specifically H&E-stained tissue sections [22-26] were in fact not analytically hypothesis-free. Although pathologists commonly use the pan-tissue H&E staining for pathologic assessment of histology specimens, it has been considered as suboptimal for fractal analysis27. This opinion was based on the theoretical consideration that only specific highlighting of malignant epithelial cells may provide structures of interest for optimal image analysis, as non-malignant components stained by H&E such as stromal cells, extracellular collagen and elastic fibres could diminish the extraction of useful prognostic clues. However, our current results indicate that image analysis aimed at cancer risk prognosis can in fact be simplified by omitting the specific staining step, without sacrifice of prognostic power. On the basis of this result it is concluded that an overall tissue structure, without any specific markings of malignant cells, provides optimal prognostic clues which can be extracted by fractal image analysis. This is a tremendous cost-advantage relative to the established immunohistochemically-stained prognostic markers and especially the expensive gene expression profiling prognostic tools such as Oncotype DX and MammaPrint [28]. Remarkably, even these expensive tests actually present cost-effectiveness through improved low-risk stratification with the consequent reduction in chemotherapy utilization [1].

Prognostic value of fractal analysis is acquired in a different way than a visual grading of histological clues by a pathologist. The underlying tumor biology that fractal analysis may reflect include yet unidentified microscopic structural patterns of a tumour, such as tissue growth, cellularity apoptosis, hypoxia, angiogenesis and distribution of mitotic cells [29-31]. Due to a stochastic nature of carcinogenesis, each tumour is a different entity with unique growth patterns that may be based in large part on fractal geometry [32], thus rendering fractal analysis a particularly relevant readout of tumour heterogeneity.

The important limitation of the regular box counting used in this study is that its methodological validity is restricted only to statistically self-similar signals. Furthermore, the requirement for signal binarization leads to a loss of information [33] adaptation of the method called differential box-counting was developed to overcome this limitation by allowing analysis of grey-scale images. A limitation in scaling is another general issue for image analysis, with a range between one pixel as the smallest scale and the size of the whole image as the largest scale. In addition to limitations in image resolution and size, natural objects are intrinsically limited in scaling due to the finite size of their structural units [12]. A limit to observable increase in detail for the biological samples becomes evident at the critical magnification of ×130.000 Furthermore, the regular box counting method is limited by its calculation of generalized parameters for the whole image without providing any insight into the regional variations. Multifractal analysis as an extended version of monofractal analysis provides regional measurements of several multifractal parameters across an image [13]. Fractal sub scans and local connected fractal dimension (LCFD) also provide regional information by calculating and visualising the distribution of local fractal dimensions within the image. All of these methods are a variation of a common calculation of the

pixel numbers for various lengths (scales) of the image. In case where fractal properties are not constant in all parts of the image, such regional analyses may perform better than fractal parameter values averaged across the whole image. The lack of full objectivity is a general important limitation of this type of image analysis based on the fact that although fractal approach enables to objectively describe irregular morphologic components and ultrastructural features, the selection of the representative tissue segments for analysis is still done subjectively to the best knowledge of a pathologist.

Based on discrimination efficiencies indicated by Kaplan Meier plots and effect sizes by hazard ratios, the prognostic performance of FD and lacunarity was comparable to established clinicopathological prognostic parameters such as pathological tumour size, estrogen receptor (ER) status and clinicopathological Adjuvant! composite score. AUCs as discrimination measures were underperformers for fractal parameters, while validated accuracies obtained in this study (64-74%) surpassed the specific prognostic factors previously investigated by us such as Cathepsin D expression [28], trefoil factor 1 and comparable to the established and costly FDA-cleared 70-genes expression Mammaprint molecular prognostic tool with the reported accuracy of 65%

Internal validation was successfully performed by bootstrap and split-sample cross-validation, suggesting that the prognostic parameters tested are generalizable. Whereas further validation is needed on external and expanded group of patients, this preliminary study indicates for the first time the potential use of fractal analysis of primary breast tumour histology as the cost-effective prognostic indicator of distant metastasis risk.

5. CONCLUSIONS

By use of fractal analysis we measured the structural characteristics of 92 breast tumour histological specimens in patients who did not receive systemic therapy. On the basis of obtained results it is concluded that histology fractal dimension and lacunarity correlate with biological properties of a tumour and present a promising strategy to assess the risk of distant metastasis. The original contribution of this study to the field of prognosis of breast cancer outcome is based on providing the first evidence that fractal characteristics of native tumour histology highlighted by the H&E pan-tissue stain, without any focus on specific structures or molecular biomarkers, can prognosticate an individual's risk of metastasis within a long median follow-up period of 150 months. The potential clinical use of FD and lacunarity is based on their prognostic performance which was comparable to standard clinicopathological prognosticators. Furthermore, the fractal prognostic biomarkers are simple and inexpensive to measure from routine archived H&E-stained histology, thus permitting both prospective and retrospective validation studies as well as possible clinical application without additional costs.

Papers of special note have been highlighted as:
*of interest
** of special interest

REFERENCES

[1]   Rouzier R, Pronzato P, Chereau E, Carlson J, Hunt B, Valentine WJ: Multigene assays and molecular markers in breast cancer: systematic review of health economic analyses. Breast cancer research and treatment 139(3), 621-637 (2013).

[2]   Mar-Aguilar F, Mendoza-Ramirez JA, Malagon-Santiago I et al.: Serum circulating microRNA profiling for identification of potential breast cancer biomarkers. Dis. Markers 34(3), 163-169 (2013).

[3]   Neumeister V, Agarwal S, Bordeaux J, Camp RL, Rimm DL: In situ identification of putative cancer stem cells by multiplexing ALDH1, CD44, and cytokeratin identifies breast cancer patients with poor prognosis. Am. J. Pathol. 176(5), 2131-2138 (2010).

[4]   Laurinavicius A, Plancoulaine B, Laurinaviciene A et al.: A methodology to ensure and improve accuracy of Ki67 labelling index estimation by automated digital image analysis in breast cancer tissue. Breast cancer research : BCR 16(2), R35 (2014).

[5]   Giordano A, Gao H, Anfossi S et al.: Epithelial-mesenchymal transition and stem cell markers in patients with HER2-positive metastatic breast cancer. Molecular cancer therapeutics 11(11), 2526-2534 (2012).

[6]   Edge SB, American Joint Committee on Cancer., American Cancer Society.: AJCC cancer staging handbook : from the AJCC cancer staging manual. (7th). Springer, New York. (2010).

[7]   Guiu S, Michiels S, Andre F et al.: Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement. Ann. Oncol. 23(12), 2997-3006 (2012).

[8]   Lonning PE, Knappskog S, Staalesen V, Chrisanthar R, Lillehaug JR: Breast cancer prognostication and prediction in the postgenomic era. Ann. Oncol. 18(8), 1293-1306 (2007).

[11] Landini G: Fractals in microscopy. J. Microsc. 241(1), 1-8 (2011).

[12] Weibel ER: Fractal geometry: a design principle for living organisms. Am. J. Physiol. 261(6 Pt 1), L361-369 (1991).

[13] Braverman B, Tambasco M: Scale-specific multifractal medical image analysis. Computational and mathematical methods in medicine 2013, 262931 (2013).

[14] Tambasco M, Eliasziw M, Magliocco AM: Morphologic complexity of epithelial architecture for predicting invasive breast cancer survival. Journal of translational medicine 8, 140 (2010).

[15] Tambasco M, Magliocco AM: Relationship between tumor grade and computed architectural complexity in breast cancer specimens. Hum. Pathol. 39(5), 740-746 (2008).

[16] Markicevic M, Kanjer K, Mandusic V, Buta M, Neskovic-Konstantinovic Z, Nikolic-Vukosavljevic D: Cathepsin D as an indicator of clinical outcome in early breast carcinoma during the first 3 years of follow-up. Biomarkers in medicine 7(5), 747-758 (2013).

[17] Markicevic M, Dzodic R, Buta M et al.: Trefoil factor 1 in early breast carcinoma: a potential indicator of clinical outcome during the first 3 years of follow-up. International journal of medical sciences 11(7), 663-673 (2014).

[18] Vujasinovic T, Pribic J, Kanjer K et al.: Gray-Level Co-Occurrence Matrix Texture Analysis of Breast Tumor Images in Prognosis of Distant Metastasis Risk. Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada, 1-9 (2015).

[19] Altman DG, Mcshane LM, Sauerbrei W, Taube SE: Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS medicine 9(5), e1001216 (2012).

[20] Smith TG, Jr., Lange GD, Marks WB: Fractal methods and results in cellular morphology-- dimensions, lacunarity and multifractals. J. Neurosci. Methods 69(2), 123-136 (1996).

[21] Karperien A, Ahammer H, Jelinek HF: Quantitating the subtleties of microglial morphology with fractal analysis. Frontiers in cellular neuroscience 7, 3 (2013).

[22] Di Ieva A, Grizzi F, Ceva-Grimaldi G et al.: Fractal dimension as a quantitator of the microvasculature of normal and adenomatous pituitary tissue. J. Anat. 211(5), 673-680 (2007).

[23] Mirza AN, Mirza NQ, Vlastos G, Singletary SE: Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. Ann. Surg. 235(1), 10-26 (2002).

[24] Cutting JE, Garvin JJ: Fractal curves and complexity. Percept. Psychophys. 42(4), 365-370 (1987).

[25] Rajkovic K, Bacic G, Ristanovic D, Milosevic NT: Mathematical model of neuronal morphology: prenatal development of the human dentate nucleus. BioMed research international 2014, 812351 (2014).

[26] Oger M, Allaoui M, Elie N et al.: Impact of tumor heterogeneity on disease-free survival in a series of 368 patients treated for a breast cancer. Diagostic Pathology 8, S43 (2013).

[27] Metze K: Fractal dimension of chromatin: potential molecular diagnostic applications for cancer prognosis. Expert review of molecular diagnostics 13(7), 719-735 (2013).

[28] Stec J, Wang J, Coombes K et al.: Comparison of the predictive accuracy of DNA array-based multigene classifiers across cDNA arrays and Affymetrix GeneChips. The Journal of molecular diagnostics : JMD 7(3), 357-367 (2005).

[29] Savage VM, Herman AB, West GB, Leu K: Using Fractal Geometry and Universal Growth Curves as Diagnostics for Comparing Tumor Vasculature and Metabolic Rate With Healthy Tissue and for Predicting Responses to Drug Therapies. Discrete and continuous dynamical systems. Series B 18(4), (2013).

[30] Normant F, Tricot C: Method for evaluating the fractal dimension of curves using convex hulls. Physical review. A 43(12), 6518-6525 (1991).

[31] Ieee Computer Society., Institute of Electrical and Electronics Engineers.: IEEE transactions on pattern analysis and machine intelligence. v.

[32] Paumgartner D, Losa G, Weibel ER: Resolution effect on the stereological estimation of surface and volume and its interpretation in terms of fractal dimensions. J. Microsc. 121(Pt 1), 51-63 (1981).

[33] Buyse M, Loi S, Van't Veer L et al.: Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. Journal of the National Cancer Institute 98(17), 1183-1192 (2006).

AUTHORS

Jelena Vasiljevic, PhD of Electrical Engineering, University of Belgrade, Institut Mihajlo Pupin, Docent for Intelligent Systems on Faculty for Computer Science Jelena Vasiljevic works as Science Assistant in Institute Mihajlo Pupin in Belgrade, Serbia and also as Docent for Intelligent Systems teaching on Faculty for Computer Science. She is teaching Intelligent Systems (Artificial Intelligence), Genetic Algorithms, Neural networks, Fuzzy Logic, Bioinformatics, Knowledge based Systems on PhD, Master and Primary studies. She is a graduate of Faculty for Electrical Engineering on University in Belgrade, Serbia. PhD research was application of multifractal analysis of digital medical images in cancer diagnostic and for it she has been several times awarded: The First Award on International Medical Fair 2014, The Special Award on International Technical Fair 2014 and many others (also many interviews for TV and newspaper, in Oxford Encyclopedia). She is also an author and co-author of many journal and conference peer reviewed.

AUTHOR INDEX