

In the Name of God

**Journal of**  
**Information Systems & Telecommunication**  
Vol. 3, No. 2, April-June 2015, Serial Number 10

**Research Institute for Information and Communication Technology**  
**Iranian Association of Information and Communication Technology**

**Affiliated to: Academic Center for Education, Culture and Research (ACECR)**

**Manager-in-charge:** Habibollah Asghari, Assistant Professor, ACECR, Iran

**Editor-in-chief:** Masoud Shafiee, Professor, Amir Kabir University of Technology, Iran

**Editorial Board**

Dr. Abdolali Abdipour, Professor, Amirkabir University of Technology

Dr. Mahmoud Naghibzadeh, Professor, Ferdowsi University

Dr. Zabih Ghasemlooy, Professor, Northumbria University

Dr. Mahmoud Moghavvemi, Professor, University of Malaysia (UM)

Dr. Ali Akbar Jalali, Professor, Iran University of Science and Technology

Dr. Hamid Reza Sadegh Mohammadi, Associate Professor, ACECR

Dr. Ahmad Khademzadeh, Associate Professor, CyberSpace Research Institute (CSRI)

Dr. Abbas Ali Lotfi, Associate Professor, ACECR

Dr. Sha'ban Elahi, Associate Professor, Tarbiat Modares University

Dr. Ramezan Ali Sadeghzadeh, Associate Professor, Khajeh Nasireddin Toosi University of Technology

Dr. Saeed Ghazi Maghrebi, Assistant Professor, ACECR

**Administrative Manager:** Shirin Gilaki

**Executive Assistant:** Behnoosh Karimi

**Art Designer:** Amir Azadi

**Print ISSN:** 2322-1437

**Online ISSN:** 2345-2773

**Publication License:** 91/13216

**Editorial Office Address:** No.5, Saeedi Alley, Kalej Intersection., Enghelab Ave., Tehran, Iran,

P.O.Box: 13145-799

Tel: (+9821) 88930150 Fax: (+9821) 88930157

Email: info@jst.ir

URL: www.jst.ir

**Indexed in:**

- |   |                  |
|---|------------------|
| - Journal of Information Systems and Telecommunication            | www.jst.ir       |
| - Islamic World Science Citation Center (ISC)                     | www.isc.gov.ir   |
| - Scientific Information Database (SID)                           | www.sid.ir       |
| - Regional Information Center for Science and Technology (RiCeST) | www.ricest.ac.ir |
| - Magiran   | www.magiran.com  |

**Publisher:**

Regional Information Center for Science and Technology (RiCeST)

Islamic World Science Citation Center (ISC)

This Journal is published under scientific support of  
Advanced Information Systems (AIS) Research Group and  
Digital & Signal Processing Research Group, ICTRC

## Acknowledgement

JIST Editorial-Board would like to gratefully appreciate the following distinguished referees for spending their valuable time and expertise in reviewing the manuscripts and their constructive suggestions, which had a great impact on the enhancement of this issue of the JIST Journal.

### (A-Z)

- Akbari Mohammad Kazem, Amirkabir University of Technology, Tehran, Iran
- Anvaripour Mohammad, University of Windsor, Windsor, Canada
- Azimzadeh Fatemeh, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Bababie Shahram, Islamic Azad University, Tabriz Branch, Tabriz, Iran
- Badie Kambiz, ITRC (Iran Telecommunication Research Center), Tehran, Iran
- Barkhordari Mohammad Hosein, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Bayat Reza, Tarbiat Modares University, Tehran, Iran
- Chamanmotlagh Abolfazl, Imam Hossein University, Tehran, Iran
- Ebrahimnezhad Hossein, Sahand University of Technology, Tabriz, Iran
- Etemad Moghadam Kamran, Industrial Management Institute, Tehran, Iran
- Farsi Hasan, University of Birjand, Birjand, Iran
- Fathian Mohammad, Iran University of Science and Technology, Tehran, Iran
- Fathi Kiadehi Elias, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Fouladi Kazem, University of Tehran, Tehran, Iran
- Ghanbari Mohammad, University of Essex, Colchester, UK
- Ghattan Kashani Zahra, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Hamidi Hojjatollah, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Heydarian Mohsen, Azarbijan Shahid Madani University, Tabriz, Iran
- Jahangiri Ehsan, Johns Hopkins University, Baltimore, USA
- Kasaei Shohreh, Sharif University of Technology, Tehran, Iran
- Lotfi Abbasali, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Marvasti Farrokh, Sharif University of Technology, Tehran, Iran
- Mahdieh Omid, University of Zanjan, Zanjan, Iran
- Mehrjoo Masoud, Sharif University of Technology, Tehran, Iran
- Mirroshandel Seyed Abolghasem, University of Guilan, Rasht, Iran
- Mirzavand Rashid, Amirkabir University of Technology, Tehran, Iran
- Mohebbi Keyvan, Islamic Azad University, Mobarakeh Branch, Tehran, Iran
- Motamed Sara, Islamic Azad University, Science and Research Branch,
- Nilforooshan Zahra, Amirkabir University of Technology, Tehran, Iran
- Ramezani Amin, Tarbiat Modares university, Tehran, Iran
- Rasoolzadegan Abbas, Ferdowsi University of Mashhad, Mashhad, Iran
- Rezaei Abdhossein, Academic Center for Education Culture and Research (ACECR), Isfahan, Iran
- Rezaei Mehdi, University of Sistan and Baluchestan, Zahedan, Iran

- Sadeghi Bigham Bahram, Institute for Advanced Studies in Basic Sciences (IASBS) ,Zanjan, Iran
- Sadeghzadeh Ramezan Ali, Khaje Nasir-edin Toosi, Tehran, Iran
- Sajedi Hedieh, University of Tehran, Tehran, Iran
- Seifori Mahmood, Shahid Rajae Teacher Training University, Tehran, Iran

## Table of Contents

- Early Detection of Pediatric Heart Disease by Automated Spectral Analysis of Phonocardiogram in Children ..... 66  
Hassan Ghassemian and Azra Rasouli Kenari
- Active Steganalysis of Transform Domain Steganography Based on Sparse Component Analysis... 75  
Hamed Modaghegh and Seyed Alireza Seyedin
- A Robust Statistical Color Edge Detection for Noisy Images ..... 85  
Mina Alibeigi, Niloofar Mozafari, Zohreh Azimifar and Mahnaz Mahmoodian
- Ant Colony Scheduling for Network On Chip ..... 95  
Neda Dousttalab, Mohammad Ali Jabraeil Jamali and Ali Ghaffari
- A Fast and Accurate Sound Source Localization Method Using the Optimal Combination of SRP and TDOA Methodologies ..... 100  
Mohammad Ranjkesh and Reza Hasanzadeh
- Better Performance of New Generation of Digital Video Broadcasting-terrestrial (DVB-T2) Using Alamouti scheme with Cyclic Delay Diversity ..... 108  
Behnam Akbarian and Saeed Ghazi Maghrebi
- Online Signature Verification: a Robust Approach for Persian Signatures ..... 115  
Mohammad Esmaeel Yahyatabar, Yasser Baleghi and Mohammad Reza Karami
- Computing Semantic Similarity of Documents Based on Semantic Tensors ..... 125  
Navid Bahrami, Amir Hossein Jadidinejad and Mozhdeh Nazari

# Early Detection of Pediatric Heart Disease by Automated Spectral Analysis of Phonocardiogram in Children

Hassan Ghassemian\*

Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran  
ghassemi@modares.ac.ir

Azra Rasouli Kenari

Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran  
azra.rasouli@gmail.com

Received: 07/ July/2014

Revised: 25/Apr/2015

Accepted: 27/Apr/2015

## Abstract

Congenital heart disease is now the most common severe congenital abnormality found in live births and the cause of more than half the deaths from congenital anomalies in childhood. Heart murmurs are often the first signs of pathological changes of the heart valves, and they are usually found during auscultation in the primary health care. Auscultation is widely applied in clinical activity; nonetheless sound interpretation is dependent on clinician training and experience. Distinguishing a pathological murmur from a physiological murmur is difficult and prone to error. To address this problem we have devised a simplified approach to pediatric cardiac scanning. This will not detect all forms of congenital heart disease but will help in the diagnosis of many defects. Cardiac auscultatory examinations of 93 children were recorded, digitized, and stored along with corresponding echocardiographic diagnoses, and automated spectral analysis using discrete wavelet transforms was performed. Patients without heart disease and either no murmur or an innocent murmur ( $n = 40$ ) were compared to patients with a variety of cardiac diagnoses and a pathologic systolic murmur present ( $n = 53$ ). A specificity of 100% and a sensitivity of 90.57% were achieved using signal processing techniques and a k-nn as classifier.

**Keywords:** Phonocardiogram (PCG); Murmur; Cardiac; K-nn Classifier; Pediatric; Wavelet.

## 1. Introduction

Acoustical vibrations produced by the mechanical action of the heart contain valuable information about the pathological condition of the cardiovascular system. Cardiac murmurs are often the first sign of pathological changes in the heart valves and they are caused by turbulent blood flow or jet flow impinging on and causing vibration of surrounding tissue. Murmurs are critical and must be detected as soon as possible. A heart murmur in pediatrics can be an indicator of congenital heart disease (CHD). Innocent heart murmurs are of no clinical consequence while pathologic heart murmurs indicate congenital heart disease is present [1].

Heart murmurs are an important feature to identify cardiac disorders in childhood, infancy, and especially in newborns. Unrecognized heart disease in newborns carries a serious risk of avoidable mortality, morbidity and handicap.

CHD is the most common congenital disorder in newborns. Unfortunately Iran is one of the countries that is not supporting the screening program for CHD at the moment.

Computer-aided auscultation would allow fast and cheap decisions by using a tool widely known by both physicians and patients: the stethoscope. Since most heart diseases are reflected to the sound that the heart produces,

stethoscopes are part of the first line of screening and diagnosis of heart pathologies.

In simple healthcare establishments the heart auscultation is the basic tool for a first screening of patients and deciding which of them should be referred to more complex and costly medical examinations and tests (e.g. based on advanced imaging techniques) and/or specialised cardiologists. Also, many heart diseases cause differentiations of heart sound in much earlier stages before they can be observed in other comparable techniques, such as the Electrocardiogram (ECG). Therefore increasing the accuracy and the whole effectiveness of heart auscultation is of critical importance for improving both the health level of the populations (by diagnosing heart diseases in their early stages) and also the economics of the health systems (by avoiding unnecessary costly medical examinations and tests due to incorrect screening). Furthermore, it should be taken into account that in some circumstances, such as in the developing countries, the auscultation is the only available tool for diagnosis of heart diseases for most of their population.

Cardiac auscultation is one of the most important physical examination and a part of the first medical diagnostic procedures. Many of the heart diseases can be recognized in the primary stage using heart sounds auscultation and therefore are treated earlier. Irregularities detected during auscultation begin the therapy [2].

\* Corresponding Author

Cardiac murmurs occur frequently in healthy children, but it can also be a feature associated to many forms of congenital heart disease. The incidence of heart murmurs in the pediatric population is reportedly as high as 77% to 95%. However, less than 1% of this population has heart disease. Given the prevalence of innocent murmurs and the relatively low incidence of actual heart disease, the primary health care provider may have difficulty determining which patients with murmurs need specialist referral, especially when there are no other signs or symptoms of heart disease. Seven types of innocent heart murmurs are reported in children, i.e. still's murmur, innocent pulmonary flow murmur, innocent pulmonary branch murmur of infancy, supraclavicular bruit, venous hum, mammary souffle, and cardiorespiratory murmur. Generally, clinical history and physical examination are diagnostic for these murmurs.

In studies of patients referred by primary care physicians because of heart murmurs either directly for echocardiography or for evaluation by the cardiologist, only 20% to 30% of the patients have pathology.

Early recognition is an important goal, and equally important is avoiding misdiagnosing a pathological heart murmur in a healthy child without heart disease. To acquire high-quality auscultation skills, requires the guidance of an experienced instructor using a sizable number of patients along with frequent practice. Unfortunately, the interpretation of auscultation findings overall remains prone to error. Imaging technologies can provide more direct evidence of heart disease; however, they are generally more costly.

There is an acute shortage of physicians in developing countries and many rural clinics are run by nurses. Given the high incidence of heart murmurs, automated screening based on electronic auscultation at clinic level would be of great benefit. The main advantages for early recognizing a cardiac disease are that newborns will be seen and assessed earlier and in better clinical conditions. Acceptance will obviously depend on the sensitivity and specificity of the system. Selection of representative data for diagnosis must be relatively simple for the system to be of practical use in rural clinics [3].

Appropriate devices allow nowadays the digitization and storage of heart sounds in digital format, their inclusion in electronic health records, their transmission to other (possibly remote) systems (e.g. using wireless technologies, the Internet, etc.), their presentation on a screen (both in the time and in the frequency domain) and their processing in order to remove noise and other undesirable components. More advanced systems can also perform intelligent processing and provide suggestions of diagnostic nature to the doctor, e.g. concerning the existence of additional sound components, such as the third heart sound (S3), the fourth heart sound (S4), various murmurs, clicks, snaps, etc., or even the existence of particular heart diseases. This combination of the 'traditional' auscultation with the modern information and communication technologies is expected to revitalise the interest in and use of auscultation in the near future [4].

The digital analysis of heart sounds has revealed itself as an evolving field of study. In recent years, numerous approaches to create decision support systems were attempted. Recent advances in digital signal processing have led to a reexamination of the potential role of spectral analysis of heart sounds in cardiac diagnosis [5,6,7,8,9]. In this study, we investigated a new technique for evaluating heart murmurs in children and young adults using automated analysis of the systolic energy content found in digital recordings of cardiac auscultations.

Heart sound features such as spatial loudness, relative amplitude, murmurs, and localization of each component may be indicative of pathology. In this work we build on the "prototypical systole" and then extract two different feature sets that characterize acoustic activity in the systolic phase. One feature set is related to physiological activity of heart sounds. The other is the first three principal components derived from principal component analysis. The objective of our study was to assist the clinician in the detection and evaluation of heart murmurs.

## 2. Methods

In this sections we will discuss the signal processing stage, signal processing is the most important element of the system, because it provides information useful for diagnosticians. Its purpose is to de-noise and make segmentation of the signal, features extraction and classification between the normal and abnormal heart sound.

### 2.1 Data Acquisition

The monitoring of sounds heard over the chest walls is known as auscultation, which is usually performed with the stethoscope. The heart sounds signal could be measured in two ways, with analog or digital stethoscope or by phonocardiography. Stethoscope is simple tool to transmit the heart sounds from the chest wall to the ear, but it is not easy to use it correctly. Firstly the human ear is not sufficiently sensitive to determine all frequencies (different level of auscultation skill) and some components of signal can be omitted. Secondly, to make of the diagnosis, a long time-practice and experience of the doctors is required. Also the signal from an acoustic stethoscope has very less sound amplification [2]. Therefore, it was a need to use a device, which could make an automatic recording and interpretation of a signal. Phonocardiography (PCG) is a main tool showing the timings and relative intensities of the heart sounds with graphic recordings on the paper. The digital stethoscopes with suitable software allowing automatic analysis of the signal could be more widely distributed diagnostic tool which has been used in this study.

Fig. 1 shows a simple diagram of a system presented in [2], which consists of an analog part to record signal and with digital to analysis and display waveform. Chest piece is a component of electronic stethoscope. Then signal is recorded by a microphone and by an IC recorder.

It is a possibility to hear it by an earphone or it might be transmitted to the computer. Signal processing is implemented in MATLAB®.

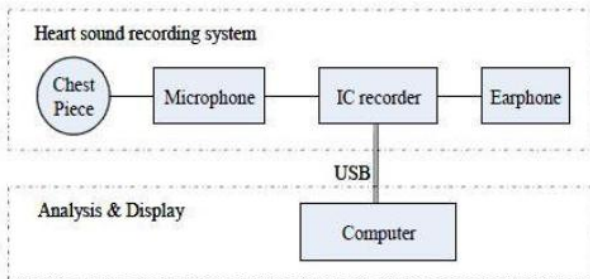


Fig. 1. Schematic diagram of recording and analyzing system [2].

The WelchAllyn Meditron Analyzer, ECG and electronic stethoscope were used during the auscultation. The digital recordings were made at a sampling frequency of 44.1 KHZ, 16 bit precision and saved via USB onto a laptop in the uncompressed WAV file format.

Cardiac auscultatory examinations of 93 children and young adults were recorded, digitized, and stored along with corresponding echocardiographic diagnoses. 40 subjects were diagnosed as normal and the other 53 subjects were found to have one of the pathological cardiac conditions namely VSD, AS and PS. For each patient a 10 second, ECG and heart sound (HS) recording was made at the Apex in the supine position. The sounds recorded on the surface of the body depend on the source location, intensity, and the acoustical properties of the surrounding tissues. It is important that the location of the microphone be specified. The recordings were made by experienced pediatric cardiologists, and in most cases, in a noisy clinical environment. Echocardiography was done on all patients to confirm patient diagnosis.

## 2.2 Pre-Processing

A signal is a means to convey information. It is sometimes generated directly by the original information source. We may then want to learn about the structure or functioning of the source from the extracted information. The signal available may not yield directly the required information. We then apply some operations on the signal in order to enhance the information needed.

Pre-processing of heart sounds is necessary to obtain consistent useful data for analysis and improve robustness in the presence of noise such as breathing noise, artifacts, voice and external noise.

In the event that the recording environment is not well controlled, noise is coupled to the PCG. To avoid unpredictable effects brought by noise, filtering out the unwanted noise becomes important for later processing. That is the reason why the first step of signal processing consists of filtering heart sounds, since the main spectrum of first heart sound (S1) and second heart sound (S2) stay within the range of 200Hz, We chose the second order filter with a cut-off frequency equals 600Hz in order to save a heart sounds and murmurs bandwidth. In sum, preprocessing was done due to two reasons, first was to

eliminate DC values and motion artifacts and the second was to omit redundant high frequency contents.

Heart sounds were down sampled to 2000 samples per second for memory optimization and processing speed. As mentioned above, the significant information of HS signals was contained below 600Hz and no useful information was lost during down sampling. All HSs were filtered using second order, low pass Butterworth filter with cut-off at 600Hz with zero phase delay.

## 2.3 Heart Sound Segmentation

The first step in the processing is usually that of segmentation. The signal may drastically change its properties during time. We then observe and process the signal only in a finite time window. The length of the time window depends on the signal source and goal of processing. We may use a single window with predetermined length, or we may require some scheme for automatically dividing the signal into varying length segments as we did for this paper.

Since sound pressure level greatly varies subject by subject, the first step in this section was to normalize all HS signals with its norm, for each vector of heart sound,  $x[n]$ , as shown in Equation (1). This could enable us to compare heart sounds of different patients.

$$x_{norm}[n] = \frac{x[n]}{norm(|x(n)|)} \quad (1)$$

Initially the heart cycle must be segmented into the S1, systolic, S2 and diastolic phases. This is not a trivial task to do in general. In normal patients there is a clear distinction between the lub and dub of the heart sound. In the presence of a systolic murmur, such as generated by aortic stenosis, the boundaries between the end of S1 and the beginning of the murmur, and the end of the murmur and the beginning of S2 can become very obscure.

Frequency differences can assist this segmentation task with S1 and S2 being in the 40 to 80Hz range and murmur being higher, 80 to 150Hz. but is not always the case. Although the fundamental resonances of S1 and S2 are at relatively low frequencies in the audible range, the lub-dub effect can be still heard in the 100 to 200Hz range. An algorithm has been devised that looks for peaks of signal around the probable region where S1 and S2 occur.

Automatic segmentation of the heart signals into individual cycles provide information such as the S1, S2, systolic and diastolic durations. Special algorithms implemented in MATLAB® enable automatic detection of these events. In this research ECG as a reference is used. Another method is to apply a Wavelet multi-resolution analysis [2].

Typically, heart sounds consist of two regularly repeated thuds, known as S1 which is generated by the nearly simultaneous closure of the mitral and the tricuspid valve, being followed by the systolic phase, and the second heart sound S2, which is generated by the nearly simultaneous closure of the aortic and the pulmonic valve, being followed by the diastolic phase. Most heart diseases

generate additional components in the heart sound, such as murmurs in the systolic or/and the diastolic phase. The time interval between S1 and S2 is the systole, while the gap between S2 and the next S1 corresponds to the diastole. Therefore a single cardiac cycle or heart beat contains components S1-systole-S2-diastole, as shown in Fig. 2.

A cycle starts just before the R-peak of the ECG and ends before the next R peak. This RR interval was used to segment the heart signal into individual beats. Segmentation of the heart beat was then done to find the first and second heart sound.

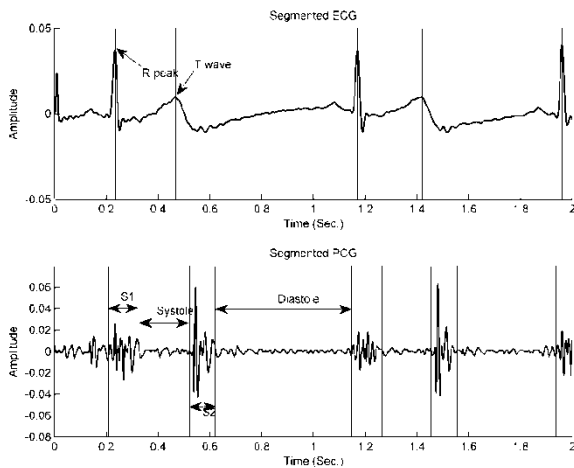


Fig. 2. A normal PCG that we have segmented it by using its corresponding ECG signal.

It should be noted that systolic length may differ in some cases. So in order to distinguish the starting time of S2, we would need to locate the time where T wave occurs in the corresponding ECG signal. S2 start time is the maximum position between the start point of T wave and 100 to 150 ms after that in the HS signal, based on the RR interval duration.

## 2.4 Feature Extraction and Selection

Not all the information conveyed by the signal is of interest. The signal may contain redundancies. When effective storing and transmission are required or when the signal is to be automatically classified, these redundancies have to be eliminated. The signal can be represented by a set of features that contain the required information. The most effective feature extraction methods are the time-frequency analysis. Research about time-frequency analysis describe different methods of signal processing from FFT to continuous wavelet transform and the latest method is recognized as the best [2], which we are planning to exploit in the next future. In this study we employ DWT in order to extract features. These features are then used for storage and classification. We construct our diagnostic tool from five functional blocks working in series, as shown in Fig. 3.

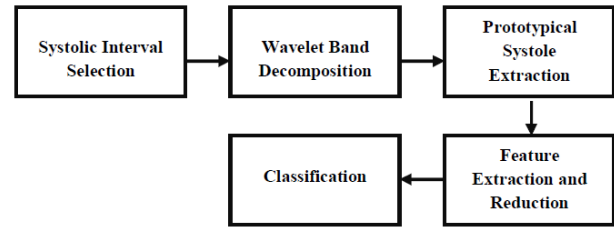


Fig. 3. Sub-Problem decomposition of automated auscultation.

### 2.4.1 Systolic Interval Selection

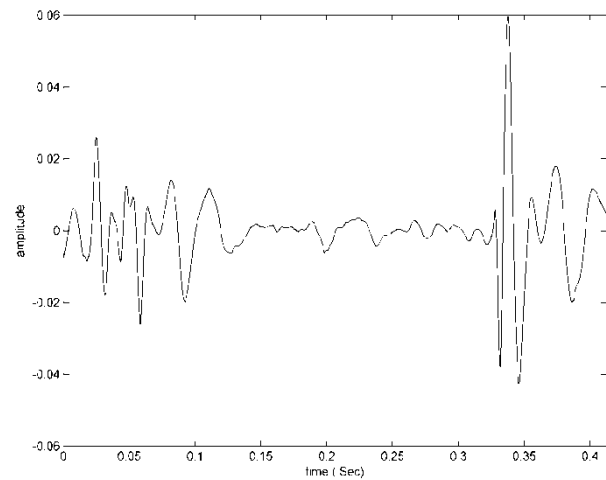


Fig. 4. A sample systole corresponding to a normal HS.

Since our Data set only consists of valvular diseases that take place in the systolic phase of the cardiac cycle, we have extracted only the systolic parts in each beat. Fig. 4 shows the most informative part extracted from the PCG signal shown in Fig. 2.

We have chosen each systole to start with the QRS complex in the corresponding ECG signal and end simultaneously with S2. The reason why we have chosen these two points was the fact that the interval between them contains the most diagnostic information and also these points are easy to locate.

### 2.4.2 Wavelet Band Decomposition

Since the amplitude of low frequency energy in the recorded signal is generally several orders of magnitude greater than the amplitude of high frequency energy, as shown in Fig. 5, we separate each systolic interval into three bands, approximately corresponding to frequency ranges of 75-150 Hz, 150-300 Hz, and 300-600 Hz.



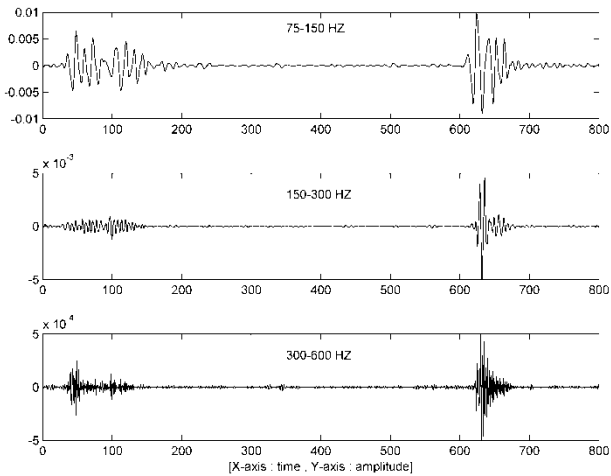


Fig. 5. Variation in amplitude of different wavelet bands.

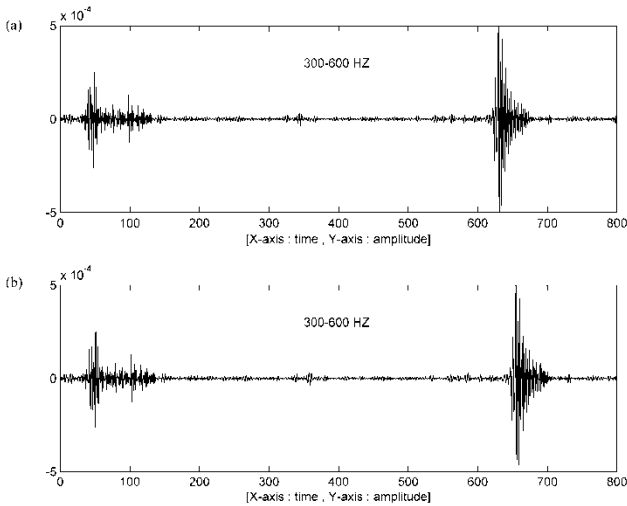


Fig. 6. a) Wavelet band 3 decomposition of a sample systolic interval  
b) the same signal after scaling

So that the energy content of any frequency band wouldn't be colored by that of adjacent bands. In order to do so, we use wavelet bands while providing high temporal resolution at high frequencies.

It should be noted that before doing any further processing we should normalize the length of all three bands of each selected systole through scaling step. All three band representations of each systole are stretched to the duration of the longest systole that has been recorded for a patient. By enforcing a standard length, timing characteristics are viewed as a percentage of systole rather than as a scalar offset from S1. Fig. 6 illustrates how scaling step performs length normalization for band three of the described wavelet bands. We have shown only band three to provide better comparison for readers.

### 2.4.3 Prototypical Systole Extraction

The time-frequency decomposition provides us with the frequency components for each selected beat. In order to observe the characteristic trends persisting amongst the majority of the beats, we also develop a mechanism to

merge information from multiple systolic segments to create a single representative heart-beat for the patient. In other words, we assimilate information from the selected beats to generate the time-frequency decomposition of a hypothetical "typical" beat for the patient.

Like [10], we assimilate information from the selected systoles to generate the time-frequency decomposition of a hypothetical "typical" systole for each patient. The wavelet band decomposition components are time-envelope characterized (see Fig. 7).

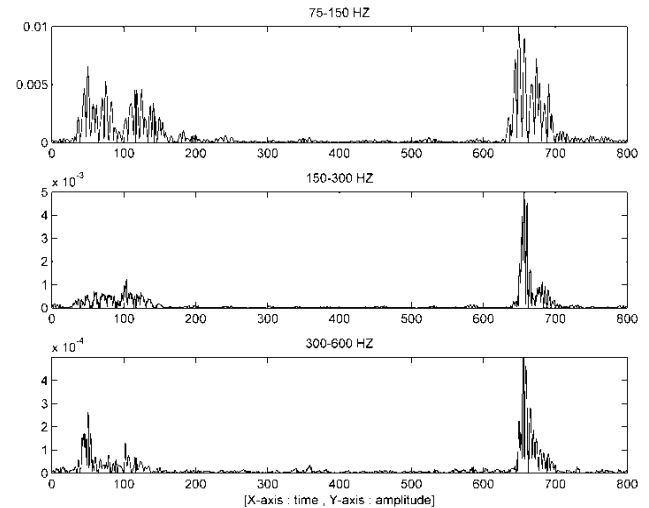


Fig. 7. Time envelope of systolic frequency bands.

We calculate the absolute value at every time instant for the component signals to avoid destructive interference whereby positive and negative values from different systoles may cancel each other. This would lead to the aggregate incorrectly indicating diminished energy content or even the absence of energy altogether, at any time instant.

The step of the prototypical systole calculation can be represented as finding the median four elements along time vector for all of the three bands as in the third block of Fig. 8.

The mean of these median amplitudes is then calculated for each wavelet band. This is illustrated on the last block in Fig. 8 Meaning the end result is a time-frequency decomposition of the prototypical systole.

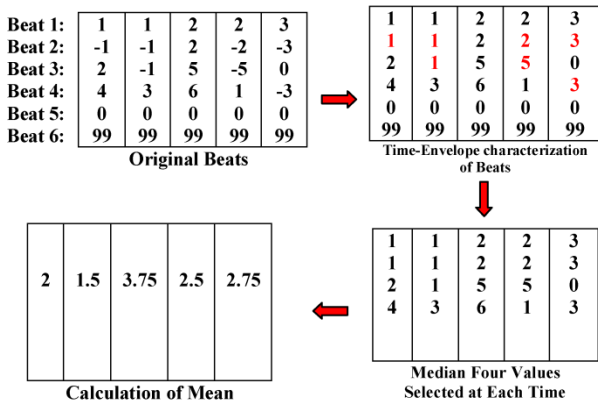


Fig. 8. Prototypical systole construction stages [10].

### 2.4.4 Feature Extraction

In each band, we record energies from one hundred adjacent bins spanning prototypical systole and normalize by the maximum feature value. This step serves as the basis for our feature set. As it's shown in Fig. 9 when these energy values are plotted in sequence, they would roughly capture the original shape of prototypical systole in a fixed and relatively small number of samples.

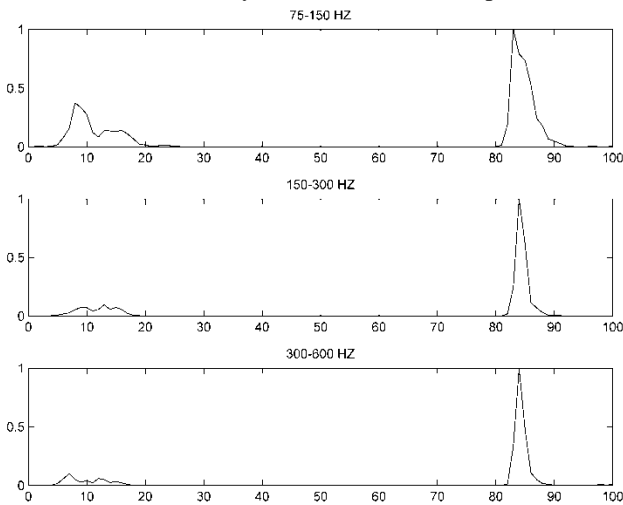


Fig. 9. Energy contents of a windowed sample prototypical systole.

The process of feature extraction would lead to extract some physiological feature from these energy contents for any of the prototype systoles and in all three bands. We have extracted 15 (5\*3) features in this step. These features are as follows: energy and width of S2, the height of a possible murmur and the temporal location of its peak and finally mid-systolic energy. In what follows we will see that only four of these features were considered for classification. If too many features are used, the performance of the classifier will decrease as well as recognition rate (due to overfitting). The reason lies in the existence of many different solutions that are consistent with the training examples, but disagree on unseen (test) examples. Hence lack of training data enforced us to choose only some of features.

### 2.4.5 Principal Components Analysis

Principal component analysis or PCA is a method that has been widely used in several signal processing studies for various reasons, such as data compression or feature extraction or reduction techniques.

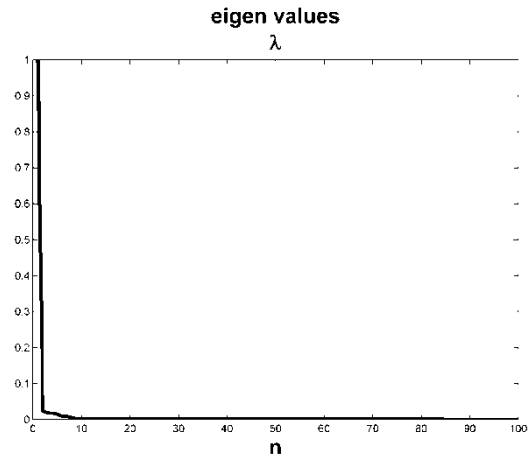


Fig. 10. Eigen values of the covariance matrix (n refers to the n<sup>th</sup> principle component).

In this study we have used PCA to extract principal components in order to use the first three principle components as our second feature set.

As shown in Fig. 10 the eigen values of the covariance matrix of the input data are in a descending order with a sharp slope after the third principle component. And this is why we have just used three of these components.

Principle component analysis was done to only band three of the signal, which measures energy contents in 100 bins for the prototype systole.

## 3. Results

In many cases it is required to do an automatic detection of normal and abnormal heart sound. It can be done using methods such as implementation of a neural network, or Hidden Markov Model (HMM). In this study the simple K-nn classifier has been employed because it proved to be efficient and accurate and in some cases better than MLP. Input data set of 93 children recorded HS signals are derived from various sources in digitized sample form, manually recorded with stethoscope on patients, CD-ROM [20], and internet sites, which comprised of 40 normal HSs and 53 patients with a variety of cardiac diagnoses and a pathologic systolic murmur present.

Selecting a wavelet then becomes a process of seeing which wavelet suits the style of the signal to be investigated rather than being constrained by the compression process. To determine that which mother wavelet would best adapt with the prototype systoles we investigated between some common mother wavelets that have been employed in many similar research, namely db4, db2, db10, sym7 and sym8.

A conclusion drawn from the Fig. 11 suggests that there is little difference from a variety of different wavelets when applied to heart sounds, especially when the wavelets visually look closely matched. In sum, there is not much difference in the various mother wavelets, with a slight preference towards the db4 that would best suit to our case.

As it's shown in Fig. 12, the classifier performance will reach to a fixed level for the case when feature vector contains one principal component and four physiological features and at least 55% of the whole dataset is used for training. These results are taken when 55% of total samples are used as the training dataset and 100% of them as the test dataset. Meaning the train and test datasets contain 51 and 93 prototypical systoles respectively.

Fig. 12 demonstrates that in order to reach to an optimum condition 55% of the whole databank should be used for training. In order to increase the performance of our system we have added two more principal components to the input feature vector and increased the percent of training data to 76%. So this new feature vector contains 7 features for each patient.

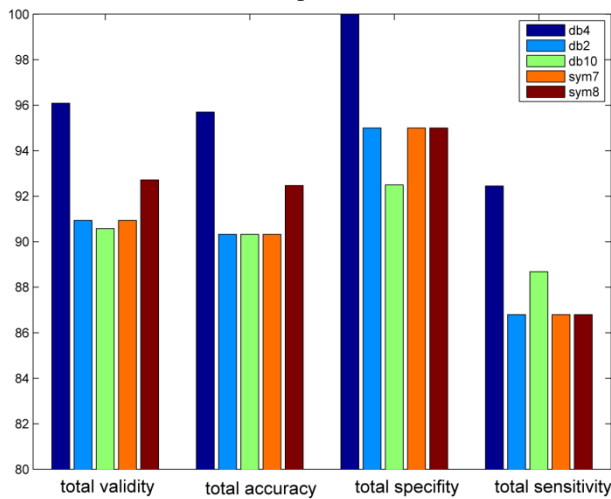


Fig. 11. Total Performane changing mother wavelets.

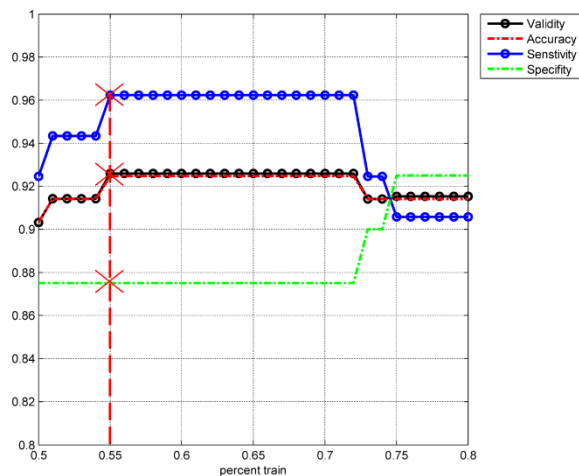


Fig. 12. percent of train databank vs. validity, accuracy, sensitivity and specificity of the classifier.

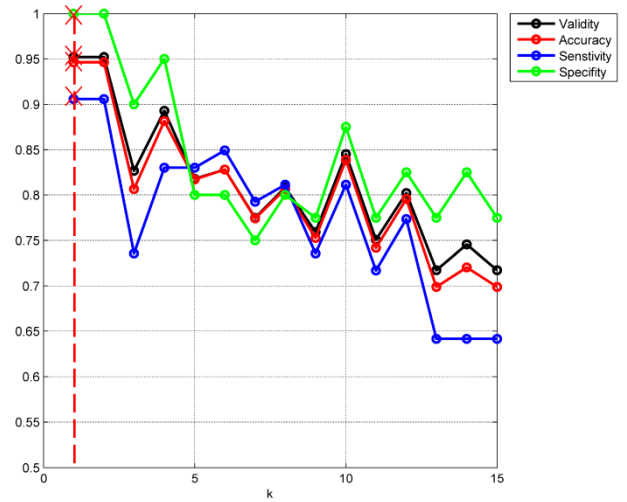


Fig. 13. Number of neighbors vs. validity, accuracy, sensitivity and specificity of the classifier.

After all, the classification was carried out in accordance to tuning of the k-nn classifier. By stepwise increasing of neighbor number, the optimal value of k is determined for the best classifier accuracy. Fig. 13 illustrates the performance of the proposed method for different amounts of k.

This is clearly obvious that the classifier performance decreases as the number of neighbors increases. Therefore, and in order to ensure the computational stability, a value of k=1 was selected. Tabel 1 summarizes the final values related to performance of the proposed method. Amounts related to the crossed points of Fig. 13 are particularly shown in Table 1. G-means metric, which is defined in [21], has been used for evaluating classifiers on imbalanced datasets. We also use this metric to evaluate our classifier. The G-means criterion is represented in Table 1 and is calculated as follows:

$$g = \sqrt{acc+ \cdot acc-} \tag{2}$$

where  $acc+$  is sensitivity and  $acc-$  is specificity.

Table 1. shows that the whole dataset contains 40 prototypical systoles for normal children and 53 for children with CHDs. We used the whole dataset to evaluate the performance of our method. The accuracy and validity of the classifier in this case is higher than 93% which is a significant result.

Table 1. performance of the proposed method.

	Actual Group	
	CHD	Normal
CHD	48	0
Normal	5	40
Total = 93	<b>Sen = 90.57%</b>	<b>Spe = 100%</b>
<b>Total Accuracy = 94.62%</b> <b>Total Validity = 95.22%</b> <b>G-means= 95.17%</b>		

## 4. Conclusions

In this study, seven features were extracted from the murmurs of four groups of patients. These features were used to train a simple knn classifier. Training set consisted of 30 non-pathological and 40 pathological HSs. The input feature vector contained three features from systole, one feature from the energy of systolic signals, and three features from the PCs which altogether summed up to seven features. Since the number of features used in this study was low, the overall computational time was only a few seconds.

The feasibility of the automatic classification method of phonocardiogram utilizing the k-nn and spectral analysis for early cardiac disease screening has been examined. Based on spectral and timing properties of a child heart sounds signal, we have developed a five step algorithm for a complete automatic detection of pediatric heart disease. In the first step, we have extracted all systolic intervals. In the second and third step, we have used wavelet analysis for the aim of prototypical systole construction. In the two final steps, feature extraction stage and classification were done.

Previous studies present excellent classification results well above 95 % when classifying a number of different heart abnormalities [3,11,12,13,14,15,16,17,18-19].

In this study the classification accuracy was at best 94.62% when classifying VSD, AS and PS. The advantage of our study is that actual clinical data was used. In practice this means that the recordings contained noise such as friction rubs, rumbling sounds from the stomach, breathing sounds from the lungs, background

noise from the clinical environment and baby crying sounds. In many other studies, the data is either provided from teaching tapes, [11,12] or from specially selected heart cycles of very high quality and with typical morphology [3,12,13]. Another reason could be the choice of the classifier. The choice of an optimal classifier was however not the aim of this paper.

Results show that the algorithm is efficient and could be used as an effective tool for a complete automatic auscultation in a computerized screening of CHDs system. A system based on this approach will be both accurate and robust, while remaining simple enough to be implemented at low cost. Also unlike many other studies the segmentation step is automatically done without the help of an operator.

Our classifier demonstrated a sensitivity of 90.57% and specificity of 100% for classification of normal and pathological murmurs which is a significant result.

A need for more data is evident for clinical validation. More patients are also needed since a rule of thumb is to use 10 times as many cases as there are features for classification, which is far from reached in the present set-up. Areas for future work include the further development of the system to encompass a broader range of symptoms and pathologies, and an evaluation of the resulting system using a larger and more diverse set of clinical data.

## References

- [1] S. R. Bhatikar, and C. G. DeGroff, "A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics", *Artificial Intelligence in Medicine*, Vol. 33, 2005, pp. 251-260.
- [2] K. Pariaszewska, and M. Młyńczak, "Digital stethoscope system: the feasibility of cardiac auscultation.", In *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*, Vol. 8903, 2013, pp. 890315-890315.
- [3] C. G. DeGroff and S. Bhatikar, "Artificial Neural Network-Based Method of Screening Heart Murmurs in Children", *Circulation*, Vol. 103, 2001, pp. 2711-2716.
- [4] I. Maglogiannis, and E. Loukis, "Support Vectors Machinebased identification of heart valve diseases using heart sounds", *computer methods and programs in biomedicine*, Vol. 95, 2009, pp. 47-61.
- [5] S. Jabbari, and H. Ghassemian, "Modeling of heart systolic murmurs based on multivariate matching pursuit for diagnosis of valvular disorders", *Comp. in Bio. and Med.* Vol. 41. No. 9, 2011, pp. 802-811.
- [6] S. Jabbari, and H. Ghassemian, "A Time-Frequency Approach for Discrimination of Heart Murmurs", *J. Signal and information Processing*, Vol. 2, No. 3, 2011, 232-237.
- [7] A. M. Amiri, and G. Armano, "Heart Sound Analysis for Diagnosis of Heart Diseases in Newborns", *APCBEE Procedia*, Vol. 7, 2013, pp. 109 – 116.
- [8] R.M. Rangayyan, and R.J. Lehner, "Phonocardiogram signal analysis: a review". *CRC Crit Rev Biomed Eng.*, Vol. 15, 211–236, 1988.
- [9] M.E. Tavel, and D.D. Brown, "Enhanced auscultation with a new graphic display system", *Arch Intern Med*. Vol. 154, 1994, pp. 893–898.
- [10] A. R. Kenari, and M. Hassan Ghassemian. "Automated spectral analysis for pediatric cardiac auscultation." *Electrical Engineering (ICEE), 2013 21st Iranian Conference on. IEEE*, 2013.
- [11] Cathers, "Neural network assisted cardiac auscultation". *Artif. Intell. Med.* 7:53–66, 1995.
- [12] S. A. Pavlopoulos, A. C. Stasis, and E. N. Loukis. "A decision tree-based method for the differential diagnosis of aortic stenosis from mitral regurgitation using heart sounds". *Biomed. Eng. Online*. 3:21.
- [13] A. Voss, A. Mix, and T. Hubner. "Diagnosing aortic valve stenosis by parameter extraction of heart sound signals". *Ann. Biomed. Eng.* 33:1167–1174, 2005.
- [14] A. N. Gupta, R. Palaniappan, et al. "Neural network classification of homomorphic segmented heart sounds." *Appl. Soft Comp.* In Press, 2006.

- [15] T. Olmez, and Z. Dokur. "Classification of heart sounds using an artificial neural network". *Pattern Recogn. Lett.* 24:617–629, 2003.
- [16] Z., M. Sharif, S. Zainal, et al. " Analysis and classification of heart sounds and murmurs based on the instantaneous energy and frequency estimations." *Proc. TENCON* 2:130–134, 2000.
- [17] I. Turkoglu, A. Arslan, and E. Ilkay. "An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks." *Comput. Biol. Med.* 33:319–331, 2003.
- [18] C. Kwak, O.W. Kwon, "Cardiac disorder classification by heart sound signals using murmur likelihood and hidden Markov model state likelihood", *J. IET Signal Process.*, Vol. 6, Iss. 4, pp. 326–334, 2012.
- [19] Castro, Ana, et al. "Heart sound segmentation of pediatric auscultations using wavelet analysis." *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE. IEEE*, 2013.
- [20] M. McConnell, *Pediatric Heart Sounds*, Atlanta, GA: Springer, 2008.
- [21] M. Kubat and S. Matwin. "Addressing the Curse of Imbalanced Training Sets: One- Sided Selection". *Proceedings of the 14th International Conference on Machine Learning*, Nashville, TN, USA, 1997.

**Azra Rasouli Kenari** received the B.Sc degree in Electrical Engineering from Babol Noshirvani University of technology, Babol, Iran and M.Sc degree in Biomedical Engineering from Tarbiat Modares University, Tehran, Iran in 2010 and 2013 respectively. She is currently working toward the Ph.D. degree in the Department of Advanced Medical Technology at the Isfahan University of Medical Sciences, Isfahan, Iran. Her research interests include medical image analysis and modeling.

**Hassan Ghassemian** received his B.Sc degree in Electrical Engineering from Tehran College of Telecommunication in 1980 and the M.Sc degree in Electrical Engineering and Ph.D degree from Purdue University, West Lafayette, USA in 1984 and 1988 respectively. Since 1988, he has been with Tarbiat Modares University in Tehran, Iran, where he is a Professor of Electrical and Computer Engineering. Dr. Ghassemian has published more than 370 technical papers in peer-reviewed journals and conference proceedings. His current research interests focus on Multi-Source Signal/Image Processing, Information Analysis and Remote Sensing.

# Active Steganalysis of Transform Domain Steganography Based on Sparse Component Analysis

Hamed Modaghegh \*

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran  
hamed.modaghegh@stu-mail.um.ac.ir

Seyed Alireza Seyedin

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran  
seyedin@um.ac.ir

Received: 20/Aug/2014

Revised: 19/Jan/2015

Accepted: 27/Jan/2015

## Abstract

This paper presents a new *active* steganalysis method to break the *transform* domain steganography. Most of steganalysis techniques focus on detecting the presence or absence of a secret message in a *cover* (*passive* steganalysis), but in some cases we need to extract or estimate a hidden message (*active* steganalysis). Despite the importance of estimating the message, little research has been conducted in this area. In this study, a new *active* steganalysis method based on Spars Component Analysis (SCA) technique is presented. Here, the sparsity property of the *cover* image and hidden message has been used to extract the hidden message from a stego image. In our method, the *transform* domain steganography is formulated mathematically as a linear combination of sparse sources. Thus, the *active* steganalysis can be presented as a SCA problem. The feasibility of the SCA problem solving is confirmed by Linear Programming methods. Then, a fast algorithm is proposed to decrease the computational cost of steganalysis and still maintains the accuracy. The accuracy of the proposed method has been confirmed in different experiments on a variety of *transform* domain steganography methods. According to these experiments, our method not only reduces the error rate, but also decreases the computational cost compared to the previous *active* steganalysis methods in the literature.

**Keywords:** Sparse Component Analysis (SCA); Active Steganalysis; Blind Source Separation (BSS); Transform Domain steganography.

## 1. Introduction

After the seminal study of Johnson and Jajodia [1], [2], steganalysis has attracted growing attention [3]–[7]. Different types of steganalysis techniques (STs), mostly *passive*, have been proposed [4]–[7]. While Steganography deals with hiding information by embedding a message in another object (*cover*) such as an image, steganalysis focuses on revealing those hidden messages from the *cover*. Steganalysis has gained prominence in the international security since the detection of hidden messages can lead to the prevention of catastrophic events, such as terrorist attacks.

Current STs focus on detecting the presence of a hidden message in the *cover* (*passive* manner). An in-depth review of *passive* STs has been presented by Nissar et al. [5]. They attempted to classify various approaches. They have categorized STs into signature and statistical techniques. Their categorization is either based on the signature of the applied technique or the image statistics which is used to detect the presence of hidden messages. Furthermore, in their classification, each category is subdivided into specific and universal approaches.

Specific steganalysis targets a particular steganographic technique [4], [7], [8]. These methods analyze the embedding operation and concentrate on some image features or statistics. As a result, it may fail if

any other steganography method is used or simply a change occurs in the steganography algorithm. Consequently, universal STs [9]–[11], were introduced to overcome the deficiency of specific STs. These methods could detect embedded messages using any type of steganographic technique even in the absence of prior knowledge of embedding technique. Most of them train a classifier with *cover* and stego images in the detection procedure.

Following the detection procedure, sometimes it is necessary to extract and determine the content of the hidden message (*active* steganalysis). In fact, by revealing the hidden messages, *active* steganalysis complements the *passive* one. Most STs deal with *passive* techniques and little attention has been paid to *active* methods [12]–[16]. In this scope, some researchers focus on *active* STs based on the blind sources separation (BSS) [3], [15]–[19]. This study focuses on this class of *active* STs and discusses advantages and disadvantages of these methods.

It is worth mentioning that most BSS-based *active* STs take advantage of the independency property of the image without using sparsity property of the hidden message to achieve better results. Moreover, all *active* STs, which use only one stego image to extract the hidden message, [15]–[17], [19] increase the computational cost. They need at least two observed signals and use a denoising algorithm to generate them. This algorithm usually

\* Corresponding Author

increases the computational complexity and making ST as time-consuming algorithm.

In this paper, a new *active* ST with only one stego image is introduced which does not require a denoising algorithm. Eliminating the denoising algorithm makes this ST more efficient than the previous BSS-based STs. Our method uses sparsity property of sources to separated *cover* and hidden message. It is enough to *cover* and hidden message be sparse in two different dictionaries. Based on sparsity property, an optimization problem is proposed and the feasibility of solving it with linear programming methods is examined. Then, a fast algorithm based on fast *transforms* is proposed to solve the problem and extract the hidden message.

To this end, the paper has been organized as follows: in section II, a brief overview of the current BSS-based *active* STs is presented and their advantages and disadvantages are discussed. Section III explains the sparse component analysis and source separation problem briefly. In the next section, the details of our *active* steganalysis method are presented. Finally, a discussion of the experimental evaluation is made in section VI and conclusions are drawn in section VII.

## 2. A Brief History on Active STs

Chandramouli [3] developed the first *active* ST based on the BSS model to challenge the linear steganography. His proposed method was based on the BSS model with a hypothesis that the *cover* image and hidden message were independent. However, his proposed method needed at least two stego images with the same message, *cover* and key but different embedding strength factor. However, these conditions are not practical since steganalyst can usually access one stego image only.

Fan et al. [15] tried to apply a method to realize *active* steganalysis when there was only one stego image copy. Their method was based on Independent Component Analysis (ICA) [20] and Hidden Markov Tree (HMT) model[21]. The former is a popular BSS technique and the latter is mainly applied to denoise an image in the *transform* domain. They adopted HMT model to obtain the second copy of stego image and then the optimized ICA was applied to achieve the *active* steganalysis.

Another study with the view of active steganalysis as BSS problem was presented in [16]. It solely used a single copy of stego image. The maximum a posteriori (MAP) estimator was adopted to provide an estimate of the *cover* image. Two *active* steganalysis schemes was introduced in this method; the first scheme was similar to [15] which considered the estimated version as another stego image. In the second scheme, besides the original stego image, two other stego images were generated from the estimated image. All the three images provide an input to the ICA algorithm. These schemes were applied to extract messages from the least significant bit (LSB) steganography in spatial, discrete cosine *transform* (DCT),

and discrete wavelet *transform* (DWT) domains. The results indicated that the second scheme has a better performance than the first one. The method proposed in [17] was similar to that of [16]. However, HMT model was applied in [17] while the MAP estimator was used in [16] to gain an estimate of the *cover* image.

All these *active* steganalysis methods (ICA-based *active* steganalysis) use ICA technique to separate message from image. This technique is an inherently high computational cost technique. Ambalavanan and Chandramouli, on the other hand, introduced another *active* steganalysis based on a different BSS technique [22] in order to reduce the computational cost and to improve the performance of message extraction [18]. However, their efforts were not successful.

Modaghegh et al. [19] introduced an *active* ST that reduced the computational cost and error rate . Their method was basically a combination of the blind source separation technique and MAP estimator. Additionally, they presented a new geometrical BSS method based on the minimum range of mixed sources which reduced the computational cost of their *active* ST. Their experiments showed that their *active* ST not only reduced the error rate, but also decreased the computational cost compared to the previous *active* STs. Nevertheless, all of these methods involve a denoising algorithm which increase the computational cost required to generate an estimated version of the *cover*.

## 3. Preliminary: Sparse Component Analysis and Blind Source Separation

A brief overview of the sparse component analysis and blind source separation problems and the possible solutions [23] are given in this section. The goal of the source separation is to retrieve an unknown source signals  $S$  from observed signals  $X$  where  $X$  and  $S$  are row vectors. The observed signals are often assumed to be a linear instantaneous mixture of source signals. Thus, it can be written as  $X = AS$ , where  $A$  denotes the unknown mixture matrix. Since  $A$  and  $S$  are unknown, some assumptions are needed to solve this problem. Basically, it is necessary to have prior knowledge of the source properties such as its independency, sparsity, bounded/unbounded states and so on. The methods based on the property of source independency are called independent component analysis [20]. On the other hand, the sparse component analyses (SCAs) are based on the property of the source sparsity [24], meaning that each source is seldom *active* and mostly (nearly) zero. Let us consider a sparsity model for the probability distribution function of the sources as follows,

$$P_{s_k}(s_k) = p_{s_k} \delta(s_k) + (1 - p_{s_k}) f_{s_k}(s_k) \quad (1)$$

Where  $p_{s_k}$  is the sparsity factor of the source and  $f_{s_k}$  denotes the distribution of  $s_k$  when the corresponding source is *active*.

In the general sparse representation framework, a signal vector  $X \in R^N$  is modeled in the dictionary  $\Phi$  as the linear combination of  $T$  elementary waveforms  $\varphi_i$  (atoms of dictionary):

$$X = \Phi\alpha = \sum_{i=1}^T \alpha_i \varphi_i \quad (2)$$

In the case of overcomplete representation, the number of waveforms or atoms  $\varphi_i (1 \leq i \leq T)$  that constitute columns of the dictionary  $\Phi$  is greater than the space dimension in which  $X$  lies:  $T > N$ , or even  $T \gg N$  for highly redundant dictionaries. The separation problem of a signal or image in  $\Phi$  is concerned with recovering the coefficient vector  $\alpha$  in Eq.(2). However, as there are  $N$  equations and  $T$  unknowns, the problem has no unique solution. The solution to the underdetermined system of linear equations  $X = \Phi\alpha$  can be achieved by reducing the space of candidate solutions. In the SCA problem, sparsity imposes constraints on the solutions, meaning that among all solutions of  $X = \Phi\alpha$ , the sparsest one is preferred (with the least number of nonzero entries  $\alpha_i$ ). In other words, the sparse decomposition problem entails solving the following minimization problem:

$$\min_{\alpha \in \mathbb{R}^T} \|\alpha\|_0 \quad s.t. \quad X = \Phi\alpha \quad (3)$$

As can be seen, Eq.(3) is a *combinatorial* optimization problem that requires enumerating all collections of atoms in  $\Phi$  to find the smallest set that synthesizes  $X$ . This made authors turn to approximations or relaxations of Eq. (3). Donoho and Huo [25] proposed a method to relax the non-convex  $\ell_0$  sparsity measure by substituting the problem in Eq.(3) with the convex problem:

$$\min_{\alpha \in \mathbb{R}^T} \|\alpha\|_1 \quad s.t. \quad X = \Phi\alpha. \quad (4)$$

This problem is called *Basis Pursuit* (BP) [26]. Unlike Eq.(3), Eq.(4) is a computationally tractable convex optimization problem that can be solved efficiently by linear programming methods [27]. However, BP is not able to find a general solution for Eq.(2). Under appropriate conditions on  $\Phi$  and  $X$ , nonetheless, BP can offer a general optimal solution of Eq.(2).

On the other hand, some authors have attempted to provide a more effective solution for problem Eq.(2) [28], [29]. The morphological diversity concept introduces a new data modeling framework that allows having both a sparse representation and a fast algorithm that exploits the structure of the dictionary. Morphological diversity assumes that the signal  $X$  can be modeled as the sum of  $K$  components  $s_k$  that are morphologically different:

$$X = \sum_{k=1}^K s_k \quad (5)$$

Where  $s_k$  is a morphological component. Each  $s_k$  is sparse in a given dictionary  $\Phi_k$ , which is associated with implicit fast analysis/synthesis *transforms* such as wavelet and DCT.

## 4. The Proposed Active Steganalysis Method

In this section, we introduce our *active* steganalysis method that can be used for breaking the *transform* domain steganography techniques [30]–[34]. In the first subsection, *transform* domain steganography is formulated mathematically as a linear combination of sparse sources. Then, *active* steganalysis is formulated as a new SCA problem. In the next subsection, the feasibility of solving the new SCA problem is discussed. Finally, with the goal of reducing computational cost, a fast algorithm for solving the new SCA problem is proposed. At this point, we are ready to illustrate the details of our proposed method.

### 4.1 SCA as Active Steganalysis

Common steganography techniques can be modeled as an additive embedding, i.e. the sum of image features and hidden message. *Transform* domain steganography methods use *transform* coefficients as image features to embed hidden message [34]. Thus, these methods can be formulated as follows:

$$Y_x = Y_c + \beta W \quad (6)$$

Where  $Y_x$  denotes the *transform* coefficients of stego image,  $\beta$  is the embedding coefficient and  $W$  and  $Y_c$  are respectively the hidden message and *transform* coefficients of *cover* image. To obtain the stego image, we need to apply inverse *transform* to  $Y_x$ , so we have:

$$X = C + \text{invTrans}(\beta W) = C + S_{\beta W} \quad (7)$$

To perform *active* steganalysis, we need to extract hidden message  $W$  from stego image  $X$ . In this paper, we formulate *active* steganalysis as a source separation problem. From a BSS viewpoint, the *cover* image and inverse *transform* of the message are sources mixed to form the stego image as an observation. In this case, since the number of observations is less than the number of sources, the problem is underdetermined, and therefore there is an infinite number of solutions. As a result, in order to select one solution among all available solutions it is necessary to impose certain additional constraints based on the prior knowledge. In previous BSS-based *active* STs, the independency of image and hidden message was used as an additional constraint [3], [15]–[19]. However, in this paper, we show that both sources in the stego image are sparse in their dictionaries and use sources sparsity as an additional constraint to solve the BSS problem.

The hidden message is sparse because imperceptibility is a fundamental requirement of steganographic methods. This means that embedding a hidden message should not significantly change the *cover* so the hidden messages remain indiscernible to the human eye. Because of this feature, a hidden message needs to have short length and low amplitude. In other words, most of elements of a hidden message vector  $W$  are zero and the embedding coefficient  $\beta$  is also small. In this regard, the message source can be viewed as a sparse source in *transform* domain ( $\Phi_1$  dictionary).



On the other hand, images and practical signals are not, in general, strictly sparse. Though they may be compressible or weakly sparse in *transform* domains ( $\Phi_2$  dictionary) such as wavelet or DCT. This means that most  $\alpha_{iS}$  in Eq.(2) are near zero, and one can neglect all but perhaps a small fraction of the coefficients without significant loss. Thus, the image source can be viewed as a sparse source in  $\Phi_2$  dictionary.

Since the two sources are sparse, the original image can be separated from the hidden message if the stego image is searched smartly for sparse component, meaning that *active* steganalysis can be viewed as a SCA problem. As stated in Section 3, we can formulate SCA problem as an optimization problem Eq.(3). As our SCA problem is underdetermined and there are an infinite number of solutions, we need to use our prior knowledge to find the accurate answer (*cover* and hidden message), including the knowledge of  $\alpha_{li}$  amplitude in dictionary  $\Phi_1$  which is corresponding to the hidden message. As  $\alpha_{li}$  is small due to imperceptibility feature of the steganography method, Eq.(3) can be written as follows:

$$\min_{\alpha_i \in \mathbb{R}^T} (\|\mathbf{a}_1\|_0 + \|\mathbf{a}_2\|_0) \quad s.t. \quad \begin{aligned} X &= \Phi_1 \mathbf{a}_1 + \Phi_2 \mathbf{a}_2 \\ |\alpha_{li}| &< \alpha_{th} \end{aligned} \quad (8)$$

$$\text{and} \quad \Phi = \bigcup_{k=1}^2 \Phi_k$$

Where  $\Phi_2$  and  $\mathbf{a}_2$  are dictionary and coefficient vector of the *cover* respectively. This optimization problem has an extra constraint  $|\alpha_{li}| < \alpha_{th}$  which helps separate hidden message from the *cover*.

## 4.2 Feasibility Problem

The extra constraint distinguishes this problem from normal SCA problems. Since it is a complex non-convex problem, we can simplify this problem by converting  $l_0$  norm to  $l_1$  norm, formulating this problem like *Basis Pursuit* as follows:

$$\min_{\alpha_i \in \mathbb{R}^T} (\|\mathbf{a}_1\|_1 + \|\mathbf{a}_2\|_1) \quad s.t. \quad \begin{aligned} X &= \Phi_1 \mathbf{a}_1 + \Phi_2 \mathbf{a}_2 \\ |\alpha_{li}| &< \alpha_{th} \end{aligned} \quad (9)$$

This extra constraint also distinguishes this problem from BP. However, this is a linear programming problem which can be solved by LP methods. To do so, we need to convert Eq.(9) to a canonical form of LP problem as stated below:

$$\min B^T \mathbf{a} \quad s.t. \quad D\mathbf{a} \leq b \quad \text{and} \quad \mathbf{a} \geq 0 \quad (10)$$

Where  $B$  and  $b$  are vectors of (known) coefficients and  $D$  is a (known) matrix of coefficients. The inequalities  $D\mathbf{a} \leq b$  and  $\mathbf{a} \geq 0$  are the constraints which specify a convex polytope over which the objective function is to be optimized. Clearly, Eq.(9) can be converted into the canonical form Eq.(10) by adding extra unknowns [27]. Then, this LP problem can be solved using simplex or interior point method[27].

However, Eq.(9) does not provide a solution for Eq.(8) in general. But under appropriate conditions on  $\Phi$  and  $X$ , BP can provide the globally optimal solution for Eq.(8).

Thus, practical algorithms can solve problems that seem computationally intractable on the surface. Many studies have focused on sufficient (and sometimes necessary) conditions under which the problem BP recovers the sparsest solution of an underdetermined system of linear equations. For instance, sufficient conditions based on the mutual coherence of  $\Phi$  were introduced by several authors (see, for example, [25], [35]–[37]). The mutual coherence  $\mu_\Phi$  of  $\Phi$  is defined as:

$$\mu_\Phi = \max_{i \neq j} \left| \langle \varphi_i, \varphi_j \rangle \right| \quad (11)$$

This quantity can be viewed as a worst-case measure of the resemblance between all pairs of atoms. Donoho and Huo [25] showed that for dictionaries with small  $\mu_\Phi$ , the solution of BP is unique and this unique solution is a point of equivalence of Eq.(3) and Eq.(4). It can be shown that these conditions are available for our *active* steganalysis problem Eq.(9). Thus, we can easily solve Eq.(9) instead of non-convex Eq.(8) and obtain the hidden message. In the case Eq.(9), if the mutual coherence between atoms of two dictionaries  $\Phi_1$  and  $\Phi_2$  is low, this condition is met. Since in steganography,  $\Phi_1$  and  $\Phi_2$  correspond to *transform* domains, the *cover* and hidden message need to be sparse in two different *transforms* to acquire a small  $\mu_\Phi$ . Under this condition, the *active* steganalysis problem is feasible, and it can be solved with LP methods.

## 4.3 Proposing Fast Algorithm

In many cases, BP-like synthesis algorithms are computationally expensive. In this subsection, an alternative to these approaches has been proposed. We have chosen an approximation to our true minimization task to find a simplified optimization problem which is computationally effective. Our proposed method is similar to Morphological Component Analysis (MCA) method, which can be seen as a kind of *Basis Pursuit* method [26] called MCA steganalysis (MCAS). The algorithm is based on the Block-Coordinate-Relaxation method[38], with some changes made by the properties of the *active* steganalysis such as the hidden message amplitude constraint imposed on the reconstructed components.

In our algorithm, we assumed that a dictionary can be built by amalgamating two subdictionaries ( $\Phi_1, \Phi_2$ ) such that for each  $k$ , the representation of  $s_k$  in  $\Phi_k$  is sparse and not sparse – or at least not as sparse – in other  $\Phi_l, l \neq k$ . In other words, the subdictionaries  $\Phi_k$  must be mutually incoherent. Thus, the dictionary  $\Phi_k$  plays the role of a discriminant between the hidden message and *cover*, preferring the component  $s_k$  over the other part. This is a key observation for the success of the separation algorithm. If this condition is not satisfied and the hidden message and *cover* sources are sparse in the same dictionary, other *active* steganalysis methods can be used [19].

Since most *transform* domain steganography methods [30]–[34] use fast *transforms*, in our algorithm, the matrix  $\Phi_k$  and its transpose  $\Phi_k^T$  corresponding to each *transform*

are never explicitly constructed in the memory. Instead, they are implemented as a fast implicit analysis and synthesis *transforms* taking a signal vector  $X$  and returning  $\Phi_k^T X = T_k(X)$  (analysis side), or taking a coefficient vector  $\alpha_k$  and returning  $\Phi_k \alpha_k = \Phi_k^{-1}(X)$  (synthesis side). In the case of a simple orthogonal basis, the inverse of the analysis *transform* is trivially  $T_k^{-1} = \Phi_k$ . The use of this transformation instead of the dictionary is what makes our method computationally efficient.

One of the important ingredients of our algorithm is the coordinate relaxation. If all component coefficients  $\alpha_l$  but the  $k$ th are fixed, then a solution can be achieved through thresholding the coefficients of the marginal residuals  $r_k = X - \sum_{l \neq k} \Phi_l \alpha_l$  in  $\Phi_k$ . The other components are relieved of these marginal residuals  $r_k$  and are likely to contain mainly the salient information of  $s_k$ . This intuition dictates a coordinate relaxation algorithm that cycles through the components at each iteration and applies a thresholding to the marginal residuals. MCAS algorithm is summarized as follows:

**Algorithm 1:**

1. Initialize iteration number  $j=1$ ,  $N_{iter}$  the number of iterations, threshold  $\delta_0$ , and step size  $\lambda = \delta_0 / N_{iter}$ .
2. Calculate the residual  $r_1 = X - s_1$ ,  $r_2 = X - s_2$ .
3. Calculate the *transform*  $T_k$  of  $r_k$  and obtain  $\alpha_k = T_k(r_k)$  for  $k=1,2$ .
4. Apply hard threshold to the coefficient  $\alpha_k$  with the  $\delta_j$  threshold and obtain  $\hat{\alpha}_k$  for  $k=1,2$ .
5. Reconstruct  $s_k$  by  $s_k = T_k^{-1}(\hat{\alpha}_k)$  for  $k=1,2$ .
6. Apply the constraint correction if  $l_0(\hat{\alpha}_1) \leq N_{msg}/K_1$ ,  $s_j = 0$ .
7. Update the threshold by  $\delta_j = \delta_{j-1} - \lambda$ .
8. If  $l_0(\hat{\alpha}_1) \leq N_{msg}$ , update the  $j=j+1$  and return to Step 2, else, finish.

Unlike BP, the MCAS is stage wise and exploits the fact that the dictionary is structured (union of *transforms*), and the atoms enter the solution by groups, rather than individually. As such, MCAS is a salient-to-fine process in which the most salient content of each morphological component is iteratively computed at each iteration. These estimates are then progressively refined as the threshold  $\delta$  evolves toward  $\delta_{min}$ . In the above algorithm, we use hard thresholding instead of soft thresholding due to our formulation of the  $l_0$  sparsity penalty term and the fact that hard thresholding provides better results [39].

Besides fast *transforms* and coordinate relaxation, another important ingredient of MCAS is iterative thresholding with varying threshold. The way the threshold is decreased along the iterations of the MCAS algorithm is significant in terms of separation quality. There are two types of decreasing threshold strategy: prefixed decreasing threshold strategy and adaptive strategy. Given the information we have about the amplitude of  $\alpha_1$ , the prefixed decreasing threshold is selected. In our study, the threshold is decreased linearly so the threshold  $\delta$  sequence is as follows:

$$\delta_j = \delta_0 - j \times (\delta_0 - \delta_{min}) / N_{iter} \quad (12)$$

The first threshold  $\delta_0$  can be set automatically to a sufficiently large value which is greater than  $\hat{\beta}$  (the

estimate of the message embedding coefficient), e.g.  $\delta_0 = K_1 \hat{\beta}$  where  $K_1$  can be a number between 1.5 and 3. For an exact representation of the data with the morphological components,  $\delta_{min}$  must be set to zero. In our algorithm, considering the information we have about the message embedding rate ( $N_{msg}$ ) the algorithm can be repeated until the number of non-zero element of  $\hat{\alpha}_1$ , i.e.  $l_0(\hat{\alpha}_1)$ , become equal to  $N_{msg}$ . Since some *passive* steganalysis methods [40]–[45] can estimate the message embedding rate, it is practical to use this prior information in our algorithm.

The key part of our algorithm is step 6. In regular SCA method [28], [29], [46] all  $s_k$  are updated in every iteration. In our algorithm, however, the hidden message source has low and almost equal amplitude. Thus, we use this prior information to update the message source in case the number of extracted hidden message is greater than  $N_{msg}/K_2$  (where  $K_2$  is a number between 3 and 5). Otherwise, the extracted data does not belong to the hidden message source and  $s_j$  is not updated. Under ideal circumstances where mutual coherence is zero, this step would be unnecessary. In practical conditions, however, the mutual coherence between  $\Phi_1$  and  $\Phi_2$  is not zero and the image source  $s_2$  also produces some large elements in  $\alpha_2$ . Under these conditions, some elements of  $\alpha_l$  acquire large amplitudes, not corresponding to hidden message  $s_l$  and should be discarded. Fortunately, since the number of these large amplitude samples in hidden message coefficient vector  $\alpha_l$  is low, it can be discarded in step 6.

Another challenge facing our *active* steganalysis method is the message dictionary selection, i.e. the detection of the dictionary used in steganography algorithm. To do so, first a proposition based on the central limit theorem is presented and then it is used to detect the embedding dictionary.

**Proposition 1:** suppose that  $\Phi_1$  and  $\Phi'_1$  are uncorrelated dictionaries which correspond to orthogonal *transforms*. Furthermore,  $\alpha_1$  and  $\alpha'_1$  are defined as follows:

$$s_1 = \Phi_1 \alpha_1 = \sum_{i=1}^T \alpha_{i1} \varphi_{i1} \quad (13)$$

$$s_1 = \Phi'_1 \alpha'_1 = \sum_{i=1}^T \alpha'_{i1} \varphi'_{i1}$$

Now if  $s_1$  is sparse in  $\Phi_1$  with sparsity factor  $p_{s1} = I - L/T$  and  $|\alpha_{i1}|_{I < i < L} = \beta$ , then the probability distribution of  $\alpha'_{i1}$  is normal and  $\hat{p}_{s1} < p_{s1}$ .

**Proof:**

Since  $\Phi_1$  and  $\Phi'_1$  are orthogonal transform, we have:

$$\begin{bmatrix} \varphi_{11} \\ \dots \\ \varphi_{1T} \end{bmatrix} = \begin{bmatrix} e_{11} & \dots & e_{1T} \\ \dots & & \\ e_{T1} & & e_{TT} \end{bmatrix} \times \begin{bmatrix} \varphi'_{11} \\ \dots \\ \varphi'_{1T} \end{bmatrix} \quad (14)$$

Thus, each atom  $\varphi_{i1}$  can be written as  $\varphi_{i1} = e_{i1} \varphi'_{11} + \dots + e_{iT} \varphi'_{1T}$ . Now, substituting  $\varphi_{i1}$  in Eq.(13):

$$\begin{aligned}
s_1 &= \Phi_1 \mathbf{a}_1 = \sum_{i=1}^T \alpha_{1i} \varphi_{1i} = \sum_{i=1}^L \pm \beta \varphi_{1i} \\
&= \beta \sum_{i=1}^L (\pm e_{i1} \varphi'_{11} \pm \dots \pm e_{iT} \varphi'_{iT}) = \sum_{i=1}^T \alpha'_{1i} \varphi'_{1i}
\end{aligned} \tag{15}$$

Thus:

$$\alpha'_{1i} = \beta (\pm e_{i1} \pm \dots \pm e_{iT}) \tag{16}$$

In the case of orthogonal *transforms*, if we suppose that  $e_{ij}$  is an independent random variable with identical distribution (i.i.d), then  $\alpha'_{1i}$  will be the sum of a large number of i.i.d random variables. Therefore, according to the *central limit theorem* in the probability theory,  $\alpha'_{1i}$  will have approximately normal distribution. Since  $\alpha'_{1i}$  is a normally distributed random variable, most of  $\alpha'_{1i}$  s are non-zero and therefore  $\hat{p}_{s1}$  will be less than  $ps1$ .

According to this proposition, if signal  $s_1$  is sparse in dictionary  $\Phi_1$  then it would be less sparse in another dictionary  $\Phi'_1$ . Thus, if a message is embedded in the *transform* domain  $\Phi_1$  and the hidden message is searched in another dictionary  $\Phi'_1$ , the extracted message will be less sparse. This property is used to detect the embedding dictionary. To do so, we need to run algorithm 1 until the condition stated in step 6 is met. Then, the corresponding threshold  $\delta$  is saved for each dictionary. This process is executed for all *transforms* that might be used in steganography and the dictionary with the greatest threshold is selected. The results of running algorithm for some *transforms* will be presented in the next section.

## 5. Experimental Results

At this point, the simulation results of our *active* steganalysis method are presented so that we can evaluate its performance. The simulation was done in MATLAB environment. To evaluate the steganalysis method, we used Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) [47], a dataset consisting of 500 natural color images. These color images, span a range of indoor and outdoor scenes, are in JPEG compressed format with a quality of 75%. We converted these images into grayscale images using only the central  $256 \times 256$  region of each image.

Quality factor 75 has been selected because the default quality setting on most digital cameras and image editors is 75, which provides a good tradeoff between the file size and the perceived quality. With this quality factor, the sparsity factor of DCT coefficient ( $p_{s2}$ ) for our dataset is about 0.8. The sparsity factor of the embedded message ( $ps1$ ) is set to 0.1 (i.e. the message is embedded in 10% of DWT coefficients) in the first simulation (Table 1), while different embedding rates are selected in the next simulation (Table 2). For example, in an image of  $256 \times 256$  pixels, the number of DWT (Haar Wavelet) coefficients is  $256 \times 256$ . Thus, when data is embedded in 10% of coefficients, the capacity of data hiding will be 0.1 bit/pixel. In other words, the number of bits embedded

in the image is  $256 \times 256 \times 0.1 = 6553.6$  with a data hiding capacity of 6553.6 bit/image. Binary message bits ( $\pm 1$ ) are randomly embedded in the DWT coefficients with  $\beta=4$ . As an example, the original, extracted and stego images with 0.1 bit/pixel embedded message rate have been shown in Fig. 1.



Fig. 1. Original Lena image (a) stego image (b) extracted image (c).

The results of extracting message from 500 stego images using the proposed steganalysis method, ICA-

based *active* steganalysis methods [15][16][17] and Modagheh et al.'s method [19] are shown in Table 1. In this table, for ease of comparison and preparation of similar conditions for all methods, ICA-based *active* steganalysis methods are simulated by "Bayesian Least Squares-Gaussian Scale Mixture" (BG) denoising algorithm [48], which has been employed in Modagheh et al.'s steganalysis method [19]. This algorithm is one of the most effective denoising algorithms for removing homogeneous additive noise from natural images.

It is important to note that the results of ICA-based *active* steganalysis are for steganography methods with the message embedded in block-DCT coefficient [15][16][17] [19] not the DWT coefficient. Nonetheless, because there is no similar study in *active* steganalysis, these methods have been compared in general.

Table 1. A comparison between the proposed *active* steganalysis (message in DWT coefficient) and the ICA-based *active* steganalysis methods (message in DCT Coefficient) for the message extracted from 500 stego images.

<i>Active</i> steganalysis method	Mean of true extracted message bits (bit/image)	Mean of false extracted message bits (bit/image)	Mean error rate
Our proposed <i>active</i> steganalysis	5521.27	1021.34	% 15.64
Modagheh et al.'s <i>active</i> steganalysis [19]	5121.96	1551.96	% 23.29
ICA-based <i>active</i> steganalysis	4785.52	1638.48	% 25.51

In Table 1, false extracted message bits are embedded message bits that their sign of detected samples is not equal to the embedded ones. Moreover, the error rate is defined as follows:

$$\text{Error rate} = \frac{\text{False extracted bits}}{\text{All embedded message bits}} \quad (17)$$

As noted earlier, in above simulations, it is supposed that the message embedding rate is known. This, however, should not be considered as a restrictive assumption because there are *passive* steganalysis methods [40], [45], [49] which can precisely estimate the message embedding rate.

As shown in Table 1, our proposed method has lower error rate and higher true detected bits compared to that of ICA-based *active* steganalysis methods.

Table 2 also shows the comparative error rate of steganalysis methods for different message embedding rates. The results confirm that our method has almost similar performance in all low embedding rates, but when the embedding rate increases, the error rate is increased gradually. It is not surprising since the embedding message was supposed to be sparse, i.e. steganography has low embedding rate.

Table 2. Mean error rate for messages extracted from 100 stego images. Message embedding rate varies from 0.02 to 0.25 bits/pixel.

Message embedding rate (bit/pixel)	0.02	0.05	0.10	0.15	0.20	0.25
Mean error rate of the proposed <i>active</i> steganalysis	11.33%	12.19%	15.64%	19.80%	25.73%	30.76%
Mean error rate of Modagheh et al.'s <i>active</i> steganalysis [19]	26.65%	24.01%	22.89%	22.29%	21.98%	21.95%
Mean error rate of the ICA-based <i>active</i> steganalysis	23.03%	24.29%	24.67%	25.35%	25.64%	26.01%

We have also drawn the comparative figure for different embedding rates (Fig. 2).

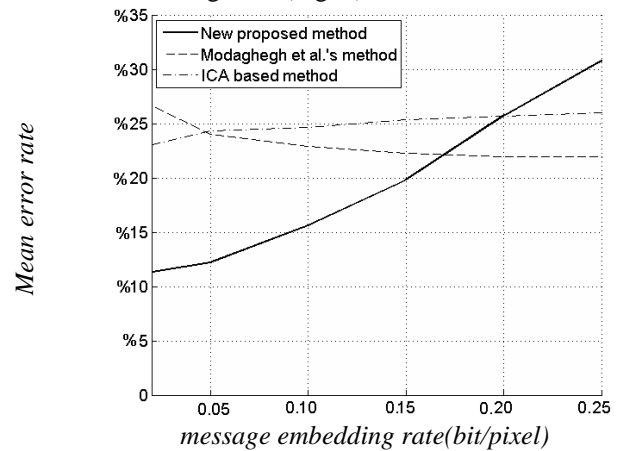


Fig. 2. Mean error rate for messages extracted from 100 stego images.

Additionally, our steganalysis method is simulated for different wavelet *transform* such as Symlet, Haar and Coiflet in Table 3. The condition of message embedding and steganalysis is similar to Table 1. The mean error rate, as shown in Table 3, is about 15% for different wavelet *transforms*.

Table 3. Mean error rate for messages extracted from 500 stego images. The message is embedded with different *transforms*.

Transform (Dictionary)	Mean error rate
Coiflet Wavelet	15.89%
Reverse Biorthogonal Wavelet	15.34%
Haar Wavelet	15.64%
Biorthogonal Wavelet	15.33%
Symlet Wavelet	15.40%

As mentioned earlier, our active ST does not need to know the *transform* domain in which the message is embedded, as it is determined by the authors. To do so, we embed the message in the DWT domain with Coiflet Wavelet, trying to extract it by another Wavelet. Then, the error rate and threshold of these Wavelets are compared (Table 4). As can be seen, the threshold of Coiflet Wavelet is greater than the other wavelets, suggesting that the message is probably embedded in this *transform*. Additionally, for two wavelets with great threshold, it can be concluded that two dictionaries are mutually coherent, so either of them can be used for extraction. For example, the results of data extracted by Meyer Wavelet correspond to Coiflet Wavelet as they are mutually coherent.

Table 4. Mean error rate and threshold for message extraction from 500 stego images. Message is embedded in Coiflet Wavelet and extracted with different dictionary.

Transform (Dictionary)	Mean error rate	Mean of threshold	Variance of threshold
Coiflet Wavelet	15.89%	3.97	0.0062
Haar Wavelet	38.77%	3.50	0.0410
Meyer Wavelet	22.87%	3.93	0.0181
Biorthogonal Wavelet	38.77%	3.5	0.0410
Symlet Wavelet	78.30%	3.2	0.0001

Finally, we compare the computational time of ICA-based *active* steganalysis methods with our proposed steganalysis method on 2.00 GHz Pentium 4 workstation. The computational time of applying steganalysis methods to stego image with different random messages have been shown in Table 5. Here, the calculated times are greater than the ones shown in [19] because the time of common parts of two methods such as denoising algorithm and DCT calculations have not been included in [19].

Table 5. the computational time of extracting message from 500 stego images by our steganalysis and ICA-based *active* steganalysis

Steganalysis method	Computational Time (second)	
	Mean	Variance
Our proposed <i>active</i> steganalysis	2.0327	0.7120
ICA-based <i>active</i> steganalysis	8.1985	1.4202

As expected, the computational time of our proposed *active* ST is lower than that of ICA-based *active* STs since our ST, does not have the denoising algorithm and works with only one image.

## References

- [1] N. F. Johnson and S. Jajodia, "Steganalysis of images created using current steganography software", in *Information Hiding*, 1998, pp. 273–289.
- [2] N. F. Johnson and S. Jajodia, "Steganalysis: The investigation of hidden information", in *Information Technology Conference*, 1998, pp. 113–116.
- [3] R. Chandramouli, "A mathematical framework for active steganalysis", *Multimedia systems*, vol. 9, no. 3, 2003, pp. 303–311.
- [4] X. Yu, T. Tan, and Y. Wang, "Reliable detection of BPCS-steganography in natural images", in *IEEE First Symposium on Multi-Agent Security and Survivability*, 2004, pp. 333–336.
- [5] A. Nissar and A. Mir, "Classification of steganalysis techniques: A study", *Digital Signal Processing*, vol. 20, no. 6, 2010, pp. 1758–1770.
- [6] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by Subtractive Pixel Adjacency Matrix", *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, 2010, pp. 215–224.
- [7] J. Fridrich and J. Kodovský, "Steganalysis of LSB replacement using parity-aware features", in *Information Hiding*, 2013, pp. 31–45.
- [8] J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive-noise modelable information hiding", in *Proceedings of SPIE*, 2003, vol. 5020, pp. 131–142.
- [9] I. Avci, N. Memon, and B. Sankur, "Steganalysis using image quality metrics", *Image Processing, IEEE Transactions on*, vol. 12, no. 2, 2003, pp. 221–229.
- [10] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics", *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 1, 2006, pp. 111–119.
- [11] Y. Yan, L. T. Li, J. B. Xue, H. G. Liu, and Q. Y. Zhang, "Study on Universal Steganalysis for BMP Images Based on Multi-Domain Features", *Applied Mechanics and Materials*, vol. 278, 2013, pp. 1906–1909.
- [12] D.-C. Lou, C.-L. Chou, T.-C. Wei, and H.-F. Huang, "Active steganalysis for interpolation-error based reversible data hiding", *Pattern Recognition Letters*, vol. 34, no. 3, 2013, pp. 1032–1036.
- [13] D.-C. Lou, C.-L. Chou, H.-K. Tso, and C.-C. Chiu, "Active steganalysis for histogram-shifting based reversible data hiding", *Optics Communications*, vol. 285, no. 10, 2012, pp. 2510–2518.
- [14] L. J. T. Guang-ming, "Active Steganalysis Based on Pixels Classification in the Image", *Journal of Electronics & Information Technology*, vol. 8, no. 34, 2012, pp. 1928–1933.
- [15] F. Fan, W. Jiazhen, L. Xiaoqin, and F. Huijuan, "An Active Steganalysis Method of Block-DCT Image Information Hiding", in *8th International Conference on Electronic Measurement and Instruments*, 2007, pp. 2–849–2–852.
- [16] B. Xu, Z. Zhang, J. Wang, and X. Liu, "Improved BSS based schemes for Active steganalysis", in *Eighth ACIS*

## 6. Conclusions

In this paper, a new *active* steganalysis method based on sparsity property of signals was proposed. Our method provided satisfactory performance on stego images in which the *cover* and hidden messages were sparse in different dictionaries. We first formulated the *active* steganalysis method as an SCA problem. Then, the feasibility of solving the SCA problem was demonstrated mathematically. Since fast *transforms* are employed in most *transform* domain steganography methods, in this study a fast algorithm was presented to solve our SCA problem.

The results of experiments showed that nearly 85% of the message bits could be estimated when the sparsity factor of message was 10%. Additionally, experiments confirmed that the computational cost of our method was approximately one fourth of the previous ICA-based *active* STs.

Overall, the comparison between our proposed method and the previous *active* steganalysis schemes revealed that the use of sparsity property of signals improved the steganalysis performance (computational cost and error rate).

- International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007, vol. 3, pp. 815–818.
- [17] L. Wenzhe, X. Bo, Z. Zhe, and R. Wenxia, "Active steganalysis with only one stego image", in Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009, vol. 5, pp. 345–348.
- [18] A. Ambalavanan and R. Chandramouli, "Blind source separation for steganalytic secret message estimation", in Proceedings of SPIE, 2007, vol. 6505, p. 650507.
- [19] H. Modaghegh and S. A. Seyedin, "A new fast and efficient active steganalysis based on combined geometrical blind source separation", *Multimedia Tools and Applications*, pp. 1–19, 2014.
- [20] C. Jutten, "Independent components analysis versus principal components analysis", in Fourth European Signal Processing Conference, 1988, pp. 643–646.
- [21] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models", *Signal Processing, IEEE Transactions on*, vol. 46, no. 4, 1998, pp. 886–902.
- [22] L. Vielva, D. Erdogmus, and J. C. Principe, "Underdetermined blind source separation using a probabilistic source sparsity model", in Proceeding of International Conference on Independent Component Analysis and Blind Source Separation (ICA), 2001, pp. 675–679.
- [23] C. Jutten and M. Babaie-Zadeh, "Source separation: Principles, current advances and applications", IAR Annual Meeting Nancy France, 2006, pp. 1–10.
- [24] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges", in 14th European Symposium on Artificial Neural Networks, 2006, pp. 323–330.
- [25] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition", *Information Theory, IEEE Transactions on*, vol. 47, no. 7, 2001, pp. 2845–2862.
- [26] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit", *SIAM journal on scientific computing*, vol. 20, no. 1, 1998, pp. 33–61.
- [27] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*, Springer, 2008.
- [28] J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. L. Donoho, "Morphological component analysis", in Optics & Photonics Conference, 2005, p. 59140Q–59140Q.
- [29] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: An adaptive thresholding strategy", *Image Processes sing IEEE Transactions On*, vol. 16, no. 11, 2007, pp. 2675–2681.
- [30] P. Liu, C. Chen, L. Ge, and Y. Luo, "Efficient Self-Adaptive Image Steganography Scheme Based on Iterative Blending and Integer Wavelet Transform", in *Unifying Electrical Engineering and Electronics Engineering*, Springer, 2014, pp. 1159–1167.
- [31] X. Wang, C. Wang, H. Yang, and P. Niu, "A robust blind color image watermarking in quaternion Fourier transform domain", *Journal of Systems and Software*, vol. 86, no. 2, 2013, pp. 255–277.
- [32] X.-Y. Wang, A.-L. Wang, H.-Y. Yang, Y. Zhang, and C.-P. Wang, "A new robust digital watermarking based on exponent moments invariants in nonsampled contourlet transform domain", *Computers & Electrical Engineering*, vol. 40, no. 3, Apr. 2014, pp. 942–955.
- [33] M. Jayamohan and K. Revathy, "A Hybrid Fractal-Wavelet Digital Watermarking Technique with Localized Embedding Strength", in *Wireless Networks and Computational Intelligence*, Springer, 2012, pp. 584–591.
- [34] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure", *Multimedia, IEEE Transactions on*, vol. 3, no. 3, 2001, pp. 273–284.
- [35] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization", in *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, 2003, pp. 2197–2202.
- [36] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases", *Information Theory, IEEE Transactions on*, vol. 48, no. 9, 2002, pp. 2558–2567.
- [37] D. L. Donoho, "Compressed sensing", *Information Theory, IEEE Transactions on*, vol. 52, no. 4, 2006, pp. 1289–1306.
- [38] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising", *Journal of computational and graphical statistics*, vol. 9, no. 2, 2000, pp. 361–379.
- [39] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation", *Image Processing, IEEE Transactions on*, vol. 16, no. 11, 2007, pp. 2662–2674.
- [40] X. Yu, Y. Wang, and T. Tan, "On estimation of secret message length in JSteg-like steganography", in *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, vol. 4, pp. 673–676.
- [41] T. Holotyak, J. Fridrich, and D. Soukal, "Stochastic approach to secret message length estimation in  $\ell_1$  embedding steganography", in *Proceeding of SPIE*, vol. 5681, 2005, pp. 673–684.
- [42] M. Jiang, E. Wong, N. Memon, and X. Wu, "A simple technique for estimating message lengths for additive noise steganography", in *Control, Automation, Robotics and Vision Conference*, 2004, vol. 2, pp. 983–986.
- [43] J. Fridrich and M. Goljan, "On estimation of secret message length in LSB steganography in spatial domain", in *Electronic Imaging*, 2004, pp. 23–34.
- [44] J. Fridrich, M. Goljan, D. Hoge, and D. Soukal, "Quantitative steganalysis of digital images: estimating the secret message length", *Multimedia Systems*, vol. 9, no. 3, 2003, pp. 288–302.
- [45] T. Zhang and X. Ping, "A fast and effective steganalytic technique against JSteg-like algorithms", in *Proceedings of the 2003 ACM symposium on Applied computing*, 2003, pp. 307–311.
- [46] P. Abrial, Y. Moudden, J.-L. Starck, B. Afeyan, J. Bobin, J. Fadili, and M. K. Nguyen, "Morphological component analysis and inpainting on the sphere: Application in physics and astrophysics", *Journal of Fourier Analysis and Applications*, vol. 13, no. 6, 2007, pp. 729–748.
- [47] "UC Berkeley Computer Vision Group - Contour Detection and Image Segmentation - Resources". [Online]. Available: <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html#bsds500>. [Accessed: 31-Mar-2013].
- [48] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain", *Image Processing IEEE Transactions On*, vol. 12, no. 11, 2003, pp. 1338–1351.
- [49] J. Fridrich, M. Goljan, and D. Hoge, "Steganalysis of JPEG images: Breaking the F5 algorithm", in *Information Hiding*, 2003, pp. 310–323.

**Hamed Modagheh** was born in Mashhad, Iran. In 2005 he received the B.Sc in Electrical Engineering from Iran University of Science and Technology, the M.Sc in Communication Engineering from Sharif University of Technology in 2007 and the Ph.D. degree in Communication from Ferdowsi University of Mashhad in 2015. His current research interests include signal processing and data hiding.

**Seyed Alireza Seyedin** was born in Iran. He received the B.S degree in Electronics Engineering from Isfahan University of Technology, Isfahan, Iran in 1986, and the M.E degree in Control and Guidance Engineering from Roorkee University, Roorkee, India in 1992, and the Ph.D degree from the University of New South Wales, Sydney, Australia in 1996. He has been an Associate Professor with the Department of Electrical Engineering, the University of Mashhad (Ferdowsi), Mashhad, Iran.

# A Robust Statistical Color Edge Detection for Noisy Images

Mina Alibeigi\*

Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran  
minaalibeigi@gmail.com

Niloofar Mozafari

Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran  
mozafari@cse.shirazu.ac.ir

Zohreh Azimifar

Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran  
azimifar@cse.shirazu.ac.ir

Mahnaz Mahmoodian

Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran  
m.mahmoodian@ut.ac.ir

Received: 02/July/2014

Revised: 27/Jan/2015

Accepted: 25/Feb/2015

## Abstract

Edge detection plays a significant role in image processing and performance of high-level tasks such as image segmentation and object recognition depends on its efficiency. It is clear that accurate edge map generation is more difficult when images are corrupted with noise. Moreover, most of edge detection methods have parameters which must be set manually. Here we propose a new color edge detector based on a statistical test, which is robust to noise. Also, the parameters of this method will be set automatically based on image content. To show the effectiveness of the proposed method, four state-of-the-art edge detectors are implemented and the results are compared. Experimental results on five of the most well-known edge detection benchmarks show that the proposed method is robust to noise. The performance of our method for lower levels of noise is very comparable to the existing approaches, whose performances highly depend on their parameter tuning stage. However, for higher levels of noise, the observed results significantly highlight the superiority of the proposed method over the existing edge detection methods, both quantitatively and qualitatively.

**Keywords:** Edge Detection; Color; Noisy Image; RRO Test; Regression; Pratt's Figure of Merit.

## 1. Introduction

One of the most important processes in low-level image processing is edge detection, and the performance of high-level image processing tasks is highly dependent on this process.

Novak and Shafer [1] claim that at most 90% of the information in color images is similar to that of gray images, which means there still remains information that may not be detected in gray images. On the other hand, in most applications, images are often corrupted by noise. Generation of accurate edge map becomes more critical and complicated in the presence of noise. Thus, the detector must be robust against noise.

In this paper, we present a robust color edge detector for automatic edge detection in the presence of additive noise. The main advantage of our detector is that the parameters are set automatically based on image content. The proposed approach adopts a statistical test called Robust Rank-Order (RRO) test [2] to detect edges. In order to detect fewer false edges in the images with higher levels of noise, lower significance levels are employed with the RRO test. Unlike the existing methods, the proposed algorithm adapts its parameters based on image content automatically. Experimental results for lower levels of noise show that the proposed edge

detector is comparable to the existing approaches; for higher levels of noise, the proposed method performs significantly better, both quantitatively and qualitatively.

The rest of this paper is organized as follows: Section 2 reviews the related works. In Section 3 the proposed algorithm for edge detection is explained. The experimental results are given in Section 4 and Section 5 presents conclusion and future work.

## 2. Related Works

The problem of edge detection saw its pioneering work at least as early as 1986 by the work of Canny [3]. Since then, various edge detection algorithms have been developed [4,5,6,7,8,9, and 10]. Most of these algorithms are based on computation of the intensity gradient vector, which are, in general, sensitive to noise. In order to reduce this problem, some spatial averaging is combined with differentiation process, such as the Laplacian of Gaussian operator and zero-crossing detection [5]. As stated before, one of the well-known methods of this category is Canny edge detector [3] which finds edges by looking for local maxima of the gradient of image. The gradient is calculated using the derivative of a Gaussian filter. Canny edge detector includes the weak edges in the

\* Corresponding Author



output only if they are connected to strong edges. This method is therefore less likely to be fooled by noise, and more likely to detect true weak edges than those detectors named above.

Also in recent years, many soft computing algorithms were proposed for edge detection [11,12,13,14,15, and 16]. In most of these algorithms, the problem of edge detection is seen as an optimization problem. For example, in [12, 13, and 14] bio-inspired optimization algorithms are used to address the edge detection problem. In [14] the authors studied the effect of different topologies in PSO-based edge detection techniques. These algorithms are more robust against noise compared to gradient based methods. However, they are much more time consuming. Also, their performance depends on implementation details and parameters settings. Therefore, their results might not be easily reproducible.

Statistical edge detection is an alternative to deterministic edge detection. Bovik et al. [17] proposed several nonparametric statistical tests for edge detection in noisy images. However, their work can only detect edges in four directions and no quantitative comparison was reported for this method. Variance tests were also applied for edge detection by Aron and Kruz [18] and Kruz and Benteftifa [19]. Recently, Lim and Jang [20] have compared three statistical two-sample tests for edge detection in noisy images. These edge operators are based on a test statistic, computed using the modified grey level values. In another study, Lim [2] proposed a new operator based on RRO test for detecting edges in all possible orientations in noisy images. The RRO detector has been designed for grey images and requires two user defined parameters: the edge height ( $\delta$ ) and the significance level ( $\alpha$ ) of statistical test, which is often set to 0.05. Since different values of these parameters may lead to different results, adaptive specification of these parameters based on image content could be interesting.

As mentioned before, color images contain more information than grey ones. In order to benefit from all the information present in color images, proposing a color edge detection method becomes indispensable. Due to importance of color information, many color edge detection methods were proposed [21, 22, 23, and 24]. Russo and Lazzari [23] proposed edge detection in color images corrupted by Gaussian noise, which adopts a multi-pass preprocessing stage that gradually reduces the amount of noise in each R, G and B channels of image, independently.

Since statistical approaches to edge detection are more robust in the presence of noise, in this study we propose a statistical color edge detector which is completely adaptive and operates robustly in the presence of Gaussian noise, the most common type of noise encountered during image acquisition [25].

### 3. The Proposed Detector: Robust Color Edge Detector

The proposed algorithm for color edge detection includes three steps: noise estimation, prefiltering, and edge detection (see Figure 1). The details of these steps are discussed in the next subsections.

#### 3.1 Noise Estimation

In the first step, noise estimation, the standard deviation of noise must be estimated. There exist several well-known methods to estimate the level of noise [23,26]. In this paper we use median noise estimation method proposed by Pizurica et al. [27], due to its efficiency and better accuracy compared to other methods. In median noise estimation, the median of wavelet coefficients in high-high band of finest scale is used to calculate the standard deviation of noise according to the following formula:

$$\sigma_n = \text{Median}(|\lambda_{h,h,1}|)/0.6745 \quad (1)$$

Where  $\lambda_{h,h,1}$  is the wavelet coefficient in high-high band of the first scale of noisy image.

#### 3.2 Prefiltering

In the second step, we adopt a prefiltering to reduce the effect of noise and outliers. Let us define noisy pixels as pixels that are corrupted by noise with amplitude not too different from their neighbors, and outliers as pixels that are corrupted by noise with amplitude much larger than their neighbors [23]. We consider two types of filters to reduce the effect of these pixels. Figure 1 illustrates the block diagram of these filters. Type A filter is to remove noisy pixels. It computes the difference of each pixel with a predefined function of its neighbors as follows:

1. Small difference between the current pixel and its neighbors is due to noise and must be decreased.
2. Large difference between the current pixel and its neighbors is due to edge and must be preserved.

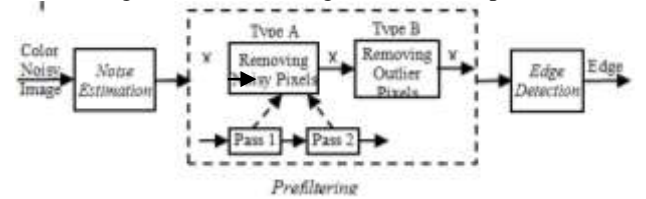


Fig. 1. The framework of the proposed edge detector.

We adopt a two-pass procedure to increase the effectiveness of the smoothing action [23]. This procedure is defined according to following equations:

$$x_{pass}^{(c)}(i, j) = x_{in}^{(c)}(i, j) + \frac{1}{8} \sum_{m=-1}^1 \sum_{n=-1}^1 \zeta_1^{(c)}(x_{in}^{(c)}(i+m, j+n), x_{in}^{(c)}(i, j)), (m, n) \neq (0, 0) \quad (2)$$

$$x_{A2B}^{(c)}(i, j) = x_{pass}^{(c)}(i, j) + \frac{1}{8} \sum_{m=-1}^1 \sum_{n=-1}^1 \zeta_n^{(c)}(x_{pass}^{(c)}(i + m, j + n), x_{pass}^{(c)}(i, j)), (m, n) \neq (0, 0) \quad (3)$$

Where,  $c = 1, 2$  or  $3$  indicates the channel number for R, G and B channels, respectively.  $\zeta_n(c)$  is a parameterized nonlinear function and is defined as:

$$\zeta_n^{(c)}(u, v) = \begin{cases} u - v & |u - v| \leq a_n^{(c)} \\ ((3a_n^{(c)} - u + v)/2)sgm(u - v) & a_n^{(c)} < |u - v| \leq 3a_n^{(c)} \\ 0 & |u - v| > 3a_n^{(c)} \end{cases} \quad (4)$$

Where  $a_n^{(c)}$  is an integer bounded at  $0 < a_n^{(c)} < 256$ , (for  $c = 1, 2, 3$ ) [26].

According to (2), the smoothing filter is first applied to each channel of the color image  $x_{in}^{(1)}$ ,  $x_{in}^{(2)}$  and  $x_{in}^{(3)}$  to produce intermediate results  $x_{pass}^{(1)}$ ,  $x_{pass}^{(2)}$  and  $x_{pass}^{(3)}$ . Then the second filtering pass is applied on these intermediate components and yields to  $x_{A2B}^{(1)}$ ,  $x_{A2B}^{(2)}$  and  $x_{A2B}^{(3)}$ , respectively.

As Figure 1 shows, the type B prefiltering is meant to remove outlier pixels, for which the difference between the processed pixel and all its neighbors is very large. Equation (5) illustrates this prefiltering:

$$x_{out}^{(c)}(i, j) = x_{A2B}^{(c)}(i, j) - (255)\Delta(i, j) \quad (5)$$

Where

$$\Delta(i, j) = MIN \left\{ \mu_{LA} \left( x_{A2B}^{(c)}(i, j), x_{A2B}^{(c)}(i + m, j + n) \right) \right\} - MIN \left\{ \mu_{LA} \left( x_{A2B}^{(c)}(i + m, j + n), x_{A2B}^{(c)}(i, j) \right) \right\} \quad (6)$$

$\mu_{LA}(u, v)$  is the membership function describing the relation of “ $u$  is much larger than  $v$ ”.

$$\mu_{LA}(u, v) = \begin{cases} \frac{u-v}{L-1} & 0 < u - v \leq L - 1 \\ 0 & u - v \leq 0 \end{cases} \quad (7)$$

Here  $L$  is the maximum possible intensity value.

After applying these filters on R, G and B channels independently, a simple averaging is performed.

### 3.3 Edge detection

In third step, to detect edge pixels we consider a small square subimage of size  $5 \times 5$  centered at each pixel. We divide the neighboring pixels centered by this square mask into two groups of size 12, as shown in Figure 2. In other words, we consider a set of  $N = m + n$  independent observations, excluding the center pixel, which are divided into  $X = (X_1, X_2, \dots, X_n)$  and  $Y = (Y_1, Y_2, \dots, Y_m)$ . It is assumed that the first set of observations comes from a continuous distribution named  $F(x - \mu_x)$  and the second one comes from another continuous distribution  $G(y - \mu_y)$ , with  $\mu_x$  and  $\mu_y$  as shift parameters. Note that the two distributions  $F$  and  $G$  are not identical. In order to create our statistical model, we define two modified sets of observations  $\{A_i\}$  and  $\{B_i\}$  as follows:

$$A_i = \begin{cases} X_i + \delta & X_i \in X \\ Y_i & Y_i \in Y \end{cases} \quad B_i = \begin{cases} X_i - \delta & X_i \in X \\ Y_i & Y_i \in Y \end{cases} \quad (8)$$

Where  $\delta$  is edge height parameter. This parameter determines the minimum grey level difference which leads to labeling the center pixel as an edge pixel, according to the following statistical test:

$$H_0^\uparrow : \mu_x + \delta \geq \mu_y \quad \text{versus} \quad H_1^\uparrow : \mu_x + \delta < \mu_y$$

$$H_0^\downarrow : \mu_x - \delta \leq \mu_y \quad \text{versus} \quad H_1^\downarrow : \mu_x - \delta > \mu_y$$

Since the distributions  $F$  and  $G$  are not identical, theoretically statistical tests such as Wilcoxon test [17] may not be appropriate [2]. Thus, in this study for testing  $H_0^\uparrow$  (or  $H_0^\downarrow$ ) against  $H_1^\uparrow$  (or  $H_1^\downarrow$ ) on  $\{A_i\}$  (or  $\{B_i\}$ ), a statistical test named RRO test [2, 18] is used.

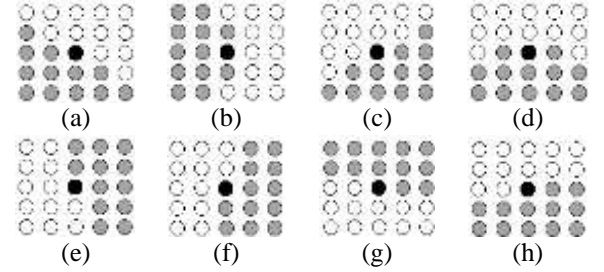


Fig. 2. Partitioning the window in eight different ways where the grey areas represent  $X$  and the white areas represent  $Y$ .

For each  $X_i + \delta$ ,  $X_i \in X$  in  $\{A_i\}$ , we count the number of lower-valued observations  $Y_i$ 's in  $Y$ . This number is denoted as  $U(Y, X_i + \delta)$ . Then the mean value of the  $U(Y, X_i + \delta)$  values is calculated using equation (9). Similarly, we find the number of lower-valued observations  $X_i + \delta$ 's in  $X$  for each  $Y_i$  and will denote this counting by  $U(X + \delta, Y_i)$ . Then the mean value,  $\overline{U(X + \delta, Y)}$ , is evaluated via equation (10), as:

$$\overline{U(Y, X + \delta)} = \frac{1}{m} \sum_{i=1}^m U(Y, X_i + \delta) \quad (9)$$

$$\overline{U(X + \delta, Y)} = \frac{1}{n} \sum_{i=1}^n U(X + \delta, Y_i) \quad (10)$$

Next, two variables are defined to demonstrate the variability of  $U(X + \delta, Y)$  and  $U(Y, X + \delta)$  as  $V_{x+\delta}$  and  $V_{y+\delta}$  according to the following formulas:

$$V_{x+\delta} = \sum_{i=1}^m [U(Y, X_i + \delta) - \overline{U(Y, X + \delta)}]^2 \quad (11)$$

$$V_{y+\delta} = \sum_{i=1}^n [U(X + \delta, Y_i) - \overline{U(X + \delta, Y)}]^2$$

Finally, the test statistics for observations  $\{A_i\}$  is given by:

$$U_A = \frac{n \cdot \overline{U(X + \delta, Y)} - m \cdot \overline{U(Y, X + \delta)}}{2 \sqrt{V_{x+\delta} + V_{y+\delta} + \overline{U(Y, X + \delta)} \cdot \overline{U(X + \delta, Y)}}} \quad (12)$$

For  $\{B_i\}$  observations, the values  $U(Y, X - \delta)$ ,  $U(X - \delta, Y)$ ,  $V_{x-\delta}$  and  $V_{y-\delta}$  are defined exactly analogous to those defined for  $\{A_i\}$  observations:

$$\overline{U(Y, X - \delta)} = \frac{1}{m} \sum_{i=1}^m U(Y, X_i - \delta)$$

$$\begin{aligned}\overline{U(X-\delta, Y)} &= \frac{1}{n} \sum_{i=1}^n U(X-\delta, Y_i) \\ V_{x-\delta} &= \sum_{i=1}^m [U(Y, X_i - \delta) - \overline{U(Y, X - \delta)}]^2 \\ V_{y-\delta} &= \sum_{i=1}^n [U(X - \delta, Y_i) - \overline{U(X - \delta, Y)}]^2\end{aligned}$$

Thus the test statistic will be given by:

$$U_B = \frac{m \cdot \overline{U(Y, X - \delta)} - n \cdot \overline{U(X - \delta, Y)}}{2 \sqrt{V_{x-\delta} + V_{y-\delta} + \overline{U(Y, X - \delta)} \cdot \overline{U(X - \delta, Y)}}} \quad (13)$$

Having determined  $U_A$  and  $U_B$ , we have defined  $U^* = \max(U_A, U_B)$ .

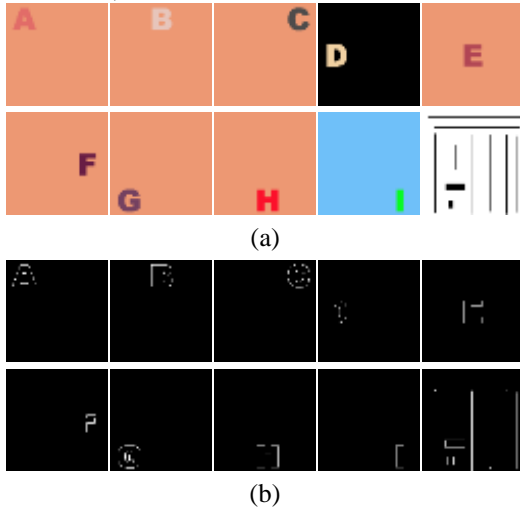


Fig. 3. (a) The synthetic images used for training parameters of RRO test  
(b) Edge maps for the above synthetic images.

Accordingly,  $H_0^1$  (or  $H_0^2$ ) will be rejected for large values of  $U^*$  [2]. If  $U^* > u_\alpha$ , for a specified threshold  $u_\alpha$  at a significance level  $\alpha$ , then an edge is detected. Flinger and Pollicello [28] present the critical values of  $U^*$  for small sample sizes up to 12. This threshold is constant in RRO detector with a pre-specified significance level which is usually set to 0.05, for all different levels of noises presents in images.

RRO test has been utilized for edge detection [2] and appropriate results have been reported. We, however, argue that for higher levels of noise, it is more desirable to detect only stronger edges in order to reduce the effects of false edges; therefore, we adjust the threshold in RRO test at different significance levels based on the standard deviation of the Gaussian noise. With higher levels of noise, we will detect edges only with higher confidence. In order to make the test completely adaptive, we also must set the edge height parameters of RRO test. The next subsection presents the procedure of adaptive tuning of these parameters.

### 3.4 Adaptive Tuning of Parameters

In order to drive the parameters, significance level ( $\alpha$ ) and edge height ( $\delta$ ), adaptively, we used a set of synthetic images as training samples. These training samples are generated artificially to cover a wide range of colors and contrasts. We used alphabets (A, B, C, D, G, H, I), lines

and squares in these images to cover a wide range of edges in different orientations which may appear frequently in a real image. Figure 3 shows the synthetic images and their corresponding true edge maps used in this study, all of the size  $256 \times 256$  pixels.

In definition, edge divides two regions. In creating true edge maps for these images we follow the idea that when moving from one region to the next, the bordering pixels we first meet in the first region are edge pixels (outer edges) but the bordering pixels in the new region (inner edges) are not considered as edge pixels.

These training images (Figure 3) are used to derive the functions which determine the optimal values of  $\delta$  and  $u_\alpha$  for various levels of noise. To find the optimal values of  $\delta$  and  $u_\alpha$ , a number of experiments was performed on these images by adding Gaussian noise with different standard deviations ( $\sigma_n = 0, 10, 20, 30, 40, 50$  and  $60$ ). After finding the edge maps of these images by our proposed edge detector for all possible values of  $\delta$  and  $u_\alpha$ , the error defined in equation (14) is computed for each value. Then for each particular noise level, those values of  $\delta$  and  $u_\alpha$  with lowest error value, will be selected to be applied in RRO test.

Error

$$= \sum_{i \in \text{Sample images}} (4 * (\text{number of missed edge points})_i + (\text{number of new faulty edge points})_i) \quad (14)$$

In (14) the missed edges (pixels which are present in true edge map but not in the edge map provided by the proposed edge detector) are 4 times more penalized than the new faulty edge points (pixels which are not in the true edge map of the image but present in the edge map provided by the proposed edge detector). This is the rational that why our edge detector is robust to noise and will remove noisy pixels from edge map in highly corrupted images, though it may lead to missing some weaker edge points, like other existing edge detectors. However, existing detectors may also detect noisy pixels as edge pixels.

Table 1 gives the optimal values calculated for  $\delta$  and  $u_\alpha$  with different levels of additive Gaussian noise. Parameters  $\delta_m$  and  $\alpha_m$  are used to refer to optimal values. As it can be seen in Table 1, these parameters are functions of the noise standard deviation ( $\sigma_n$ ). Thus, by knowing the standard deviation of the noise, the optimal threshold values can be determined according to estimation formulas for  $\tilde{\delta}_m$  and  $\tilde{\alpha}_m$ , which are defined as:

Table 1. Optimal values founded for each of the parameters,  $\delta$  and  $u_\alpha$

Standard Deviation ( $\sigma_n$ )	Edge height ( $\delta$ )	Threshold ( $u$ )
0	9	6
10	11	1
20	13	1
30	17	1
40	14	3
50	19	3
60	15	10

$$\tilde{\delta}_m(\sigma_n) = \text{round}(0.1250 * \sigma_n + 10.25) \quad (15)$$

$$\tilde{\alpha}_m(\sigma_n) = \max(1.704, (0.0058 + 0.0599 * \sigma_n)) \quad (16)$$

Where  $\sigma_n$  is the estimated standard deviation of noise computed according to (1).

In order to estimate the continuous functions, all of the optimal values at different noise levels are used as sampling points to find the relationships between both  $\delta_m$  and  $\alpha_m$  with  $\sigma_n$  using regression methods, as shown in Figures 4(a) and 4(b). With enough sample points, both functions can be estimated effectively.

We use two first-order polynomials to estimate  $\tilde{\delta}_m$  and  $\tilde{\alpha}_m$  according to (15) and (16). We use robust regression [29], which has the minimum mean-square error (LMMSE) [21] in true values of sample points in comparison with other regression methods considered.

It must be noted that for proper edge detection, according to [28] the lowest value of significance level is allowed to be 0.05 which leads to  $u_\alpha = 1.704$ . Thus the final value for  $u_\alpha$  will be the maximum of 1.704 and the one estimated via (16). In other words, in highly corrupted images, only those pixels are labeled edge that we obtain at least 95% confidence about them.

Once the estimated functions are calculated, given an image corrupted by Gaussian noise, the standard deviation of the noise is estimated according to (1). Then, using the estimated functions which are presented in Equations (15) and (16), the proper threshold values are calculated.

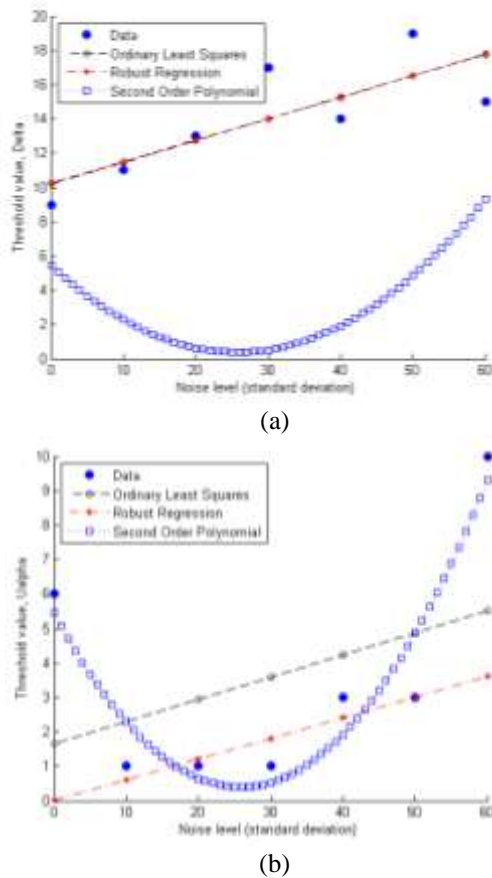


Fig. 4. Regression plot for (a)  $\delta$  and (b)  $u_\alpha$ .

## 4. Experimental Results and Discussions

In this section, the performance of our proposed method is compared with five successful detectors, particularly RRO detector [2] which uses the same statistical test as we considered in our study. Since Lim [2] has reported better performance for RRO detector in comparison with two other existing statistical detectors, T detector [20] and Wilcoxon detector [17,20], we only compared our algorithm with RRO detector. We use the  $5 \times 5$  difference-of-boxes as was shown in Figure 2 for our method as well as for RRO detector. Moreover, we compare our method with well-known edge detectors, Canny [3] and Sobel [5] as well as the recent color edge detector proposed by Russo and Lazzari [23]. Both synthetic and natural images are used in these experiments. The synthetic images allow the use of Pratt's figure of merit [30] (PFOM) as a quantitative evaluation measure, while the natural images might be more feasible for real applications.

We used a synthetic ideal step image that makes the use of PFOM evaluation measure possible; a square synthetic image of the size  $256 \times 256$  with two vertical and two horizontal edges, as shown in Figure 5. To evaluate the performance of the proposed detector in noisy images, seven different levels of noise (zero-mean Gaussian noise with standard deviations 0, 10, ..., 60) are added to the synthetic image. Some of these images are shown in Figure 5 (first column).

Figure 5 also shows the results of applying four different edge detectors to these synthetic images. The results illustrate that the proposed detector performs better in highly corrupted images. In absence of noise, it produces thicker edges compared to Canny and Sobel detectors, the filters that perform poorly in the presence of noise.

It is also observed that our edge detector produces less speckle noise in comparison with RRO detector [2]. From Figure 5, it is also clear that our algorithm performs better than the method of Russo and Lazzari [23], when tested with all noise levels. Our filter not only produces thinner edges but also removes noise and creates much less artifacts.

We adopt the following definition [2] to compute PFOM:

$$R = \frac{1}{\max(I_I, I_A)} \sum_{i=1}^{I_A} \frac{1}{1 + \beta e(i)^2} \quad (17)$$

where  $I_I$  and  $I_A$  are the number of ideal and actual edge points, respectively,  $e(i)$  is the pixel miss distance of the  $i$ th pixel which was detected as edge, and  $\beta = \frac{1}{9}$  [31] is a scaling constant.

Table 2 illustrates PFOM values of all five edge detectors for a variety of Gaussian noise distributions. We reported the PFOMs of RRO detector for different values of  $\delta = 10, 15$  and  $20$ .

The PFOM values show that without any noise, the performance of our detector is the same as Russo and Lazzari's and better than Canny and Sobel detectors, while it is worse than RRO's for all three height parameter values. However, it can be seen that as the standard deviation of noise increases, the performance of our detector improves.

In another experiment, we considered five different images (Lenna, Peppers, F16, Building and Baboon) which are commonly used in literature. Figures 6-10 show the results of applying the above edge detectors on these images for different noise levels. In all of these experiments, the edge height parameter of the RRO detector was set to  $\delta = 15$ .

As the results indicate, for original noiseless images, Canny detector is the best choice due to its thin and connected results. Considering the quality of detected edges, the second best detector is Sobel; while its performance in terms of detecting connected edges, is quantitatively lower than all other detectors. For original noiseless images, the performance of RRO, Russo and Lazzari's and our detector are not distinguishable.

For higher levels of noise, Canny detector nearly fails; as most of the present noisy pixels are detected as edge pixels. Sobel will also fail, but in a different way; by not detecting most of the true edges. However, the other three edge detectors can find most of the true edges in these situations. Albeit RRO and Russo and Lazzari's detectors will detect most of the noisy pixels as edge pixels as well, and this problem will be emphasized for higher levels of noise. On the contrary, our proposed method performs well on both original and severely noisy images. Our algorithm classifies fewer noisy pixels as edge pixels and focuses on detecting the edges for which we obtain more confidence according to (16).

Table 2. The measured values of PFOM for the edge detectors

Noise STD	Edge Detectors					
	Canny	Sobel	Russo and Lazzari's detector	$\Delta$	RRO	The Proposed
0	0.36869	0.37154	0.70708	10 <b>0.87139</b> 15 <b>0.87139</b> 20 <b>0.87139</b>	0.7070	
20	0.74238	0.38126	0.74653	10 0.80045 15 0.83835 20 0.86129	<b>0.87042</b>	
40	0.74196	0.62123	0.75389	10 0.76462 15 0.79103 20 0.81360	<b>0.83688</b>	
60	0.74120	0.80033	0.69580	10 0.74559 15 0.77764 20 0.79784	<b>0.80394</b>	

### 5. Conclusion and Future Work

We proposed a new edge detector based on the RRO test, which is a useful alternative to the Wilcoxon test. In order to detect edges in noisy images, we used a multi-pass prefiltering stage to reduce the effect of additive Gaussian noise in each channel of a color image independently. Afterwards, we applied RRO test using a window of size  $r \times r$  to detect edges in all possible orientations. Since the RRO detector is not adaptive and it needs user specified parameters for detecting edges, we purposed to make this detector adaptive in a systematic approach based on image content.

We investigated the performance of our proposed detector in comparison with well-known Canny and Sobel edge detectors as well as RRO and Russo and Lazzari's detectors for both synthetic and natural images. From the experimental results, it was observed that in highly corrupted images, the proposed edge detector performs much better than other edge detectors both quantitatively and qualitatively in a completely adaptive manner.

For future works, we can modify the RRO test statistic to handle R, G and B channels, simultaneity. Furthermore, it is interesting to introduce a prefiltering which is able to reduce the effect of different types of noise, such as impulsive noise. However, the current proposed detector can handle all types of noisy images but its performance is significantly better for Gaussian noise. Also, we intend to investigate the performance of proposed method in HIS color space. Our preliminary results on HIS color space are promising.

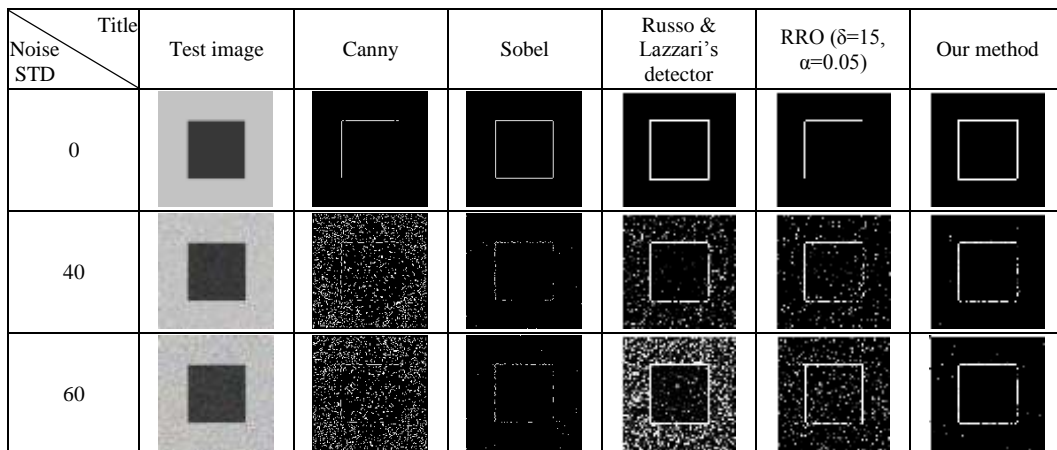


Fig. 5. Edge detector results for square synthetic images.
















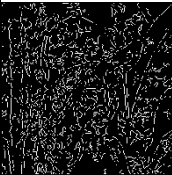





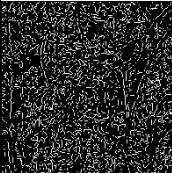


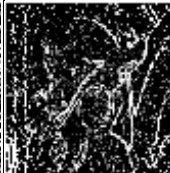

Noise STD	Lenna	Canny	Sobel	Russo&Lazzari's	RRO detector	Our method
0						
20						
40						
60						

Fig. 6. Edge detection results for Lenna image.

























Noise STD	Peppers	Canny	Sobel	Russo&Lazzari's	RRO detector	Our method
0						
20						
40						
60						

Fig. 7. Edge detection results for Peppers image.

Noise STD	F16	Canny	Sobel	Russo&Lazzari's	RRO detector	Our method
0						
20						
40						
60						

Fig. 8. Edge detection results for F16 image.

Noise STD	Building	Canny	Sobel	Russo&Lazzari's	RRO detector	Our method
0						
20						
40						
60						

Fig. 9. Edge detection results for Building image.



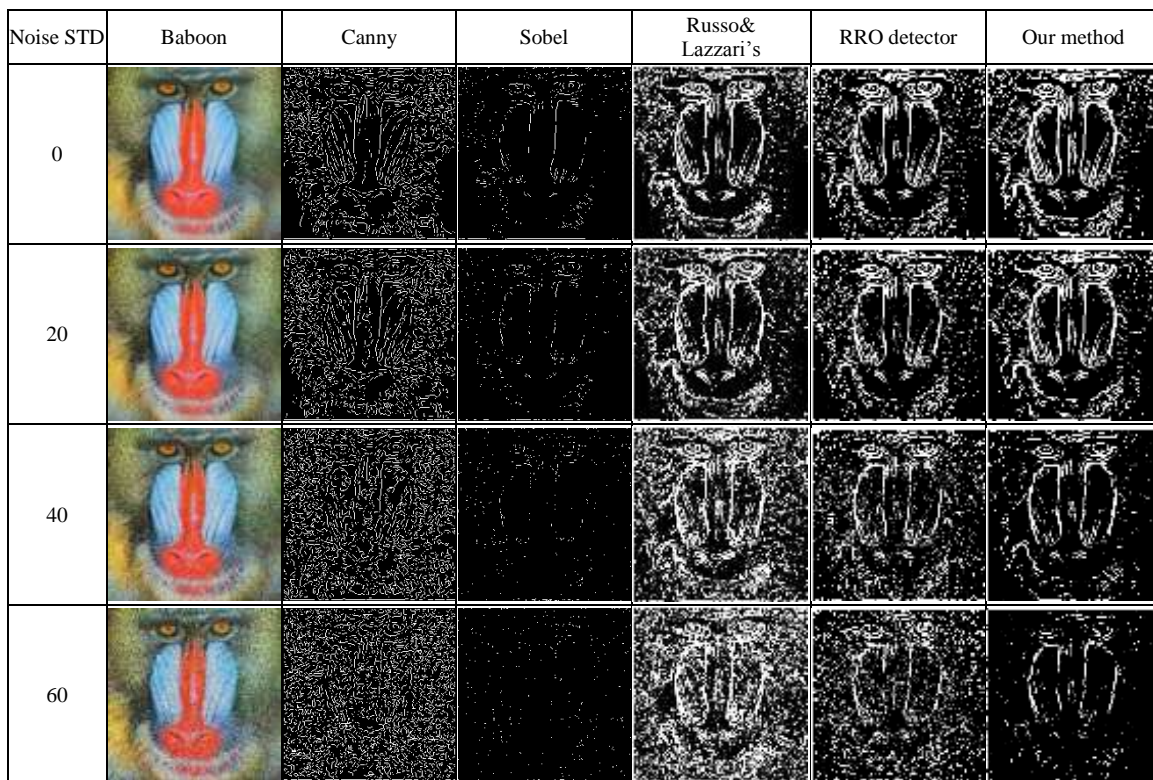


Fig. 10. Edge detection results for Baboon image.

## References

- [1] C. L. Novak, and S. A. Shafer, "Color edge detection", In proc. of DARPA image understanding workshop 1, 1987, pp. 35-37.
- [2] D. H. Lim, "Robust edge detection in noisy images", Computational Statistics and Data Analyses, Vol. 50, No. 3, 2006, pp. 803-812.
- [3] J. Canny, "A computational approach to edge detection", IEEE Trans. Pattern Anal. Machine Intell. PAMI, vol. 8, No. 6, 1986, pp. 679-698.
- [4] M. Sharifi, and M. Fathy, and M. T. Mahmoudi, "A classified and comparative study of edge detection algorithms", in Proceedings of International Conference on Information Technology: Coding and Computing, 2002, pp. 117-120.
- [5] R. C. Gonzalez, and R.E. Woods, Digital Image Processing, third ed., Addison-Wesley, NewYork, 1992.
- [6] A. Rosenfeld, and A. Kak, Digital Picture Processing, Second ed., Academic Press, New York, 1982.
- [7] M. D. Heath, and S. Sarkar, and T. A. Sanocki, and K. W. Bowyer, "Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 19, No. 12, 1997, pp. 1338-1359.
- [8] Q. Ying-Dong, and C. Cheng-Song, and C. San-Ben, and L. Jin-Quan, "A fast subpixel edge detection method using Sobel-Zernike moments operator", Image and Vision Comput., Vol. 23, No. 1, 2005, pp. 11-17.
- [9] M. Basu, "Gaussian-Based Edge-Detection Methods—A Survey", IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 32, No. 3, August 2002.
- [10] E. Nadernejad, and S. Sharifzadeh, and H. Hassanpour, "Edge detection techniques: evaluations and comparisons". Applied Mathematical Sciences, Vol. 2, No. 31, 2008, pp. 1507-1520.
- [11] N. Senthilkumaran, and R. Rajesh, "Edge detection techniques for image segmentation - a survey of soft computing approaches", International Journal of Recent Trends in Engineering, Vol. 1, No. 2, 2009, pp. 250-254.
- [12] O. P. Verma, et al. "A novel fuzzy system for edge detection in noisy image using bacterial foraging", Multidimensional Systems and Signal Processing, Vol. 24, No. 1, 2013, pp. 181-198.
- [13] M. Setayesh, and M. Zhang, and M. Johnston, "Edge detection using constrained discrete particle swarm optimisation in noisy images", IEEE Congress on Evolutionary Computation (CEC), 2011.
- [14] M. Setayesh, and M. Zhang, and M. Johnston, "Investigating particle swarm optimisation topologies for edge detection in noisy images", AI 2011: Advances in Artificial Intelligence. Springer Berlin Heidelberg, 2011, pp. 609-618.
- [15] H. Li, , X. Liao, Ch. Li, H. Huang, and Ch. Li. "Edge detection of noisy images based on cellular neural networks." Communications in Nonlinear Science and Numerical Simulation vol. 16, no. 9, 2011, pp. 3746-3759.



- [16] M. A. El-Sayed, Y. A. Estaitia, and M. A. Khafagy. "Automated Edge Detection Using Convolutional Neural Network." *International Journal of Advanced Computer Science & Applications (IJACSA)* vol. 4, no. 10, 2013, pp. 11-17.
- [17] A.C. Bovik, and T.S. Huang, and D.C. Munson, "Nonparametric tests for edge detection in noise", *Pattern Recogn.*, Vol. 19, No. 3, 1986, pp. 209–219.
- [18] J. Aron, and L. Kurz, "Edge detection using ANOVA techniques", *Int. Symposium on Inf. Theory, Israel*, 1973.
- [19] L. Kurz, and M.H. Benteftifa, *Analysis of variance in statistical image processing*, Cambridge University Press, Cambridge, 1997.
- [20] D. H. Lim, and S. J. Jang, "Comparison of two-sample tests for edge detection in noisy images", *J. R. Stat. Soc. Ser. D-Statist.*, Vol. 51, No. 1, 2002, pp. 21–30.
- [21] A. Koschan, "A Comparative study on color edge detection", In *Proc. 2nd Asian Conf. on Computer Vision*, Vol. 3, 1995, pp. 574-578.
- [22] A. Mittal, S. Sofat, and E. Hancock. "Detection of edges in color images: A review and evaluative comparison of state-of-the-art techniques." *Autonomous and Intelligent Systems*, Springer Berlin Heidelberg, 2012, pp. 250-259.
- [23] F. Russo, and A. Lazzari, "Color edge detection in presence of Gaussian noise using nonlinear prefiltering", *IEEE Trans. On Instrum. And Meas.*, Vol. 54, No. 1, 2005, pp. 352-358.
- [24] A. N. Evans, "Nonlinear Edge Detection in Color Images", in *Advances in Nonlinear Signal and Image Processing*, edited by: Stephen Marshall and Giovanni L. Sicuranza. *EURASIP Book Series on Signal Processing and Communications*, Vol. 6, 2006, pp. 329-356.
- [25] F.V. Heijden, *Image-Based Measurement Systems*, Wiley, New York, 1994.
- [26] F. Russo, "A method for estimation and filtering of Gaussian noise in images", *IEEE Trans. On Instrum. and Meas.* Vol. 52, 2003, pp. 1148-1154.
- [27] A. Pizurica, and W. Philips, and I. Lemahieu, and M. Achery, "A joint interand intrascale statistical model for Bayesian wavelet based image denoising", *IEEE Trans. On Image Processing*, Vol. 11, No. 5, 2002, pp. 545-557.
- [28] M. A. Fligner, and G. E. Pollicello, "Robust rank procedures for the Behrens–Fisher problem", *J. Am. Stat. Assoc.*, Vol. 76, 1981, pp. 162–168.
- [29] P. W. Holland, and R. E. Welsch, "Robust regression using iteratively reweighted least-squares", *Commun. in stat.: theory and methods*, Vol. 6, pp. 813-827.
- [30] W. Pratt, *Digital Image Processing*, Wiley, New York, 1978.
- [31] D. Malah, and T. Peli, "A study of edge detection algorithms", *Computer Graphics Image Process.*, Vol. 20, 1982, pp. 1-21.

**Mina Alibeigi** received her B.Sc degree in Software Engineering (2008) from Shiraz University and M.Sc degree in Artificial Intelligence (2010) from that university. She is currently a Ph.D student in Machine Intelligence and Robotics at University of Tehran. Her research interests include imitation learning, human-computer interaction, data mining and image processing.

**Niloofer Mozafari** received the B.Sc degree in Software Engineering (2007) from Shiraz University and M.Sc degree in Artificial Intelligence (2010) from that university. She is currently Ph.D student in CSE and IT department of Shiraz University, Iran. Her area research interests include social network analysis, image processing and data mining.

**Zohreh Azimifar** received the B.Sc degree in Computer Science and Engineering from Shiraz University, Shiraz, Iran, in 1994, and her Ph.D degree in Systems Design Engineering from the University of Waterloo, Waterloo, Canada, in 2005. In 2005, she was a postdoctoral fellow in medical biophysics at the University of Toronto, Toronto, Canada. Since 2006 she has been a faculty member and director of computer vision and pattern recognition lab in computer science and engineering at Shiraz University. Currently, Dr Azimifar is chair of the Shiraz university research center for intelligent image and vision processing. Her research interests include statistical pattern recognition, machine vision and cognitive vision.

**Mahnaz Mahmoodian** received B.Sc degree in Computer Engineering (2008) from Shiraz University and M.Sc degree in Computer Engineering, Artificial Intelligence (2012) from University of Tehran. She is currently working at Traffic and Transportation Research Center (TTRC), Shiraz University as a researcher. Her current fields of interest are Machine Vision, Image Processing and Intelligent Transport Systems (ITS).

# Ant Colony Scheduling for Network On Chip

Neda Dousttalab\*

Department of Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran  
n.dousttalab@gmail.com

Mohammad Ali Jabraeil Jamali

Department of Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran  
m\_jamali@itrc.ac.ir

Ali Ghaffari

Department of Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran  
a.ghaffari@iaut.ac.ir

Received: 08/Sep/2013

Revised: 03/Jan/2015

Accepted: 04/Jan/2015

## Abstract

It is undeniable that scheduling plays an important role in increasing the network quality on chip. If experts realize the significance of mapping and scheduling in getting rid of delays and increasing performance of these systems, they will ponder over these activities much more scrupulously. The operation scheduling problem in network on chip (NoC) is NP-hard; therefore, effective heuristic methods are needed to provide modal solutions. In this paper, ant colony scheduling was introduced as a simple and effective method to increase allocator matching efficiency and hence network performance, particularly suited to networks with complex topology and asymmetric traffic patterns. The proposed algorithm was studied in torus and flattened-butterfly topologies with multiple types of traffic pattern. For evaluating the performance of the proposed algorithm, specialized simulator network on chip entitled by BookSim working under Linux operation system was used. Evaluation results showed that this algorithm, in many causes, had positive effects on reducing network delays and increasing chip performance compared with other algorithms. For instance, for a complex topologies, this algorithm under maximum injection\_rate of up to (10%) increasing throughput have been observed, injection rate, on average, compared to other existing algorithms.

**Keywords:** On-Chip Interconnection Networks; Switch Allocator; Ant Colony; Scheduling.

## 1. Introduction

Network-on-chip is an approach for design communication subsystems between IP cores in a system-on-a-chip (SoC). NoCs can measure synchronous and asynchronous clock domains or use uncloaked a synchronous logic. Networking theory and methods are applied by NoC for on-chip communication and thus notable improvement is made in conventional bus and crossbar interconnections. NoC promotes the scalability of SoCs and power efficiency of complex SoCs compared with other designs [1].

A main agent in specifying network efficiency to network-level design includes topology, routing function, flow control, and router microarchitecture. One parameter which affects the performance of a NoC router is selection of switch allocator and VC allocator.

In general, role of switch allocators is scheduling flits that are buffered at the router's input to pass through crossbar thus; in order to achieve maximum router performance, output quality of matching between resources and requests should be high [2].

This paper is structured as follows: in Section 2, the related works are expressed. Section 3 covers the background including switch allocator, VC allocator, flattened butterfly topology, and ant colony optimization. The proposed scheduler is presented in Section 4. In

Section 5, the methodology is explained. In Section 6, simulation of the proposed algorithm is discussed. Conclusion is given in Section 7.

## 2. Related Works

Max matching method proposed by Hopcroft and Karp [3], due to maximum size matching, does not assure 100% throughput on non-uniform traffic. McKeown et al. performed a more complex maximum weight matching [4]. Various techniques have since been presented to approximate maximum weight matching and reach 100% throughput in [5, 6]. In this approach, although this algorithm guarantees finding a maximum match, it is difficult to parallelize or pipeline and too slow for latency-sensitive application. Nick McKeown et al. [7] presented a scheduling algorithm for input-queued switches which improved RRM by reducing the synchronization of the output arbiters. Anderson et al. [8] proposed a technique called parallel iterative matching (pim), which rapidly detected a conflict-free pairing of inputs to outputs. Wavefront allocator was also developed by Tamir [9]. Owing to the advantage of the router of the Alpha 21364, incremental allocation was used [10,11].

\* Corresponding Author

### 3. Background

In this section, a brief review of the basic concepts required for this paper is presented.

#### 3.1 Switch Allocator

Switch allocation attempts to match the set of requests from the input to the set of available crossbar slot. The aim is that at most one VC at each input port could receive a grant. The block diagram for switch allocator implementation is demonstrated in Fig. 1.

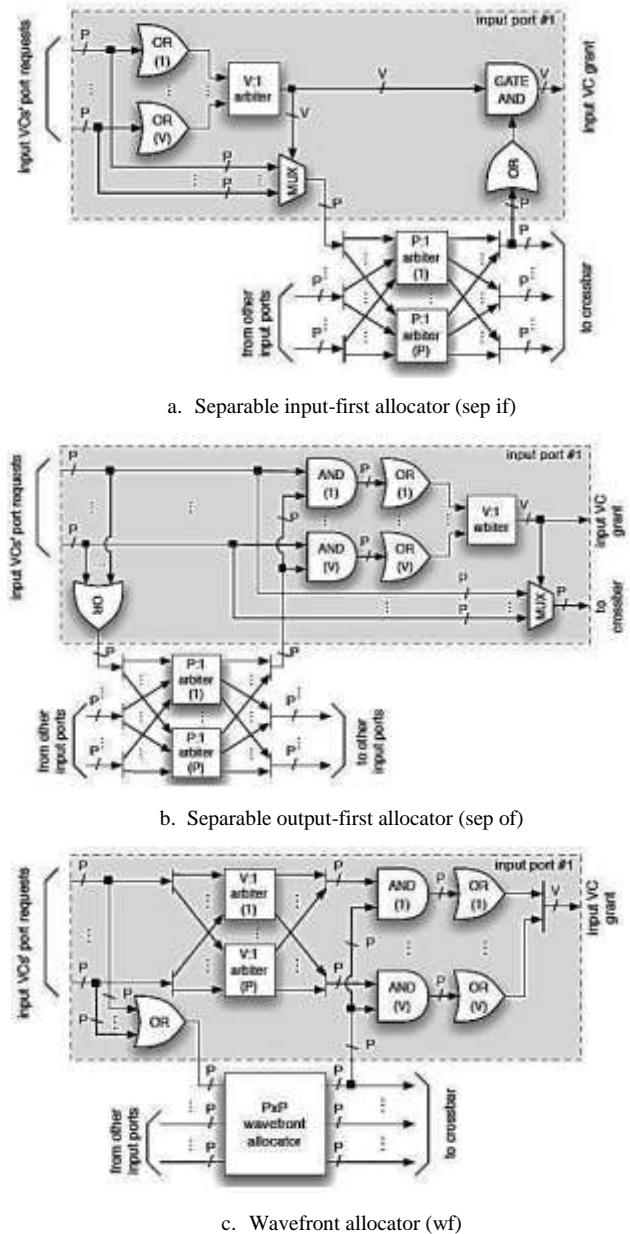


Fig.1. Switch allocator block diagram [2]

Separable input-first architecture has three stages: pending the first stage one of these active VCs selected as a winner at each input port by arbiter. In the two stages, the winning virtual channel requests are sent to the

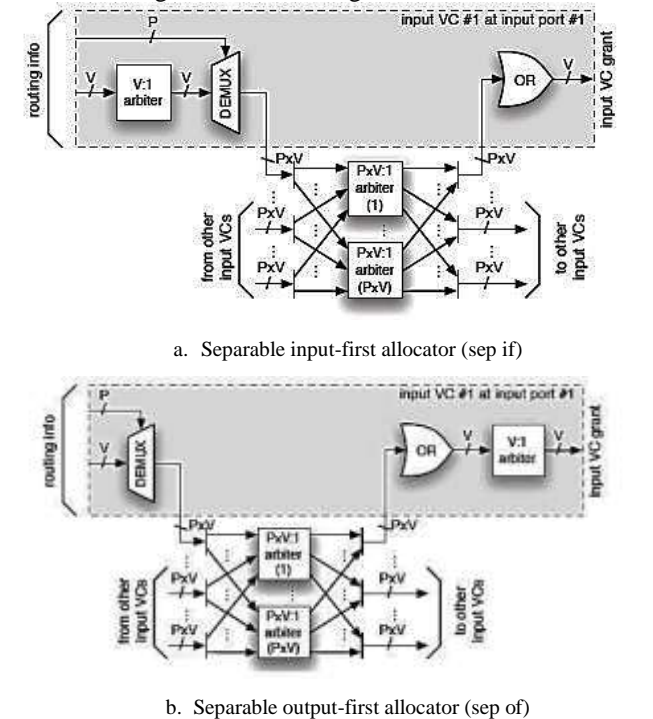
assigned output ports. Finally, output arbiters generate a grant signal.

For separable output-first allocation, as illustrated in Fig. 1(b), rearbiration is performed among all the requesting input ports that are forwarded to the associated output ports. The probability of selecting a given input by multiple output arbitrations is performed twofold. Will be conducted by one or more ports in the first referee is awarded to an input port and then arbitration performed among all wining virtual channels.

Finally, the wavefront-based implementation, shown in Fig. 1(c), Wavefront allocation, unlike the separable simultaneously does the arbitration between the requests for inputs and outputs. No justice assurances are provided outside, which ensure that all requests are eventually answered [2].

#### 3.2 VC Allocator

Head flits buffered at any of the router's input VCs must be allocated an appropriate output VC at the output port which is selected by the routing function before proceeding through the router pipeline. Hence, the VC allocator implements a matching between  $P \times V$  requesters and  $P \times V$  resources under the constraint that any output VCs requested by a given input VC at any given time shares the same output port. Block diagrams for the VC allocator designs is shown in Fig. 2.



b. Separable output-first allocator (sep of)

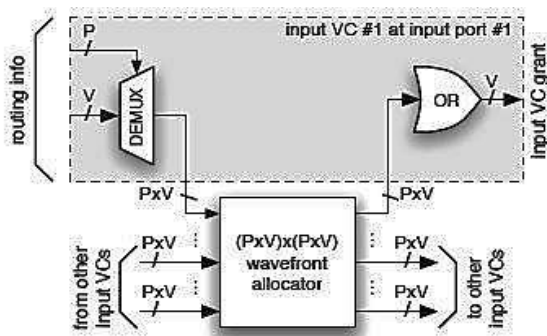


Fig. 2. VC allocator block diagrams [2]

In the separable input-first implementation (which is given in Fig. 2(a)), each input VC first determines the output VC at the destination output port for suggestion. Requests are sent to the stage of  $P \times V$ -input arbiters at the output VCs as in the canonical implementation; these large  $P \times V$ -input arbiters can be implemented as three arbiters: A stage of  $P \times V$ -input arbiters in parallel to a single  $P$  input arbiter selected among them in order to reduce delay. Finally, the grants of each input VC are categorized and reduced to a  $V$ -wide vector, which indicates the granted output VC.

The separable output-first implementation is demonstrated in Fig. 3(b). In this figure, each input VC sends requests to all the candidate output VCs at the destination port, in which arbitration is again done between all of the incoming requests. As a given input, VC requests might win the arbitration at multiple output VCs and an additional stage of arbitration is required after categorizing and reducing the grants for selecting a single winning VC.

Finally, the wavefront-based implementation (as depicted in Fig. 3(c)) includes a canonical  $P \times V$ -input wavefront allocator with an additional logic for generating the  $P \times V$ -wide request vector for each input VC, like the separable output-first case, and reduction of  $P \times V$ -wide grant vectors to a  $V$ -wide vector like the input-first case [2].

### 3.3 Flattened Butterfly Topology

Fig. 3 shows flattened butterfly topology that can be derived by combining (or flattening) the routers in each row of a contractual butterfly topology, while securing the inter-router connections. It is structurally like the generalized hypercube; the flattened butterfly significantly reduces the wiring complexity of the topology and more scalable because it condensate in the routers [12].

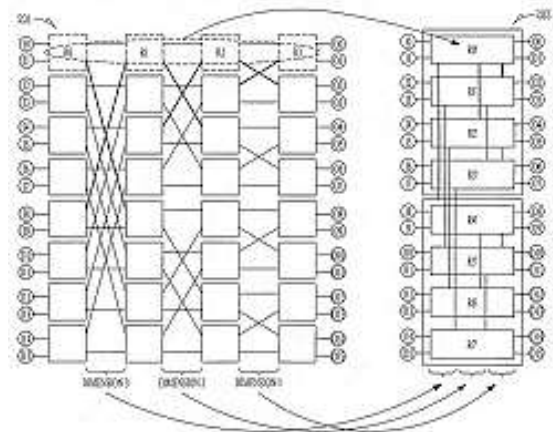


Fig. 3. Block diagram of the flattened butterfly

### 3.4 Ant Colony Optimization Algorithm

Ant colony optimization algorithm is a probabilistic technique which aimed to solve computational problems; it can be also reduced to finding good paths through graphs. This algorithm is a member of the family of ant colony algorithms in swarm intelligence methods, which constitutes some meta-heuristic optimizations. It was first proposed by Marco Dorigoin 1992 [13, 14].

In the natural world, ants (initially) go randomly around and, after finding the food, they return to their colonies while leaving some pheromone trails. If such a path is found by other ants, they probably do not keep on random travelling and instead follow the same trail, return, and reinforce it in case of eventually finding the food. However, the pheromone trail starts to evaporate over time and, thus, its attractive strength is reduced. The more the time for an ant to travel down the path and come back, the more the time for the evaporation of pheromones would be. By comparison, short paths got marched over more frequently; therefore, the pheromone density becomes higher on shorter paths than longer ones. Pheromone evaporation is advantageous due to avoiding convergence to a locally optimal solution. In the case of no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive for the following ones. In such a case, exploration of the solution space would be limited. Therefore, when one ant finds a good (and short) path from the colony to a food source, other ants probably follow that path and the positive feedback causes all the ants to eventually follow a single path. The idea of ant colony algorithm is to imitate this behavior by "simulated ants" which walk around the graph while showing the problem to be solved [15].

## 4. Proposed Ant Colony Scheduling Algorithm

In this section, the proposed ant colony algorithm is introduced for solving the scheduling problem. The purpose of this scheduling algorithm is to rapidly find a conflict-free pairing of inputs to outputs, which use a

queued cell to transmit between them. This pairing specifies which inputs transmit cells over the crossbar to which outputs in a determined time slot. Ant colony scheduling is a commonly used heuristic, since it generates "good" quality results for large complex problems. The proposed algorithm used separable allocation as a switch allocation. Additionally a dense allocator was applied as a VC allocator. The proposed allocator was based on a basic separable allocator and acted iterant. At each iterance, three steps are performed:

1. Each input port sends a request for using the queue cell.
2. Each output port selects one among all the received requests according to ACO that uses ant colony algorithm for computing priorities.
3. Each input receives multiple grants and accept desirable grant according to ACO and inform to output port after grant by output ports.

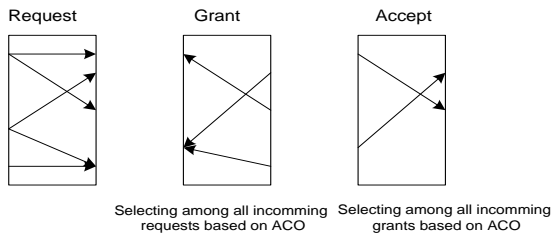


Fig. 4. ACO scheduling: One iteration

To prevent the same selection during each arbitration and eventually eliminate any possible starvation, we perform referee randomly throughout arbitrations.

### 5. Methodology

Evaluation is performed using BookSim simulator. A 2D torus and flat-fly was used. Each router was connected to one network terminal and all the channels had a single cycle of latency. Deterministic dimension-order routing (DOR) was used for torus and ran-min routing for flat-fly topology. Also, 4 VCs with 8 buffer slots statically assigned to each were applied. Uniform random, random permutation, bit complement, and asymmetric traffic patterns were also employed. Injection rate was in flits.

### 6. Simulation

As illustrated in Fig. 5, in flat-fly topology with uniform and asymmetric traffic patterns, compared to iSLIP, ant algorithm increased throughput under maximum injection rate up to (12% on average). Ant algorithm offered (12.77% on average) more throughput at maximum injection rate than wavefront and comparable throughput (7.35% on average) more pim and (10.94% on average) more loa.

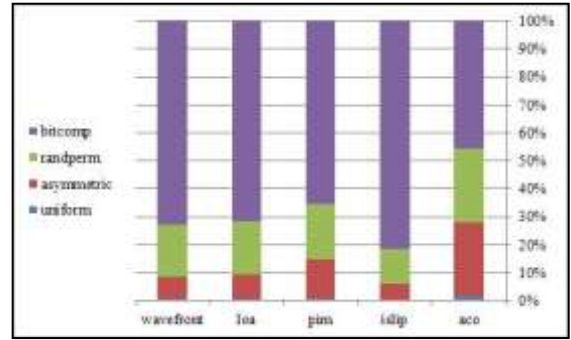


Fig. 5. Throughput comparison of flat-fly for uniform, asymmetric, randperm, and bit complement traffic patterns

Fig. 6 compares throughput by traffic pattern in torus topology. Compared to iSLIP, ant algorithm offered (0.1% on average) higher throughput. Compared to wavefront, throughput increased up to (0.17% on average). Ant algorithm can provide up to 0.14% increase in throughput compared to pim and (0.5% on average) more loa. Although the ant algorithm can provide higher throughput for some traffic patterns, it provides less throughput for bitcomp (bit complement) and randperm, because bitcomp creates some continuous flows of traffic which starve other flows.

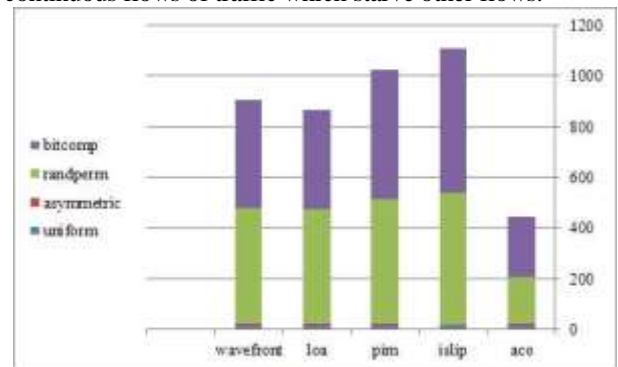


Fig. 6. Throughput comparison of torus for uniform, asymmetric, randperm, and bit complement traffic patterns

In Figs. 7 and 8, throughput of the proposed algorithm is investigated in flat-fly and torus topology with asymmetric and uniform patterns, respectively.

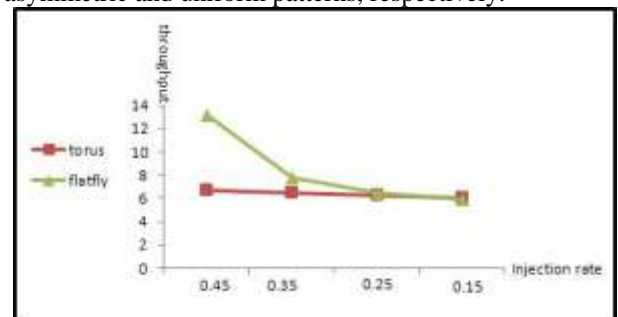


Fig. 7. Throughput comparison of torus and flat-fly across injection in asymmetric traffic

Considering this figure, it can be observed that, in the uniform pattern, the circular mesh had better performance for low and medium traffic; however, flat-fly had higher throughput for high traffics.



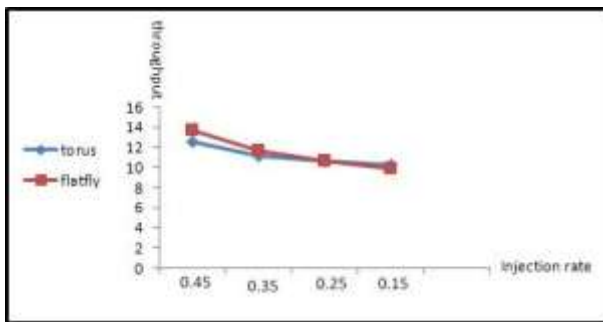


Fig. 8. Throughput comparison of torus and flat-fly across injection rate in uniform traffic pattern

## 7. Conclusions

Inspired by the nature and collective intelligence of ants, ant colony-based scheduling algorithm in the network was investigated and applied to the chip. We used ant colony approach in our scheduling because of two reasons: first of all the, possibility of employing

initial information of the problem is unique in ant colony algorithm the second reason is that fat convergence in obtaining an optimal response and its resistant parameters in initial valuing. The results of this implementation showed that ant colony algorithm is destroyed due to the evaporation of extra pheromone pathways, which prevents ant trapping in local minimums and leads to better algorithm efficiency than the existing scheduling algorithms. Moreover, this algorithm performed at a very good level in most cases due to being smart in the traffic patterns close to real traffic-like asymmetric pattern.

## Acknowledgment

This article was extracted from the thesis prepared by Neda Dousttalab to fulfill the requirements for earning the Master's degree. We thank Dr. Bahman Arasteh (Department of Computer Engineering, Islamic Azad University, Tabriz Branch) and Nan Jiang (Electrical Engineering, Stanford University, Stanford) for providing technical support.

## References

- [1] B. T. W.J. Dally, "Principles and Practices of Interconnection Networks," Morgan Kaufmann Publishers, San Francisco, CA, 2004.
- [2] D. U. B. W.J.Dally, "Allocator Implementations for Network-on-Chip Routers," In Proceedings of the 2009 ACM/IEEE Conference on High Performance Computing, Networking, Storage and Analysis, 2009.
- [3] J. E. H. a. R. M. Karp, "An  $n^2/2$  algorithm for maximum matching in bipartite graphs," SIAM Journal on Computing, vol. 2, pp. 225-231, 1973.
- [4] A. M. N.McKeown, V.Anantharam, and J.Walrand, "Achieving 100% throughput in an input-queued switch," In Proc. of IEEE INFOCOM, pp. 296–302, March, 1996.
- [5] L.Tassiulas, "Linear complexity algorithms for maximum throughput in radio networks and input queued switches," In Proc. of IEEE INFOCOM, pp. 533–539, 29 Mar-2 Apr 1998, 1998.
- [6] B. P. P. Giaccone, and D.Shah, "Towards simple high performance schedulers for high-aggregate bandwidth switches," In Proc. Of IEEE INFOCOM, pp. 1160–1169, June, 2002.
- [7] N. McKeown, "The iSLIP Scheduling Algorithm for Input-Queued Switches," IEEE/ACM TRANSACTIONS ON NETWORKING, vol. 7, April, 1999.
- [8] S. S. O. T. E. Anderson, J.B.Saxe, and Ch.P. Thacker, "High speed switch scheduling for local area networks," ACMTransactions on Computer Systems, pp. 3, 352-19, November, 1993.
- [9] Y. T. a. H.C.Chi, "Symmetric crossbar arbiters for VLSI communication switches," IEEE Transactions on Parallel and Distributed Systems, pp. 13–27, January, 1993.
- [10] F. S. S. S. Mukherjee, P.Bannon, J.Emer, S.Lang, and D. Webb, "A comparative study of arbitration algorithms for the Alpha 21364 pipelined router," In Proc of. Architectural Support or Programming Languages and Operating Systems (ASPLOS), pp. pages 223–234, octobr, 2002.
- [11] P. B. S.S. Mukherjee, S.Lang, A.Spink, and D.Webb, "The Alpha 21364 network architecture," In Proc. of the Symposium on Hot Interconnects, pp. 113–117, august, 2001.
- [12] J. B. J. Kim, and W. Dally, "Flattened Butterfly Topology for On-chip Networks," In International Symposium on Microarchitecture, pp. 172–182, December, 2007.
- [13] M.Dorigo, "ant colony optimization".
- [14] G. D. C. a. L. G. M.Dorigo, "Ant Algorithm for Discrete Optimization," Artificial Life, vol. 5, pp. 137-172, 1999.
- [15] S. S. a. S. N. S. P.Mathiyalagan "Modified Ant Colony Algorithm for Grid Scheduling," IJCSE) International Journal on Computer Science and Engineering, vol. 2, 2010.

**Neda Dousttalab** received the B.Sc degree in Computer Engineering from the Islamic Azad University, Shabestar Branch, Shabestar, Iran, in 2008 and the M.Sc degree in Computer Engineering from Islamic Azad University, Tabriz Branch, Tabriz, Iran, in 2013. Her research interests include multiprocessor systems-on-chip, networks on chip and Fault tolerance

**Mohammad Ali Jabraeil Jamali** received B.Sc degree in Electrical Engineering from Urmia University, Urmia, Iran, the M.Sc degree in Electrical Engineering from Tabriz University, Tabriz, Iran, the M.Sc degree in Computer Engineering from Islamic Azad University, Science and Research Branch, Tehran, Iran and the Ph.D degree in Computer Engineering from Islamic Azad University, Science and Research Branch, Tehran, Iran, in 1994, 1997, 2003 and 2009, respectively. He is an assistant professor of Computer Engineering at Islamic Azad University, Shabestar Branch. He is the author/co-author of more than 50 publications in technical journals and conferences. His current research interests include processor and computer architectures, chip multiprocessors, multiprocessor systems-on-chip, networks on chip, ad hoc and sensor networks.

**Ali Ghaffari** received the B.Sc degree in Computer Engineering from the University of Tehran, Tehran, Iran, in 1994 and the M.Sc degree in Computer Engineering from the University of Tehran, Tehran, Iran, in 2002 and the Ph.D degree in Computer Engineering from the Islamic Azad University, Science and Research Branch, Tehran. His research interests include information security, wireless networks, and ad hoc networks.

# A Fast and Accurate Sound Source Localization Method Using the Optimal Combination of SRP and TDOA Methodologies

Mohammad Ranjkesh

Department of Electrical Engineering, University Of Guilan, Rasht, Iran  
m.ranjesh@khi.ac.ir

Reza Hasanzadeh\*

Department of Electrical Engineering, University Of Guilan, Rasht, Iran  
hasanzadehpak@guilan.ac.ir

Received: 30/Aug/2014

Revised: 09/Apr/2015

Accepted: 17/May/2015

## Abstract

This paper presents an automatic sound source localization approach based on a combination of the basic time delay estimation sub-methods namely, Time Difference of Arrival (TDOA), and Steered Response Power (SRP) methods. The TDOA method is a fast but vulnerable approach for finding the sound source location in long distances and reverberant environments and is so sensitive in noisy situations. On the other hand, the conventional SRP method is time consuming, but a successful approach to accurately find sound source location in noisy and reverberant environments. Also, another SRP-based method, SRP Phase Transform (SRP-PHAT), has been suggested for the better noise robustness and more accuracy of sound source localization. In this paper, based on the combination of TDOA and SRP based methods, two approaches were proposed for sound source localization. In the first proposed approach called Classical TDOA-SRP, the TDOA method is used to find the approximate sound source direction and then SRP based methods were used to find the accurate location of sound source in the Field of View (FOV) which is obtained by the TDOA method. In the second proposed approach which called Optimal TDOA-SRP, for more reduction of computational processing time of SRP-based methods and better noise robustness, a new criterion has been proposed for finding the effective FOV which is obtained through the TDOA method. Experiments were carried out under different conditions confirming the validity of the purposed approaches.

**Keywords:** Steered Response Power; Time Delay Estimation; Steered Response Power Phase Transform; Sound Source Localization; Time Difference Of Arrival; Field Of View.

## 1. Introduction

Distributed microphone systems have been considered for various applications including human computer/machine interfaces, talker tracking, Robotic domain and beam-forming for signal-to-noise ratio (SNR) enhancements [1,2]. Many of these applications require detecting and localizing the sound sources. Therefore, proposed methods for sound source localization problems with distributed microphone arrays are usually very important. In some practical sound source localization (SSL) applications, the source should be automatically detected for computer driven analyses of the auditory scene [1]. SSL algorithms can be broadly divided into indirect and direct schemes [3]. Indirect algorithms usually follow a two-step procedure. In the first step, the time delay of arrival between each microphone pairs is computed and in the second step, they estimate sound source position based on the estimated delay and the geometry of arrays. The direct algorithm performs time delay of arrival and sound source location estimations in one single step by scanning a set of candidate source locations and selecting the most likely position as an estimated sound source location [4,5]. There are several algorithms for SSL applications categorized in a similar

manner. But, two most successful and recently proposed methods which are well-known as Steered Response Power (SRP) and Time Difference of Arrival (TDOA) have been considered in the recent years for direct and indirect approaches, respectively [5,6]. The basic principle of SRP methodology is based on the filter-and-sum (delay-and-sum) beam-forming operation, which leads to noise power reduction proportional to the number of uncorrelated microphone channels used in the operation [6,7]. Although SRP methods have been used properly for applications such as intrusion detection and gunfire location, this kind of SSL method is time consuming which makes real time applications inappropriate [8]. On the other hand, TDOA is another popular SSL method, which is more appropriate for practical and real time applications [9]. This method is nonlinear in its nature, but it has significant computational advantages over any other SSL methods. However, this kind of SSL methods is only able to estimate the direction of the corresponding sound source location in long distances. Therefore, this problem makes TDOA method inappropriate for applications in which the precise detection of the SSL is necessary [8,10].

In this paper, a combination approach has been proposed to estimate the sound source direction using a

\* Corresponding Author

basic TDOA method and then SRP method has been used to find the final sound source location in the estimated direction.

The experimental results in this paper show that because of the pre-estimation of the sound source direction in the proposed methods, we have a valuable reduction of the computational time and more noise robustness relative to the conventional SSL methods.

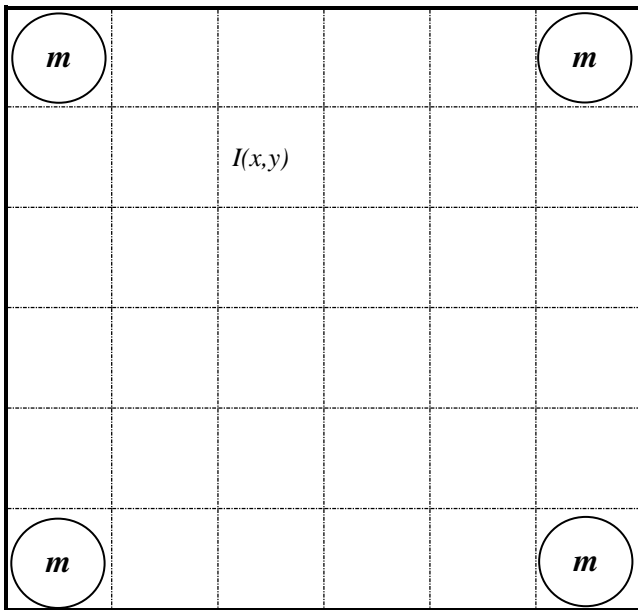


Fig. 1. FOV for SRP setup with four microphones.  $I(x, y)$  is a typical grid point

## 2. Steered Response Power (SRP) Method

The SRP methods use sound's power and create a SRP image to show the sound source location. The SRP method can be affected by different type of uncorrelated and correlated noises [6]. The uncorrelated noise typically results from the independent noise on each microphone channel and the correlated noise, on the other hand, results from coherent noise sources such as sources outside the Field of View (FOV), multiple targets and reverberations [6]. In the SRP method, the correlated noise creates greater challenges for beam-forming compared to the uncorrelated noise [6] and will be used in the experimental results of this paper. In order to reduce the impact of noise on the sound source location estimation, several filters for the SRP method have been proposed for improving performance, such as Maximum Likelihood (ML)[11], Smooth Coherence Transforms (SCOT)[12], Phase Transform (PHAT)[13] and the Roth Processor[14]. The experimental results show that PHAT has a better performance than others in noisy and reverberant environment [15].

### 2.1 Mathematical Methodology of SRP

Fig. 1 shows a simple fundamental structure of SRP methods in 2-dimensional case such that Sound source ( $I$ ) and microphones ( $m$ ) are at the same  $Z$  coordinate. In the

SRP method, a microphone array is used to make the beam-form for each point in the FOV [16]. For each grid point of interest, the SRP delays each microphone signal to result in a coherent addition for a sound source traveling from the point of interest. For each point in the region of interest, the received signals are delayed accordingly and summed together coherently, and finally the power of each point in the region of interest is computed, respectively. The detection and location of the sound source is based on value of the estimated power at each point. Also, the power estimation maybe corrupted by noise sources, reverberation and the finite distributions of microphones [15].

As shown in fig. 1, for finding the location of sound source, it can be assumed that the FOV is formed as grid points i.e.  $I(x,y)$ . By defining a 2-dimensional FOV (assuming that the sound source and microphones are in the same horizontal place e.g.  $xy$  plane) and  $N$  microphones and also considering the output from  $q$ 'th microphone is  $m_q(t)$ , the SRP at the spatial point  $X=[x,y]$  for a time frame  $n$  of length  $L$  can be defined as

$$P_n(X) = \int_{nL}^{(n+1)L} \left| \sum_{q=1}^N m_q(t - \tau(X, q)) \right|^2 dt \quad (1)$$

In this equation,  $\tau(X, q)$  is the direct time of travel from location  $X$  to microphone  $q$ . In [17], it is shown that the SRP can be computed by summing the General Cross-Correlation (GCC) for all possible pairs of the set of microphones. The GCC for a microphone pair ( $k, l$ ) is computed as

$$R_{m_k m_l}(\tau) = \int_{-\infty}^{+\infty} M_k(w) M_l^*(w) e^{jw\tau} dw \quad (2)$$

where  $\tau$  is the time lag,  $*$  denotes complex conjugation, and  $M_l(w)$ ,  $M_k(w)$  are the Fourier transform of the microphone signals  $m_l(t)$ ,  $m_k(t)$ , respectively.

Taking into account the symmetries involved in the computation of (1) and removing some fixed energy term, the part of  $P_n(X)$  that changes with  $X$  is isolated as[5]

$$P_n'(X) = \sum_{k=1}^N \sum_{l=k+1}^N R_{m_k m_l}(\tau_{kl}(X)) \quad (3)$$

where  $\tau_{kl}(X)$  is the microphone time delay function of each pair is given by,

$$\tau_{kl}(X) \equiv \frac{\|X - X_k\| - \|X - X_l\|}{c} \quad (4)$$

where  $X_k$ ,  $X_l$  are the microphone locations and  $c$  is the speed of sound which is calculated by[21],

$$c = 20.05\sqrt{273.15 + T} \quad (5)$$

In this equation,  $c$  is sound's propagation speed in (m/s) and  $T$  is environmental temperature ( $^{\circ}\text{C}$ ).

In the SRP method,  $P_n'(X)$  is evaluated on the FOV to find the sound source location,  $X_s$  which provides the maximum value [5,7,17].

$$X_s = \arg \max P_n'(X), X \in FOV \quad (6)$$



## 2.2 SRP Phase Transform (SRP-PHAT) Method

The basic principle of SRP-PHAT is similar to SRP method, but in this method, a weighting function has been used to increase the accuracy of finding the sound source delays beside the advantage of its simplicity in the implementation [5]. In this method, the weighting function works as a normalizing factor which relates to the phase spectrum information of sound source. Regarding this term, equation (1) can be formalized as follows [5],

$$P_n(X) = \int_{nT}^{(n+1)T} \left| \sum_{q=1}^N w_q m_q(t - \tau(X, q)) \right|^2 dt \quad (7)$$

where  $w_q$  is weighting factor and  $\tau(X, q)$  is the direct time of travel from location  $X$  to microphone  $q$ .

Therefore, SRP can be computed by summing the GCCs for all possible pairs of the set of microphones [5]. The GCC for a pair  $(k, l)$  is computed as

$$R_{m_k m_l}(\tau) = \int_{-\infty}^{+\infty} \varphi_{kl}(w) M_k(w) M_l^*(w) e^{jw\tau} dw \quad (8)$$

where  $\tau$  is the time lag, \* denotes complex conjugation,  $M_l(w)$  is the Fourier transform of the microphone signal  $m_l(t)$  and  $\varphi_{kl}(w)$  is a combined weighting function in the frequency domain [5]. In the SRP-PHAT, the weighting function for a reverberant environment is defined as [5],

$$\varphi_{kl}(w) \equiv \frac{1}{|M_k(w) M_l^*(w)|} \quad (9)$$

In SRP-PHAT, GCC is computed using (8) instead of (2) to obtain  $P_n(X)$  which is mentioned in (3). Finally, the sound source location can be evaluated by finding the point source location  $X_s$  that provides the maximum value in (6) [5].

## 3. Time Difference of Arrival (TDOA) Method

The TDOA is one of the time delay estimation (TDE) sub-methods that is used in low noise or noise free environments, which leads to a considerable reduction of computational complexity. In this method, at least two microphones should be used to find sound source direction ( $\theta$ ). For finding  $\theta$ , we need to calculate the time delay between received signals of each microphone, respectively. An approach to estimate the time delay between the received signals at two microphones is cross-correlation [18]. The computed cross-correlation values give the point at which the two signals from separate microphones have their maximum correlations. The cross-correlation of sound signals  $s_i$  and  $s_j$  received in microphones  $i$  and  $j$  respectively is given by [8],

$$R_{ik}(i) = E\{s_i^* [j] s_k [i + j]\} \quad (10)$$

where  $E$  denotes the expectation operator,  $i$  is discrete time shift,  $j$  points samples of each sound source signal and \*denotes complex conjugate operation. As shown in

(11), the discrete time delay between received signals,  $\tau_{kl}$ , can be obtained by finding argument of the maximum value of cross-correlation, where the signals are best aligned[8],

$$\tau_{kl} = \arg \max_i (R_{kl}(i)) \quad (11)$$

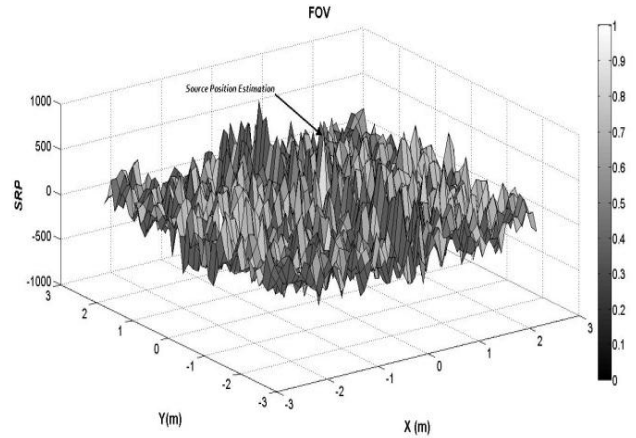


Fig. 2. SRP image for  $5 \times 5 \text{ m}^2$  FOV in the presence of noise with  $\text{SNR}=10\text{dB}$ .

The time delay between two typical microphones is also given by [8],

$$t_{kl} = \frac{\tau_{kl}}{f_s} \quad (12)$$

Where  $f_s$  is sampling frequency rate of sound source. Therefore, the sound source direction,  $\theta$  can be given by,

$$\theta = \sin^{-1} \left( \frac{c \times t_{kl}}{d} \right) \quad (13)$$

Where  $d$  and  $c$  are distances between two microphones and sound's propagation speed, respectively. In this approach. It is assumed that  $d$  should not be larger than sound wave length [18].

Fig. 3 shows a typical setup of TDOA method. As shown in Fig. 3, two candidates of  $\theta$  can be mentioned for sound source direction [18-20]. For solving this problem, two pairs of microphones can be used to find the accurate sound source direction [21,22].

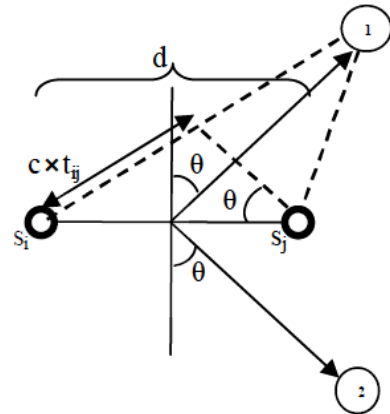


Fig. 3. Calculating the angle of sound source

One of the suggestions for setup of microphone pairs which can be aligned together is shown in fig. 4 and used in this paper.

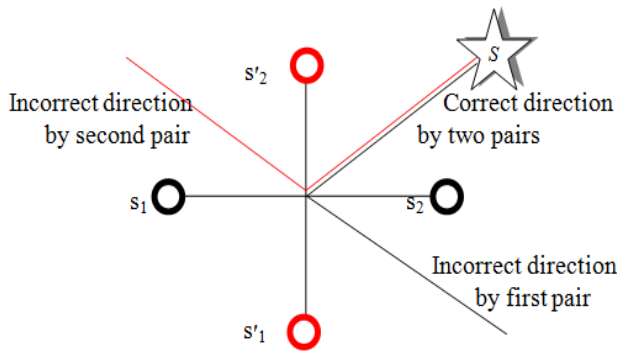


Fig.4. A typical TDOA microphone array ( $s_1$  and  $s_2$  are the first microphone pairs,  $s'_1$ ,  $s'_2$  are second microphone pairs,  $S$  is the sound source position)

### 4. Combination of SRP/SRP-PHAT and TDOA

As mentioned in section 1, although the SRP method can find sound source location, it is time consuming. On the other hand, although the TDOA is a low computational time method, it is noise effective. A suggestion can be derived using a combination of these methods to decrease the computational time as well as more robustness in the presence of noise. As shown in fig.5, two TDOA setups such as fig. 4 are used at the center of FOV [8] and three additional microphones are also utilized for each quarter [2,5,6,8]. For each quarter, these additional microphones with the central microphone can be used for SRP methodology. Therefore, in this paper, 13 microphones have been used in order to have a symmetric structure.

#### 4.1 Classical Combination of TDOA and SRP/SRP-PHAT

First, based on TDOA method mentioned in section 3, the sound source direction can be determined using four microphones placed in the center of FOV.

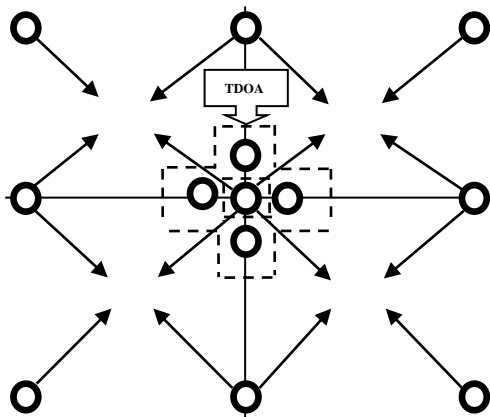


Fig. 5. Combination FOV (circles are microphone positions)

The next step is to find which quarter contains the accurate sound source location. Finally, one of the SRP or SRP-PHAT methods can be used to find the actual location

of sound source in the selected quarter. For a better discrimination, the two proposed methods are briefly named as TDOA-SRP and TDOA-SRP-PHAT, respectively.

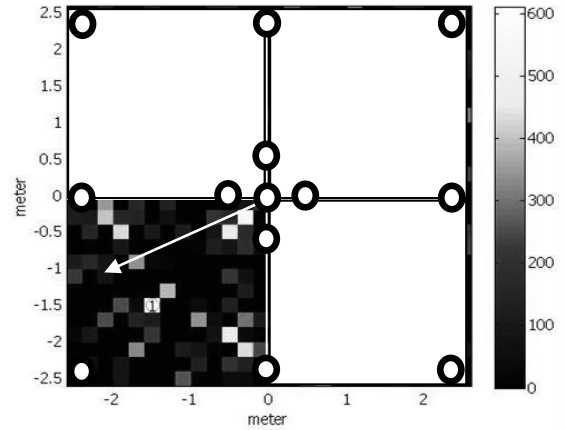


Fig. 6. SRP/SRP-PHAT method computes sound source location in the quarter in the FOV selected using TDOA method.(Circles are microphone positions)

Fig. 6 shows a general FOV that TDOA method has been proposed to find sound source direction. Afterward, this direction can be recognized to find the quarter contained sound source location. Each quarter containing the direction arrow is the goal quarter in the first step.

As shown in fig. 7, the selected quarter can be used by SRP or SRP-PHAT grid search methods to find the actual sound source location. In this selected quarter, the SRP or SRP-PHAT search in grid points and find sound source location.

#### 4.2 Optimal Combination of TDOA and SRP/SRP-PHAT

The classical combination of TDOA and SRP/SRP-PHAT mentioned in subsection 4-1 can reduce the search area to a quarter of grid points. But, it should be noted that for noise-free or low-noise environments, the SRP/SRP-PHAT methods just need to span the grid points along the direction which has been estimated by TDOA method. On the other hand, in the heavy noise environment and based on the noise effective nature of TDOA method, the SRP/SRP-PHAT methods should span nearly all of the grid points mentioned in the selected quarter to find the actual sound source location. Considering the time consuming nature of SRP-based methods, it seems for usual environmental noise that it is over qualified to seek all grid points of a quarter to find actual sound source location.

Our experimental results show that for the successful detection of sound source location in the real environment with a different noise level, SRP-based methods can be computed in a region with a deviation,  $\delta$ , around the direction obtained by TDOA method. Our empirical results indicate that this parameter can be selected proportional to  $\delta=\sigma$ , where  $\sigma$  is noise standard deviation.

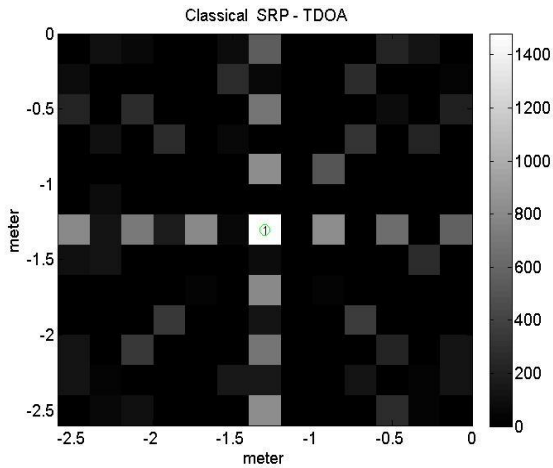


Fig. 7. Sound Source Localization in the selected quarter of Fig.6 (Classical TDOA-SRP/SRP\_PHAT)

Fig. 8 shows a typical example of this approach. For better discrimination, two proposed methods optimized through the new structure are briefly named O-TDOA-SRP and O-TDOA-SRP-PHAT, respectively.

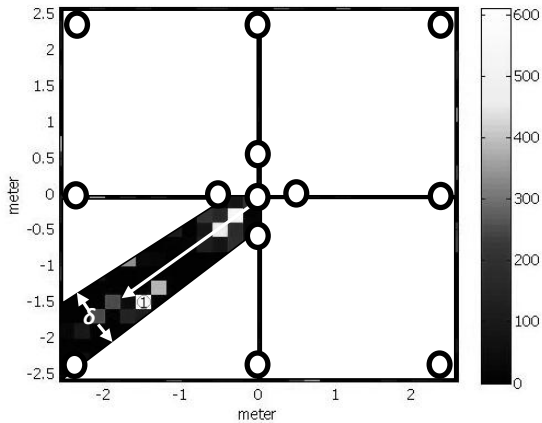


Fig. 8. Optimal TDOA-SRP/SRP\_PHAT (Circles are microphone positions)

### 5. Experimental Results

To obtain the experimental results, a PC with the following software and hardware specifications has been used. Software: MATLAB R2013a, Hardware: PC Core(TM)i7-3632QM, CPU 2.20 GHz, RAM 8 GB.

In this experiment, the resolution of grid points is assumed to be 100 and 200 mili meters in 5-1 subsection and is assume as 200 mili meters in other parts. The dimensions of FOV for sound source location are 5x5x2 meters in length, width and height, respectively. The sound source used for this experiment is Chainsaw sound in wav format with the time spectrum mentioned in fig. 9. It's number of bits per sample is 16. Maximum frequency of the sound is 21.956 kHz, and the sound source sampling frequency is 44.1 kHz according to Nyquist sampling theorem. The processing was carried out using a sampling rate of 44.1 kHz, with the time windows of 4096 samples of length and 50 % overlap.

The fundamental of the work presented here is based on SRP methodology and TDOA method only has been used to reduce FOV's area by detecting sound source direction (not location). Therefore, experimental results of the proposed methods have been compared with the other SRP base methods which can find both of sound source direction and location.

#### 5.1 Comparison of the Proposed Methods in the Presence of Noise

In this section, the deviation between actual and estimated sound source location of proposed methods is evaluated. In this comparison, for each level of noise, sound sources were degraded  $h$  times by noise. Then, the accuracy of the proposed SSL methods have been computed  $h$  times using the deviation of estimated and actual SSL methods in terms of Root Mean Square Error (RMSE) [20] as follows,

$$RMSE = \sqrt{\frac{1}{h} \sum_{k=1}^h (r_{ref} - r_k)^2} \tag{14}$$

where  $r_{ref}$  is the distance between actual sound source location and center of FOV, and  $r_k$  is the distance between estimated sound source location and center of FOV. We use  $h=10$ , and also several SNR = 40, 25, 10, 0 and -10dB are considered to evaluate the performance of proposed methods.

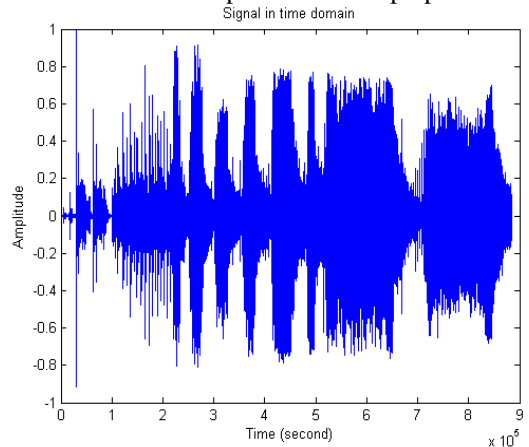


Fig. 9. time spectrum of Chainsaw sound source

Fig. 10 (a) and (b) show the SSL performance of proposed methods in the presence of different levels of noise for grid resolution 100 and 200 millimeter, respectively. As shown in fig. 10, for the all ranges of SNR, we have a significant difference between SRP-based methods and the proposed methods through classical and optimal combination of TDOA and SRP-based methods. In all methods, by increasing SNR till SNR= 0dB, the RMSE will be reduced and also the classical and optimal TDOA-SRP and TDOA-SRP-PHAT methods have better performances than SRP and SRP-PHAT methods.

On the other hand, due to the reduction of searching region of true sound source location, O-TDOA-SRP and O-TDOA-SRP-PHAT methods can successfully eliminate the similar sound source locations and led to a better performance than C-TDOA-SRP and C-TDOA-SRP-PHAT

methodologies, respectively. An overall evaluation can also show that O-TDOA-SRP-PHAT has the best robustness and accuracy in the presence of different levels of noise.

By comparing diagrams (a) and (b) in fig. 10, it can be concluded that decreasing the grid point resolution from 200 millimeters to 100 millimeters can reduce significantly the value of RMSE and improve all methods performance, respectively.

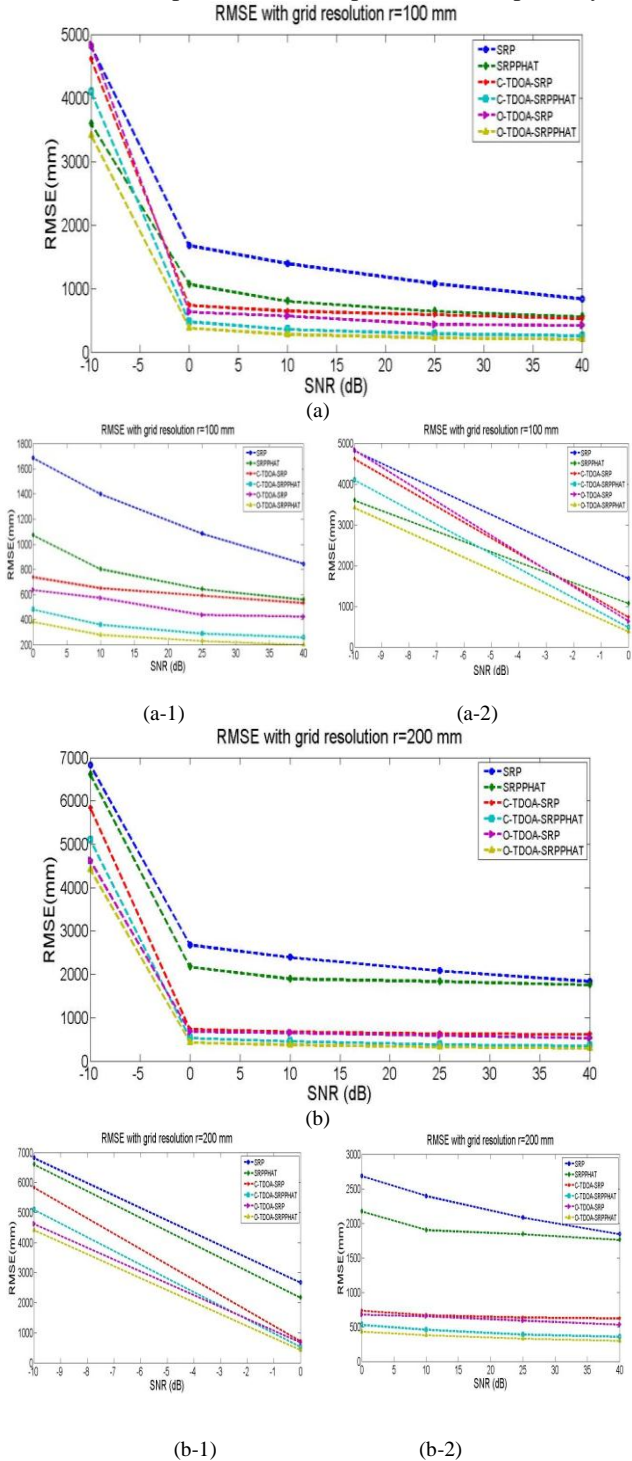


Fig. 10. Sound Source localization performance in terms of RMSE for proposed methods when different SNR are applied, (a) Grid resolution  $r = 100\text{mm}$  (a-1: -10 to 0 dB), (a-2: 0 to 40 dB) and (b) Grid resolution  $r = 200\text{mm}$  (b-1: -10 to 0 dB), (b-2: 0 to 40 dB).

As shown in fig. 10 (a-1, b-1), for SNR less than 0dB, RMSE is increased abruptly and the performance of methods is reduced effectively. Furthermore, it can be seen in SNR=-10dB that SRP and SRP-PHAT methods have a better performance than their classical combination methods. It is due to this fact that the proposed classical combination methods seek all of the selected quarter to find the sound source location. Therefore, it may lead to several outputs that satisfy the true conditions of real sound source direction. This problem has been solved in the optimized version of proposed methods due to limitation of the seeking area of sound source location.

### 5.2 Comparison Speed of Proposed Methods

In tables I and II, the computation times of the proposed methods are compared for SNR=10,-10dB in three different dimensions of FOV (these dimensions are in meter). As mentioned, for this entire situation, FOV's height is 2 meters. To calculate these times, each method runs ten times and the mean value of the processing time is reported.

Table 1. Comparison of the Time of Processing for SNR=10dB

Proposed Methods		Dimensions (m2)		
		5×5	10×10	20×20
SRP	Processing Time (Sec)	49	186	728
SRP-PHAT		47	193	781
C-TDOA-SRP		13	50	187
C-TDOA-SRP-PHAT		15	55	194
O-TDOA-SRP		8	24	90
O-TDOA-SRP-PHAT		8	25	100

Table 2. Comparison of the Time of Processing for SNR= -10dB

Proposed Methods		Dimensions (m2)		
		5×5	10×10	20×20
SRP	Processing Time (Sec)	52	189	735
SRP-PHAT		48	193	784
C-TDOA-SRP		11	51	188
C-TDOA-SRP-PHAT		14	54	189
O-TDOA-SRP		7	25	92
O-TDOA-SRP-PHAT		8	26	102

As shown in tables I and II, because of the ability of quarter selection in C-TDOA-SRP and C-TDOA-SRP-PHAT, they have lower computational time than SRP and SRP-PHAT. Also, because of the limitation in the region of process (grid point search), O-TDOA-SRP and O-TDOA-SRP-PHAT have less computational time than any other methods brought in this experiment. Therefore, for a real-time SSL, the optimal combination methods e.g. O-TDOA-SRP and O-TDOA-SRP-PHAT can have better abilities.

The comparison between tables I and II shows that there are no comparable differences between computation time of proposed methods in the presence of high and low SNR (respectively 10 , -10 dB). Therefore, it can be assumed that the computational time of proposed methods can be considered independent relative to SNR variations and have similar performance.

### 5.3 Stability Comparison

Another comparison can be mentioned through evaluating the stability of the proposed methods in the presence of several level of noise. The standard deviation (SD) between actual and estimated sound



source location can be used as a suitable objective manner to compare the stability of the proposed methods [5]. As shown in fig. 11, the SD of the proposed methods has been computed for three different SNRs (40, 25, 10 dB) and in three different FOV dimension. A lower SD points to the more stability of each method.

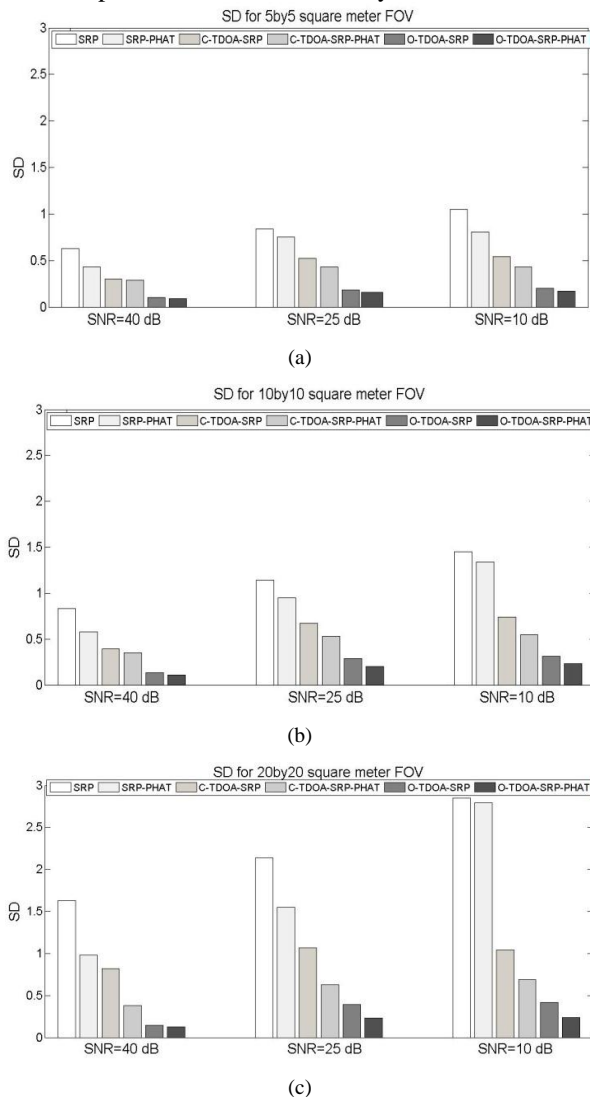


Fig. 11. Stability Comparison with Standard Division examination (each of the six columns is relevant to a SNR)  
(a):5\*5m<sup>2</sup> FOV; (b):10\*10m<sup>2</sup> FOV; (c):20\*20m<sup>2</sup> FOV

## References

- [1] K. D. Donohue, K. S. McReynolds, A. Ramamurthy, "Sound Source Detection Threshold Estimation using Negative Coherent Power," Proceeding of the IEEE, Southeast Con., pp. 575-580, April 2008.
- [2] K. D. Donohue, S. M. Saghalian Nejad Esfahani and Jingjing Yu, "Constant False Alarm Rate Sound Source Detection with Distributed Microphones," EURASIP Journal on Advances in Signal Processing, Vol.2011, no.1, pages 12, Article ID 656494, doi:10.1155/2011/656494, Mar. 2011.
- [3] N. Madhu and R. Martin, "Acoustic source localization with microphone arrays," in Advances in Digital Speech Transmission. Hoboken, NJ: Wiley, pp. 135-166, 2008.
- [4] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," EURASIP J. Appl. Signal Process., vol. 2006, pp. 1-19, 2006.
- [5] M. Cobos, Amparo Marti, Jose J. Lopez, "A modified SRP\_PHAT functional for robust real time sound source localization with scalable spatial sampling," IEEE Signal Processing Letters, vol. 18, no. 1, pp. 71-74, 2011.

Based on fig. 11, each sub-fig shows increasing the level of noise can lead to lower stability (more SD) for all proposed methods.

But, in different level of noise (different signal to noise ratio), O-TDOA-SRP-PHAT has the best stability. Furthermore, for different level of SNR, Optimal combination methods (O-TDOA-SRP and O-TDOA-SRP-PHAT) have fewer variations in SD level and more stability compared to other methods.

Comparison of figs 11 (a, b and c) show although increasing FOV's dimension reduce stability of methods but, optimal combination methods remain more stable than others.

## 6. Conclusion and Future Work

Although SRP-based methods are practical and suitable ways for sound source localization in the noisy and reverberant environment, they need valuable processing time. On the other hand, although TDOA is a low computational approach for sound source localization, this method is very noise effective.

In this paper, experimental results show a combination approach based on TDOA and SRP/SRP-PHAT methodologies optimized and simplified by reducing the initial search region, and can decrease the time of processing as well as the better suppression of noise effect. Results indicate that the proposed sound source localization methods have better robustness and lower computational time relative to the simple SRP method. This reduction has been shown to be sufficient for the development of real-time sound source localization applications. Also, results show that SRP-PHAT method can have a better performance than SRP, even when combined with a basic TDOA. Moreover, the combination of the SRP and SRP-PHAT with the TDOA method increases their stability in different signals to noise ratio level. The limitation of the proposed methods (combined methods) is the number of microphones which may make these approaches inappropriate for some practical applications. The next research challenge for authors is how the number of microphones can be reduced besides keeping the appropriate performances of the proposed methods.

- [6] K. D. Donohue, J. Hannemann, H. G. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments," *Signal Processing*, vol.87, no.7, pp.1677-1691, July 2007.
- [7] Unnikrishnan, Harikrishnan. "Audio Scene Segmentation Using a Microphone Array and Auditory Features," M.s Thesis, 2010.
- [8] N. M. Kwok, J. Buchholz. "Sound source localization: microphone array design and evolutionary estimation," *IEEE International Conference Industrial Technology (ICIT)*, pp.281-286, December 2005.
- [9] D-H. Kim, Y. Park, "Development of sound source localization system using explicit adaptive time delay estimation," *International Conference on Control Automation and Systems (ICCAS)*, Muju Resort, Jeonbuk, Korea, 2002.
- [10] I. A. Mc Cowan, "Robust Speech Recognition Using Microphone Arrays," Ph.D. dissertation, Univ. Technology, Queensland, Australia, pp.1-38, April 2001.
- [11] J.H. DiBiase, H.F. Silverman, M.S. Brandstein, "Robust localization in reverberant rooms," *Microphone Arrays, Signal Processing Techniques and Applications*, Springer, Berlin, pp. 157-180, 2001.
- [12] C. Knapp, G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoustic Speech Signal Process*, Vol.24, pp. 320-327, Aug. 1976.
- [13] J. Kuhn, "Detection performance of the smooth coherence transform (SCOT)," *IEEE International Conference on Acoustics, Speech, and Signal Processing '78 (ICASSP '78)*, Hartford, CT, Vol.3, pp.678-683, Ap., 1978.
- [14] T. Qiu, H.Wang. "An Eckert-weighted adaptive time delay estimation method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vo.44, No.9, pp.2332 - 2335, Sep., 1996.
- [15] A. Ramamurty, H.Unnikrishnan, K. D. Donohue, "Experimental Performance Analysis of Sound Source Detection with SRP PHAT  $\beta$ ," *IEEE Trans. Southeast con.*, pp. 422-427. March, 2009.
- [16] K. D. Donohue, "Audio systems array processing toolbox," (for MATLAB), Audio Systems Laboratory, Department of Electrical and Computer Engineering, University of Kentucky, (Updated: 27-10-2009), <http://www.Engr.uky.Edu/~Donohue/audio/Arrays/MAT Toolbox. Html>
- [17] J.H. Dibiase "A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays," Ph.D. Thesis, Brown University, Providence, RI, May 2001.
- [18] J. C. Murray, H. Erwin and S. Wermter, "Robotic sound-source localization and tracking using interaural time difference and cross-correlation," *Proceedings of Neurobotics Workshop*, Germany, pp.89-97, September 2004.
- [19] M. Brandstein, H. Silverman, "A Practical Methodology for Speech Source Localization with Microphone Arrays," *Computer, Speech, and Language*, Vol.11, no.2, pp.91-126, April 1997.
- [20] A. Brutti, M. Omologo, P. Svaizer, " Comparison between different sound source localization techniques based on a real data collection," in *proceedings of the Hands-free Speech Communication and Microphone Arrays (HSCMA)*, pp.69-72, May 2008 *Proceedings of the IEEE, Southeast Con.*, pp. 575-580, April 2008.
- [21] A. Pourmohammad, S. M. Ahadi, " TDE-ILD-based 2D half plane real time high accuracy sound source localization using only two microphones and source counting," *Electronics and Information Engineering (ICEIE) ,IEEE*, vol. 1, pp. 566-572, Aug 2010.
- [22] S. Astapov, J. Berdnikova, and J. S. Preden, "A method of initial searchregion reduction for acoustic localization in distributed systems," in *Proc. 20th Int. Conf. Mixed Design of Integrated Circuits and Systems (MIXDES)*, pp. 451-456, June 2013.

**Mohammad Ranjkesh Eskolaki** was born in Rasht. He received the B.Sc degree in Electronic Engineering from the department of electrical engineering, the University of Guilan, Rasht, Iran, in 2013. He is currently pursuing the M.Sc degree in Communication System from the department of communication system, KHI, Mashhad, Iran.

His current research interests include statistical signal processing and applications, digital audio processing, different sound source localization methods, MRI image processing, wireless communication channel estimation, 4G and 5G mobile systems.

**Reza Hasanzadeh** was born in Rasht, Iran, in 1978. He received the B.Sc degree (Hons) in Electrical Engineering from the University of Guilan, Rasht, Iran, in 2001 and the M.Sc and Ph.D degrees (Hons) in Electrical Engineering from the Amirkabir University of Technology (Tehran Poly-technique), Tehran, Iran, in 2003 and 2008, respectively.

He is currently an Associate Professor in the Department of Electrical Engineering, and has been the Director of the Digital Signal Processing (DSP) Research Laboratory, University of Guilan, Iran, since 2008. From 2005 to 2007, he was the Researcher of the Niroo Research Institute (NRI) in communication group, Tehran, Iran. Since 2008, he has been the Supervisory member of the Guilan Science and Technology Park. His current research interests include fuzzy logic and signal/image processing techniques for industrial and medical applications.

# Better Performance of New Generation of Digital Video Broadcasting-terrestrial (DVB-T2) Using Alamouti scheme with Cyclic Delay Diversity

Behnam Akbarian\*

Department of Electrical Engineering, Yadegar-e- Imam Khomeini (RAH) Branch, Islamic Azad University, Tehran, Iran  
be.akbarian@yahoo.com

Saeed Ghazi Maghrebi

Department of Electrical Engineering, Yadegar-e- Imam Khomeini (RAH) Branch, Islamic Azad University, Tehran, Iran  
s\_ghazi2002@yahoo.com

Received: 10/Sep/2014

Revised: 12/Mar/2015

Accepted: 17/Mar/2015

## Abstract

The goal of the future terrestrial digital video broadcasting (DVB-T) standard is to employ diversity and spatial multiplexing in order to achieve the fully multiple-input multiple-output (MIMO) channel capacity. The DVB-T2 standard targets an improved system performance throughput by at least 30% over the DVB-T. The DVB-T2 enhances the performance using improved coding methods, modulation techniques and multiple antenna technologies. After a brief presentation of the antenna diversity technique and its properties, we introduce the fact of the well-known Alamouti decoding scheme cannot be simply used over the frequency selective channels. In other words, the Alamouti Space-Frequency coding in DVB-T2 provides additional diversity. However, the performance degrades in highly frequency-selective channels, because the channel frequency response is not necessarily flat over the entire Alamouti block code. The objective of this work is to present an enhanced Alamouti space frequency block decoding scheme for MIMO and orthogonal frequency-division multiplexing (OFDM) systems using the delay diversity techniques over highly frequency selective channels. Also, we investigate the properties of the proposed scheme over different channels. Specifically, we show that the Alamouti scheme with using Cyclic Delay Diversity (CDD) over some particular channels has the better performance. Then, we exemplarily implement this scheme to the DVB-T2 system. Simulation results confirm that the proposed scheme has lower bit error rate (BER), especially for high SNRs, with respect to the standard Alamouti decoder over highly frequency-selective channels such as single frequency networks (SFN). Furthermore, the new scheme allows a high reliability and tolerability. The other advantages of the proposed method are its simplicity, flexibility and standard compatibility with respect to the conventional methods.

**Keywords:** Alamouti Coding; DVB-T2; MIMO; OFDM.

## 1. Introduction

OFDM is known as an attractive modulation scheme because of its robustness to frequency selective fading channels [1], [2]. Moreover, the MIMO is considerable to achieve diversity for combating the channel fading. Combining OFDM with MIMO technology is vital to overcome the multipath distortion, to simplify equalization and to increase the data rate [3], [4]. Thus, the Alamouti space-frequency coded was offered as a suitable multiple antenna technology. However, it requires processing at both transmitter and receiver [5], [6]. Further, some technologies like space-time coding, are constrained in the choice of number of antennas or the applicable code rate [7], [8]. The Delay Diversity (DD) is known as simple method. However, CDD is more efficient. The Alamouti method with transmit antenna selection in flat Rayleigh fading channels is presented in [9]. However, this result is valid only when the channel does not change within the same Alamouti pair [10]. Correspondingly, this means that the simple Alamouti

decoding suffers from time-variation of the channel, which is not necessarily flat over the entire Alamouti block code. Therefore, using the simple Alamouti decoding will cause performance degradation in such channels [1], [11]. On the other hand, many of the multiple antenna technologies require processing at both transmitter and receiver [12], [13]. In practice, the delay DD methods have attracted as a simple and efficient method in the past few years [14], [15]. The other diversity technique is CDD, which is a special cyclic delay diversity scheme [16].

To combat the performance degradation of the conventional Alamouti scheme, we introduce an enhanced Alamouti standard Alamouti space-frequency block code (SFBC) using CDD which is compatible with the DVB-T2 [17] system. In the proposed method we obtain a significant performance enhancement in BER without increasing the number of antenna. The other advantages of the proposed method are its simplicity, flexibility, compatibility, high reliability and tolerability with respect to the standard Alamouti scheme.

\* Corresponding Author

This paper is organized as follow. In Section 2, antenna diversity techniques which are included cyclic delay diversity, channel properties and determination of cyclic delay value is described. In Section 3, combination of the enhanced Alamouti decoding scheme and CDD is investigated. Section 4 provides simulation results for the proposed scheme with respect to two types of delay diversity techniques for MIMO-OFDM with CDD. Section 5 concludes the paper and mentioned the possible future work.

## 2. Antenna Diversity Techniques

Transmit antenna diversity play an important role to increase the robustness and reliability over wireless fading channels. There are simple approaches to achieve diversity, such as Delay Diversity which is used in our research.

### 2.1 Cyclic Delay Diversity

The Delay Diversity is a simple diversity scheme which was proposed by Witteneben in 1993 [12]. The DD techniques have regained attraction recently as simple transmit antenna diversity methods for establishing a reliable link in fading environments. In principle, the DD transmits delay diversity of a signal over further TX-antenna. Because of linearity, implementing the DD at the receiver side is possible, but it causes the channel delay spread increment. In order to overcome this drawback, the CDD method is used in which the TX-antenna specific delays are replaced with cyclic shifts [18]. The CDD scheme enhances the frequency selectivity of the channel transfer function without increasing the observable time domain channel delay spread at the receiver. Figure 1 shows the front-end of a generic MIMO-OFDM transmitter using the CDD [19],[20]. In this figure we have two data streams.

Alamouti space-time block code, was primarily proposed for MIMO system using two transmit and two receive antennas. Using STBC to the OFDM is very straightforward because individual subcarriers can be considered as independent flat fading channels. Alamouti space-time code can be represented by the following matrix:

$$s = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} \quad (1)$$

Each row indicates the transmitting antenna and each column indicates a time interval. Thus, in the first time interval  $s_1$  symbol will be transmitted from the first antenna and  $s_2$  from the second one, while in the second time interval  $-s_2^*$  will be transmitted from the first antenna and  $s_1^*$  from the second one.

In the DVB-T2 case, simple Alamouti code is used in the transmitter which enhances spectral efficiency and link reliability by maximizing the diversity in the receiver while using low complexity equalization based on

maximal ratio combining (MRC) [10]. In the first data stream, after the IFFT transform, cyclic prefix (CP) is added to each symbol by repeating the end of the same symbol. Therefore, the obtained signal  $S_0(k)$ , which is a sequence of  $N_{FFT}$  data symbols of  $S_b$ , for  $l=0, \dots, N_{FFT}-1$ , is transmitted via TX-antenna zero without delay. In the other data stream, the CDD scheme is implemented for providing additional diversity; therefore the system performance will be improved. In order to keep the average transmission power independent of the number of related TX-antennas ( $N_T$ ), the data stream is normalized by  $(N_T-1)^{0.5}$  factor. Before inserting the CP, the symbol is shifted cyclically, which results in the antenna specific TX-signal as below

$$S_i(k) = \frac{1}{\sqrt{N_T-1}} \tilde{S}(K - D_i^{cyc} \text{ mod } N_{FFT}) \quad (2)$$

for  $i=0, \dots, N_T-1$  and  $k=-N_G, \dots, N_{FFT}-1$ , where  $N_G$  denotes the length of the CP. Also,  $\tilde{S}(k)$  is the sample of an Alamouti symbol, which is modulated in time domain and then the signal is shifted cyclically by  $D_i^{cyc}$  before the CP is added. After guard interval removing at the receiver, as long as TX-antenna specific delays  $D_i$  is equal to cyclic shifts  $D_i^{cyc}$ , both CDD and DD yield the same signal. The OFDM symbols of the DD signal partly overlap the guard interval of the subsequent OFDM symbol at about  $D$  which is a restriction in the choice of  $D$ . To avoid inter symbol interference (ISI), the  $N_G$  must be

$$N_G \geq \begin{cases} L_{max} + \max_i D_i & \text{for DD} \\ L_{max} & \text{for CDD} \end{cases} \quad (3)$$

where  $L_{max}$  is the maximum channel delay samples. Based on (2) the minimum length of the CDD guard interval does not depend on the cyclic delays  $D_i^{cyc}$ , and we can choose shorter guard interval [4]. In other words, the CDD scheme does not depend on the number of TX-antenna and there is no overlapping of OFDM symbols. However, when the received signal is free of the ISI, the DD has the same performance as the CDD. Note that, the MIMO channel is changed to a single-input multiple-output (SIMO) channel via CDD method, i.e. the spatial diversity is transformed into frequency diversity [4].

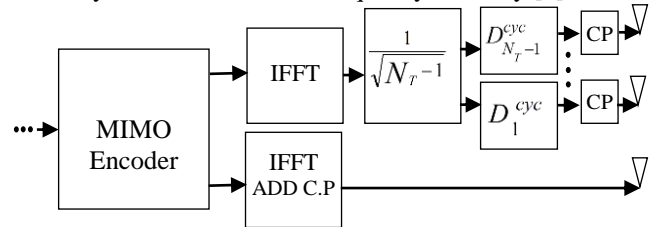


Fig. 1. Front-end of a generic MIMO-OFDM transmitter using CDD

### 2.2 Channel Properties of Cyclic Delay Diversity

In order to analyze the properties of CDD, we consider the received time domain signal as

$$r(k) = \frac{1}{\sqrt{N_T}} \sum_{i=0}^{N_T-1} \sum_{p=0}^{N_{FFT}-1} s(k-p-D_i^{cyc} \text{ mod } N) h_i(k,p) \quad (4)$$

where  $k = 0, \dots, N-1$  and  $h_i(k,p)$  denotes the time domain fading process (time index is  $k$ ) for TX-antenna  $i$



and path delay  $p$  (in samples) [21]. Note that, to simplify the equation (3), we neglect additive white Gaussian noise (AWGN). We desired to obtain effective channel transfer function subcarrier fading correlation. Thus, equation (3) is transformed into the frequency domain and is given by ...

$$R(l) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} r(k) e^{-j2\pi kl/N} = \frac{1}{\sqrt{N \cdot N_T}} \sum_{v=0}^N S(v). \quad (5)$$

$$\sum_{i=0}^{N_T-1} \sum_{p=0}^{N_{\max}} H_i(l - v \bmod N, p) \cdot e^{-j2\pi(p+D_i^{\text{cyc}})v/N}$$

Where

$$H_i(q, p) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} h_i(k, p) \cdot e^{-j2\pi kq/N} \quad (6)$$

is the  $N \times N$  discrete Fourier transform (DFT) of the time variant channel fading process  $h_i(k, l)$  with respect to time index  $k$ . We can use this expression as the transferred channel at the receiver and it can help us to analyze the properties of channel in order to modify it. Note that, if the fading processes  $h_i(k, p) = h_i(p)$  are constant for the duration of an OFDM symbol we have ICI-free transmission over subcarriers [16]. Then if we consider a quasi-static fading, the received signal will be

$$R(l) = S(l) \cdot \frac{1}{\sqrt{N_T}} \cdot \sum_{i=0}^{N_T-1} e^{-j2\pi \cdot D_i^{\text{cyc}} \cdot l/N} \cdot \sum_{p=0}^{N_{\max}} h_i(p) \cdot e^{-j2\pi pl/N} \quad (7)$$

Then, we rewrite equation (6) and simplify as

$$\tilde{H}(l) = \sqrt{\frac{N}{N_T}} \sum_{i=0}^{N_T-1} H_i(l) \cdot e^{-j2\pi \cdot D_i^{\text{cyc}} \cdot l/N} \quad (8)$$

which is the equivalent channel transfer function and it can clearly specify the CDD effects. Also,  $H_i(l)$  are the frequency domain channel fading coefficients for subcarrier  $l$  of the OFDM symbol from TX-antenna  $i$  to the receiver [21]. Note that,  $H_i(l)$  is the DFT of  $h_i(k, p)$  with respect to the path delay  $p$  and the quasi-static fading assumption (i.e.  $h_i(k, p) = h_i(p)$ ). Finally,  $H_i(q, p)$  is the DFT of  $h_i(k, p)$  with respect to the time variable  $k$  and path delay  $p$  as a parameter.

### 2.2.1 Correlation Properties

The expectation  $E\{H^*(l) \cdot H^*(l+v \bmod N)\}$  yields the correlation properties of the frequency domain channel fading, where  $(\cdot)^*$  means complex conjugate. We assume these processes are stationary. In other word, identical statistical channel properties from TX-antenna  $i$  to the receiver,  $\sigma_p$  does not depend on TX-antenna index  $i$ . Then, we obtain the fading correlation as

$$\phi_{(v)} = E\{\tilde{H}^*(l) \cdot \tilde{H}(l+v \bmod N)\} \quad (9)$$

$$= \frac{1}{N_T} \sum_{i=0}^{N_T-1} e^{-j2\pi \cdot D_i^{\text{cyc}} \cdot v/N} \sum_{p=0}^{N_{\max}} \sigma_p^2 \cdot e^{-j2\pi \cdot p \cdot v/N}$$

That means,  $\sigma_p^2 = E\{|h_i(p)|^2\} \cdot \phi^{\text{Ch}}(v)$  is the fading correlation function in frequency direction for the component channels, i.e. the channel observed from the TX-antenna  $i$  to the receiver antenna. Therefore, equation

(8) allows the design of roots for the correlation function which leads to find the value  $v$  independent of the multipath channel properties  $\phi^{\text{Ch}}(v)$  [16].

### 2.3 Determination of Cyclic Delay Value

In this section we investigate the impact of the value of  $D_i^{\text{cyc}}$ . By increasing cyclic delays with a constant increment ( $D$ ) as

$$D_i^{\text{cyc}} = D \cdot i, \quad i = 0, \dots, N_T - 1 \quad (10)$$

we get

$$\phi^{\text{CDD}}(v) = \frac{1}{N_T} \cdot e^{-j\frac{\pi \delta (N_T-1)v}{N}} \cdot \frac{\sin(\frac{\pi \delta N_T}{N} \cdot v)}{\sin(\frac{\pi \delta}{N} \cdot v)} \quad (11)$$

Then, it can be considered for two cases:

i)  $N_T$  is a divider of  $N$ :

we set  $\delta = N/N_T$  and (10) simplifies to

$$\phi^{\text{CDD}}(v) = \delta(v \bmod N_T) \quad (12)$$

Thus, a subcarrier fading process is uncorrelated to its  $N_T - 1$  neighbors.

ii) we choose  $\delta$  such that  $\delta \cdot N_T$  is a divider of  $N$ :

with  $\eta = N/\delta \cdot N_T$

$$\phi^{\text{CDD}}(v) = \frac{1}{N_T} \cdot e^{-j\frac{\pi (N_T-1)v}{\eta \cdot N_T} \cdot \frac{\sin(\pi v/\eta)}{\sin(\pi v/\eta N_T)}} \quad (13)$$

## 3. Combination of Enhanced Alamouti Decoding Scheme and CDD

In the latest wireless broadcast systems, such as DVB-T2, in order to increase bandwidth efficiency, multiple antenna technologies are used [17]. The block diagram in Figure.2 illustrates the proposed scheme which is a genetic MIMO-OFDM DVB-T2 system using CDD. At the DVB-T2 transmitter side, after error protection, mapping and interleaving, the MIMO coding is performed [17]. Then each independent data stream together pilots and transmission parameter signaling (TPS) data are arranged in an OFDM frame. In this paper, we focus on using of Alamouti STBC, that was primarily introduced for MIMO scenario with two transmit and two receive antennas. With assumption of  $S_k, S_{k+1}$  as two successive subcarriers, the MIMO encoder outputs two pairs,  $[S_k, -S_{k+1}^*]$  for the first data stream, and  $[S_{k+1}, S_k^*]$  for the second data stream. In this paper, we combine the enhanced Alamouti decoding and the CDD scheme. For this propose, considering one MIMO-OFDM symbol in (1), the  $S_0(k)$  is transmitted as un-shifted ( $D_0^{\text{cyc}} = 0$ ) signal over TX-antenna 0. The other

TX-antenna signal is shifted cyclically by  $D_i^{\text{cyc}}$  before CP is added due to Fig. 2. In the receiver side, after removing the CP, the remaining OFDM time domain symbol is transformed into frequency domain by an FFT which yields to

$$Y = S\tilde{H} + N \quad (14)$$

where

$$\tilde{H} = \frac{1}{\sqrt{N_T}} \sum_{i=0}^{N_T-1} \sum_{k=0}^{N_{FFT}-1} h_i e^{\frac{j2\pi \cdot kl}{N_{FFT}}} e^{-\frac{j2\pi \cdot D_i^{cyc} \cdot l}{N_{FFT}}} + N \quad (15)$$

where  $N$  is complex Gaussian noise which is mutually uncorrelated between different subcarriers. Also,  $h_i$  are the taps of the channel impulse response and  $\tilde{H}$  denotes an equivalent channel transfer function. It means that a receiver cannot distinguish whether a propagation path results from CDD or the channel itself [4]. In the channel estimation, known pilots are used for estimation of the complex valued channel fading coefficients for each subcarrier. After FFT transformation and pilot extraction, the received signal will be:

$$\begin{cases} Y_1^K = S_k h_{1,1}^k + S_{k+1} h_{1,2}^k + N_1^K \\ Y_2^K = S_k h_{2,1}^k + S_{k+1} h_{2,2}^k + N_2^K \\ Y_1^{K+1} = -S_{k+1}^* h_{1,1}^{k+1} + S_k^* h_{1,2}^{k+1} + N_1^{k+1} \\ Y_2^{K+1} = -S_{k+1}^* h_{2,1}^{k+1} + S_k^* h_{2,2}^{k+1} + N_2^{k+1} \end{cases} \quad (16)$$

where  $h_{i,j}^k$  is the channel frequency response at the carrier  $k$ th, between the  $i$ th receive antenna and  $j$ th transmit antenna, and  $N_i^k$  is AWGN noise at the  $k$ th carrier of the  $i$ th receive antenna. Also,  $*$  denotes conjugate operator. In the standard Alamouti algorithm, it is assumed that the channel frequency response is constant over two consecutive carriers, i.e.  $h_{i,j}^k = h_{i,j}^{k+1}$ . However, the performance of the standard Alamouti decoding degrades in highly frequency-selective channels since quasi static fading for adjacent subcarriers is required. To tackle this drawback, we propose an efficient Alamouti decoding scheme using CDD. Let us consider the changes in the channel matrix over adjacent OFDM carriers which are described as:

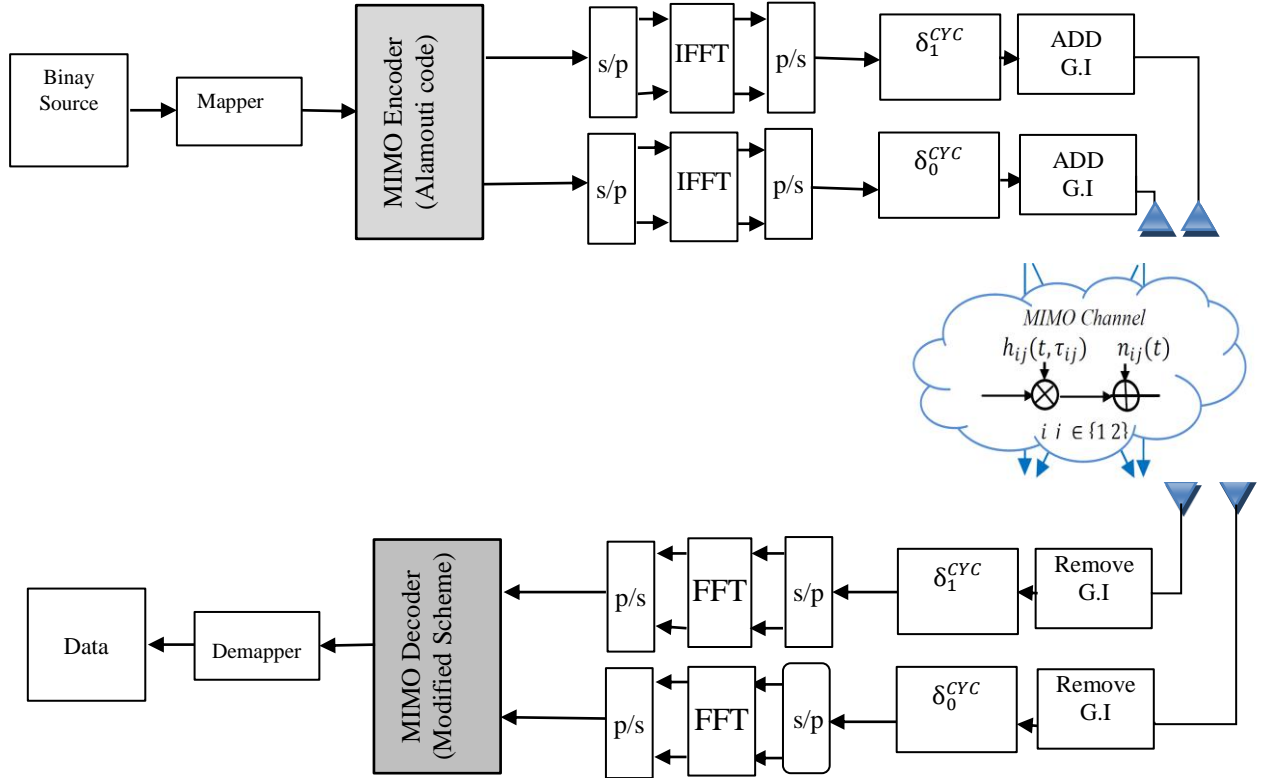


Fig. 2. Transmitter and receiver sides of proposed scheme

$$\Delta h^T = \begin{bmatrix} 0 & 0 & (h_{1,2}^{k+1} - h_{1,2}^k)^* & (h_{2,2}^{k+1} - h_{2,2}^k)^* \\ 0 & 0 & -(h_{1,1}^{k+1} - h_{1,1}^k)^* & -(h_{2,1}^{k+1} - h_{2,1}^k)^* \end{bmatrix} \quad (17)$$

Where, the  $T$  refers to the conjugate transpose. Thus, there is not any necessity to consider the channel constant over two consecutive subcarriers [1]. Then, the estimated transmitted symbol in minimum mean square error (MMSE) criterion will be:

$$\tilde{S} = \frac{1}{\det(I_2 + B\Delta h)} \text{adj}(I_2 + B\Delta h) \cdot B \cdot Y \quad (18)$$

where,

$$B = (h^T h)^{-1} h^T \quad (19)$$

Where  $\det()$  and  $\text{adj}()$  denote the determinant and the adjoint of matrix of  $(I_2 + B\Delta h)$ , respectively. Then, the transmitted data can be recovered at the receiver through (9). It is considerable that the new scheme has a lower complexity since in (9) does not consist of direct inverse of the matrix  $(I_2 + B\Delta h)$ .

#### 4. Simulation Results

In this paper, we applied DD and CDD with different  $D_i$  to the MISO systems ( $2 \times 1$ ). We use the  $2k$  mode (2048 points IFFT) with 4-QAM modulation. The guard interval length is  $N_G = 1/4$ . At the receiver sides we assume that the channel estimation is perfect. For our investigations, we use 9-path multipath Rayleigh channel models which is similar to Indoor Commercial-channel B models large open centers, such as shopping malls and airports. Its power-delay profile is  $[0-8-17-19-21-23-25]_{\text{dB}}$ . The fading processes for the several propagation paths are statistically independent. Figure 3 compares the performance of MISO with DD and CDD. As can be seen for short guard interval, DD and CDD perform the same such as  $D_i = 5$  samples. In other word, the performance of DD and CDD is equal if the transmission is ISI free. However, by increasing guard interval, using the MISO-DD method, causes noticeable performance degradation at the receiver. For instance, at  $\text{BER} = 10^{-4}$  a MIMO-CDD system with  $D_i = 1$  requires an SNR of 17.2. But,  $D_i = 3$  and  $D_i = 5$  require SNRs of 23.5 and 24.5 respectively. There is a fact to take into account that if the guard interval is increased for DD, we obtain a huge amount of interference.

To study the property compatibility of MIMO scheme (Fig.2) with the DVB-T2 system, a MIMO-OFDM system is compared to a MIMO-OFDM with CDD system. For this purpose, we implement the proposed scheme (Fig.2). It worth mentioning that in [22] both methods of CDD and Alamouti are used separately for the channel DRM+ system at a velocity. In [22], the author shows that the CDD performs better than simple Alamouti in fast fading channels [22]. We use 15-path multipath Rayleigh fading channel model which is similar to the pervious channel model. We applied 4-QAM, 8k length with a guard interval of  $1/4$ . Channel estimation is assumed to be perfect.

We observe that the performance of the proposed method is much better than the Alamouti scheme in DVB-T2 system (Fig. 4). In this simulation, we use enhanced decoding Alamouti which is introduced in [1]. Compared to a MIMO-OFDM system which requires an SNR of 21 dB at  $\text{BER} = 10^{-3}$ , a MIMO-OFDM system using CDD requires 16.5 dB, which provides a gain of 4.5 dB for the considered DVB-T2 parameter set (Fig. 4).

In order to verify the better performance of the proposed enhanced Alamouti decoder using CDD with respect to the Alamouti decoder in [1], we consider the MIMO-OFDM system based on the DVB-T2 standard. An 8k mode where 8192 sub-carriers with 4-QAM are used. We applied 10000 OFDM symbols with a guard interval of  $1/4$  in a MIMO DVB-T2 system ( $2 \times 2$ ). We used the P1 multipath Rayleigh fading channel model, which is described in [5]. The channel has 20 taps without any Doppler effects. We also evaluate the performance of uncoded scenario; thus the coding and interleaving are bypassed in the system simulation. We applied four profiles as MIMO channels with the same power and different phase, i.e. they are randomly chosen between 0

and  $2\pi$  in order to make the uncorrelated channel. In order to simulate highly frequency selective channels, random phase with different values for  $a$  and  $b$ , as the power of the delayed channel and the delayed spread respectively, are used. In principle, increment in the delay spread and the power of delay channel converts the channel to a highly frequency-selective channel [1]. Fig. 5 shows the results of the efficient Alamouti decoding scheme using CDD versus the introduced method in [1] and the standard one. We chose  $D_i^{\text{cyc}} = 2$ , because as shown in [4], the SNR does not increase further when  $D_i^{\text{cyc}} > 1.5\mu\text{s}$ . In this simulation channel estimation is assumed to be perfect. Fig. 5 shows that for  $a=0.5$ , the proposed scheme performs approximately the same as the scheme in [1]. On the other hand, the performance of standard decoder degrades. For  $a=0.9$ , we observe that the performance of the proposed scheme is better than the scheme in [1]. This improvement is a fact because the proposed scheme increases the degree of diversity. The proposed scheme allows a significant performance enhancement that can be achieved without increasing the number of antenna. However, we decrease BER by using an enhanced Alamouti-CDD. In other word, even for such channels the performance of the proposed scheme is still much better than the other schemes. For instance, this figure shows that the proposed scheme outperforms the scheme in [1] by 3.2dB at an average BER of  $10^{-3}$ .

#### 5. Conclusions

In this paper, we have shown that the performance of well-known Alamouti decoding degrades in the highly frequency selective channel. To tackle this problem, we proposed an enhanced Alamouti decoder using CDD scheme. The proposed modification in the standard Alamouti decoding caused a high tolerability of the system in the frequency selective channels such as the SFN channels. Simulation results confirm that the new introduced scheme has significant performance improvement. Therefore, the new scheme is useful in MIMO-OFDM systems such as the DVB-T2 standard which allows a high reliability and capacity enhancement.

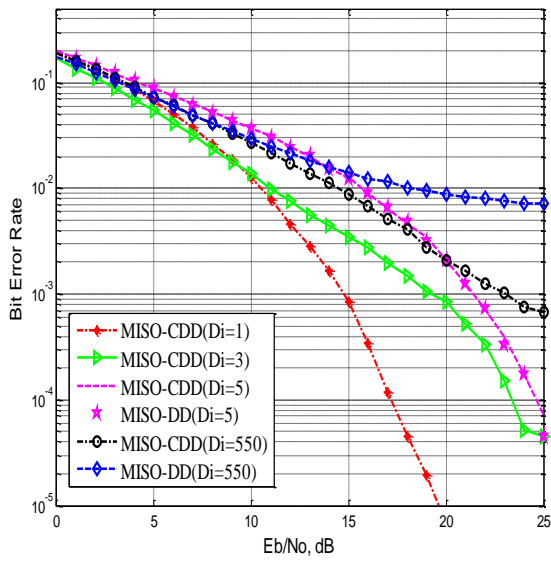


Fig. 3. a comparison of SNR in CDD and DD respect to  $D_i$

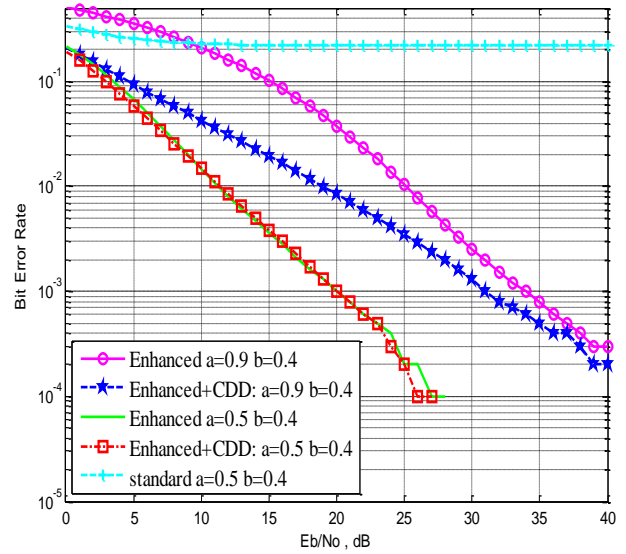


Fig. 5. The performance of different schemes in BER vs. SNR.

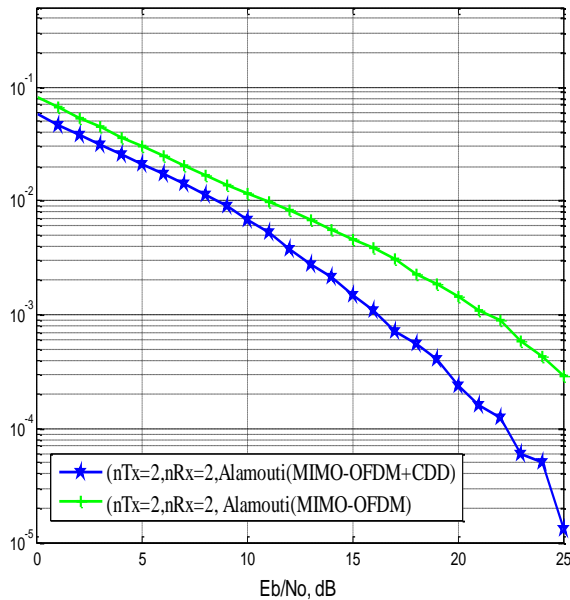


Fig. 4. The performance of DVB-T2 system by using Alamouti scheme and CDD.

### References

- [1] A. Omri, R. Hamila, A. Hazmi, R. Bouallegue, A. Al-Dweik, "Enhanced Alamouti Decoding Scheme for DVB-T2 Systems in SFN Channels", IEEE Conf. PIMRC, Toronto, 2011, pp.1626–1630.
- [2] G. Bauch, "Aspects of Delay Diversity in OFDM", African Journal of Information and Communication Technology, vol.2, no.1, 2006, pp.12-20.
- [3] J.S. Baek, J.S. Seo, "Effective Symbol Timing Recovery Based on Pilot-Aided Channel Estimation for MISO", IEEE Transaction on Broadcasting, 2010, vol. 56, no. 2.
- [4] E. Biglieri, R. Calderk, A. Constantinides, A. Goldsmith, A. Paulra, A. Poor, "MIMO Wireless Communications", Cambridge University Press, New York, 2005.
- [5] S.H. Ahmed, S.M. Umar Talha, A. Khan, "Performance Evaluation of DVB-T Based OFDM over Wireless Communication Channels", Proceedings of the International Multi Conference of Engineer and Computer Scientists, IEEE, Hong Kong, vol.1, 2012, pp.1-4,.
- [6] M. Tormos, C. Tanougast, A. Dandache, D. Masse, P. Kasser, "Modeling and performance evaluations of Alamouti technique in a single frequency network for DVB-T2", EURASIP Journal on Wireless Communications and Networking, no.1, 2013, pp.1499-2013.
- [7] N. Cornillet, M. Crussi'ere, J.F. H'elard, "Performance of the DVB-T2 System in a Single Frequency Network:

- Analysis of the Distributed Alamouti Scheme", BMSB, IEEE, Nuremberg, 2011, pp.1-4.
- [8] A. Dammann, S. Kaiser, "Low complex standard conformable antenna diversity techniques for OFDM systems and its application to the DVB-T system", ITG FACHBERICHT, 2002, pp. 253- 260.
- [9] Z. Chen, J. Yuan, B. Vucetic, Z. Zhou, "Performance of Alamouti scheme with transmit antenna selection", *Electronic Letters*, 2003, vol. 39, pp.1666-1668.
- [10] S M. A. "Alamouti, Simple Transmit Diversity Technique for Wireless Communications", *IEEE Journal, Selected Areas in Communications*, 1998, vol.16, pp.1451 – 1458.
- [11] D. Gozálviz, D. Gómez- Barquero, D. Vargas, N. Cardona , "Time Diversity in Mobile DVB-T2 Systems", *IEEE Transaction on Broadcasting*, 2011, vol.57, no.3, pp.617-628.
- [12] S. Kaiser, "Spatial Transmit Diversity Techniques for Broadband OFDM Systems", *IEEE. GLOBECOM*, 2002, vol.3, no.1, pp.1824-1828.
- [13] Y. Zhang, J. Cosmas, K. Loo, M. Bard, R.D. Bari, "Analysis of Cyclic Delay Diversity on DVB-H Systems over Spatially Correlated Channel", *IEEE Tra. Broadc*, 2007, vol. 53, no. 1, pp- 247 – 255.
- [14] S. Plass, A. Dammann, "Channel Correlation Properties in OFDM by using Time-Varying Cyclic Delay Diversity", *Journal of Communication*, 2008, vol. 3, no1.3, pp.19-26.
- [15] M. Liu, M. Crussi`ere, M. H´elard, J.F. H´elard, "Enhanced Mobile Digital Video Broadcasting with Distributed Space-Time Coding", *ICC Workshop on Telecommunications: From Research to Standards*, Ottawa Canada, 2012, vol.1, pp.6971 - 6976.
- [16] S. Plass, A. Dammann, S. Sand, "An Overview of Cyclic Delay Diversity and its Applications", *Vehicular Technology Conference (VTC)*, IEEE 68th, Calgary, BC, 2008, pp.1 – 5.
- [17] ETSI, 'Digital Video Broadcasting (DVB); "Implementation guidelines for a second generation digital terrestrial television broadcasting system (DVB-T2)", ETSI, TS 102 831 V1.2.1, 2012.
- [18] A. Dammann, R. Raulefs, S. Plass, "Soft Cyclic Delay Diversity and its Performance for DVB-T in Ricean Channels", *GLOBECOM '07. IEEE*, 2007, pp . 4210 -4214.
- [19] A. Dammann, S. Plass, S. Sand, "Cyclic Delay Diversity — A Simple, Flexible and Effective Multi-Antenna Technology for OFDM", *Spread Spectrum Techniques and Applications*, IEEE 10th International Symposium, Bologna, Italy, 2008, pp.550 - 554.
- [20] C.G. Calero, L.C. Navarrete, L.de Haro, R. Martínez, "A 2x2 MIMO DVB-T2 System: Design, New Channel Estimation Scheme and Measurements with Polarization Diversity", *IEEE Transaction on Broadcasting*, 2011, vol. 57, no. 2, pp. 184 -192.
- [21] A. Dammann, S. Plass, "Cyclic delay diversity: Effective channel properties and applications." *Proceedings IEEE International Conference on Communications (ICC)*, Glasgow, Scotland, vol. 1, pp. 1-4.
- [22] H.Schulze, "A Comparison between Alamouti Transmit Diversity and (Cyclic) Delay Diversity for a DRM+ System." *proceeding of international OFDM workshop*, 2006, vol 1. Pp.1-6

**Behnam Akbarian** received the B.Sc degree in Electrical Engineering from Islamic Azad University (Majlesi Branch) in 2008 and the M.sc degree in Telecommunication Systems from Islamic Azad University (Yadegar- e- Imam Khomeini) in 2013. His main research interests include digital communications with emphasis on MIMO-OFDM, antenna diversity techniques for broadcast systems and space-time coding.

**Saeed Ghazi Maghrebi** is a professor of Communication Engineering department of Azad University, Yadegar-e-Imam Khomeini (RAH) Share-rey Branch, Tehran, Iran. He was born in Iran in 1963. He received the B.S degree in Electrical Engineering from Kerman University, Iran, in 1988, and the M.Sc degree from the Khajeh Nasir-edin- toosi University of Technology, Iran, in 1995. He received the PhD degree in Digital Communications from Islamic Azad University, Iran in 2010. His current research interests are digital communication, signal processing and adaptive filtering.

# Online Signature Verification: a Robust Approach for Persian Signatures

Mohammad Esmaeel Yahyatabar

Department of Electrical and Computer Engineering, Babol University of Technology, Babol, Iran  
me.yahyatabar@stu.nit.ac.ir

Yasser Baleghi\*

Department of Electrical and Computer Engineering, Babol University of Technology, Babol, Iran  
y.baleghi@nit.ac.ir

Mohammad Reza Karami

Department of Electrical and Computer Engineering, Babol University of Technology, Babol, Iran  
mkarami@nit.ac.ir

Received: 22/Jun/2014

Revised: 24/Dec/2014

Accepted: 17/Jan/2015

## Abstract

In this paper, the specific trait of Persian signatures is applied to signature verification. Efficient features, which can discriminate among Persian signatures, are investigated in this approach. Persian signatures, in comparison with other languages signatures, have more curvature and end in a specific style. An experiment has been designed to determine the function indicating the most robust features of Persian signatures. To improve the performance of verification, a combination of shape based and dynamic extracted features is applied to Persian signature verification. To classify these signatures, Support Vector Machine (SVM) is applied. The proposed method is examined on two common Persian datasets, the new proposed Persian dataset in this paper (Noshirvani Dynamic Signature Dataset) and an international dataset (SVC2004). For three Persian datasets EER value are equal to 3, 3.93, 4.79, while for SVC2004 the EER value is 4.43.

These experiments led to identification of new features combinations that are more robust. The results show the overperformance of these features among all of the previous works on the Persian signature databases; however, it does not reach the best reported results in an international database. This can be deduced that language specific approaches may show better results.

**Keywords:** Online Signature Verification; Support Vector Machine; Robust Feature Extraction; Online Signature Dataset.

## 1. Introduction

Nowadays, biometric methods are more considered for identification. These methods strongly depend on inherent characters of people, thus they are highly reliable. A specific signature such as other biometric features is exactly associated to a specific person and this unique feature is applied to identification and verification. IBG1[1] reported that the signature modality is the second behavioral trait in commercial importance just after voice biometrics. Applications of online signature verification in legal (document authentication), medical (record protection), and banking sectors (cheque and credit card processing) are so common and increasing [2].

Signature verification consists of two types including static and dynamic verification. Shapes of signatures are available in static signature verification, i.e. recognition has to be done on a two-dimensional shape and the final decision is based on signature appearance. However, dynamic features are considered as well as appearance features in dynamic verification. In this method, Pressure sensitive tablet records the 2D coordination, pressure, Azimuth and Altitude of signatures in specific intervals of time.

Generally, feature extraction in signature verification is categorized into parametric and functional types [3]. In functional type, time sequences describing local properties of the signature are used for recognition[2], whereas other features such as global and shape-based features are included in parametric category.

### 1.1 Parametric Features

Many researchers have worked with different methods on parametric features. Pippin[4] used global features such as average pressure, average velocity, pen tip and number of curves. Curves in signatures are extracted and compared with Dynamic Time Warping (DTW) to find similarity between reference signature and input signature. With specifying a threshold value, similarity is analyzed and decision-making is done. Deghani [5] presents a two-phases method for Persian signatures. The first phase consists of feature extraction based on fractal vector of signature pressure. Then the Adaptive Network based Fuzzy Inference System (ANFIS) is applied to classify Persian signatures, which passed first phase. Alizadeh [6] extracts 62 parametric features such as total signing duration, signature height, maximum of x and y and associated time. Extracted features were compared with two threshold values in two-stages in classification.

---

<sup>1</sup> International Biometric Group

\* Corresponding Author



## 1.2 Functional Features

As mentioned before, some functions can be achieved from the tablet data package. Nanni [7] implements discrete wavelet transform on functions extracted from signature and decreases its dimensions using Discrete Cosine Transform (DCT). Gained vectors are considered as classifier inputs. One-dimensional signals of  $X(t)$  and  $Y(t)$  have been processed in parallel in [8]. The signature has been pre-processed by Mellin transform being scale invariant. Feature vectors have been extracted using Mel Frequency Cepstral Coefficient (MFCC) and feature dimensions have been decreased using Principal Component Analysis (PCA). Finally, decision is made using linear classifier. Muhammad Khan [9] proves that parts including middle velocity can be suitable criteria for proper verification. Middle velocity is categorized into low middle velocity and high middle velocity. Classifier inputs are considered using middle velocity features.

Marianela et al. [10,11] studied the discriminative power of combinations of most commonly time functions related to signing process. A consistency factor is defined to quantify the discriminative power of these different feature combinations. They presented experimental results that show there is a good correlation between the consistency factor and the verification errors, suggesting that consistency values could be used to select the optimal feature combination.

In [12], a new partitioning method is proposed for online signature verification. The partitions represent areas of high and low speed of signature and high and low pen's pressure. The method is performed on SVC2004 and BioSecure databases.

The One-Class Support Vector Machine (OC-SVM) based on independent parameters is used in [13]. This method is proposed for the situation when the forgery signatures are lack as counterexamples. In order to reduce the misclassification, a modification of decision function used in the OC-SVM is suggested.

In [14] a minimum distance alignment between the two signatures is made using dynamic time warping technique that provides a segment to segment correspondence. Fuzzy modelling of the extracted features is carried out in the next step. The experiments are carried out on SVC2004 and SUSIG databases.

The rest of the paper is organized as follows. Datasets are briefly described in the next section. In section 3, proposed method is presented. The results are described in section 4. Section 5 discusses about the proposed method and the conclusion and future works are given in the last section.

## 2. Datasets

To analyze the proposed method for Persian signatures, two datasets are used from [5] and [15] and a new own dataset named Noshirvani Dynamic Signature Dataset (NDS) is produced and will be published. The

method that is used for gathering data is the same as what is applied in SVC2004 [16] international dataset.

Datasets from [5] are the first used datasets in which signatures are generated from 40 people as well as 10 signatures per person are involved. The second dataset [15], the next used dataset in this paper, involves dynamic data that is generated from 50 signatures. Each person has registered 25 signatures and there are 15 forgery signatures per sample.

Specific features of Persian signatures are tested to identify and analyze in this study. This method has been applied for international dataset called SVC2004. The mentioned dataset includes 1600 signatures generated from 40 people in which 20 forgery and 20 genuine signatures are involved per person [16].

### 2.1 Data Acquisition (NDS)

The signatures of NDS database -newly introduced in this work- in are acquired by WACOM INTUOS4 digitizing tablet. The tablet sent a data package including pen tip coordination, pressure, azimuth and altitude angles (see figure 1). Data are sent every 10 milliseconds and signers sign in a plate with size of  $129 \times 96$  mm. The pen senses 2048 levels of pressure. The interface software was programmed by visual basic software and the tablet was connected to computer with USB port.

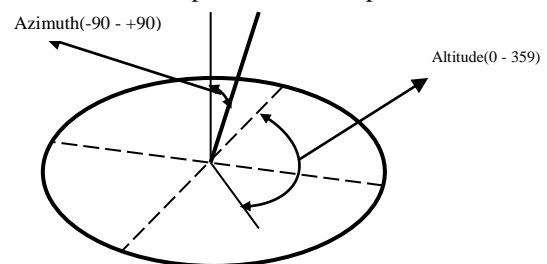


Fig. 1. Azimuth and Altitude angles.

NDS dataset was prepared in Digital Signal Processing Laboratory of Babol Noshirvani University of technology. 55 students of computer and electrical engineering department participated in producing the dataset. Each person signed 65 signatures in two different times with more than 3 days interval. People signed in two situations of standing and sitting. The dataset users were in range of 18 to 40 years old. Seven signers of them were left handed and 23 signers were female.

Fifteen professional forgers forged all signatures. Two types of professional forging were performed. In the first type, the signers could see just the shape of the genuine signatures of people and had enough time to practice for forging. In this type of forging, forger person should guess the signature path and other dynamic features. In the second type, Forgers had dynamic information of signatures and tried to forge shapes and dynamic information of signatures. The signatures path and pen tip velocity were animated for forgers and the signatures intensities were proportional with the pen pressure on the tablet. In both types, forgers tried at least 15 times to

forge signatures before recording the dataset signatures. Ultimately, forty forgery signatures were recorded from four forgers assigned for each genuine signature. In figure 2 some genuine and forgery signatures from NDS dataset are illustrated.

NDS dataset are illustrated.

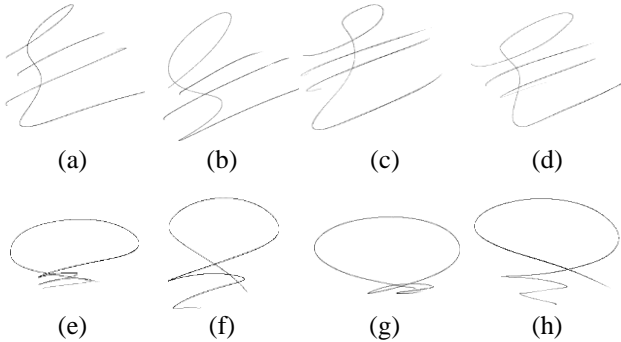


Fig. 2. figures 2.(a,b) , 2.(e,f) are genuine signature and 2.(c,d) , 2.(g,h) are their forgery signatures respectively.

### 3. Proposed Method

The proposed biometric verification system in this paper can be sketched as in figure 3. Signatures acquisition from the set, extracting features and making decision are three main stages of the method.

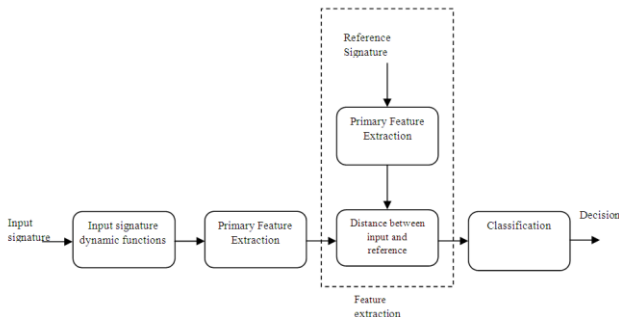


Fig. 3. Block Diagram of the proposed method.

In the proposed method, at first, specified features of Persian signatures are considered. The stability experiment of dynamic features has been carried out on Persian available signatures. Robust and reliable features of all Persian signatures are recognized. Besides, the results obtained from the experiment as well as two other features are the base of primary feature extraction. SVM classifier inputs are the distance between the input signature primary features and the reference signature features. The classification will be improved due to assigning small numbers to all genuine signatures and large numbers to forgeries through this process.

In the following sections, the proposed algorithm is described in detail:

#### 3.1 Persian Signatures

Persian signatures are significantly different from other language signatures. In other languages, the shape of signatures is close to the names whereas Persian signatures

are made by some lines, curves, and signs and almost different from the people's names. Some of the most important features in signatures of some languages such as Persian language can be pointed as following features:

- Using more curves in signatures
- More discrete lines than other languages
- Distribution on length and width (against some languages that are on a straight line)

Figure 3 (a-d) illustrates some Persian signatures from NDS dataset. Some international signatures from SVC2004 are shown in figure 4 (e-h). The different characteristic mentioned above can be seen in the figure.

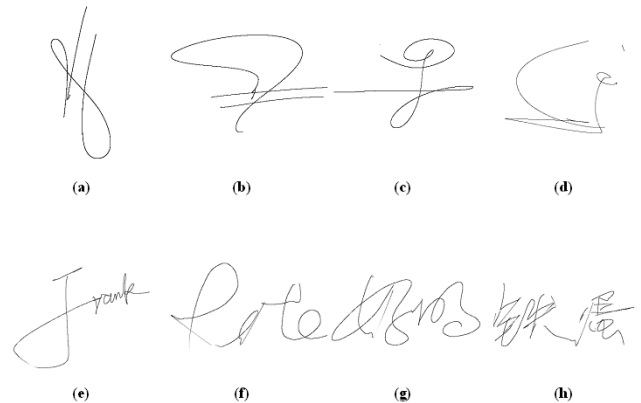


Fig. 4. Signatures 4.a to 4.d are Persian signatures from NDS Persian dataset and in 4.e to 4.h includes some samples of SVC2004 international dataset.

Figure 4 illustrates more curves in Persian signatures that are mentioned before. While signing, the velocity of pen has special state and usually the velocities in these curves are more than the others. In addition, it is deemed that a Persian signer is moving his/her wrist and fingers more than other languages. This additional motion may lead to more discrepancy in various iterations. However, considering the smooth motion of hand, dynamic features of signatures are persistent enough in the specific zone of signature (i.e., curves). The experiments of this study indicate that all signatures with the motion on vertical and horizontal directions have close dynamic features in specific curves. This issue triggers the authors to do more experiments regarding dynamic features of signatures.

The general idea of this work is to find velocity, acceleration and pressure functions of signatures, segmenting the functions to different ranges and finally a comparison between segmented functions of corresponding curves of input signatures and the reference signature.

However, some questions should be answered. What ranges of these functions can be selected for this experiment? How to find the reference signature? Scale and recording angle variant are another hazard in this work. These issues are to be discussed in the rest of the paper

#### 3.2 Dynamic Features Stability Experiment

The introduced device is used to record the dynamic signatures outputs including the coordination, pressure,

Azimuth and altitude in the specified interval. In this work, coordination and pressure of points are used.

Pressure: two resolution levels of 1024 or/and 2048 for pressure of points are directly available.

Velocity: considering the constant interval of time for signature record, as in equations (1) it is possible to calculate their velocity by calculating the difference between length and width.

$$\left. \begin{aligned} V_x(i) &= x(i) - x(i-1) \\ V_y(i) &= y(i) - y(i-1) \end{aligned} \right\} V(i) = \sqrt{V_x^2(i) + V_y^2(i)} \quad \begin{aligned} x(0) &= x(1) \\ y(0) &= y(1) \end{aligned} \quad (1)$$

In the above equation,  $x(i)$  and  $y(i)$  are defined as  $i$ th samples coordination.  $V_x$  and  $V_y$  indicate the velocities in the direction of  $x$  and  $y$  respectively.  $v(i)$  represents the velocity in points  $i$ . The measure of  $v$  has the same size of  $x$  and  $y$ . All values of vector  $v$  are positive.

Acceleration: as velocity, acceleration is calculated by the difference between velocities. Equations (2) and (3) are the associated formulas.

$$a(i) = V(i) - V(i-1) \quad (2)$$

$$V(0) = V(1) \quad (3)$$

$a(i)$  is the acceleration of  $i$ th point and  $V(i)$  is the velocity of  $i$ th point.

### 3.2.1 Length Equalization

One of the problems ahead is that the signatures recorded by a person do not have the same length even in small sequential times. In other words, many factors such as standing or sitting of a person may affect on signing. It is necessary for a signature verification system to consider these factors. These are typical in realistic scenarios. Therefore, the best way to consider these factors is equalization to a reference.

Because of equal time interval of samples in signing and velocity vector independence from pen tip direction, the length can be calculated by sum of all points' velocities. Using equations (4) to (7) is the way to reach signature length.

$$X(i) = V(i)t + X(i-1) \quad (4)$$

$$X(i+1) = V(i+1)t + X(i)$$

⋮

$$X(N) = V(N)t + X(N-1) \quad (5)$$

$$\Rightarrow X(N) = \sum_{i=1}^N V(i)t \quad (6)$$

$$t = cte \Rightarrow X(N) = \sum_{i=1}^N V(i) \quad (7)$$

In equations (4) to (7),  $X(i)$  indicates length of signature from initiation to the point  $i$ .  $V(i)$  is the velocity of the point and  $t$  is defined as time. Considering the relative velocity and position, the constant value  $t$  is removed from the equations. All signatures recorded by each person follow this procedure and curve lengths of all signatures are obtained. The objective is to equalize the lengths. So the average length is considered as a reference for the person's signatures. This average value is assumed

for all signatures. New  $x$  and  $y$  are computed using following equations.

$$r = \frac{L}{\ell} = \frac{\sqrt{X^2 + Y^2}}{x^2 + y^2} \quad (8)$$

Substituting  $X=r.x$ ,  $Y=r.y$ , equation (8) will change to equation (9):

$$r = \frac{\sqrt{r^2 \cdot x^2 + r^2 \cdot y^2}}{\sqrt{x^2 + y^2}} \quad (9)$$

$r$  is defined as reference length in proportion to current signature length.  $L$ ,  $X$  and  $Y$  are also defined as curve length, reference length and reference width respectively that  $r$ ,  $x$ ,  $y$  are transformed to them. So if  $r$  is multiplied by  $x$  and  $y$  functions, all signatures of a person will have same length.  $r$  is different for each user signature.

### 3.2.2 Rotation Normalization

The angle of recording signature is another issue. Identical signing angle is essential for correct verification. All signatures are matched with binary image that the signature pixels are depicted with white colour. First, a signature is considered as a reference randomly and angles of all signatures are equalized to the angle of this reference signature. All signatures are rotated from  $-90$  to  $+90$ . In each step, cross correlation of rotated signature and reference is calculated and analyzed. Angle with the maximum correlation amount is considered to rotate the signature. This rotation is done around the centroid of signature. Figure 5 shows rotated signatures of a sample signature for 7 different angles. In figure 5 the intensity of signatures are proportional with their cross correlation with the reference signature.



Fig. 5. Rotating signatures around their centroid. The intensity of signatures are proportional to their cross correlation with the reference signature.

### 3.2.3 Functional Segmentation

Velocity, pressure and acceleration histogram are closely similar to normal distribution. Histogram of sample signature acceleration is shown in figure 6.

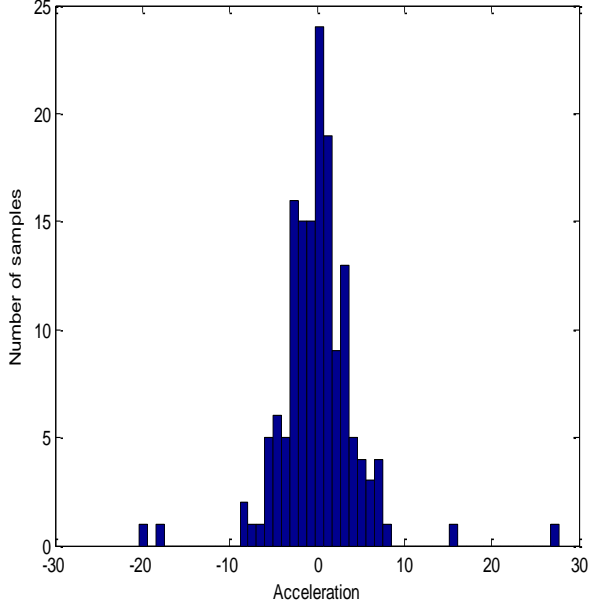


Fig. 6. acceleration histogram of a signature

Therefore, mean and standard deviation of each dynamic parameter are calculated from formulas (10) and (11):

$$m_f^i = \frac{1}{N} \sum_{j=1}^N f(j) \quad (10)$$

$$\sigma_f^i = \left( \frac{1}{N} \sum_{j=1}^N (m_f^i(j) - f(j))^2 \right)^{\frac{1}{2}} \quad (11)$$

As a result, all segmentation zones of functions are calculated by two above-mentioned parameters. Boundary values, i.e.  $m^i - \sigma^i, m^i$  and  $m^i + \sigma^i$  divide a signature into four zones. This way is done for all three functions, i.e., velocity, pressure, and acceleration. To improve the visualization of these segments, four colours are assigned to each segmented zone. Table 1 shows the colours associated areas.

Table 1. Colours used in signature segmentation and associated areas.

Segmented colour	Function area
Red	$f > m^i - \sigma^i$
Green	$m^i - \sigma^i < f < m^i$
Blue	$m^i < f < m^i + \sigma^i$
white	$f > m^i + \sigma^i$

Parameter  $f$  is the considered function. For instance, if the velocity is the intended function, all zones of the signature that their velocity is less than  $m^i - \sigma^i$  are shown in red, values between  $m^i - \sigma^i$  and  $m^i$  are shown by green, values between  $m^i$  and  $m^i + \sigma^i$  are in blue and values more than  $m^i + \sigma^i$  are shown by white. Background pixels are black. Figure 7 shows two genuine signatures coloured by this way.

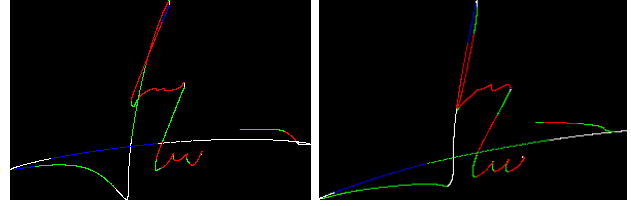


Fig. 7. Signature segmentation with respect to its acceleration

Now this data includes signatures with identical curve length, identical record angle, and four coloured types of zones. The conformity between forgery signature and genuine signature is analyzed in the following.

### 3.2.4 Conformity

All signatures recorded by a person are conformed. More conformity of the zones with same colour leads to more stability of the selected boundary for the specific feature.

Due to better observation and regardless of trivial changes of lines, signatures are thickened by morphological dilation. Images of the signatures are separately conformed according to the colours. The quantities of genuine signatures lines pixels that conformed on other signatures lines are counted for each colour and indicated by  $C_{g,color}^{i,function}$ . The area of all conformed genuine signatures are calculated in pixel and defined as  $A_{g,color}^{i,function}$  ( $i$  indicates the signers number,  $function$  is function names like pressure, velocity and acceleration and  $g$  shows the parameters are calculating for genuine signatures). It is obvious that more

$R_{g,color}^{i,function} = \frac{C_{g,color}^{i,function}}{A_{g,color}^{i,function}}$  leads to more stability of range and type of dynamic function for genuine signature. For more confidence, it is done on the forgery signatures existing in the dataset. The values of  $C_{f,color}^{i,function}$  and  $A_{f,color}^{i,function}$  are calculated. As expected, the value of  $R_{f,color}^{i,function} = \frac{C_{f,color}^{i,function}}{A_{f,color}^{i,function}}$  is not large in the latter set.

$S_{color}^{i,function} = \frac{R_{g,color}^{i,function}}{R_{f,color}^{i,function}}$  expresses the conformity of considered feature and separation between forgery signature and genuine signature for one person, e.g., when twenty third person of recorded signatures considered, the value of  $S_{green}^{23,p}$  indicates the proportion of genuine signatures and separation between genuine signatures and forgery signatures via the pressure between  $[m_p^{23} - \sigma_p^{23}, \sigma_p^{23}]$  of twenty third person.

These values are computed for all people and finally in equation (12):

$$S_{color}^{function} = \frac{1}{I} \sum_{i=1}^I S_{color}^{i,function} \quad (12)$$

Where  $I$  is total number of signers and  $S_{color}^{function}$  expresses final parameter that is defined as measure of suitability for the selected feature and this value is

obtained from average of  $S_{color}^{i,function}$  for all signers of the dataset.

Table 2 shows the results of experiment.

Table 2. Results of “Dynamic features stability experiment”

Datasets Parameters	Persian Signatures Datasets			International Dataset
	Dataset 1 (NDS)	Dataset 2 (Dehghani) [5]	Dataset 3 (Zoghi) [15]	Dataset 4 (SVC2004) [16]
$S_{red}^V$	1.7248	0.3699	0.9651	1.1480
$S_{green}^V$	1.9499	1.3940	1.9425	1.2528
$S_{blue}^V$	1.9599	1.5659	1.9941	1.2245
$S_{white}^V$	0.0291	1.3048	1.9542	0.0225
$S_{red}^a$	1.8576	1.3977	1.8470	1.1194
$S_{green}^a$	1.4356	0.0232	0.4384	1.1344
$S_{blue}^a$	0.4320	0.0253	0.4559	1.1375
$S_{white}^a$	1.7898	1.3665	1.9820	0.0321
$S_{red}^p$	1.3813	0.0382	0.0134	1.2626
$S_{green}^p$	1.9458	1.4281	1.8343	1.1862
$S_{blue}^p$	1.7097	1.4230	1.8439	1.2027
$S_{white}^p$	0.2330	1.0725	1.8233	0.1233

### 3.3 Feature Extraction

Considered features for a sample generally include the difference between one or several parameters based on a single template signature in this paper. In fact, during the process described as follows, a signature is specified as a reference signature. The closer signature to the reference signature results in more possibility to be the genuine signature.

Pressure, velocity, acceleration and angular velocity are analyzed in the following. So, extracted features are considered as functional features. Features are classified into three categories that are described in the following.

#### 3.3.1 Critical Samples

As mentioned before, Persian signatures have distinctive features in specific areas, e.g. these signatures often have more curves compared to other signatures. Velocity, acceleration and pressure have specific state in these areas. Therefore, it is possible to compare the curves of the signature associated to specific area of triple functions (i.e. velocity, acceleration, and pressure) with the curves extracted from reference signature.

As for this experimental result, each of triple functions is analyzed in four zones. The best range of the best function was explored. Based on this observation, pen tip velocity in range of  $m$  and  $m + \sigma$  is the best criterion that is more stable in genuine Persian signatures.

Figure 8 illustrates the velocity diagram of signature. Suppose that the goal is separation of the samples that their

velocity is more than average velocity and less than its standard deviation. After calculation of velocity and sample separation from signature, the comparison is applied among samples associated to the considered curves.

In spite of the experiment, the samples were used instead of associated signature curves. Because considering samples, cover another point that is important in signature verification. The point is total samples of signature that indicates total time of recording a signature.

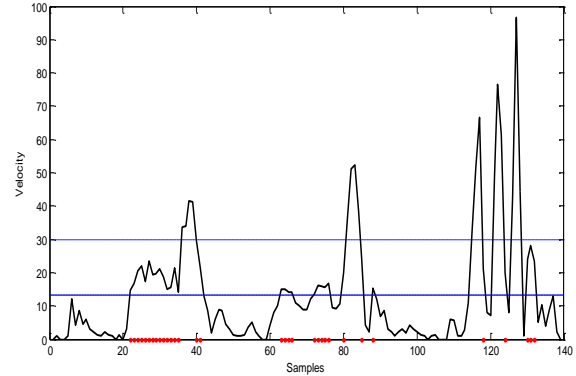


Fig. 8. The black curve is velocity curve of a sample and blue dashed lines are for illustrating boundary points.

#### 3.3.2 Maximum Velocity Area

As noted before (section 3.1), the maximum value of velocity and its position in signature could be essential and play an important role in Persian signatures. Therefore, by windowing the velocity functions and shifting them in time axis, a number is assigned to each window.

The window with maximum number is recorded for the considered signature. The difference between these numbers and recorded number of reference signature specify the next feature. “Maximum velocity area” feature formulas are represented in equation (13) to (17).

$$rect(a,b) = \begin{cases} 1 & a - \frac{b}{2} < n < a + \frac{b}{2} \\ 0 & o.w \end{cases} \quad (13)$$

$$T_{i,j}^k = rect(j \times (N - C), N) \times f_i^k \quad (14)$$

$$S_{i,j}^k = \frac{1}{N} \sum_{j=1}^N T_{i,j}^k \quad (15)$$

$$S_i^k = \arg \max_{j=1:N} (S_{i,j}^k) \quad (16)$$

$$feature2_i^k = distance(S_i^k, S_{2\_Template}^k) \quad (17)$$

As noted in equation (13), the function  $rect$  is a rectangular function centralized in  $a$  and including  $b$  samples. In equations (14) to (17),  $j$  indicates window number.  $N$  and  $C$  indicate length and the overlapping respectively.  $T_{i,j}^k$  represents separated window from  $i$ th signature related to  $k$ th person.  $S_{i,j}^k$ ,  $S_i^k$ ,  $2\_Template$  and  $feature2$  indicate average value of samples in window, window number with the maximum  $S_{i,j}^k$ , the second

extracted feature for reference signature and finally the second extracted feature respectively.

### 3.3.3 Relative Angular Velocity

A forger might do a dynamic forge if he/she knows dynamic features or can do shape based forgery. However, forging dynamic features and signature shape simultaneously is too hard even if the forger has all signature information and signature shape. In fact, relative angular velocity is changing signature line for two sequential samples.

This feature is calculated with the formulas that are given in equations (18) to (21).

$$\omega' = \frac{\Delta\theta}{\Delta t} \xrightarrow{\Delta t = cte} \omega = \Delta\theta = \frac{\Delta y}{\Delta x} \quad (18)$$

$$\omega_{i,n}^k = \frac{y_{i,n}^k - y_{i,n-1}^k}{x_{i,n}^k - x_{i,n-1}^k} \quad (19)$$

$$\omega_{Template,n}^k = \frac{y_{3\_Template,n}^k - y_{3\_Template,n-1}^k}{x_{3\_Template,n}^k - x_{3\_Template,n-1}^k} \quad (20)$$

$$feature3_i^k = distance(\omega_i^k - \omega_{3\_Template}^k) \quad (21)$$

In equations(18) to (21),  $\omega'$  is angular velocity, x and y are samples coordination,  $\omega_{i,n}^k$  is relative angular velocity of  $i$ th signature of signer number  $k$  in point number  $n$ . In addition, distance function is distance between signatures.  $3\_template$  and  $feature3$  are reference signature of third category and third category feature.

Figure 9 illustrates extracted features of Zoghi Dataset's signatures respectively by blue dots and red triangles.

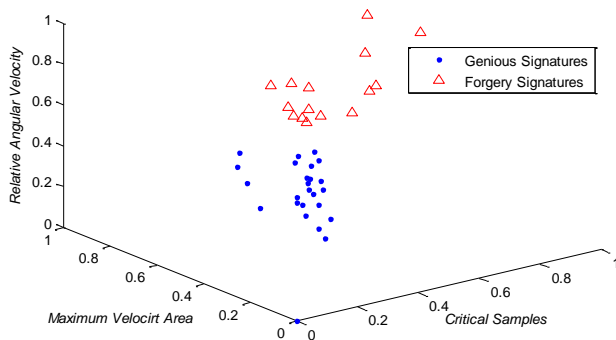


Fig. 9. Spatial illustration of features extracted from signatures. Blue dots are spatial representation of genuine signatures and red triangles illustrate the forgery signatures.

Two points are important in the proposed feature extraction.

- The first two categories (critical samples and maximum velocity area) are applied to express dynamic features of Persian signatures. Relative angular velocity is complementary for two previous categories. These three categories show both signature features (dynamic features and shape-based) and type (Persian signature) behavior.

- Calculation of difference between input signature samples and reference signature sample is required for all three categories feature extraction. Therefore, a similarity criterion is used. Dynamic Time Warping and Euclidean distance were options for this work. Because of less time consuming and good result, Euclidean distance is selected. Therefore, distance in all parts of this paper is Euclidean distance.

## 3.4 Reference Selection

### 3.4.1 Reference Signature Selection

A signature is selected as a reference in extraction of each feature and this signature is considered the best signature sample according to that feature.

To select reference signature, several genuine signatures are randomly selected as template signatures. Then the distances between each two template signatures are computed using the specified feature. The signature that sum of its distance to other signatures is less than other signatures is selected as specific reference for the feature [9].

### 3.4.2 Reference Vector Selection

As mentioned before, signatures of one person may extremely vary in different tries. It is reasonable to expect that a signature with great difference may be recorded as a genuine signature. Since selecting template signatures is done randomly, these unreliable signatures have chance to be selected as reference signature. Unreliable Reference Signatures (URS) cause poor verification. In order to eliminate URS another method is represented. In reference vector selection, all of the template signatures participate in producing a reference signature. In this method after selecting templates signatures, the difference between each signatures pair is computed. Sum of difference between a signature and other templates is calculated. Uniformly, a number as a total difference is assigned to each template signature. The smaller number shows more similarity to other templates and should have more effect on reference selection. Differences between the maximum values are assigned to templates and each signature achieves impact weights of templates. Weighted average with these weights indicates to the reference vector for each feature.

## 3.5 Classifier

### 3.5.1 Support Vector Machine

Support vector machine<sup>1</sup> is a new tool to pattern recognition. Basically, SVM uses a hyper plane to separate two classes[17]. Support vector machine leads to decrease structural risk. Whereas artificial neural network decreases experimental risk [18]. This point causes to increase generalization and better training with few train samples.

<sup>1</sup> SVM



As shown in figure 8, a classifier with linear separation is needed. Support vector machine separates data into genuine and forgery data.

### 3.5.2 Decision Making

Support vector machine is applied as a classifier indecision-making phase. Seven genuine samples and seven forgery samples regarding Zoghi [15] dataset, NDS and SVC and six genuine samples and six forgery samples regarding Dehghani's [5] dataset are used to train classifier. Genuine data used in training stage are the features extracted from template signatures as noted (in reference signature selection section). As a result eight samples (4 genuine and 4 forgery signatures) from dataset1, 26 samples (18 genuine and 8 forgery signatures) from dataset2, 91 samples (58 genuine and 33 forgery signatures) from NDS and 26 signatures (13 genuine and 13 forgeries) from SVC are used to test the proposed algorithm.

## 4. Results

Three criteria are used to verify the performance of signature verification algorithms, i.e. False Acceptance Rate (FAR), that is the error rate that the classifier incorrectly claims acceptable and False Rejection Rate (FRR) that is rate of incorrect rejections. Equal Error Rate (EER) is defined as error percentage when FRR and FAR are equal. EER is considered as the main criteria to study the performance of the algorithms. FAR and FRR change by the variation in classifier threshold.

A hyper plane is defined to separate two class patterns. In SVM, to achieve verification error rates as algorithms measure, a threshold value is applied to classification results before final decision making (sign function). Actually, the hyper plane bias is used as the threshold to change error rates (FAR and FRR). The Euclidian distance between patterns and separating hyper plane are the class membership degrees.

Two types of threshold (General and Individual) are applied to data to achieve error rates. A discussion on both types of thresholds and their associated verification results will come in the next section.

As stated before in section "Dynamic features stability experiment" and "Feature Extraction", three categories of features have been extracted from each signature sample. Each feature expresses three major characteristic of Persian signatures. One of triple functions and range of the function amount used in "critical samples" is selected in this experiment. Section "maximum velocity area" extracts special behavior of curves in Persian signatures and third category is shape-based feature.

Table 3 represents the result of verification.

Table 3. Result of the proposed algorithm on three Persian dataset and international SVC2004 dataset and comparison with best previous result

	Datasets	Best previous EER (%)	Proposed algorithm EER (%)
Persian Datasets	First Dataset (Dehghani)	3.95[5]	3.12
	Second Dataset (Zoghi)	11.2[15]	3.98
	NDS	-	4.26
International Dataset	SVC2004	3[8]	4.58

## 5. Discussion

In previous section, the results of signature verification with seven genuine and seven forgery signatures (except for first dataset with six genuine and six forgery signatures) are illustrated. This method was implemented with different numbers of training samples on Persian datasets. Results (Table 4) show that the algorithm performance is relatively acceptable for few training data.

The top of this paragraph illustrates a sub-subheading.

Table 4. Results (EER) of verification with different number of training samples. Half of training numbers on each row are from genuine signatures and other signatures are from forgery group.

		Dataset1 (Dehghani)	Dataset2 (Zoghi)	NDS
Number of training sample	4	8.26	7.66	10.83
	8	6.41	5.11	7.14
	12	3.12	4.51	5.13
	14	3.04	3.98	4.26
	20	-	2.32	4.01
	40	-	-	3.17

Of course, the verification by few training data is unreliable. Nevertheless, it can show the reliability of features and capability of leading classifier to discriminate the classes. Large variety of people signatures and especially Persian signatures may cause lower rate of verification with few number of training samples. As noted in "selecting reference" section, if the template signatures are selected from limited number of signatures, the probability of selecting bad references will increase. Hence the algorithm process should run repeatedly for achieving correct performance rate.

As illustrated in Table 2, the selected zone and function ( $m^i < V < m^i + \sigma^i$ ) of all three Persian signatures datasets is equal, while it is different from the international one. Since all three categories are based on Persian signatures datasets, the presented method is not significant for international SVC2004 dataset. This method was applied on SVC2004 dataset and as expected, the result of verification was not better than previous works. In Table 2 it can be seen that the  $S_{blue}^V$  is the parameter selected for Persian dataset and  $S_{green}^V$  is for international one. Since the signature samples are not enough, this criterion is not reliable. However both two features (selected features from "critical samples" category) with two other feature categories are applied to

SVC2004 dataset. The results (see Table 5) show that SVC2004 selected parameter with two other Persian-based features leads to better results than the other one.

The results of  $S_{color}^{function}$  in section 3.2.4 show that  $m^i < V < m^i + \sigma^i$  is the best option for extracting critical samples of Persian signatures. This result was tested on three available Persian datasets and triple function and different zones selected for extracting critical samples feature in the test.

Table 5. Verification results of SVC2004 with two different parameters (Persian and SVC2004's based) of first "critical samples" category

Parameter	EER (%)
$S_{blue}^V$ (Persian based)	4.58
$S_{red}^P$ (For SVC)	4.39

Figure 10 illustrates the verification results for different options of critical samples feature and two other constant features mentioned in feature extraction section.

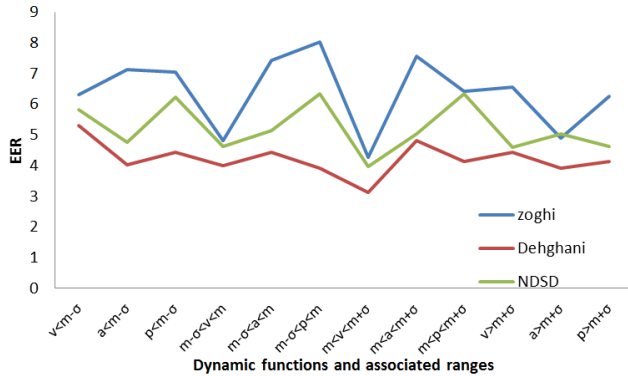


Fig 10. Verification result with different critical sample feature

As mentioned in previous section the threshold can be chosen for all writers or set individually one for each signer. A common threshold is used for the entire enrolment data from all the signers. This threshold is applied to a set containing all data.

To adopt the verification process to the single signers' properties, a signer dependant threshold should be applied. In Table 6 the results of the two threshold types are listed.

Table 6. Comparison between applying general and individual thresholds

	Datasets	General Threshold EER (%)	Individual Threshold EER (%)
Persian Datasets	First Dataset (Dehghani)	3.78	3.12
	Second Dataset (Zoghi)	4.52	3.98
	NDSD	4.64	4.26
International Dataset	SVC2004	5.55	4.58

In "Dynamic features stability experiment" critical curves have been used and because of dependency to time factor, critical samples are used instead. However, both types of critical samples and critical curves were used as features and their results are illustrated in Table 7.

Table 7. Comparison between critical curves and critical samples as third used feature

Datasets	Critical Curves Feature EER (%)	Critical Samples Feature EER (%)
First Dataset (Dehghani)	6.02	3.12
Second Dataset (Zoghi) NDSD	6.25	3.98
	6.53	4.26

Another key feature in an identification system is its usability. Actually real forgery signatures are not available and an identification system must be independent of forgery signatures. The purpose of the system is to use genuine signature for verification. The proposed algorithm is based on simulating forgery signatures by random patterns. As illustrated in figure 9 forgery signatures are at large distance from genuine with high scatters. The random points with normal distribution and equal mean and standard deviation are used as representative of forgery signatures. Table 8 shows mean and standard deviation of three independent features explained in section 3.3 of forgery signatures in three Persian datasets.

Table 8. Normal distribution parameters of forgery signatures in three Persian datasets

Datasets	Forgery Signatures Features Mean			Forgery Signatures Features Standard Deviation		
	1	2	3	1	2	3
First Dataset (Dehghani)	0.20	0.33	0.62	0.24	0.28	0.22
Second Dataset (Zoghi)	0.45	0.31	0.45	0.27	0.24	0.20
NDSD	0.41	0.53	0.72	0.27	0.30	0.21
Average	0.36	0.39	0.6	0.26	0.28	0.21

Their average values are used to produce random patterns independently. In the last row of Table 8 final normal distribution parameters for Persian signatures are shown. Verification results of the proposed algorithm with only three genuine signatures are illustrated in Table 9.

Table 9. Verification results with three genuine signature and producing random features instead of forgeries

Datasets	EER (%)
First Dataset (Dehghani)	16.67
Second Dataset (Zoghi)	12.14
NDSD	5.17

Due to independence of signers and forgers, and low computational complexity, the method can be practical for real world applications.

## 6. Conclusion and Future Works

An online signature verification based on special properties of Persian signatures is presented. Usually Persian signers move their wrist and fingers more than signers of other languages do and these motions cause variation in dynamic features. An experiment has been designed to explore robust features and the best one has been selected. Two dynamic features and relative angular velocity are extracted from signatures and the distance from reference signatures are used as the input to classifier. A linear SVM is used to classify signatures. The results of verification illustrated that acceptable EER was achieved.

Signature is the behavioral biometric that changes in different tries. In the proposed algorithm, distance between reference and input signature is the verification basis. Selecting bad reference leads to bad verification. In the algorithm, the probability of selecting improper signature as reference is not zero. It is expected that an intelligent method that specifies best representation of genuine signature cause less EER. Also using only two or three

genuine signatures that make verification system more practical will be possible if the mentioned method works.

### Acknowledgments

The author would like to thank the anonymous reviewers and editors for the useful and constructive comments which have improved the article.

### References

- [1] International Biometric Group (IBG). Available: <http://www.biometricgroup.com>
- [2] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. USA: Springer, 2008.
- [3] K. R. Radhika, G. N. Sekhar, and M. K. Venkatesha, "Pattern recognition techniques in on-line hand written signature verification - A survey," in *Multimedia Computing and Systems, 2009. ICMCS '09. International Conference on, 2009*, pp. 216-221.
- [4] C. E. Pippin, "Dynamic Signature Verification using Local and Global Features," 2004 July 2004.
- [5] E. Deghani and M. E. Moghaddam, "On-line signature verification using ANFIS," in *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on, 2009*, pp. 546-549.
- [6] T. A. A. Alizadeh, Z. Daei, "Optimal Threshold Selection for Online Verification of Signature," presented at the *IMECS2010, Hong Kong, 2010*.
- [7] L. Nanni, E. Maiorana, A. Lumini, and P. Campisi, "Combining local, regional and global matchers for a template protected on-line signature verification system," *Expert Systems with Applications*, vol. 37, pp. 3676-3684, 2010.
- [8] A. Fallah, M. Jamaati, and A. Soleamani, "A new online signature verification system based on combining Mellin transform, MFCC and neural network," *Digital Signal Processing*, vol. 21, pp. 404-416, 2011.
- [9] M. A. U. Khan, M. K. K. Niazi, and M. A. Khan, "Velocity-Image Model for Online Signature Verification," *Image Processing, IEEE Transactions on*, vol. 15, pp. 3540-3549, 2006.
- [10] M. Parodi and J. C. Gómez, "Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations," *Pattern Recognition*, vol. 47, pp. 128-140, 2014.
- [11] M. Parodi, Go, x, J. C. mez, M. Liwicki, and L. Alewijnse, "Orthogonal function representation for online signature verification: which features should be looked at?," *Biometrics, IET*, vol. 2, pp. 137-150, 2013.
- [12] K. Cpałka, M. Zalasiński, and L. Rutkowski, "New method for the on-line signature verification based on horizontal partitioning," *Pattern Recognition*, vol. 47, pp. 2652-2661, 2014.
- [13] Y. Guerbai, Y. Chibani, and B. Hadjadj, "The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters," *Pattern Recognition*, vol. 48, pp. 103-113, 1// 2015.
- [14] A. Q. Ansari, M. Hanmandlu, J. Kour, and A. K. Singh, "Online signature verification using segment-level fuzzy modelling," *Biometrics, IET*, vol. 3, pp. 113-127, 2014.
- [15] M. Zoghi and V. Abolghasemi, "Persian signature verification using improved Dynamic Time Warping-based segmentation and Multivariate Autoregressive modeling," in *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on, 2009*, pp. 329-332.
- [16] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, et al., "SVC2004: First International Signature Verification Competition," ed. <http://www.cse.ust.hk/svc2004/results.html>, 2004.
- [17] C. Gruber, T. Gruber, S. Krinninger, and B. Sick, "Online Signature Verification With Support Vector Machines Based on LCSS Kernel Functions," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, pp. 1088-1100, 2010.
- [18] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, pp. 307-319, 2003.

**Mohammad Esmaeel Yahyatabar** received M.Sc degree from Babol University of Technology, Babol, Iran and B.Sc from University of Mazandaran, Babolsar, Iran in Electronics engineering. His research interests are: signature verification and Machine Learning.

**Yasser Baleghi** is an Assistant Professor of Electronic Engineering at Babol University of Technology. He holds a Ph.D degree in Electronic Engineering from Iran University of Science & Technology. His research interests are evolvable and adaptive hardware, image processing and fault tolerant system design.

**Mohammad Reza Karami** received the B.Sc degree in Electronic Engineering in 1992, M.Sc. degree in Signal Processing in 1994, and the Ph.D degree in Biomedical Engineering from I.N.P.L d'Nancy of France. He is now the Associate at the Department of Electrical & Computer Engineering, Babol University of Technology. Since 1998 his research is in signal and speech processing. He has published 80 articles in journals and conferences.

# Computing Semantic Similarity of Documents Based on Semantic Tensors

Navid Bahrami

Department of Electrical, Computer and IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran  
bahrami.navid@gmail.com

Amir Hossein Jadidinejad\*

Department of Electrical, Computer and IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran  
amir.jadidi@qiau.ac.ir

Mozhdeh Nazari

Department of Engineering, Guilan Science and Research Branch, Islamic Azad University, Rasht, Iran  
mozhdeh\_nazary@yahoo.com

Received: 14/Dec/2014

Revised: 17/Mar/2015

Accepted: 06/Apr/2015

## Abstract

Exploiting semantic content of texts due to its wide range of applications such as finding related documents to a query, document classification and computing semantic similarity of documents has always been an important and challenging issue in Natural Language Processing. In this paper, using Wikipedia corpus and organizing it by three-dimensional tensor structure, a novel corpus-based approach for computing semantic similarity of texts is proposed. For this purpose, first the semantic vector of available words in documents are obtained from the vector space derived from available words in Wikipedia articles, then the semantic vector of documents is formed according to their words vector. Consequently, semantic similarity of a pair of documents is computed by comparing their corresponding semantic vectors. Moreover, due to existence of high dimensional vectors, the vector space of Wikipedia corpus will cause curse of dimensionality. On the other hand, vectors in high-dimension space are Usually very similar to each other. In this way, it would be meaningless and vain to identify the most appropriate semantic vector for the words. Therefore, the proposed approach tries to improve the effect of the curse of dimensionality by reducing the vector space dimensions through random indexing. Moreover, the random indexing makes significant improvement in memory consumption of the proposed approach by reducing the vector space dimensions. Additionally, the capability of addressing synonymous and polysemous words will be feasible in the proposed approach by means of the structured co-occurrence through random indexing.

**Keywords:** Information Retrieval; Natural Language Processing; Random Indexing; Semantic Similarity; Semantic Tensor.

## 1. Introduction

How similar are “Cat flu” and “Feline influenza”? Humans have initiate ability to compute semantic similarity due to their background knowledge about words and their interpretation ability. However, computing semantic similarity of words with multiple meanings is still remained as an obstacle. It must be noted that the meaning of each word is expressed according to the context that it appears and humans can interpret the meaning of a word according to its context. The main challenge refers to machines and how they deal with natural language and interpret concepts. In order to behave same as human, machines require human knowledge. Majority of natural language processing approaches leverage encyclopedias to transform knowledge and train machines. Moreover, there are many drawbacks in using encyclopedias. One of the obstacles is deep recognition of destination language for considering its syntax structure in processing. Another issue refers to extracting the meaning of words from encyclopedia. This problem is addressed here by considering the meaning of a word according to a given context.

The proposed approach is capable of extracting concepts from encyclopedia directly without any manual control. Whereas, encyclopedias contain wide range of documents, the meaning of each word can be expressed in high dimensional vector space using texts of documents. The most important achievement of the proposed approach is considering synonymy and polysemy. Indeed, it is able to disambiguate ambiguous and polysomic words.

The main characteristic of the proposed approach refers to employing simple texts of encyclopedias. In addition, it can limit the deep understanding of destination language to particular language structures such as punctuations, separators and etc. The main object of this approach is to compute semantic similarity of documents by extracting concepts from hierarchical structure of Wikipedia [1] and creating a semantic vector for each document and finally compare them. Due to Wikipedia's structure, the meaning of words can be expressed in different categories. Therefore, a three-dimensional vector space is created as a vector of words in various topics, which is organized by three-dimensional tensor structure. As an example, consider the meaning of the word “apple”. The fruit “apple” will be the first concept that is inspired

\* Corresponding Author

in readers' mind. Nevertheless, if such a word is used with words such as "Ipad", "computer" and "corporation", then the meaning of fruit "apple" will not be visualized and imagined. A method like bag of words does not pay attention to the relations among words in the texts and considers the texts merely as a set of words without order and relation. Thus, in this approach the word "apple" has only one meaning and that is deducted from the repetition of this word in the text. However, as it was previously mentioned, the main idea of the proposed approach of this paper is based on considering the meaning of the words in different texts. The possibility of extracting the best concept from a large corpus (the Wikipedia corpus in this paper) and forming the semantic vector of the word "apple" can be done by determining the meaning of "apple" with the help of its neighboring words in the text. Thus, the meaning convergence to the word "apple" will be provided by Wikipedia corpus, if the word "apple" has the same meaning in the two different documents but its neighboring words are different. In other word, the semantic vector of the document will be obtained, if all the semantic vectors of available words in the document are gained. Therefore, the possibility to compare the documents will be created due to their semantic vectors.

On the other hand, finding the words' meaning in a high dimensional space is the neglected issue in this approach, which is doomed to failure due to curse of dimensionality [2]. Whereas, spaces of vectors in high dimensional space of Wikipedia corpus are very similar to each other, achieving the best semantic vector for each word would be a meaningless and vain task [3]. Although there is no final solution for this problem, reducing the dimensions of vector space is the most appropriate and acceptable method. The random indexing [4,5] is used in the proposed approach to extract the semantic vector of words from Wikipedia corpus and reduce the dimensions of vector space. In other word, this method is capable of computing the meaning of words and reducing the dimensions of vector space simultaneously, which can reduce the processing load and memory consumption.

This approach has widespread applications in natural language processing. Finding the most relevant documents to a query, classifying documents based on their semantic content and computing semantic similarity of documents in order to compare them are the most notable applications. Moreover, according to reduced dimensions of vector space, the proposed approach can efficiently be used in large-scale systems.

The reminder of the paper is organized as follows: In section 2 the state of the art in computing semantic similarity are described. Way of constructing semantic space using word co-occurrences is presented in section 3. Key notions of the proposed approach such as extracting the meaning of words from Wikipedia and way of using them are indicated in section 4. Empirical experiments of the proposed approach for determining the effectiveness, Analyzing memory consumption and processing time are presented in section 5.

## 2. Related Work

Computing semantic similarity is one of the well-known agents in many fundamental tasks of computational linguistics such as word sense disambiguation, information retrieval and error correction [6]. Previous studies in this field can be classified into three main categories:

According to first category, texts are compared based on their common words using binary [7] and bag of words methods [8]. These methods are simple but whereas texts may contain many common words and express a concept with synonym words, they do not indicate any remarkable results.

On the other hand, knowledge-based methods leverage semantic relations of concepts defined in lexical resources such as WordNet [9] or Roget thesaurus [10] or network of concepts of Wikipedia [11] for computing semantic similarity and other applications. Then, the characteristics of graph structure of a lexicon are used for computing semantic similarity [6], such as method proposed by Resnik [12], Jiang and Conrath [13] and Lin [14]. These methods are confronted with some drawbacks. Noteworthy, they can only cover a limited range of vocabularies of a language and they do not include information of a particular filed either. Furthermore, knowledge-based methods are inherently limited to words and complex metrics are required for comparing texts. In contrast, these approaches are able to consider the contexts of the words. However, due to limitations of words in knowledge sources, considering the context of words and word sense disambiguation are also limited.

Other existing approaches employ statistical occurrences of words in a large corpus of unlabeled data. Latent Semantic Analysis (LSA) [15] is one of these approaches trained by word-document co-occurrence matrix. Vector space dimensions of this matrix are reduced using Singular Value Decomposition (SVD). This approach attempts to identify the most effective data known as implicit data in co-occurrence matrix in order to reduce dimensions. Therefore, interpreting its concepts is difficult and most of them are not commonly used by humans.

Another existing method which employs large corpus for computing semantic relatedness is Explicit Semantic Analysis (ESA) [16]. This approach leverages Wikipedia corpus as training set. Consequently, it uses Wikipedia concepts for considering concepts of words and documents. These concepts are directly defined by humans and are also consistent with natural concepts. Another advantage of determining concepts and their relations in Wikipedia corpus is presenting the related keywords to each concept, which has particular application in online advertisements of search engines [17].

The prominent idea of Explicit Semantic Analysis and Latent Semantic Analysis approaches is based on semantic kernel concept. Moreover, the aim of kernel methods is mapping data objects ( $o$ ) of  $D_{n \times m}$  matrix from a semantic space ( $\emptyset(o)$ ) to a more comparable vector

space ( $\tilde{\mathcal{O}}(o)$ ) [18]. It means that data are transformed to a new computational semantic space and accordingly the result of calculation will have more accuracy and less complexity [19].

The rows of  $D_{n \times m}$  present a set of data objects  $o = \{o_1, o_2, \dots, o_n\}$  and its columns present semantic features  $f = \{f_1, f_2, \dots, f_n\}$ . Based on classic models (bag of words), data objects of matrix are corresponded to documents and the words are its features. Noteworthy, this approach does not take into consideration semantic relations among data objects. Consequently, to fill this lacuna, data objects must be mapped into a similar vector space which considers semantic relations among data objects. Therefore, for considering semantic content in vector space, transformation is employed as  $\tilde{\mathcal{O}}(o) = \mathcal{O}(o) K_m$ , where  $K_m$  is a semantic matrix. Different choices of the matrix  $K_m$  lead to different variants of vector space semantic kernels [18]. Such as creating  $K_m$  matrix explicitly (Wikipedia Semantic Kernel [20]) or implicitly (Latent Semantic Kernel [21]).

Despite the precision of these methods in computing semantic relatedness and similarity, they are confronted with some disadvantages. One of the fundamental obstacles of both ESA and LSA methods refers to their high processing time in encountering a new training document. Whereas the weighting method used in these approaches requires the computation of the probability of each word in all documents of a corpus, by adding a new document all weighting process must be recalculated practically.

Temporal Latent Semantic Analysis [22] is an extended form of LSA which contains time elements. It is capable of organizing weight of words in different time intervals using tensor structure. Moreover, this method has decreased the computational complexity of adding new training document to a time interval. Following the similar line of research, Temporal Semantic Analysis [23] is an extended form of ESA at words level where time is added as new component. Based on this method, meanings of words are expressed according to Wikipedia concepts at different states. This method has efficiently reduced processing time of ESA in encountering a new training document.

### 3. Semantic Space

Semantic space models are based on distributional hypothesis [24,25[24],[25]. This hypothesis indicates that semantic similarity of a pair of words is computing the similarity of co-occurrence distribution among them. Therefore, distributional hypothesis and vector space model are related to each other because the distributional hypothesis emphasizes on co-occurrences of words corresponded to word frequency in vector space model. Consequently, creating semantics for word co-occurrences depends on how an algorithm presents semantic alternations [5].

### 3.1 Semantics Using Word Co-occurrence

Due to semantic space, the meanings of words are mapped to a multidimensional space. Space dimensions represent the differences between the meanings of words. Therefore, semantically similar words specify close vector representation. To understand the meaning of semantic co-occurrence obviously, consider the word "apple". Suppose "Calories in red delicious apple" is a document describing the word apple that inspires the meaning of apple as a fruit in readers' mind. This meaning is obtained by co-occurrences of apple by words such as red, color and delicious. On the other hand, consider apple in "the Ipad is an apple product" document. In this document apple is used to represent the company that manufactures computer products. Consequently, the meaning of word "apple" is changed according to its co-occurrence with words such as Ipad and product. Simple two-dimensional representation of word "apple" in co-occurrence with other words is illustrated in Fig.1.

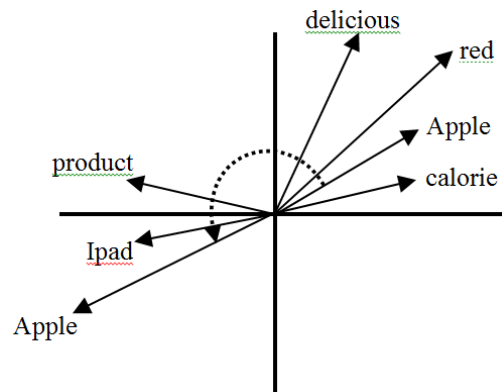


Fig. 1. Representation of word "apple"

### 3.2 Random Indexing

Simple co-occurrence can be efficiently used in a large corpus. Based on this model, each word attempts to assign its dimensions to words which are co-occurred. The results can be hundreds, thousands or perhaps millions of dimensions. Fundamentally, considering the number of dimensions according to the number of unique words (considering one repetition of each word) can be a complex issue in a large corpus and considerable efforts have been made to reduce space dimensions. Random indexing is one of these approaches which employs random projection of co-occurrence matrix to a space with less dimensions. Based on this technique, an index vector is assigned to a unique word. This vector is a random vector of numbers of 1, 0 and -1 in a space with constant dimensions (e.g. 500). The size of random index vector indicates the number of dimensions in semantic space.

Index vectors are constructed in such a way that any two arbitrary index vectors are orthogonal to each other with high probability. This feature is essential for accurate approximation from a word co-occurrence matrix to a low dimensional matrix. The meaning of each word is computed by summation of random index vector of co-occurred words in a small window of text. Random



indexing method can work efficiently in reducing the dimensions of a corpus that has already been processed [4]. Formally, consider word  $w$ , then  $w_i$  represents the co-occurrences of current word to a word with distance of  $i$  and  $index(w_i)$  is the index vector of co-occurred word. For a word  $w$ , a window size  $n$  is defined which is considered as the number of co-occurred words. Therefore, the meaning of word  $w$  is presented as:

$$Semantics(w) = \sum_{o \in D} \sum_{-n \leq i \leq n} index(w_i) \quad (1)$$

#### 4. Where $o$ is All Occurrence of Word $w$ in the Corpus $D$ ( $\forall o \in D$ ). The Proposed Approach

The prominent goal of the proposed approach is using Wikipedia corpus in order to map documents into a high dimensional vector space. Constructed vector space

comprises semantic discrimination and consequently the meanings of documents and words can be expressed based on topics.

The proposed approach is divided into two main phases. Training phase contains extracting documents from Wikipedia corpus based on specific categories and creating discriminative semantic space for each word of documents in each category. The created semantic space is then used in second phase. This stage is called test phase where semantic vectors of existing words in input documents are specified based on specific category and semantic vectors of input documents are computed based on them. The architecture of the proposed approach is presented in Fig. 2.

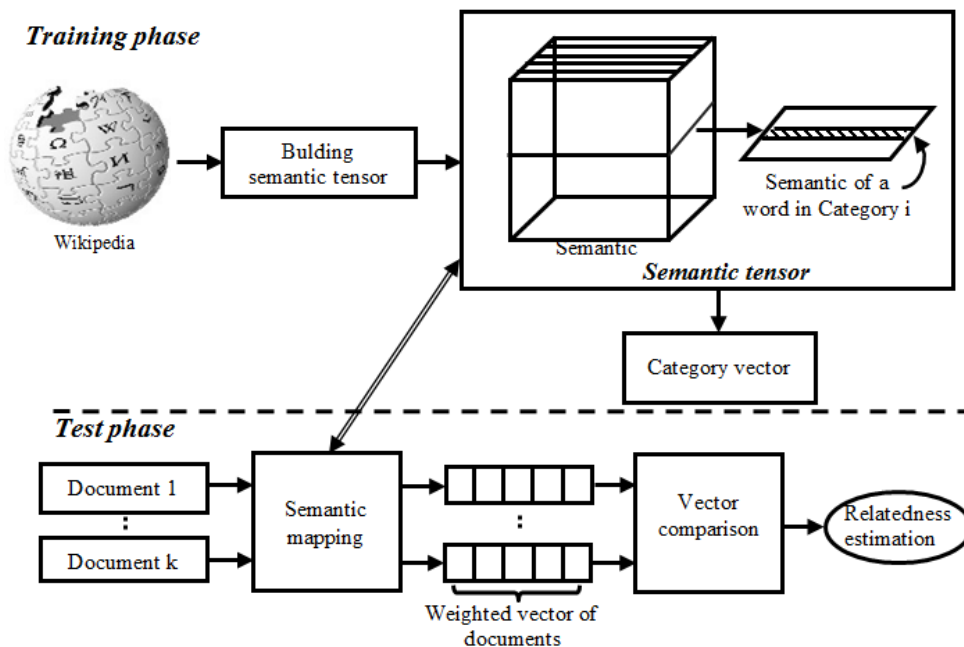


Fig. 2. The proposed approach

#### 4.1 Category of Semantic Space

Adding categories to a semantic space causes word semantic discrimination according to different thematic areas. Therefore, the meaning of each word is recognized with respect to the topic of a document. The main part of training phases is focused on creating this separated space. According to test phase on Fig. 2, in the beginning documents are extracted from Wikipedia corpus based on categories. Categories must be selected in such a way that the meaning of each category is reasonably discriminative towards another. Considering directed acyclic graph of Wikipedia, it is possible to obtain categories with high degree of semantic discrimination and high level of access.

The next step in training phase is organizing the meanings of words according to the documents existing in

each category, which is done by a particular structure called semantic tensor. Semantic tensor adds categories to vector space. Therefore, instead of using two-dimensional matrix of word×semantic, a three-dimensional matrix of word×semantic×category can be used. Two-dimensional vector of word  $w$  and three-dimensional representation of category vector are illustrated in Fig. 3.

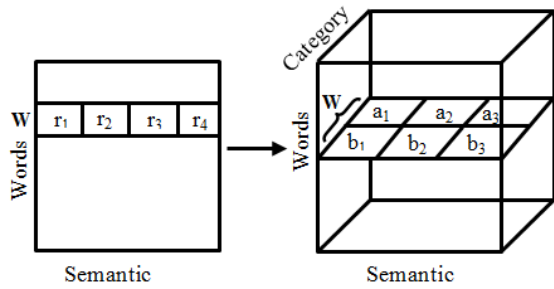


Fig. 3. Two and three dimensional vector representation of word  $w$

Semantic tensor based on this model has three main advantages:

1. It is possible to add new documents to each category.
2. Semantic tensor representation enables semantic meaning of a word to be compared in different categories.
3. Random indexing method used for weighting words in a document and reducing dimensions of vector space can perform processing on each document separately. Consequently, time and memory can be saved efficiently.

In order to add categories to semantic tensor, addition operation for weighted vector of each word cannot be added to Eq. (1) immediately. In other word, addition must be done at different categories separately. The summation of categories defines a semantic part for each word. The semantic part of word "apple" is presented in Fig. 4.

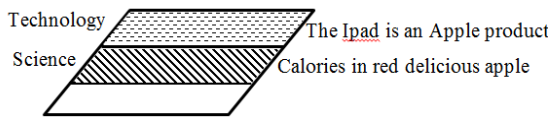


Fig. 4. Semantic slice of word "apple"

Therefore, the meaning of each word in a category is equivalent to the results of random indexing where all its values in that category are summed together. Finally, a single vector is obtained for each word in each category.

Fig. 4 presents how a word such as apple can have two different meanings. First semantic vector is constructed from summation of word apple vectors in category of technology that introduces apple as a manufacturer of computer components and accessories and the second

semantic vector in category of science expresses the calories in apple.

Therefore, the semantic tensors can be defined more formally. The input is a set of documents as follows:

$$D = (CT_0, d_0), (CT_1, d_1), (CT_2, d_2), \dots, (CT_i, d_i) \quad (2)$$

Where  $d_i$  is the set of documents occurring at category  $CT_i$ . If  $W_D$  is considered as a set of unique words in the collection ( $w \in W_D$ ), a unique index ( $index(w)$ ) is assigned to each word in the set. Consequently, the meaning of each word in each category is defined as follows:

$$Semantics(w, CT) = \sum_{o_{CT} \in d_i} \sum_{-n \leq i \leq n} index(w_i) \quad (3)$$

Where  $o_{CT}$  is the context for an occurrence of word  $w$  at category  $CT$ . Moreover,  $index(w_i)$  is an index vector of co-occurred words with distance of  $i$  to the main word. The semantic slice can be defined as follows:

$$Slice(w) = \{(CT_i, Semantics(w, CT_i)) \mid w \in d_i, i = 1, k\} \quad (4)$$

The final step of training phase is creating weighted vectors for categories of semantic tensor. This operation is done by summing the existing words vector in each category of semantic tensor. Therefore, category vector  $CT$  is equal to:

$$Semantics(CT) = \sum_{1 \leq i \leq k} index(w_i) \quad (5)$$

Where  $i$  is index identifier of all words in the category  $CT$ .

According to this, input documents with semantic tensor category are allowed to be compared in test phase based on their weights.

### 4.2 Computing Semantic Similarity

Mapping word semantic vector of semantic tensor to the corresponding word of input document and creating a semantic vector for each document is the basis of test phase. Consequently, vectors can be compared for computing semantic similarity.

As it is indicated in Fig. 2, the initial step in testing phase is receiving the input documents, which will be compared for computing semantic similarity. These documents contain simple and explicit text, which can be easily interpreted by humans. Nevertheless, the most important step in test phase is semantic mapping. According to Fig. 5, the semantic mapping for each input document performs the three following functions:

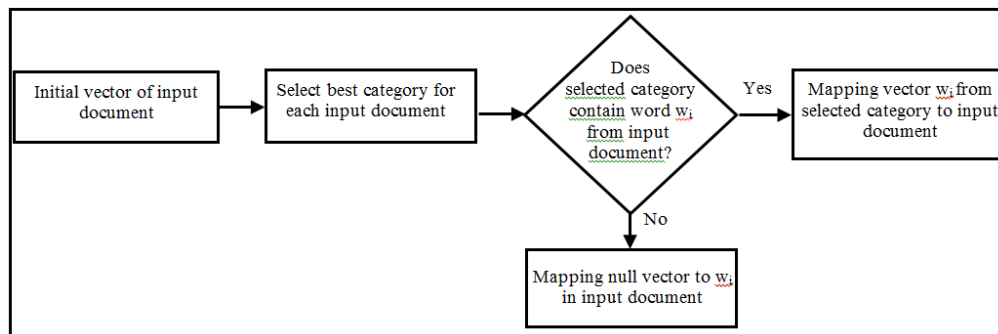


Fig. 5. Flowchart of semantic mapping

1. Creating initial index vector for input documents. Therefore, the preliminary weighting of input documents is accomplished using random indexing method by considering existing words of each document. As a result, the primary index vector for each document is computed by summation of words' weights. This method is simple but logical because words are components of sentences and documents. Index vector created for each document is leveraged in order to be compared to index vector of categories.
2. Detecting target category from semantic tensor for each input document is done by cosine comparison of weighted vector of documents and weighted vector of semantic tensor categories. Consequently, the most appropriate category is determined for each document.

$$\text{Cosine}(\text{index}(d_{\text{input}}), \text{index}(CT_i)) \quad (6)$$

This action provides a background for weighting of input documents using semantic tensors.

3. Mapping weights of words existing in target category of semantic tensor to their corresponding words in input document. Therefore, the weight of word  $w$  for input document of category  $k$  is defined as follows:

$$\text{Semantics}(d_{\text{input}}(w)) = \text{Semantics}(w, CT_k) \quad (7)$$

It should be noted that if a word of input  $k$  document does not exist in selected category, it will be considered null and it actually will have no effect on weighting of input document.

According to Fig. 2, after semantic mapping step, words of each input document are weighted based on target category of semantic tensor and only a single vector is remained for each document. Consequently, weighted vector of each document can be obtained by summing the weight of its words. Indeed, it can be mentioned that the main purpose of test phase is using semantic tensor in order to determine the semantic meaning of input documents.

The complementary step of testing phase is comparing documents' semantic vector to compute the degree of their similarity done by comparing documents in vector space using cosine similarity measure. Accordingly, the proposed measure is a corpus based approach which is capable of comparing documents using input documents and creating quantity values in a high dimensional vector space.

## 5. Empirical Experiments

In order to evaluate the proposed method, various experiments are carried out to reveal the efficiency of the proposed measure in comparison to other existing semantic similarity measures. These experiments contain two fundamental approaches. Experiments are performed to determine the potentiality of the proposed approach in computing semantic similarity of documents. The evaluations of these experiments are done by computing

the Pearson correlation coefficient between empirical results and human judgments on Lee benchmark dataset.

Other experiments include the analysis of memory consumption of the proposed method in comparison to other commonly used semantic similarity measures. The effect of increasing the number of training documents and unique words on memory consumption of the proposed method and other existing measures are also highlighted in these experiments.

The required processing time for executing the proposed method is presented in the following of this section using two various experiments.

### 5.1 Corpus

The proposed method is capable of using a corpus where hierarchical structure of categories and related documents to each category are specified. The reason of this issue refers to the possibility of extracting required documents based on semantic tensor's categories.

The proposed method is implemented using 2011 Wikipedia version containing 3573789 articles which are organized in 739980 categories. The English version of Wikipedia has been employed in our experiments but other languages can be also used.

Before using Wikipedia corpus for constructing semantic tensors, preprocessing is accomplished on documents of different categories as follows:

1. Removing bookmarks
2. Removing stop words
3. Stemming using Porter stemmer

These processes efficiently eliminate documents' disorders. Therefore, high frequently words which do not express a particular meaning are removed and in order to have uniform text, other words are transformed to their basic forms.

### 5.2 Benchmark Dataset

For comparing the precision of the proposed method in computing semantic similarity of texts, lee dataset [26] has been used, which contains a collection of 50 documents from Australian Broadcasting Corporation's new mail service. The length of these documents is between 51 to 128 words and they include large number of topics. Judgment had been done by 83 students of Adelaide University of Australia. These documents were paired in all possible ways and each of the 1225 pairs has 8-12 human judgment. Finally, the average of obtained values was considered as degree of semantic similarity of each pair of documents.

### 5.3 Empirical Result

Two following tools were employed for empirical experiments of this paper:

1. Wikipedia miner [27] based on Java for extracting documents from Wikipedia corpus
2. S-Space [28] library based on Java for leveraging implementations of random Indexing, LSA and

ESA existing in this library and applying the required process of the proposed approach.

Experiments are divided into two main categories. The main reason for choosing categories in each set of documents is the ability of document semantic discrimination of each category to the others. The first set of experiments are accomplished using extracted documents of four categories of matter, life, concept and society. These four categories exist in depth one of Wikipedia's hierarchical structure after fundamental category and all of documents are organized in subcategories of these four main categories. The second set of experiments is performed on seven categories of arts, biography, geography, history, mathematics, science and society. These categories are chosen from Wikipedia's documents classification in English Wikipedia website (at the time of experiments). The results of experiments present acceptable semantic discrimination among these seven decided categories. It is due to that if semantic discrimination was not correctly obtained, result of experiments would confront with significance decrease.

In both sets of experiments the same number of documents are extracted from each category (500 documents) and the length of random index vector of semantic tensor has been considered 120 with a window size of  $\pm 3$ . Then unequal number of documents has been extracted from each category (more than 3000 documents for each category) and the length of random index vector in semantic tensor has been set 150 with word window size of  $\pm 3$ . Empirical experiments conducted by Karlgren and Sahlgren [29] indicated that short length word window often provides better functionality. This issue also seems reasonable because sentences with length less than eight words provide high readability and are able to explain the meaning of a word clearly.

The results of experiments are presented in table 1. For the aim of comparison, the results of experiments conducted by the Bag of Word approach, Basic random indexing, LSA and ESA on Lee dataset are also shown in this table.

Table 1. Pearson correlation coefficient between various semantic similarity measures and human judgments on Lee dataset

Algorithm	Pearson correlation with human judgments
Bag of words	0.50
Random indexing	0.52
Latent Semantic Analysis (LSA)	0.60
Explicit Semantic Analysis (ESA)	0.72
Our approach (4 category, 500 documents)	0.64
Our approach (4 category, different number of documents)	0.61
Our approach (7 category, 500 documents)	0.60
Our approach (7 category, different number of documents)	0.62

Table 1. Pearson correlation coefficient between various semantic similarity measures and human judgments on Lee dataset

Comparison results indicate - the effectiveness of the proposed method in comparison to other semantic similarity measures. According to the results obtained by the proposed approach, it can perform better than Bag of Words, random indexing and LSA methods and it only presents lower performance than ESA method.

Although four conducted experiments are not very different from each other, the amount of difference can express some points. Considering the results obtained by four categories, by increasing the number of documents the results are decreased. This is due to negative impact of added documents to each category. These documents had not only negative impact on weighting of documents' words, but also made some difficulties for selecting a target category for each document. This issue is probably due to discrimination reduction of the general meaning of each category by adding new documents to them. As it was noted, reducing efficiency by increasing the amount of documents in these four categories is negligible. However, it presents that increasing the number of documents of each category requires a lot of precision.

On the other hand, in experiments containing seven categories, by increasing the number of training documents, the results have been improved. It seems that it refers to inadequate number of training documents towards covered topics in text documents. It cannot be definitely expressed that how many categories would present better results. However it must be noted that training documents in each category must be selected in such a manner to cover a wide range of topics in that field. In addition, the semantic discrimination of each category must be observed towards another one. On the other hand, it must also be mentioned that the large number of categories is one of the main factor of increasing error detection of target category because semantic discrimination process and appropriate document selection for each category is difficult and error probability in selecting related documents to each topic would be increased.

#### 5.4 Memory Analysis

In this section the memory consumption of the proposed method is compared to ESA and LSA methods by considering a set of specific documents and their unique words. The proposed method contains seven categories in this experiment and the length of random index vector is 100. These two factors along with the number of documents and their unique words are the main factors affecting the memory size of the proposed method. Two sets of experiments have been carried out in order to examine the effectiveness of increasing the number of documents and words on memory consumption increment. These two sets of experiments have some specific features as follows:

1. The first set contains 26989 documents and 159947 unique words.
2. The second set contains 228312 documents and 501436 unique words.

3. Number of extracted documents from search in hierarchical structure of Wikipedia to specific depth considering determined topic categories. In order to increase the number of documents in second experiments, depth of search is also increased.
4. The rate of document increment from first series to second series is equal to 45.8 times and the rate of word increment is equal to 3.13 from first series to second series.

Memory consumption of the proposed method, ESA and LSA are illustrated in Fig. 6 according to mentioned features.

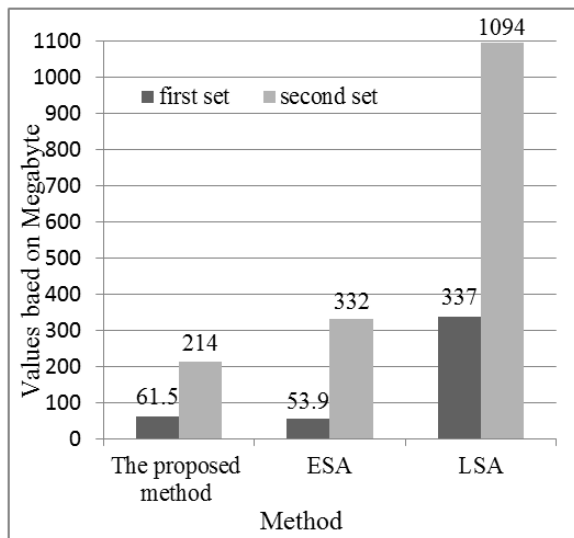


Fig. 6. Histogram of memory consumption of the proposed method, LSA and ESA

By analyzing the results, the rate of memory consumption increment of the proposed method from first series to second series of experiment is equal to 3.47. Moreover, the rate of memory consumption increment of ESA is equal to 6.15 and this rate is equal to 3.24 for LSA. The growth rate of the proposed method largely depends on the growth rate of words, whereas based on the proposed method a wide range of words can be initialized using a vector with the length of 100 (with respect to unique and orthogonal vectors) and the number of vectors is only increased by escalating the number of words. However, this rate highly depends on the number of considered categories.

On the other hand, ESA has a particular structure with various numbers of posting which are dependent on the number of documents and unique words. By increasing the number of documents and unique words, the probability of word occurrence in documents is increased and the length of postings in inverted index is also increased frequently. The rate of memory consumption of this method confirms this issue. Consequently, LSA depends on the number of words; whereas by decreasing dimensions, vectors with the same length are created for each word (vector with length of 100 in this particular example). The memory consumption of this method refers memory consumption of each cell in each word vector according to the required decimal precision. Accordingly,

it can be stated that the proposed method is more optimal in comparison to LSA and ESA in memory consumption.

### 5.5 Time Analysis

The required time for creating the final vector for each input document based on the proposed method is presented in this section. Effective steps for estimating the required time are illustrated in Figure 7.

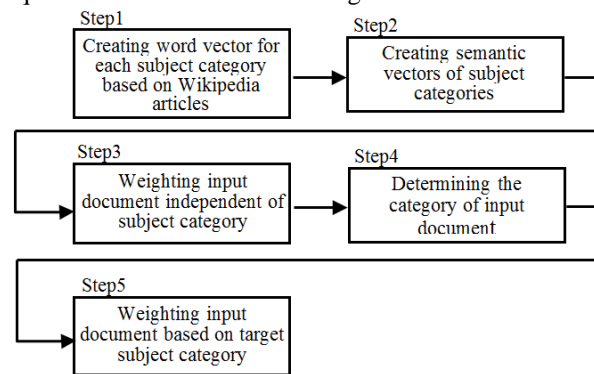


Fig. 7. Effective steps in time analysis of the proposed method

As an example, consider two datasets of section 5-4. Time analysis of the first set containing 26989 documents and the second set containing 228312 documents based on time steps of the proposed method (Figure7) are presented in Table 2. It must be noted that the characteristics of computer hardware that experiments were conducted on it is CPU Core2Duo E4600 and 3GB of RAM.

Table 2. The required time for each step of the proposed method

	First set (Second)	Second set (Second)
Step1	25.11	154.561
Step2	9.183	36.651
Step3	0.097	0.097
Step4	0.19	0.146
Step5	1.238	3.709
Total time	35.818	195.164

Time analysis with mentioned hardware for ESA approach on the first set is equal to 110.447 seconds and on the second set is equal to 1401.952 seconds. Moreover, time analysis for LSA approach on the first set is equal to 126.295 seconds and on the second set is equal to 1329.941 seconds. According to empirical experiments, ESA approach on first set requires 3.08 more processing time in comparison to the proposed method and this value is equal to 7.183 on second set. Furthermore, the processing time of LSA approach is 3.52 more in comparison to the proposed method on first set and this value is equal to 6.81 on the second set. The results represent the significant improvements in time consumption of the proposed method in comparison to LSA and ESA approaches.

Implementing the ESA and LSA approaches on mentioned hardware was done using S-Space package. English Porter Stemmer was employed for stemming step of both approaches. Additionally, vector space dimensions of LSA approach are considered 300. This

approach reduces the dimensions of word-document matrix using SVD method. Wikipedia version 2011 has been employed as primary dataset of these approaches.

Long processing time of LSA and ESA approaches in comparison to the proposed method is due to their complexity in constructing training vector using Wikipedia corpus. ESA approach creates  $m \times n$  table of words and concepts. Each element of this table presents the weight  $tf.idf$  of a word in a particular concept. Noteworthy, computing  $tf.idf$  for each word is a time consuming task. Moreover, in order to reduce the dimensions of vector space, ESA approach requires inverse vector algorithm, which subsequently increases the processing time. LSA approach is confronted with the same drawbacks. The created word-document matrix is weighted using entropy measure and the dimensions are reduced using SVD algorithm which requires long processing time. The reason of this issue is due to the complexity of this algorithm, whereas the time complexity of the fastest implemented SVD algorithm is equal to  $O(m.n^2)$  [5].

On other hand, whereas the proposed method employs random indexing method it has not only significant reduction in memory consumption, but also considerable reduction in processing time. Random indexing method considers random vectors with constant length for words existing in documents. These random vectors contain normal numbers and only require meeting unique vectors condition. Moreover, vectors with constant length cause reduction in vector space dimensions. Therefore, weighting and reducing dimensions of vector space are done simultaneously and simply. As result, the proposed method prospers significant time reduction in creating semantic vectors of words.

## 6. Discussion

The most important achievement of this paper refers to determining the effectiveness of the proposed method and improving its results in comparison to basic random indexing methods. This is due to the nature of random indexing method where indexing employs neighbor words to express the meaning of a key word. It requires large sets of documents to identify key concepts and balance their weights. This issue is clearly marked by comparing the results of experiments because by adding a large corpus of documents to basic random indexing method in the proposed method, the results have been significantly improved.

One of the major advantages of the proposed method in comparison to ESA and LSA is that it does not impose significant processing load during adding or changing the training set. In the view of fact it leverages random indexing method. Moreover, by employing this method, words in new documents are weighted independently and these new weighted values are added to previous weights. Since the used weighting method in LSA and ESA (usually  $TF.IDF$ ) requires to consider the weight of a word in all training set documents and by adding even a new training

document the weight of all words must be recalculated and therefore high processing load is imposed based on them.

Although the proposed method employs three-dimensional tensor, it uses less memory than approaches with two-dimensional structure. This indicates the high potentiality of random indexing method in reducing the dimensions of vector space towards other existing methods. Moreover, it can significantly reduce the memory consumption of the proposed method.

Besides the advantages mentioned for the proposed method, it confronts with some limitations. Initially, it requires a rich corpus which contains many words. This limitation is due to random indexing method. In other word, according to random indexing method if a word does not exist in the corpus, no weight will be considered for it. Therefore, word would not have any effect in computing semantic similarity. The second limitation refers to determining the number of documents in each category. If the documents of each category are semantically close to each other and their semantic discrimination is low, error probability in choosing the best category will increase. Furthermore, having large number of categories with various topics causes complexity in identifying the appropriate category for extracting the meaning of words. Accordingly, if a target category is not selected properly, the overall performance of the proposed method will decrease.

## 7. Conclusions

In this paper a novel method based on semantic tensor is proposed for computing semantic similarity of texts. This is a semantic technology for natural language processing. The proposed method is based on Wikipedia corpus where articles are categorized in different topics and documents are extracted from these categories. The most important aspect of the proposed method is its ability for identifying synonymy and polysemy, which are one of the most important issues in natural language processing. Therefore, this method does not merely rely on common word frequency in texts and it can identify the value of association between two texts that express a topic with various texts.

The evaluation results revealed acceptable performance of the proposed method in computing semantic similarity according to optimal memory consumption. Consequently, the Pearson Correlation coefficient of the proposed method and human judgments is between 0.54 and .064. Although ESA has better performance than the proposed method, the other existing methods show lower performance.

According to experiments, memory consumption of the proposed method is 80% less than the memory required by LSA and 30% less than the amount of memory required by ESA. By increasing the amount of documents of a corpus, this value would improve. Consequently, the efficiency is simultaneously improved by decreasing memory consumption.



## References

- [1] Medelyan O., D. Milne, C. Legg, and I.H. Witten, "Mining meaning from Wikipedia", *International Journal of Human-Computer Studies*, Vol. 67, No. 9, 2009, pp. 716-754.
  - [2] Kriegel H.P., P. Kröger, and A. Zimek, "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering", *ACM Transactions on Knowledge Discovery from Data*, Vol. 3, No.1, 2009, pp. 1-58.
  - [3] Assent I., "Clustering high dimensional data", *WIREs Data Mining Knowl Discov*, Vol. 2, 2012, pp. 340-350.
  - [4] Chatterjee N., and S. Mohan. "Extraction-based single-document summarization using random indexing", *19th IEEE International Conference on Tools with Artificial Intelligence*, Vol. 2, IEEE, 2007, pp.448-455.
  - [5] Jurgens D., and K. Stevens, "Event detection in blogs using temporal random indexing", *Proceedings of the Workshop on Events in Emerging Text Types(eETTs)*, Association for Computational Linguistics, 2009, pp. 9-16.
  - [6] Budanitsky A., and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness", *Computational Linguistics*, Vol. 32, No. 1, 2006, pp. 13-47.
  - [7] Pincombe B., "Comparison of human and latent semantic analysis (LSA) judgements of pairwise document similarities for a news corpus", *DTIC Document*, 2004.
  - [8] Baeza-Yates R., and B. Ribeiro-Neto, "Modern information retrieval", *ACM press New York*, Vol. 463. 1999.
  - [9] Fellbaum C., "WordNet: An Electronic Lexical Database", *MIT Press, Cambridge*, 1998.
  - [10] Roget P., "Roget's Thesaurus of English Words and Phrases", *Longman Group Ltd*, 1852.
  - [11] Jadidinejad A.H., and F. Mahmoudi, "Unsupervised Short Answer Grading Using Spreading Activation over an Associative Network of Concepts", *Canadian Journal of Information and Library Science*, Vol. 38, No. 4, 2014, pp. 287-303.
  - [12] Resnik P., "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language", *Journal of Artificial Intelligence Research*, 1999. Vol. 11, No. 1, 1999, pp. 95-130.
  - [13] Jiang J.J., and D.W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy", *Proceedings of International Conference Research on Computational Linguistics (ROCLING X)*, 1997, pp. 19-33.
  - [14] Lin D., "An Information-Theoretic Definition of Similarity", *Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 296-304.
  - [15] Landauer T.K., P.W. Foltz, and D. Laham, "An introduction to latent semantic analysis", *Discourse Processes*, Vol. 25, 1998, pp. 259-284.
  - [16] Gabrilovich E., and S. Markovitch, "Wikipedia-based Semantic Interpretation for Natural Language Processing", *Journal of Artificial Intelligence Research*, 2009. Vol. 34, 2009, pp. 443-498.
  - [17] Jadidinejad A.H., and F. Mahmoudi, "Advertising Keyword Suggestion Using Relevance-Based Language Models from Wikipedia Rich Articles", *Journal of Computer & Robotics*, Vol. 5, No.1, 2014, pp.29-35.
  - [18] Jadidinejad A.H., F. Mahmoudi, and M.R. Meybodi, "Clique-based semantic kernel with application to semantic relatedness", *Natural Language Engineering*, Cambridge University Press, Vol. 1, No. 1, (To appear), pp.1-18;
  - [19] Jadidinejad A.H., and Marza V., "Building a Semantic Kernel for Persian Text Classification with a Small Amount of Training Data", *Journal of Advances in Computer Research*, Vol. 6, No. 1, 2014, pp.125-136.
  - [20] Wang P., and C. Domeniconi, "Building semantic kernels for text classification using wikipedia", *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 713-721.
  - [21] Cristianini N., and J. Shawe-Taylor, and H. Lodhi, "Latent Semantic Kernels", *Journal of Intelligent Information Systems*, Vol. 18, No. 2-3, 2002, pp. 127-152.
  - [22] Wang Y., and E. Agichtein, "Temporal latent semantic analysis for collaboratively generated content: preliminary results", *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, ACM, 2011, pp. 1145-1146.
  - [23] Radinsky K., E. Agichtein, E. Gabrilovich, and S. Markovitch, "A word at a time: computing word relatedness using temporal semantic analysis", *Proceedings of the 20th international conference on World wide web*, ACM, 2011, pp. 337-346.
  - [24] Baroni M., and A. Lenci, "Distributional memory: A general framework for corpus-based semantics", *Computational Linguistics*, Vol. 36, No. 4, 2010, pp. 673-721.
  - [25] Turney P.D., and P. Pantel, "From frequency to meaning: vector space models of semantics", *Journal of Artificial Intelligence Research*, Vol. 37, 2010, pp. 141-188.
  - [26] Lee M.D., B. Pincombe, and M. Welsh, "An Empirical Evaluation of Models of Text Document Similarity", *Proceedings of the 27th annual meeting of the Cognitive Science Society (CogSci'05)*, Erlbaum: Mahwah, NJ, 2005, pp. 1254-1259.
  - [27] Milne D., and I.H. Witten, "An open source toolkit for mining wikipedia", *Artificial Intelligence*, Vol. 194, 2012, pp. 222-239.
  - [28] Jurgens D., and K Stevens, "The S-Space package: An open source package for word space models", *Proceedings of the ACL 2010 System Demonstrations*, Uppsala, Sweden, 2010, pp. 30-35.
  - [29] Karlgren J., and M. Sahlgren, "From Words to Understanding", *Foundations of real-world intelligence*, CSLI Publications, 2001.
- Navid Bahrami** received the B.Sc degree in Software Engineering from Ghiasodin Jamshid Kashani University of Abyek, Qazvin, Iran in 2010 and the M.Sc degree in Software Engineering from Islamic Azad University of Qazvin, Qazvin, Iran in 2014. His research interests are Machine learning, Information retrieval and Data mining.
- Amir Hossein Jadidinejad** is faculty member of Islamic Azad University of Qazvin (QIAU). His research interest include Information Retrieval, Machine Learning and Statistical Data Analysis. He is also interested in applying state-of-the-art models of Information Retrieval and Machine Learning to very large collections, such as WWW.
- Mojdeh Nazari Soleimandarabi** received the B.Sc degree in Computer Software Engineering from Guilan University and M.Sc degree in the same field of study from Science and Research University, Guilan, Iran in 2012 and 2015, respectively. Her research interests include information retrieval, semantic relatedness, semantic web and data mining.