In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 3, No. 3, July-September 2015, Serial Number 11

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

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

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@jist.ir
URL: www.jist.ir

Indexed in:

- Journal of Information Systems and Telecommunication www.jist.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)

- Abdollahi Azgomi Mohammad, Iran University of Science and Technology, Tehran, Iran
- Abutalebi Hamid Reza, Yazd University, Yazd, Iran
- Azimzadeh Fatemeh, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Darmani Yousef, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Daneshpour Negin, Shahid Rajaee Teacher Training University, Tehran, Iran
- Fouladi Kazem, University of Tehran, Tehran, Iran
- Ghanbari Mohammad, University of Essex, Colchester, UK
- Ghassemian Hassan, Tarbiat Modares University, Tehran, Iran
- Haghighi Hasan, Shahid Beheshti University, Tehran, Iran
- Haji Mohammadi Zeinab, Abrar Institute of Higher Education, Tehran, Iran
- Hasanzadeh Reza, University Of Guilan, Rasht, Iran
- Jalali Aliakbar, Iran Science & Technology of University, Tehran, Iran
- Jamili Oskouei Rozita, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran
- Lotfi Abbasali, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Malekzadeh Gholamreza, Ferdowsi University of Mashhad, Mashhad, Iran
- Marvi Hossein, Shahrood university of Technology, Shahrood, Iran
- Mirroshandel Seyed Abolghasem, University of Guilan, Rasht, Iran
- Moallem Peyman, University of Isfahan, Isfahan, Iran
- Mollaei Mahmoud, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Mousavirad Seyed Jalaleddin, University of Kashan, Kashan, Iran
- Nilforooshan Zahra, Amirkabir University of Technology, Tehran, Iran
- Rezaei Abdlhossein, Academic Center for Education Culture and Research (ACECR), Isfahan, Iran
- Sadeghzadeh Ramezan Ali, Khaje Nasir-edin Toosi, Tehran, Iran
- Sadegh Mohammadi Hamidreza, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Saeb Armin, Islamic Azad University, Yadegar-e-Emam Branch, Tehran, Iran
- Safaei Ali Asghar, Tarbiat Modares university, Tehran, Iran
- Safarinejadian Behrooz, Shiraz University, Shiraz, Iran
- Sajedi Hedieh, University of Tehran, Tehran, Iran
- Shadi Nazanin, Islamic Azad University, Borujerd Branch, Borujerd, Iran
- Shirvani Moghaddam Shahriar, Shahid Rajaee Teacher Training University, Tehran, Iran
- Yassoubi Farshid, National Informatics Corporation, Tehran, Iran
- Zakeri Bijan, Babol Noshirvani University of Technology, Babol, Iran

Table of Contents

Opinion Mining in Persian Language Using Supervised Algorithms

Saeedeh Alimardani*
Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran
sa1.alimardani@gmail.com
Abdollah Aghaei
Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran
aaghaie@kntu.ac.ir

Received: 30/Jan/2015 Revised: 29/Jun/2015 Accepted: 12/Jul/2015

Abstract

Rapid growth of Internet results in large amount of user-generated contents in social media, forums, blogs, and etc. Automatic analysis of this content is needed to extract valuable information from these contents. Opinion mining is a process of analyzing opinions, sentiments and emotions to recognize people's preferences about different subjects. One of the main tasks of opinion mining is classifying a text document into positive or negative classes. Most of the researches in this field applied opinion mining for English language. Although Persian language is spoken in different countries, but there are few studies for opinion mining in Persian language. In this article, a comprehensive study of opinion mining for Persian language is conducted to examine performance of opinion mining in different conditions. First we create a Persian SentiWordNet using Persian WordNet. Then this lexicon is used to weight features. Results of applying three machine learning algorithms Support vector machine (SVM), naive Bayes (NB) and logistic regression are compared before and after weighting by lexicon. Experiments show support vector machine and logistic regression achieve better results in most cases and applying SO (semantic orientation) improves the accuracy of logistic regression. Increasing number of instances and using unbalanced dataset has a positive effect on the performance of opinion mining. Generally this research provides better results comparing to other researches in opinion mining of Persian language.

Keywords: Opinion Mining; Persian; Supervised Algorithm; SentiWordNet.

1. Introduction

Nowadays, rapid growth in the number of Internet and social networks users, has paved the way of accessing people's opinions. Recognizing orientation of this opinion is easy for a human being, but since number of these opinions is increasing massively, it is impossible to analyze all of them manually.

Therefore opinion mining is required to automatically analysis these opinions and extract useful information from them.

"Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes"[1].

Opinions contain worthy information that are useful for both customers and organizations. For example, people express pros and cons of different aspects of a product in their opinions, so by analyzing them, others can be aware of different aspect of products before buying them. Opinion mining also lets companies improve products, resolve their weaknesses and acquire useful information about their rivals. The purpose of opinion mining is to obtain such information.

Different approaches have been developed for opinion mining. Most of these approaches were applied in English [2-4], Spanish [5-6] and Chinese [7-10]. Although Persian language is spoken in different countries, but there are few researches that consider opinion mining in Persian. Hence, in this article opinion mining in Persian language is investigated by applying standard machine learning techniques naive Bayes, SVM and logistic regression. For this purpose, opinion mining is experimented in different conditions. Also different features are incorporated to accomplish this task in a hotel domain. Up to now there are few researches in opinion mining for Persian language. So we decided to do a research in this field, studied different methods that are introduced in other languages and observed that a new lexicon can be produced for our purpose to do opinion mining in Persian language. The main reason that we conduct this research is to build Persian SentiWordNet and also to investigate the behavior of opinion mining in different conditions using this lexicon.

This article seeks to investigate:

1. The effect of weighting features by their semantic orientation.
2. The effect of instance number in Persian opinion mining.
3. The effect of unbalanced data set in Persian opinion mining.
4. Persian opinion mining using supervised algorithm.

The paper is organized as follows. Next section, presents basic concept of opinion mining. Section 3 presents related works on sentiment analysis. Lexicon creation and data preparation are described in section 4. Section 5 expresses experimental results. In conclusion section we present a summary of the article and results.

* Corresponding Author

2. Basic Concept

With rapid growth of social networks, forums, blogs and websites, the produced data is increasing rapidly. A part of this data is in a form of text. This text contains valuable information about different subjects. But raw text without analyzing hasn't any value and also the amount of this data is too huge. So this text data should be analyzed automatically to extract valuable information. Opinion mining is a field of study that used for this purpose.

Opinion is people's belief, idea and view point about different subjects.

Bing Liu [1] defines three general categorizations for opinion mining: Document-level, sentence-level, and phrase-level. In document-level a whole document is analyzed. In a sentence-level, sentence should be recognized and then be analyzed. In phrase-level, a phrase is identified and then its orientation is determined.

Researchers have introduced different methods for opinion mining. In general these methods are categorized in two categories: supervised and unsupervised methods. In supervised methods, there are Training and testing data to assign an appropriate class to given review and it requires labeled instances. Naïve Bayes, SVM and other supervised algorithm can be used here. But in unsupervised methods there are not any labeled instances. For example Turney [4] performs classification based on syntactic patterns.

Also different feature weighting method can be used to weight features in supervised algorithms such as Terms and their frequency, Part of speech and Sentiment words and phrases.

In this research opinion mining in document level is experimented and supervised methods are applied. We also used Sentiment word as feature weighting method.

3. Related Works

Up to now, different approaches are developed in the field of opinion mining. There are two main approaches, unsupervised and supervised approaches.

One of the first approaches in opinion mining applied average of phrases' SO in a document for classifying document as recommended or not recommended[4]. In this approach semantic orientation of a phrase is defined as a difference of its dependency with "excellent" and "poor". This dependency is number of hits calculated from a search engine.

Pang and Lee applied different Machine learning algorithms SVM, maximum entropy and naive Bayes with different features [3]. They showed that using unigram and present-absent as features achieve better results and SVM performs better than other algorithms. Zhang et al. [11] examined opinion mining by applying SVM, naive Bayes and character based N-gram model. Their findings showed that SVM and character based approaches outperform naive Bayes approach.

There are some approaches that applied lexicon based opinion mining. Taboada et al. [12] proposed a lexicon based approach that called semantic orientation calculator (SO-CAL). In this method, they used dictionary of words that contains words and their orientation. SO-CAL considers negation and strength too. They showed that SO-CAL has a consistent performance across different domains. Hung and Lin [13] used SentiWordNet and SVM for sentiment analysis. They observed that using objective words, can improve the performance of opinion mining, In Their method, if an objective word appears more in positive sentences, it got a positive score and if it appears more in negative sentences it got a negative score.

Martina and Finin [14] introduced new feature, Delta TF-IDF[1]. This feature is the difference of word's TF-IDF scores in the positive and negative training corpora. Using Delta TF-IDF has improved accuracy.

Tan and Zhang [15] applied four different feature selection methods and five learning method for sentiment analysis of Chinese language. They showed that IG (information gain) performs best for feature selection and SVM outperforms other algorithms for sentiment classification.

Beside this approaches, some researches use clustering-based approach for classification of document [16-17,2]. "The clustering-based approach is able to produce basically accurate analysis results without any human participation, linguist knowledge or training time" [17].

Basari et al. [18] introduced a method for improving SVM in sentiment analysis. Their method is combination of SVM and Particle swarm optimization (PSO). PSO is used for improving SVM parameters. Vinodhini and Chandrasekaran [19] applied Principal Component Analysis to decrease dimensionality problem in SVM.

There are few researches that have been conducted on sentiment analysis for Persian language. Shams et al. [20] proposed a method that is combination of unsupervised LDA [2] -based approach and PersianClues lexicon. This approach applied machine translation to translate MPQA[3] lexicon [21] and used Latent Dirichlet allocation for opinion mining. Their method achieved about 80 percent of accuracy.

Hajmohammadi and Ibrahim [22] used SVM for sentiment analysis in Persian. The experimental results showed that, SVM performs better than naive Bayes and unigram outperform bigram and trigram for feature selection.

Saraee and Bagheri [23] proposed a new feature called Modified Mutual Information (MMI) for Persian opinion mining. They examined different features and applied naive Bayes as a classifier. The results showed that MMI outperform MI (mutual information) as a feature selection method.

[1] Term Frequency–Inverse Document Frequency
[2] Latent Dirichlet Allocation
[3] Multi-Perspective Question Answering

4. Methodology

This section expresses different phases of opinion mining in this research as is shown in figure 1. Since in this article, opinion mining is performed using a lexicon, the first phase dedicate to create a lexicon. In next phase, required data for opinion mining is collected and then collected data are pre-processed. In the classification phase, reviews are classified in positive or negative classes. In last phase, different problems are investigated. Detailed descriptions of these phases are expressed as follows.

4.1 Persian Sentiwordnet Creation

There are different lexicons that can be used for annotating word's semantic orientation, but most of them are developed for other languages especially for English. Each of these lexicons has different mechanism for showing semantic orientation. For example, in SentiWordNet [24], each word contains three scores between [-1,1] and each score stands for positive, negative and objective orientation. Bing Liu [25] is a lexicon that contains two lists of positive and negative words. MPQA is a subjectivity lexicon that shows word's SO as positive or negative and also its subjectivity.

In Persian language, there is not such a lexicon. Shams et al [20] used machine translation to translate existing MPQA lexicon to Persian and developed a new lexicon named Persianclues.

We use existing English SentiWordNet for creating Persian SentiWordNet. Since each word in SentiWordNet has positive, negative and objective score, it didn't need to calculate this score. These scores were used for our Persian lexicon.

In this article, opinion mining is performed by applying combination of SentiWordNet and supervised algorithms. For this purpose Persian SentiWordNet is developed. English SentiWordNet has created by attaching positive and negative sentiment scores to WordNet synsets[1] [26]. In fact each word in SentiWordNet has an equivalent in WordNet. Therefore for creating Persian SentiWordNet a WordNet is required. Second version of Persian WordNet [27] is applied for developing a Persian SentiWordNet. Persian WordNet contains words in the form of noun, verb, adverb and adjective. Each synset contains some information such as part of speech, glossary and example of word usage. Beside this information, each synset has an equivalent in English WordNet. Since each word in the English SentiWordNet has an entry in English WordNet, this equivalent can be used to map each word in the Persian WordNet to English SentiWordNet. By considering this equivalent, Persian SentiWordNet is developed in three steps:

1. All words in Persian WordNet doesn't have equivalent in English WordNet, so only those words that have English equivalent are chosen. By searching automatically in Persian WordNet, 15904 synsets that have an equivalent were extracted. The problem with Persian WordNet is that some synsets have more than one English equivalent. For example word "خوب" (good) has five equivalent. For solving this problem, average of all equivalents for each synset, is calculated for its SO.
2. In the second step, all extracted synsets from the first step are mapped to their equivalent in English SentiWordNet and SO of them are retrieved. Each word may have multiple SOs in SentiWordNet. So average of these entire SOs was defined as the word's SO. In table 1 example of extracted words are shown. As you can see some words such as rain have only one equivalent and some words such as bad have more than one equivalent so we use average sentiment of these equivalents as a final sentiment. So final positive, negative and objective score of "bad" is 0.0416, 0.75 and 0.2083 respectively.
3. Retrieved verbs from Persian WordNet are in the infinitive form. So stemming is required and NLP[2] tool [28] is applied for stemming these words.

4.2 Data Description

To assess the opinion mining performance, data are collected from Persian website www.hellokish.com using Mozenda [3] web crawler. The extracted reviews from Hellokish are related to hotel domain. 1805 negative and 4630 positive reviews about hotels were collected and each of them contains an opinion about hotel, its date, writer, an option that shows it is recommend or not and percentage of satisfaction. For our purpose only opinion and its recommendation option are collected. Since each opinion contains recommendation option, this option is used as a class indicator.

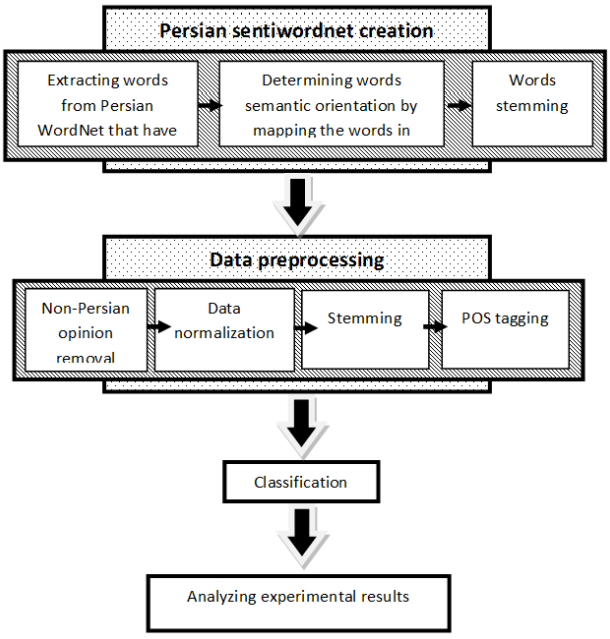


Fig. 1. Different phases of Persian opinion mining

[1] Sets of cognitive synonyms

[2] Natural Language Processing

[3] Mozenda is web crawler software

4.2.1 Data Pre-processing

Pre-processing is the process of cleaning and preparing the text for classification [29]. Keeping extra sections may increase dimensionality of classification. There are different pre-processing tools that can be applied according to our requirement. In this paper, the pre-processing has four steps. We use NLP tool [28] for step two to four.

Table 1. Example of extracted words from Persian WordNet and SentiWordNet

Synonym in WordNet	Part of speech	Synonym in SentiWordNet	ID	Positive	negative	objective
دل‌خوش, دلخوش - بشاش - خوشحال, خوش حال - شاد	Adjective	happily, merrily, mirthfully, gayly, blithely, jubilantly	50297	0.5	0.25	0.25
تماشایی - دیدنی - زیبا - قشنگ	Adverb	beautifully, attractively	242006	0.375	0	0.625
باران, بارون - مطر	Noun	rain, rainfall	11501381	0	0	1
زباله - خاکروبه - آشغال	Noun	rubbish, trash, scrap	14857497	0	0.125	0.875
نامرغوب - بد - زشت - ناشایست - چرت - مزخرف - ناشایسته - ضعیف	Adjective	bad	1125429	0	0.625	0.375
		atrocious, abominable, awful, dreadful, painful, terrible, unspeakable	1126291	0	0.875	0.125
		inapprop riate	135718	0.125	0.75	0.125

- In the first step, all opinions that were written in English are detected and removed.
- In the second step, data are normalized. Informal words, intra-word spacing problems are corrected and Arabic letters are replaced by Persian ones. There are some letters in Persian that can be written with Arabic letters. For example letter "ى" is incorrectly written in Arabic as "ي".
- In third step, words are stemmed.
- In the last step, all words are tagged with their POS[1]. Some words have more than one POS in different situations. Each POS may have different SOs in SentiWordNet. Therefore, POS tagging is required for determining words' POS.

5. Experimental Results

This research is accomplished in document level and each opinion is considered as a document. We experiment opinion mining with three standard algorithms, naive Bayes, support vector machines and logistic regression for classifying documents in two positive and negative classes. Logistic regression and naive Bayes are run by default parameters.

LibSVM with linear kernel is applied for SVM. 5-fold cross-validation is performed for the experiments reported in this study. All three algorithms are run in weka[2].

Collected data are converted to a vector of words to be operative for classifier by using StringToWord filter in weka. Three feature weighting methods (TF[3], TF-IDF and present- absent) are applied to produce three different feature sets by setting feature's frequency to three. At last, 1196 features are produced.

After that, SO of words are determined by using Persian SentiWordNet. Each word with its POS is looked up in Persian SentiWordNet. In SentiWordNet, each word has a positive, a negative and an objective score. Greater score was chosen so if the negative score is greater than the positive one, negative score is selected and vice a versa. For words that have equal positive and negative scores, zero is set as their SO. For those words that have more than one result in SentiWordNet, the average of all results as their SO is calculated.

To examine the result of opinion mining in different conditions, five hypotheses are defined. In each hypothesis, two cases are examined. The first case is related to classification of reviews before weighting by SO and the second case is related to classification after weighting by SO. In all hypotheses, a document is a vector of features and each feature is represented by a numeric value. Features have different values in different dataset according to the method of feature weighting. Four indexes were applied for evaluating classification. All these indexes were calculated in weka. Weka produces some indexes for evaluating a classification's result. So we choose four of these indexes to analysis the classification results.

These four indexes are computed as follows in Weka:

Accuracy: number of Correctly Classified Instances

Precision = TP[4] / (TP + FP[5])

Recall = TP/ (TP+ FN[6])

F- Measure = 2 * Precision * Recall / (Precision + Recall)

Details of each hypothesis are expressed bellow:

1. Present-absent: When a word exists in a document it is represented by one and if it doesn't exist it is represented by zero. Then this value is multiplied by word's SO to create a new weight. Table 2 represents the results of these two cases. As it is shown in table 2, Using SO improves accuracy of logistic regression by 3.3 percent. But accuracy of SVM and NB decreases by 0.4 and 1.7 percent respectively. This reduction for SVM is not

[1] Part of Speech
[2] Weka is an open source machine learning software
[3] Term Frequency
[4] True Positive
[5] False Positive
[6] False Negative

significant. In both cases SVM gets better results comparing to other algorithms.

2. Term frequency: In this hypothesis, Term frequency is investigated. Each number in the vector represents word's frequency in the document. Table 3 shows the results of classification. TF is defined as:

TF=log (1+f_{ij}) where f_{ij} is the frequency of the word i in document j. Using SO improves accuracy of logistic regression regressing by 3.4 percent.

Table 2. Classification results of present-absent feature

	results of classification after weighting by SO			results of classification before weighting by SO		
Result	SVM	NB	Logistic regression	SVM	NB	logistic regression
Accuracy	**85.9**	83.3	82	**85.5**	81.6	85.3
Precision	85.9	85.4	82.4	85.1	83	85
Recall	85.9	83.3	82	85.5	81.6	85.3
f-measure	85.9	83.9	82.2	85	82	85.1

Table 3. Classification results of term frequency feature

	results of classification after weighting by SO			results of classification before weighting by SO		
Result	SVM	NB	logistic regression	SVM	NB	logistic regression
Accuracy	**87**	80.1	82.4	**85.9**	79.7	85.8
Precision	86.9	82.1	82.8	85.6	80.8	85.5
Recall	87	80.1	82.4	85.9	79.7	85.8
f-measure	87	80.7	82.5	85.3	80.1	85.6

But accuracy of SVM and NB decreases by 0.4 and 1.1 percent respectively. This reduction for NB is not significant. In both cases SVM gets better results comparing to other algorithms.

3. Frequency–Inverse Document Frequency: This feature is defined as:

f_{ij}*log(number of documents /number of documents that has word i) where f_{ij} is the frequency of word i in document j. Table 4 shows the results of classification by applying this feature. Accuracy of Logistic regression improves by 3.2 percent when SO is considered. But accuracy of SVM and NB decreases by 1.6 percent. Like cases 2 and 3 SVM performs better than other algorithms.

In three above hypothesis, SVM and logistic regression achieve better results than naïve Bayes in most cases.

Accuracy of logistic regression improves about 3 percent by Appling SO. So it means that using SO has a positive influence on the performance of this algorithm.

4. Effect of number of instances: In this hypothesis, we try to investigate the effect of number of instances in the classification performance. The effect of this hypothesis is examined for term frequency feature. Table 5, shows the accuracy of classification. In both cases, number of positive and negative instances is equal.

By increasing number of instances, performance of classification increases too. Among these algorithms, SVM performs better even by decreasing number of instances. Observations show that, by decreasing the number of instances the accuracy of all algorithms decreases. But in most of cases, decrease in the accuracy of logistic regression is more than SVM and NB. Hence logistic regression is more sensitive to the number of instances.

It can be concluded that number of instances can affect classification performance and increasing number of instances has a positive effect on the performance of classification. SVM outperforms other algorithms by decreasing the number of instances. So Among these algorithms, SVM is more robots to the low number of instances.

5. Effect of unbalanced dataset: This Hypothesis investigates the effect of different number of positive and negative instances in the performance of opinion mining. In all cases total number of instances is equals 3600. Table 6 represents the accuracy of classification by using unbalanced dataset.

As table 6 shows, It is obvious that, unbalanced dataset, results in better performance for classification.

By increase in the difference between number of positive and negative instances, classification's accuracy improves too and it has a positive effect in opinion mining performance. In both cases SVM and logistic regression perform better than Naïve Bayes. Improvement in the results of SVM is more than other algorithms. This result confirm that when most of the instanced are belong to one class, the classifier assign a correct class to given sample.

Table 4. Classification results of TF-IDF feature

	results of classification after weighting by SO			results of classification before weighting by SO		
Result	SVM	NB	logistic regression	SVM	NB	logistic regression
Accuracy	**83.9**	80	82.1	**85.5**	81.6	85.3
Precision	84	80.8	82.4	85.1	83	85
Recall	83.9	80	82.1	85.5	81.6	85.3
f-measure	83.9	80.3	82.2	85	82	85.1

Table 5. Effect of number of instances on the performance of classification

Results of classification after weighting by SO			Results of classification before weighting by SO			
Number of instances	logistic regression	NB	SVM	logistic regression	NB	SVM
83.0556	76.6111	72.6389	**81.6667**	3600	80.1389	75.5
83.4118	77	73.6176	**81.7941**	3400	80.1176	76.0882
83.25	77.375	71.4375	**82**	3200	79.9063	76.6563
83.4	77.0667	71.4667	**81.7**	3000	80.6667	77.2667
82.5714	76.75	69.75	**81.7857**	2800	79.3214	76.1786
82.3462	76.7308	67.7308	**81.8077**	2600	76.6154	75.9615
82.7083	77.5	67.4167	**80.9583**	2400	76.8333	75.9167
82.0909	76.7727	65.7273	**81.3636**	2200	76.2727	76.5909
83.4	77.65	66.65	**82.8**	2000	75.7	77.3
82.7778	77.7778	65.5556	**82.6111**	1800	74.7778	77.7222
83.125	76.5	67.25	**82.625**	1600	74.625	75.125
81.2857	78.2857	61.5714	**80.5**	1400	70.1429	76.1429
80.6667	76.25	66.25	**79.5**	1600	67.9167	73.5833
78.5	73.9	70.6	**76.8**	1000	64.6	74.1
78.25	78	77.125	**77.5**	800	65.875	77.125
79.1667	78.1667	78.8333	**77**	600	61	75.3333
78.25	70.5	73	**74.75**	400	66.75	69.5
63	**67**	66.5	**67.5**	200	63	64.5

Table 6. Effect of unbalanced dataset on the performance of classification

Results of classification after weighting by SO			Results of classification before weighting by SO				
SVM	NB	logistic regression	SVM	NB	logistic regression	Number of negative instances	Number of positive instances
83.8611	76.8889	73.9444	**82.4444**	76.1389	81.4167	1700	1900
84.1944	78.2222	74.4444	**82.0833**	76.6111	81.1667	1600	2000
84.6389	78.25	76.3333	**82.4167**	77.5833	81.5556	1500	2100
83.6667	79.25	75.25	**83.3333**	78.4722	81.9167	1400	2200
84.5556	79.5833	74.9722	**83.8611**	77.9722	81.5278	1300	2300
83.6667	79.25	75.25	**84.9722**	79.8611	83.0278	1200	2400
84.5556	79.5833	74.9722	**85.1944**	80.5278	83.7778	1100	2500
85.7778	79.8056	76.6389	**85.7222**	81.4167	84.5833	1000	2600
87.25	82.4722	79.1389	**86.5278**	81.9444	84.6944	900	2700
88.0556	83.1944	78.3611	**87.0833**	82.3333	85.6944	800	2800
89.5278	84.2222	78.25	**87.7778**	82.8333	85.6944	700	2900
89.6111	83.9444	78.9722	**88.6667**	82.5	86.0556	600	3000
86.1111	83.6667	81.1667	**90.1944**	83.1667	87.25	500	3100
91.7222	84.1667	82.7222	**91.5833**	82.1944	87.0833	400	3200

6. Conclusions

In this article, a comprehensive experiment of opinion mining was conducted in Persian language. Opinion mining was performed by applying combination of a Persian SentiWordNet and supervised algorithms SVM, logistic regression and Naive Bayes. Opinion mining is not performed in Persian language using SentiWordNet. So this lexicon is created to investigate its effect on opinion mining.

Therefore at the first phase a Persian SentiWordNet was created by applying Persian WordNet. For our purpose, the reviews from hotel domain were collected and defined some hypothesis to investigate the Persian opinion mining in different conditions. In the first three hypotheses, three feature weighting methods (present-absent, TF and TF-IDF) were applied. In the all hypotheses, features value was multiplied by their SOs. The results of classification by using these features were compared to results of classification before considering SO. SVM and logistic regression performed better than Naïve Bayes. By considering SO the accuracy of logistic regression improved by about 3 percent. By examining the effect of number of instances, we observed that increasing number of instances has a positive effect on the performance of opinion mining.

By decreasing number of instances, performance of all algorithms decreases too, but decrease in the performance of logistic regression was more than other algorithms and also SVM is more robust to low number of instances comparing Naïve Bayes and logistic regression. Using unbalanced dataset improved the classification results. In the future work, we want to investigate Persian SentiWordNet performance in other domains and also assess the performance of other approaches in the Persian opinion mining. By considering all of these results we can observe that SentiWordNet can gain acceptable accuracy.

References

[1] Liu, B (2012). Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies), Morgan & Claypool. doi:10.2200/S00416ED1V01Y201204HLT016

[2] Albornoz, J, Plaza, L, Gerv´as, P , D´ıaz, A (2011). A joint model of feature mining and sentiment analysis for product review rating In 33rd European Conference on IR Research. pp. 55–66. Doi: 10.1007/978-3-642-20161-5_8

[3] Pang, B, Lee, L (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In ACL-02 conference on Empirical methods in natural language processing 10: 79–86. 10.3115/1118693.1118704
[4] Turney, P.D (2001). Thumbs up or thumbs down? In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02. Morristown, NJ, USA: Association for Computational Linguistics: 417-424. Doi:10.3115/1073083.1073153
[5] Brooke, J, Tofiloski, M, Taboada, M (2009) Cross-linguistic sentiment analysis: From English to Spanish. In Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing, pp 50–54.
[6] Molina-González M.D, Martínez-Cámara, E, Martín-Valdivia, M(2013). Semantic orientation for polarity classification in Spanish reviews. Expert Systems with Applications, 40(18):7250–7257. DOI: 10.1016/j.eswa.2013.06.076
[7] Wei, B, Pal, C (2010). Cross lingual adaptation: an experiment on sentiment classifications. In Proceedings of the ACL 2010 Conference Short. Pp. 258-262 .
[8] Wan, X (2008). Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. In EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 553–561.
[9] Wan, X (2009). Co-Training for Cross-Lingual Sentiment Classification. In ACL '09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1. pp. 235–243.
[10] Yu, L, Ma, J (2008). Opinion mining: A study on semantic orientation analysis for online document. In 7th World Congress on Intelligent Control and Automation. Ieee, pp. 4548–4552. Doi: 10.1109/WCICA.2008.4594529
[11] Ye, Q, Zhang, Z, Law, R(2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Systems with Applications, 36(3):6527–6535. DOI: 10.1016/j.eswa.2008.07.035
[12] Taboada, M, Brooke, J, Tofiloski, M, Voll, K, Stede, M(2011). Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics: 37(2), pp.267–307. doi>10.1162/COLI_a_00049
[13] Hung, C, Lin, H.K (2013). Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification. IEEE Intelligent Systems, 28(2), pp.47–54. DOI: 10.1109/MIS.2013.1
[14] Martineau, J, Finin, T (2009). Delta TFIDF: An Improved Feature Space for Sentiment Analysis. In Third AAAI International Conference on Weblogs and Social Media.
[15] Tan, S, Zhang, j (2007). An empirical study of sentiment analysis for chinese documents, Expert Systems with Applications 34: 2622–2629.
[16] Li, G, Liu, F (2012). Application of a clustering method on sentiment analysis. Journal of Information Science 38:127–139. Doi:10.1177/0165551511432670
[17] Li, G, Liu, F (2013). Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. Applied Intelligence 40: pp.441–452. Doi: 10.1007/s10489-013-0463-3
[18] Basari, A.S.H, Hussin, B, Ananta, G.B, Zeniarja, J (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Procedia Engineering, 53, pp.453–462. DOI: 10.1016/j.proeng.2013.02.059
[19] Vinodhini, G, Chandrasekaran, R. M (2014). Sentiment Mining Using SVM-Based Hybrid Classification Model. In G. S. S. Krishnan et al., eds. Proceedings of ICC3, volum 246: 155–162. Doi: 10.1007/978-81-322-1680-3_18
[20] Shams, M, Shakery, A, Faili, H (2012). A non-parametric LDA-based induction method for sentiment analysis. In The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012). IEEE, pp. 216–221. Doi: 10.1109/AISP.2012.6313747
[21] Wiebe, J, Wilson, T, Cardie, C (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 39: 165-210. Doi: 10.1007/s10579-005-7880-9
[22] Hajmohammadi, M.S, Ibrahim, R (2013). A SVM-based method for sentiment analysis in Persian language. In Z. Zhu, ed. international Conference on Graphic and Image Processing, p. 876838. doi:10.1117/12.2010940
[23] Saraee, M, Bagheri, A (2013). Feature Selection Methods in Persian Sentiment Analysis. In 18th International Conference on Applications of Natural Language to Information Systems: 7934: 303–308. Doi: 10.1007/978-3-642-38824-8_29
[24] Baccianella, S, Esuli, A, Sebastiani, S (2010). SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2200-2204. European Language Resources Association.
[25] Hu, M, Liu, B (2004). Mining and summarizing customer reviews. In Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04. New York, New York, USA: ACM Press, pp 168.
[26] Miller, G.A (1995). WordNet: A Lexical Database for English, Communications of the ACM Vol. 38, No. 11: 39-41
[27] Shamsfard, M, Hesabi, A, Fadaei, H, Mansoory, N, Famian, A, Bagherbeigi, S, Fekri, E, Monshizadeh, M, Assi, S.M (2010). Semi Automatic Development of Farsnet; the Persian Wordnet. Proceedings of 5th Global WordNet Conference (GWA2010). Mumbai, India
[28] Estiri, A, Kahani, M, Hoseini, M, Asgarian, E (2012) Designing Persian language parser tool, International Conference on Asian Language Processing (IALP 2012)
[29] Haddi, E, Liu, X, Shi, Y (2013). The Role of Text Pre-processing in Sentiment Analysis. In Procedia Computer Science. pp. 26–32. DOI: 10.1016/j.procs.2013.05.005

Saeedeh Alimardani received the B.Sc degree in Information Technology from Payamenoor university, Qazvin, Iran in 2011 and the M.sc degree in Information Technology from Khajeh Nasir Toosi University of Technology, Tehran, Iran in 2014. her main research interests include data mining, text mining and opinion mining.

Abdollah Aghaei is a professor of Industrial Engineering department of Khajeh Nasir Toosi University of Technology, Tehran, Iran. He has B.Sc. degree in Industrial Engineering from Sharif University of Technology and M.Sc. degree in Industrial Engineering from New South Wales University, Australia. He received Ph.D. degree in ndustrial Engineering from Loghbrough University, UK. His field of Experiences are Modeling and computer Simulation, Quality systems and Standards, Knowledge Management, Innovation Management, Project Management, Ergonomics, Risk Management, Business Planning and e-Marketing.

Application of Curve Fitting in Hyperspectral Data Classification and Compression

Seyed Abolfazl Hosseini
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
abolfazl.hosseini@modares.ac.ir
Hassan Ghassemian*
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
ghassemi@modares.ac.ir

Received: 10/May/2015 Revised: 24/Jun/2015 Accepted: 30/Jul/2015

Abstract

Regarding to the high between-band correlation and large volumes of hyperspectral data, feature reduction (either feature selection or extraction) is an important part of classification process for this data type. A variety of feature reduction methods have been developed using spectral and spatial domains. In this paper, a feature extracting technique is proposed based on rational function curve fitting. For each pixel of a hyperspectral image, a specific rational function approximation is developed to fit the spectral response curve of that pixel. Coefficients of the numerator and denominator polynomials of these functions are considered as new extracted features. This new technique is based on the fact that the sequence discipline - ordinance of reflectance coefficients in spectral response curve - contains some information which has not been considered by other statistical analysis based methods, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) and their nonlinear versions. Also, we show that naturally different curves can be approximated by rational functions with equal form, but different amounts of coefficients. Maximum likelihood classification results demonstrate that the Rational Function Curve Fitting Feature Extraction (RFCF-FE) method provides better classification accuracies compared to competing feature extraction algorithms. The method, also, has the ability of lossy data compression. The original data can be reconstructed using the fitted curves. In addition, the proposed algorithm has the possibility to be applied to all pixels of image individually and simultaneously, unlike to PCA and other methods which need to know whole data for computing the transform matrix.

Keywords: Hyperspectral; Feature Extraction; Spectral Response Curve; Curve Fitting; Classification.

1. Introduction

Hyperspectral (HS) data contain hundreds of spatially coregistered images in the form of image cubes; each image corresponds to a specific narrow spectral band usually in the wavelength range of 400-2500 nm. In other words, pixels of an N bands HS data set can be considered as N-dimensional vectors containing sequential intensities that are measured by the HS sensor. For each pixel, the plot of bands intensity values ($y= [y_1, y_2 \dots y_N]^t$) vs. band numbers ($x=[1, 2,\dots, N]^t$) is named as spectral signature or spectral response curve (SRC) of pixel. Some theoretical and practical problems appear in supervised classification of this type of data because of high dimensional spaces specifications. The most important problem in is Hughes phenomenon [1]. There are four strategies to combat Hughes phenomenon in hyperspectral images classification: semi-supervised classification [2,3], combination of spatial and spectral information [4-6], utilizing classifiers like SVM which are less sensitive to the number of training samples [7], and feature reduction (extraction/selection) [1].

Statistical analysis based transformations for performing unsupervised or supervised feature reduction approaches are powerful methods in classification and visualization of remote sensing data [8,9]. In feature reduction we find a transformation that maps data to a lower dimensional space, mainly preserving essential discriminative information. Great efforts have been made to develop advanced feature reduction and classification methods to improve classification accuracy rate. Principal component analysis (PCA) [1], decision boundary feature extraction (DBFE), non-parametric weighted feature extraction [10], wavelet transform [11], maximum margin projection (MMP) [12], linear discriminant analysis (LDA) [13], and independent components analysis (ICA) [14] are examples of feature extraction (F.E.) techniques for data redundancy reduction in remotely sensed data. Nonlinear extensions of PCA (KPCA) have been proposed by using kernel trick [15]. Kernel Fisher discriminant analysis (KFD) [16] and Generalized discriminant analysis (GDA) [17] have been developed independently as kernel-based nonlinear extensions of LDA. Also optimal selection of spectral bands has been extensively discussed in the literature by utilizing approaches like sequential forward/backward methods and information theoretical based methods [1]. After feature reduction step, the features are fed to the classifier. A widely used parametric classifier in hyperspectral data classification is the maximum likelihood classifier [18].

* Corresponding Author

Although methods like PCA, LDA and many other statistical analysis based methods have simple structure and relatively good results, but there are some deficiencies when they are used in hyperspectral Image classification. An important deficiency is that these methods do not consider the geometric aspects of SRCs and the ordinance of original features which is a rich source of information. For each pixel of a hyperspectral image we have a vector of measured quantities corresponding to reflection coefficients of consecutive wavelengths (SRC). Therefore the ordinance of measured data might have some information that could be useful in classification process. In another word, SRC, as a geometrical curve has some useful information. A feature extraction techniques based on the fractal nature of SRC was introduced by [19,20] that depends on the ordinance of samples. Other disadvantage of many F.E. methods is that they cannot be applied to individual pixels independently and first we need to transform whole data to new space and then produce new features. The main contribution of the present letter is introducing another F.E. method which considers the geometrical nature of the SRCs and the ordinance information existing in the SRC that yields improvement of correct classification rate and some other evaluation factors. Indeed we try to fit a rational function with polynomial numerator and denominator to the SRC of each pixel. Then the coefficients of these polynomials are considered as new feature vectors and are fed to an ML classifier. Results are compared to PCA and LDA as two basic and classic F.E. methods with relatively good performance for HS data classification. Unlike PCA and LDA, our proposed method is applied pixel by pixel and does not need to transform whole data to a new space. Therefore, a parallel implementation of the algorithm is possible. Moreover, since the proposed transform is invertible, it could be used as a lossy compression algorithm for HS data, too.

2. Curve Fitting and its Discriminating Ability

In many measurement problems, curve fitting is a traditional approach to find the mathematical relationship between observed values and independent variables. It tries to fit an appropriate curve to the observed values. In addition, curve fitting can be used for noise reduction and data smoothing purposes, and data interpolation / extrapolation [21,22]. The aim of curve fitting is to find a function $f(\lambda)$ in a pre-specified class of functions for the data $\{(\lambda, I_\lambda)\}$ where λ= 1,2,…,N, which minimizes the residual (the distance between the data samples and $f(\lambda)$) under the weight W= (w_λ) [23].

Least squares (LS), least absolute residuals (LAR), and bisquare fitting method are some of fitting criteria used to perform linear or nonlinear fittings to find the function $f(\lambda)$ [24]. In the LS method $f(\lambda)$ is found by minimization of the following weighted mean squared error:

$$\frac{1}{N}\sum_{\lambda=1}^{N} w_\lambda (f(\lambda) - I_\lambda)^2 \qquad (1)$$

There are different curve fitting models like polynomial, linear, spline, etc. In this letter, a special model of nonlinear curve fitting using rational functions is utilized in order to extract new features for classification and compression purposes. Therefore, it is important that $f(\lambda)$ has fewer parameters than the number of data samples.

Consider a function f and the two integers L ≥ 0 and M ≥ 0, the rational function approximant of order (L,M) for f is defined as:

$$[L/M]_f(\lambda) \triangleq \hat{f}(\lambda) = \frac{\sum_{j=0}^{L} c_{j+M+1}\lambda^j}{1+\sum_{j=1}^{M} c_j \lambda^j} \qquad (2)$$

Since, dissimilarities of curves yields differences in the coefficients of the approximants of the corresponding curves, it seems that these coefficients could be used as discriminating features for the curves. For example, in figure 1(a) two different families of curves are plotted, and the histogram of the coefficients of their rational function approximants are demonstrated in figure 1(b)-(f). In this curve fitting problem, we fitted a rational function with L=0 and M=4 to the both families of curves. Therefore, each curve can be expressed with its own five coefficients. As can be seen, the histograms of some coefficients completely separate the two families. This fact was our motivation for using the coefficients of the rational function fitted curve of the SRCs as discriminating features in HS data classification tasks.

3. The Proposed Feature Extraction Method

We introduced the spectral response curve (SRC) of each pixel of a hyperspectral image as the plot of its measured intensities in different wavelengths versus wavelengths or band numbers.

In other words, each SRC can be considered as a plot of a function $f(\lambda)$. As a matter of fact, we do not know the exact mathematical expression of $f(\lambda)$, but we have the amounts of its N consecutive points. We show that an approximation of $f(\lambda)$ in the form of a rational function with polynomial numerator and denominator can be developed through an LS method with uniform weights in (1). Then, we demonstrate that the coefficients of these polynomials can be applied as new features for maximum likelihood classifier with satisfactory results. Also these features can be used for reproducing the original data.

To provide equal condition for all hyperspectral images, the variable λ is considered in the form λ/N as normalized band number. The rational approximant of $f(\lambda/N)$ for the pixel located at (x,y) is given by:

$$\hat{f}_{(x,y)}\left(\frac{\lambda}{N}\right) = \frac{\sum_{j=0}^{L} c_{j+M+1}(\frac{\lambda}{N})^j}{1+\sum_{j=1}^{M} c_j(\frac{\lambda}{N})^j} \qquad (3)$$

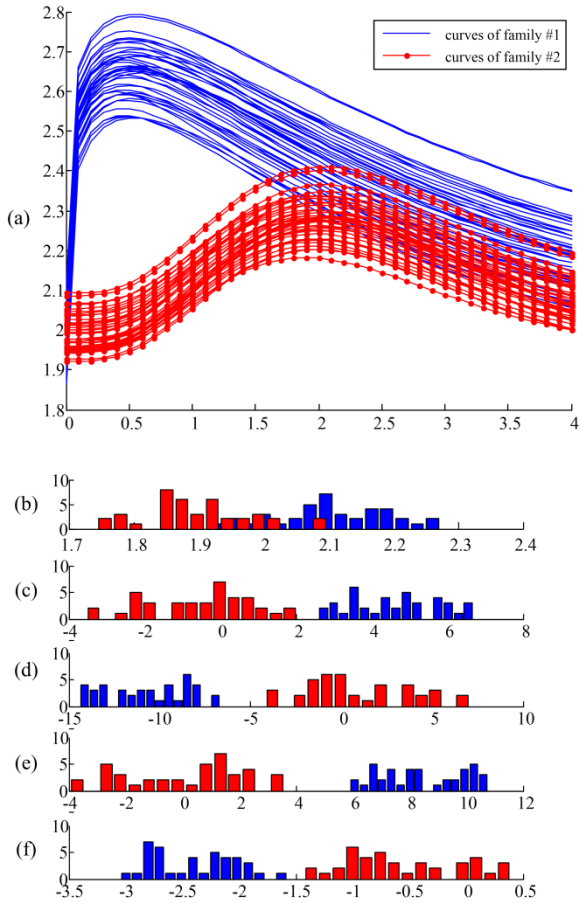


Fig. 1. Two families of curves; and the histograms of the coefficients of their corresponding Padé approximants, (a) Curves, blue (solids) and red (dotted) lines (b) numerator coefficient distribution, (c)-(f) denominator coefficients distribution

We want to determine coefficients vector $\boldsymbol{c}=[c_1\ c_2\ \ldots c_{M+L+1}]^t$ to minimize:

$$E = \frac{1}{N}\sum_{\lambda=1}^{N}\left(\hat{f}(\lambda/N) - f(\lambda/N)\right)^2 \qquad (4)$$

By computing partial derivatives of E with respect to the coefficients and setting them to zero, a system of nonlinear equations is obtained. A sufficient but not necessary condition for solving this system is to find **c** such that:

$$\sum_{j=0}^{L} c_{j+M+1}\left(\frac{\lambda}{N}\right)^j - f\left(\frac{\lambda}{N}\right)\sum_{j=1}^{M} c_j\left(\frac{\lambda}{N}\right)^j = f\left(\frac{\lambda}{N}\right); \qquad \lambda = 1,\ldots,N \qquad (5)$$

Now we have a system of linear equations with $M+L+1$ unknowns and N equations which can be rewritten in matrix form as:

$$A_{N\times(M+L+1)}\,c_{(M+L+1)\times 1} = B_{N\times 1} \qquad (6)$$

where,

$$\begin{cases} A_{N\times(M+L+1)} = [a_{\lambda j}] \\ a_{\lambda j} = \begin{cases} -f\left(\frac{\lambda}{N}\right).\left(\frac{\lambda}{N}\right)^j & j = 1,\ldots,M \\ \left(\frac{\lambda}{N}\right)^j & j = M+1,\ldots,M+L+1 \end{cases} \\ B = \left[f\left(\frac{1}{N}\right), f\left(\frac{2}{N}\right),\ldots,f\left(\frac{N}{N}\right)\right]^t \end{cases}$$

For feature reduction purposes, we want $M+L+1<<N$, therefore the matrix **A** is not square and the system of linear equations in (6), does not have a unique solution and sometimes may not have any solution. Instead, we use Moore-Penrose pseudo inverse of **A** and find **c** such that the norm of **Ac-B** be minimized.

The above procedure must be performed for all pixels of the hyperspectral data set and the vectors **c** replace the original data and new image cube is developed. Therefore the third dimension of data is changed to $M+L+1$, achieving to resize data for a rate of $N/(M+L+1)$. The procedure can be performed for all pixels, simultaneously.

This rational function curve fitting feature extraction (RFCF-FE) method is an unsupervised feature extraction method. The new features are applied to Maximum Likelihood (ML) classifiers and results are compared to PCA as a traditional unsupervised feature extraction method and LDA as a supervised feature extraction method. As it is demonstrated in next section, the RFCF-FE method results are more accurate than its competing methods for both urban and agricultural data sets. Also, it can be used as a coding algorithm for hyperspectral data compression.

4. Experimental Results

4.1 Hyperspectral Data Sets

The first data set used in our experiments is a mixed forest/agricultural 145×145 pixels image from Indian Pine Site (IPS) in Indiana [25]. It is captured by the Air-borne Visible/Infrared Imaging Spectrometer (AVIRIS). The spatial resolution of this data set is 20m. The picture contains 220 spectral bands in the wavelength range from 400 to 2500 nm with 10nm resolution. After removing twenty water absorption bands, N=200 bands were left. The class map of data contains 16 different land covers.

The other data set was gathered over the urban area of the University of Pavia (UP), by the Reflective Optics System Imaging Spectrometer (ROSIS) [25]. The image size is 610×340 pixels, the spatial resolution of 1.3m and the number of spectral bands is 115 in the wavelength range from 430 to 860 nm. Discarding some noisy bands yields N=103 bands, finally. Its ground truth map contains 9 different classes of land cover.

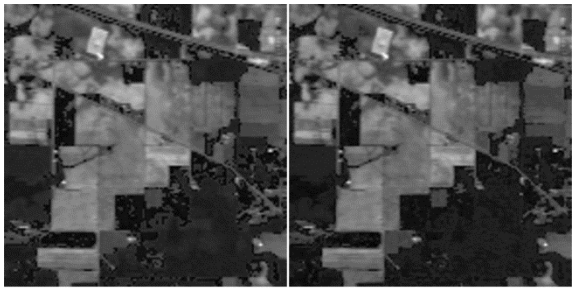
As an example, Images of band 18 of these data sets are demonstrated in Figures 2 (a) and 3 (a), respectively.

4.2 Results and Discussion

The proposed feature extraction method has been applied to both hyperspectral data sets IPS and UP, and the extracted features have been fed into an ML classifier. The classification results have been compared to those of PCA and LDA features, the two basic and traditional feature reduction approaches. For a given number of features, D, the parameter L has been changed in the range of 0 to D-1. Then M is selected regarding to the constraint: $M+L+1 = D$. For each value of D, the values

of L and M producing best results have been selected for comparison to PCA and LDA with the same dimensions. Since, 10 percent of the whole data volume with a minimum of 15 and a maximum of 50 samples per class is used for training the classifier, D is changed in the range $2 \leq D \leq 14$. Figures 2-(b) and 3-(b) show the images of 18th band of input data sets, reconstructed from extracted features by the RFCF-FE method (L=0, M=13 for IPS and L=12, M=1 for UP, as two typical amounts of M and L). Reconstruction process has been performed using (4) for λ=1, 2,…, N. Examples of real SRCs and their rational fitted curves of these two data sets are plotted in figure 4 where (L,M) are equal to (0,13) and (12,1) for IPS and UP, respectively.

The above discussions about Figures 2-4 show that the RFCF-FE method has high ability to preserve spatial characteristics of data as well as the spectral characteristics. Although, in a few points as it is illustrated in Figure 4, there may be a large difference between the original and the fitted curve, our results show that these exceptional points do not have severe impact on the classification performance. This is because: 1) these exceptional points in each SRC are rare, if any; 2) the locations of these points are not the same for different SRCs, i.e. this phenomenon does not destroy any specific band completely; 3) for each SRC, the coefficients of the corresponding fitted curve are used as the features, not the curve itself. As illustrated in Figure 4, the RFCF-FE method performs a smoothing operation in spectral domain. Moreover, it can be seen that by applying this method for IPS data some spatial domain smoothing has occurred without destroying edges of regions.



(a) (b)

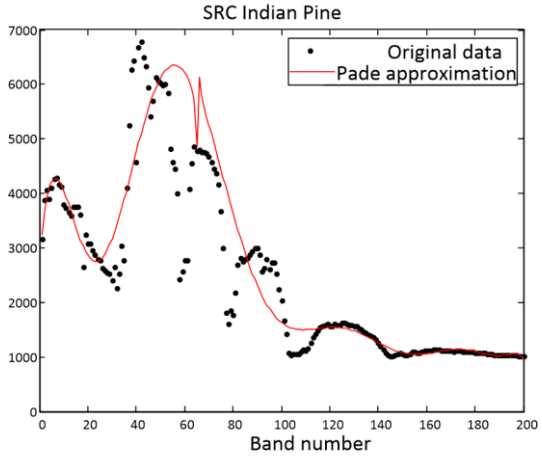
Fig. 2. Comparing the original and the reconstructed image of IPS (band 18) (a) Original data and (b) RFCF-FE method

(a) (b)

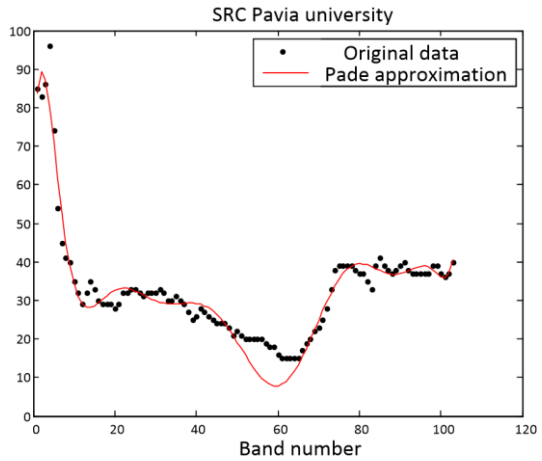
Fig. 3. Comparing the original and the reconstructed image of UP (band 18) (a) Original data, and (b) RFCF-FE method

Figures 5 and 6 demonstrate the accuracy assessment measures for IPS and UP data sets, respectively. These results are obtained by averaging the results of ten Monte Carlo runs (the standard deviations are illustrated by error bars). Also, Table 1 contains the comparison results in terms of parameter Z (see appendix A). The optimum values of L (and so M) parameters of the RFCF-FE method differ for different values of D and different iterations of the algorithm, but in the most cases, the best classification results have occurred when values of L are 0, 1, D-2, and D-1. Note that, despite PCA and the RFCF-FE method, maximum number of extracted features in LDA method is equal to N_c-1, where N_c is the number of classes.

The superiority of the RFCF-FE method compared to PCA and LDA methods is apparent from this table and Figures 5 and 6. As it can be seen, all measures (average accuracy, average validity, overall accuracy, kappa statistics, and Z) have been dramatically improved by RFCF-FE algorithm in comparison to both PCA and LDA. In terms of the first four accuracy measures, this improvement in IPS is more than in UP because of its larger ground pixels and agricultural nature with relatively large uniform areas and less details. Whereas, in terms of Z the improvement in UP is much more obvious. Note that average accuracy, average validity, overall accuracy, and kappa statistics focus on pure average accuracies despite Z parameter which focuses on disagreements between two algorithms and determines the superior one. Therefore, our method excels PCA and LDA for agricultural scenes as well as urban ones.



(a)



(b)

Fig. 4. SRC of pixel (20,20) and its approximation (a) IPS data and (b) UP

Table 2 contains PSNR (see appendix B) values for reconstructed images from PCA and the proposed method for IPS and UP, respectively when D varies from 3 until 15. PSNR for the proposed method corresponds to the values of L and M that yields to the best result. These optimum values of L and M have been shown in these tables. The superiority of the proposed method with respect to PCA based compressing method is apparent from these tables. As it is demonstrated in Table 2, in most cases for UP data set, the best result happens when M=0. It implies that McLaurin series has better performance in these cases.

Another advantage of the proposed method is that unlike the competitive methods, it is applied pixel by pixel and does not need to transform the data to the projection space as a whole. Therefore, real-time or parallel implementation of the algorithm is possible. However, the complexity of the method might be higher than the competitive algorithms. Moreover, since the proposed transform is invertible, it could be used as a compression algorithm for HS data.

Table 1. Parameter Z (mean± standard deviation of 10 MonteCarlo runs) of ML classifier for RFCF features vs. PCA and LDA

No. of Features	IPS data set		UP data set	
	RFCF-FE vs. PCA	RFCF-FE vs. LDA	RFCF-FE vs. PCA	RFCF-FE vs. LDA
2	5.87 ± 4.96	5.50 ± 1.99	0.10±0.30	0.07±0.23
3	7.60 ± 0.94	8.47 ± 1.36	15.54 ± 6.07	16.15±5.91
4	2.87 ± 1.12	5.27 ± 2.30	7.15±4.54	8.38±3.53
5	0.37 ± 0.40	1.86 ± 1.70	12.51±5.24	13.55±5.19
6	5.32 ± 4.33	5.56 ± 3.76	13.91±4.86	18.89±3.94
7	2.40± 1.79	2.50 ± 2.63	19.41±4.24	22.90±3.44
8	3.19 ± 3.12	4.11 ± 4.05	18.40±5.25	20.98±5.40
9	1.04 ± 1.25	3.07 ± 3.33	24.03±6.10	-
10	0.07 ± 0.21	0.27 ±0.49	23.82±4.80	-
11	0.26 ± 0.61	1.27 ± 1.07	21.62±4.70	-
12	3.67 ± 1.07	4.87 ± 1.91	23.50±6.23	-
13	0.67±0.81	1.36±1.50	36.25±6.90	-
14	4.12±1.74	4.04±1.87	36.21±6.57	-

Table 2. Comparing PSNR of the proposed method and inverse PCA for IPS and UP data sets in terms of compression rate (N/D)

Rate *(N/D)*	IPS			Rate *(N/D)*	UP		
	Best *(L,M)*	PSNR	IPCA PSNR		Best *(L,M)*	PSNR	IPCA PSNR
200/3	(0,2)	34.95	24.39	103/3	(2,0)	31.85	29.76
200/4	(0,3)	27.42	25.74	103/4	(3,0)	20.25	16.44
200/5	(0,4)	58.70	24.20	103/5	(2,2)	32.92	23.54
200/6	(0,5)	29.63	23.28	103/6	(5,0)	21.35	16.17
200/7	(4,2)	24.17	24.46	103/7	(5,1)	25.29	19.27
200/8	(0,7)	22.78	25.13	103/8	(2,5)	24.06	19.52
200/9	(6,2)	55.11	25.21	103/9	(1,7)	43.77	22.52
200/10	(2,7)	34.38	25.90	103/10	(2,7)	38.98	21.77
200/11	(6,4)	28.46	27.03	103/11	(6,4)	68.19	22.74
200/12	(8,3)	28.17	28.99	103/12	(11,0)	34.11	23.65
200/13	(5,7)	31.47	29.43	103/13	(12,0)	27.86	22.91
200/14	(2,11)	41.01	31.47	103/14	(13,0)	26.05	24.08
200/15	(10,4)	42.38	30.14	103/15	(14,0)	35.47	23.98

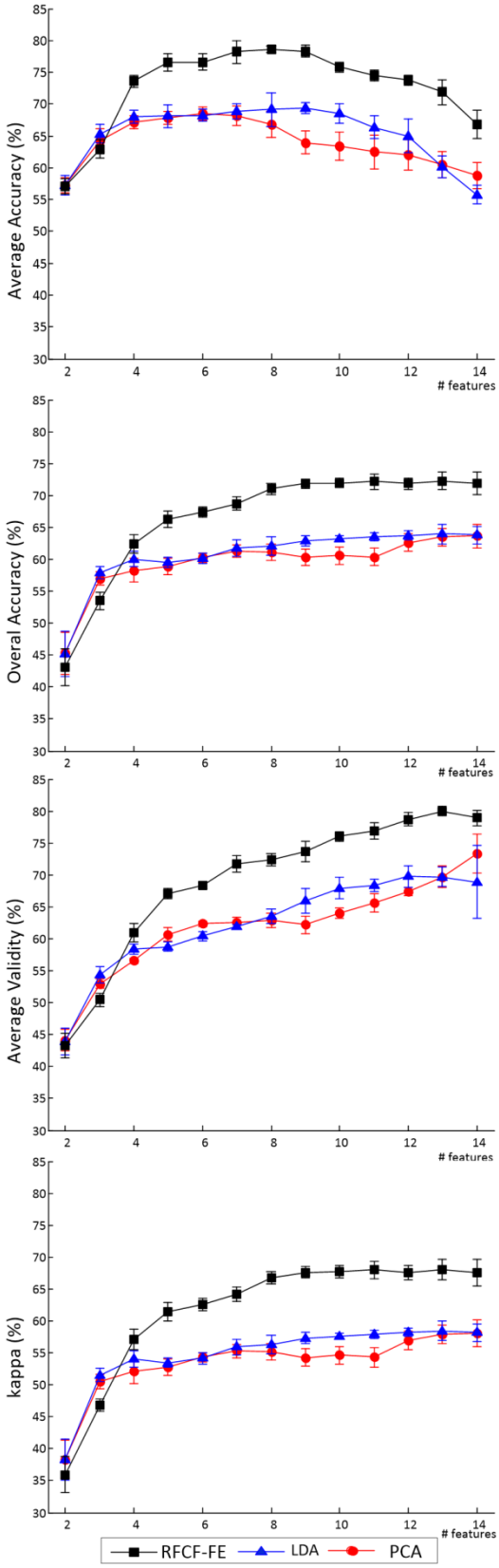


Fig. 5. Comparison of the RFCF-FE method with PCA and LDA for IPS data

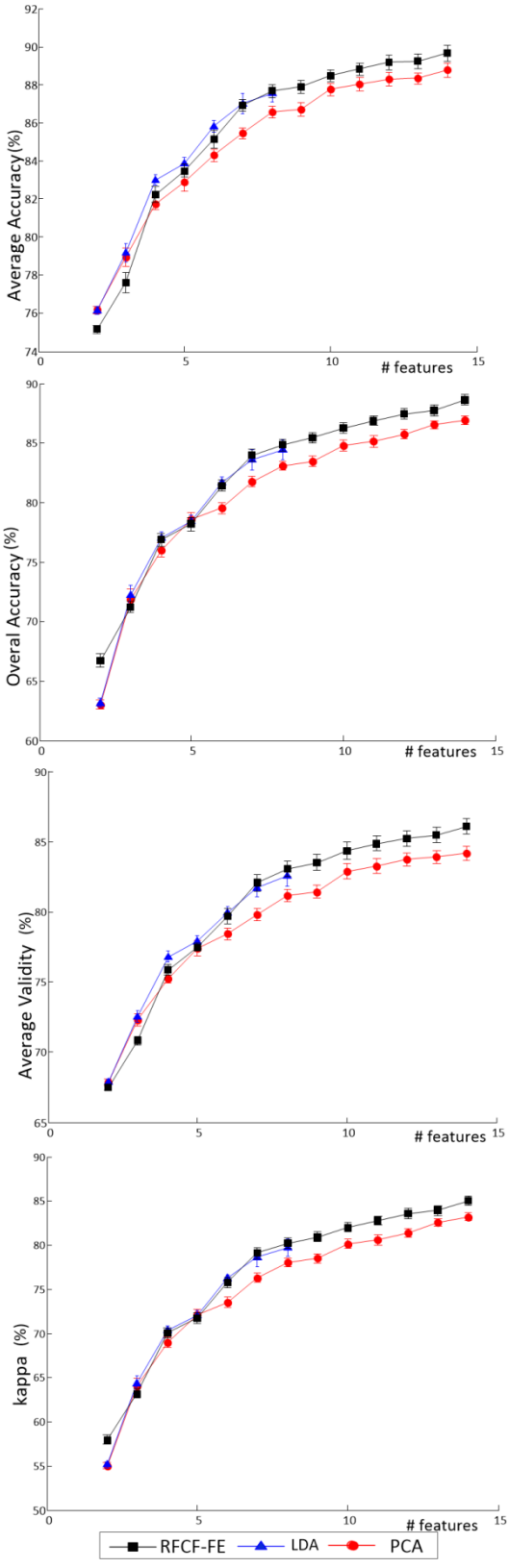


Fig. 6. Comparison of the RFCF-FE method with PCA and LDA for UP data

5. Conclusions

A new feature extraction method for hyperspectral data was proposed based on rational function curve fitting. The main motivation for using curve fitting approach for hyperspectral data feature extraction is the utilization of the information that exists in the sequence of original features (ordinance of reflectance coefficients in SRC) that are neglected by competing methods. The coefficients of SRCs approximants are calculated through analytical operations. These extracted features are then fed into an ML classifier. The classification performance was compared to PCA and LDA as two classic and traditional F.E. methods. The results show the superiority of the proposed method. Also, it has been shown that this technique has satisfactory results for signal visualization and signal representation, and can be considered as a good coding algorithm for lossy compression of HS data. The proposed method is applied pixel by pixel and does not need to transform whole data to a new space simultaneously. Therefore, it can be applied to all pixels in a parallel procedure. In addition, this method is a novel approach which can be used as a powerful base for developing more efficient feature extraction methods.

6. Acknowledgement

This research is done by support of Iran communication research centre under contract number 18133/500 T by Identification code: 90-01-03. The authors gratefully acknowledge that organization for its support.

Appendices

A. McNemar's test:

In a classification assessment procedure, the statistical significance of differences can be computed using McNemar's test. This test is based upon the standardized normal test statistic $Z = (n_{12} - n_{21})/\sqrt{n_{12} + n_{21}}$ where n_{12} is the number of samples classified correctly by classifier 1, but wrongly by classifier 2. At the 5% level of significance, if the difference in accuracy between classifiers 1 and 2 is statistically significant. When classifier 1 is more accurate than classifier 2, *Z* is positive or vice versa [26].

B. Average Classification Accuracy (AA)

Classification accuracy for each individual class is calculated as ACC(c) = nc/Nc, in which, nc is the number of the pixels of class *c* correctly classified, and N_c is the number of test pixels in that class. Average classification accuracy (AA) is then defined as follows:

$$\mathrm{AA} = \frac{1}{\mathrm{C}}\sum_{\mathrm{c}=1}^{\mathrm{C}} \mathrm{ACC(c)} \qquad (7)$$

where, C is the number of all classes.

C. Average Classification Validity (AV)

Classification validity (or reliability) for each individual class is calculated as $\mathrm{VAL(c)} = n_c/m_c$ in which, n_c is the number of the pixels of class *c* correctly classified, and m_c is the number of all the pixels labelled as class *c* in output class map. Average classification validity (AV) is then defined as:

$$\mathrm{AV} = \frac{1}{\mathrm{C}}\sum_{\mathrm{c}=1}^{\mathrm{C}} \mathrm{VAL(c)} \qquad (8)$$

D. Overall Classification Accuracy (OA)

This measure is similar to average classification accuracy, except that classes are not considered individually, but as a whole:

$$\mathrm{OA} = \frac{\mathrm{n}}{\mathrm{N}} \qquad (9)$$

in which, n is the number of all the pixels correctly classified, regardless of the class label, and N is the number of all pixels in the test set.

E. Kappa coefficient

This coefficient is the overall accuracy (OA) corrected by the amount of agreement that is expected due to chance:

$$kappa = \frac{OA - P_e}{1 - P_e} \; ; in\ which\ \ P_e = \frac{\sum_{c=1}^{C} m_c \,.\, N_c}{N^2} \qquad (10)$$

in which, OA is the overall accuracy mentioned above, m_c is the number of all the pixels labelled as class *c* in the output class map.

F. PSNR:

PSNR is defined as: 10 log*(S/N)*, where *S* is the energy of original signal and *N* is the energy of difference between original and decompressed signal. Also, compression rate is defined as the ratio of the original to the compressed signal volume.

References

[1] D. Landgrebe, "Hyperspectral image data analysis as a High Dimensional Signal Processing Problem," *Signal Processing Magazine, IEEE*, vol. 19, pp. 17-28, 2002.

[2] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, pp. 1087-1095, 1994.

[3] M. Marconcini, G. Camps-Valls, and L. Bruzzone, "A Composite Semisupervised SVM for Classification of Hyperspectral Images," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, pp. 234-238, 2009.

[4] H. Ghassemian and D. A. Landgrebe, "Object-oriented feature extraction method for image data compaction," *Control Systems Magazine, IEEE*, vol. 8, pp. 42-48, 1988.

[5] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-Spectral Remote Sensing Image Classification With Graph Kernels," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, pp. 741-745, 2010.
[6] F. Mirzapour and H. Ghassemian, "Improving hyperspectral image classification by combining spectral, texture, and shape features," *International Journal of Remote Sensing*, vol. 36, pp. 1070-1096, 2015/02/16 2015.
[7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, pp. 1778-1790, 2004.
[8] J. Xiuping, K. Bor-Chen, and M. M. Crawford, "Feature Mining for Hyperspectral Image Classification," *Proceedings of the IEEE*, vol. 101, pp. 676-697, 2013.
[9] G. Lianru, L. Jun, M. Khodadadzadeh, A. Plaza, Z. Bing, H. Zhijian, et al., "Subspace-Based Support Vector Machines for Hyperspectral Image Classification," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, pp. 349-353, 2015.
[10] G. H. Halldorsson, J. A. Benediktsson, and J. R. Sveinsson, "Source based feature extraction for support vector machines in hyperspectral classification," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, 2004, p. 539.
[11] R. Pu and P. Gong, "Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping," *Remote Sensing of Environment*, vol. 91, pp. 212-224, 5/30/ 2004.
[12] H. Xiaofei, C. Deng, and H. Jiawei, "Learning a Maximum Margin Subspace for Image Retrieval," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, pp. 189-201, 2008.
[13] D. Qian, "Modified Fisher's Linear Discriminant Analysis for Hyperspectral Imagery," *Geoscience and Remote Sensing Letters, IEEE*, vol. 4, pp. 503-507, 2007.
[14] W. Jing and I. C. Chein, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, pp. 1586-1600, 2006.
[15] G. Yanfeng, L. Ying, and Z. Ye, "A Selective KPCA Algorithm Based on High-Order Statistics for Anomaly Detection in Hyperspectral Imagery," *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, pp. 43-47, 2008.
[16] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.*, 1999, pp. 41-48.
[17] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation*, vol. 12, pp. 2385-2404, 2000.
[18] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sensing of Environment*, vol. 86, pp. 554-565, 8/30/ 2003.
[19] A. Hosseini and H. Ghassemian, "Classification of hyperspectral and multispectral images by using fractal dimension of spectral response curve," in *Electrical Engineering (ICEE), 2012 20th Iranian Conference on*, 2012, pp. 1452-1457.
[20] S. A. Hosseini and H. Ghassemian, "A new hyperspectral image classification approach using fractal dimension of spectral response curve," in *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, 2013, pp. 1-6.
[21] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. I. Theory," *Signal Processing, IEEE Transactions on*, vol. 41, pp. 821-833, 1993.
[22] D. A. Wagenaar and S. M. Potter, "Real-time multi-channel stimulus artifact suppression by local curve fitting," *Journal of Neuroscience Methods*, vol. 120, pp. 113-120, 10/30/ 2002.
[23] L. Fang and D. C. Gossard, "Multidimensional curve fitting to unorganized data points by nonlinear minimization," *Computer-Aided Design*, vol. 27, pp. 48-58, 1// 1995.
[24] H. Motulsky and A. Christopoulos, *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*: Oxford University Press, 2004.
[25] M. Kamandar and H. Ghassemian, "Linear Feature Extraction for Hyperspectral Images Based on Information Theoretic Learning," *Geoscience and Remote Sensing Letters, IEEE*, vol. 10, pp. 702-706, 2013.
[26] G. M. Foody, "Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy," P*hotogrammetric Engineering & Remote Sensing*, vol. 70, pp. 627-633, 05/01/ 2004.

Seyed Abolfazl Hosseini received the B.Sc. degree in 1993 and M.Sc. degree in 1997 in Electrical Engineering from Sharif University of technology and K.N.Toosi University of technology, Tehran, Iran, respectively. He is currently working toward the Ph.D. degree in the Faculty of Electrical and Computer Engineering at the Tarbiat Modares University, Tehran, Iran. His research interests include the hyperspectral image analysis, feature extraction and pattern recognition in remote sensing applications.

Hassan Ghassemian received his B.Sc. degree from Tehran College of Telecommunication in 1980 and the M.Sc. and Ph.D. degree from Purdue University, West Lafayette, USA in 1984 and 1988 respectively. Since 1988, he has been with Faculty of Computer and Electrical Engineering at 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.

Acoustic Noise Cancellation Using an Adaptive Algorithm Based on Correntropy Criterion and Zero Norm Regularization

Mojtaba Hajiabadi*
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
mhajiabadifum@gmail.com

Received: 25/May/2015 Revised: 14/Jul/2015 Accepted: 26/Jul/2015

Abstract

The least mean square (LMS) adaptive algorithm is widely used in acoustic noise cancellation (ANC) scenario. In a noise cancellation scenario, speech signals usually have high amplitude and sudden variations that are modeled by impulsive noises. When the additive noise process is nonGaussian or impulsive, LMS algorithm has a very poor performance. On the other hand, it is well-known that the acoustic channels usually have sparse impulse responses. When the impulse response of system changes from a non-sparse to a highly sparse one, conventional algorithms like the LMS based adaptive filters can not make use of the priori knowledge of system sparsity and thus, fail to improve their performance both in terms of transient and steady state. Impulsive noise and sparsity are two important features in the ANC scenario that have paid special attention, recently. Due to the poor performance of the LMS algorithm in the presence of impulsive noise and sparse systems, this paper presents a novel adaptive algorithm that can overcomes these two features. In order to eliminate impulsive disturbances from speech signal, the information theoretic criterion, that is named correntropy, is used in the proposed cost function and the zero norm is also employed to deal with the sparsity feature of the acoustic channel impulse response. Simulation results indicate the superiority of the proposed algorithm in presence of impulsive noise along with sparse acoustic channel.

Keywords: Adaptive Filter; LMS Algorithm; Sparse Acoustic Channel; Zero Norm; Impulsive Noise; Correntropy.

1. Introduction

Adaptive Filters are used in large applications to endow a system with learning and tracking abilities, especially when the signal statistics are unknown and are expected to vary with time. Over the last several years, a wide range of adaptive algorithms has been developed for diverse demands such as channel equalization, spectral estimation, target localization, system identification and noise cancellation. One group of the basic adaptive algorithms is gradient-based algorithms such as the LMS algorithm. The well-known LMS algorithm is perhaps one of the most familiar and widely used algorithms because of its good performance in many circumstances and its simplicity of implementation [1],[2].

In many scenarios such as speech echo cancellation, parameters of the acoustic channel impulse response can be assumed to be sparse [3]-[5]. When the system changes from a non-sparse to a highly sparse one, conventional algorithms like the LMS based adaptive filters can not make use of the priori knowledge of system sparsity and thus, fail to improve their performance both in terms of transient and steady state. Using such prior information about the sparsity of acoustic channel can be helpful to improve LMS

Algorithm performance. In the past years, several algorithms have been proposed for sparse adaptive filtering using LMS, which was motivated by recent progress in compressive sensing [6]. The basic idea of these techniques is to introduce a penalty into the cost function of the standard LMS to exploit the sparsity of the system impulse response and achieve a better performance [7].

Many approaches for signal processing problems have been studied when the additive noise process is modeled with Gaussian distribution. However, for many real-life situations, the additive noise of the system is found to be dominantly nonGaussian and impulsive. One example of nonGaussian environments is the acoustic noise in speech processing applications [8]-[10]. When the additive noise process is nonGaussian or impulsive, LMS algorithm has a very poor performance [11]. In [12],[13] it was shown that for some environments with nonGaussian noise, maximum correntropy criterion (MCC) algorithm outperforms LMS algorithm.

In order to modify LMS algorithm performance in sparse conditions and nonGaussian noises, Wentao Ma, proposed a hybrid algorithm in [14] based on MCC and correntropy induced metric (CIM), for robust channel estimation problem. Specifically, MCC is utilized to mitigate the impulsive noise while CIM is adopted to exploit the channel sparsity.

Based on ANC recent works, it is clear that in this field of research, we need to deal with two important features, sparse acoustic channels [3]-[5] and nonGaussian acoustic noises [8]-[10]. Thus, in order to address this problem, we propose a novel adaptive algorithm in this paper which is mathematically different

* Corresponding Author

with [14] in order to exploit sparsity. The proposed algorithm is obtained by combination of maximum correntropy criterion and zero norm regularization. The zero norm is utilized in the cost function to deal with sparsity feature of acoustic channel and the correntropy criterion is used to eliminate nonGaussian noises from speech signal. Computer simulation results show that the proposed adaptive algorithm achieves better performance compared to the conventional adaptive LMS algorithm.

This paper is organized as follows. After the introduction, adaptive noise cancellation configuration is expressed in section II. In order to cancelling nonGaussian noise from speech signal, a novel adaptive sparse MCC algorithm is developed in section III. Finally, simulation and comparison results are given in section IV, followed by conclusions in section V.

2. Adaptive Noise Cancellation

An important application of adaptive filters is in acoustic noise cancellation [15]. Fig. 1 shows the configuration of a noise cancellation system. Assume that signal x(n) is the acoustic noise which passes through an acoustic channel, with impulse response:

$$h(n) = \sum_{i=0}^{N-1} w_i \delta(n-i) \quad (1)$$

By sorting the channel coefficients w_i into a column vector, the acoustic channel impulse response can be expressed as follows,

$$W^o = [w_0, w_1, \dots, w_{N-1}]^T \quad (2)$$

that $(.)^T$ represents the transpose operator. An observation of the desired signal which is sensed by the first microphone is denoted by,

$$d(n) = X(n)^T W^o + s(n) \quad (3)$$

where $s(n)$ is the speech signal (speaker, music or etc) and $X(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$ denotes a vector of delayed input signal which is sensed by the second microphone. Given a desired signal $d(n)$ and acoustic noise $X(n)$, adaptive filter tries to replicate colored noise by exactly modeling the sparse acoustic channel between the noise source and the desired signal. The difference between the desired signal $d(n)$ and the output of adaptive filter $y(n)$ is in fact the noise-free signal (cleaned speech).

The objective of an adaptive algorithm is to identify the sparse channel W^o using the signals $X(n)$ and $d(n)$. Let $W^o = [w_0, w_1, \dots, w_{N-1}]^T$ be the estimated vector of the adaptive filter at iteration n. In the standard LMS, the cost function $J_{MSE}(n)$ is defined as

$$J_{MSE}(n) = e^2(n) \quad (4)$$

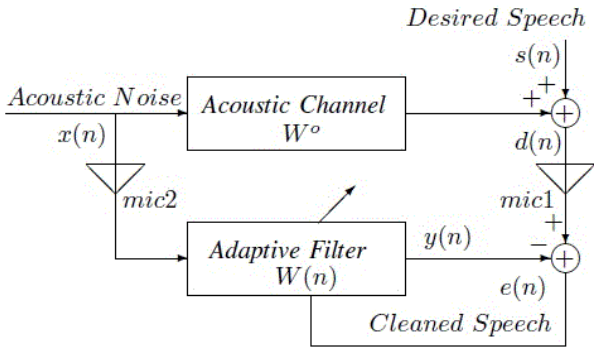


Fig. 1. The block diagram of a noise cancellation system.

where $e(n)$ is the instantaneous error determined as

$$e(n) = d(n) - y(n) \quad (5)$$

in which $y(n)$ is the output of adaptive filter and it is equal to $X(n)^T W(n)$. The filter coefficients vector is then updated by stochastic gradient descent equation [1]:

$$W(n+1) = W(n) - \mu \nabla J_{MSE}(n) \quad (6)$$

in which ∇ represents gradient and can be calculated by,

$$\nabla J(n) = \left[\frac{\partial J(n)}{\partial w_0}, \frac{\partial J(n)}{\partial w_1}, \dots, \frac{\partial J(n)}{\partial w_{N-1}}\right] \quad (7)$$

According to the equations (4), (6) and (7), the LMS algorithm is obtained,

$$W(n+1) = W(n) - \mu \nabla J_{MSE}(n) = W(n) + 2\mu e(n) X(n) \quad (8)$$

where μ is the step size and controls the convergence rate and steady state error.

More recently, there have been concerns about the effects of nonGaussian noise on adaptive algorithms [11]–[14]. This has led a number of authors to investigate adaptive algorithms which reduce the bad effects of nonGaussian noise. On the other hand, the sparse nature of such an impulse response causes standard adaptive algorithms like LMS to perform poorly [5].

In this work, we design a novel adaptive algorithm based on maximum correntropy criterion and zero norm regularization, in order to improve LMS weak performance in ANC application, in which the acoustic channel is sparse and the system noise is nonGaussian and impulsive. The next section developes the proposed algorithm that aim to give improved performance when these two important features exist in speech data.

3. Adaptive l_0 MCC Algorithm

The Mean Square Error (MSE) criterion may perform poorly in nonlinear and nonGaussian situations, especially when the data are disturbed by impulsive noises. To improve the performance in these situations, the maximum correntropy criterion (MCC), which is a robust criterion for non-Gaussian signal processing, has recently been successfully applied in adaptive filtering [12],[13]. The correntropy is a nonlinear measure of similarity

between two random variables. Given two random variables X and Y, the correntropy is:

$$V(X,Y) = E[k_\sigma(x-y)] \quad (9)$$

where $k_\sigma(x-y)$ is a positive definite kernel with the kernel width σ. The MCC has recently been applied to adaptive filtering algorithm to improve the tracking performance in impulsive interference [12], while MSE-based algorithms perform poorly [11]. The most widely used kernel in correntropy is the Gaussian kernel:

$$k_\sigma(x-y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2\sigma^2}} \quad (10)$$

By using a Taylor series expansion of the exponential function in the Gaussian kernel,

$$V(X,Y) = E\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2\sigma^2}}\right]$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\, 2^n \sigma^{2n}} E[(x-y)^{2n}] \quad (11)$$

it can be seen that the correntropy criterion involves all the higher even order statistical moments of the error random variable $(x-y)$ .On the other hand, mean square error (MSE) criterion just contain the second order statistical moment. Thus the MCC included more information of the error random variable $(x-y)$ and it should be very useful for cases when the measurement noise is nonzero mean, non-Gaussian, with large outliers.

Under the MCC criterion, the optimal weight vector of the adaptive filter can be obtained by maximizing:

$$J_{MCC}(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}} \quad (12)$$

A stochastic gradient ascent based adaptive algorithm, namely the MCC algorithm can be easily derived [12],

$$W(n+1) = W(n) + \mu\nabla J_{MCC}(n)$$
$$= W(n) + \mu\nabla\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}}\right]$$
$$= W(n) + \mu_{MCC} e^{-\frac{e^2(n)}{2\sigma^2}} e(n)X(n) \quad (13)$$

in which μ_{MCC} is equal to $\frac{\mu}{\sigma^3\sqrt{2\pi}}$. By choosing the kernel width so large, the MCC algorithm will simplify to the LMS algorithm:

$$\sigma \to \infty \;\Rightarrow\; e^{-\frac{e^2(n)}{2\sigma^2}} \to 1$$
$$W(n+1) = W(n) + \mu_{MCC} e^{-\frac{e^2(n)}{2\sigma^2}} e(n)X(n)$$
$$W(n+1) = W(n) + \mu_{MCC} e(n)X(n) \quad (14)$$

Comparing the MCC (13) with the LMS (8) weight update rule, we see that the weight update equation at each iteration in (13) just contains an extra scaling factor which is an exponential function of the instantaneous error $e(n)$. This factor rejects the impulsive and nonGuassian noise. As $\to \infty$, the exponential function goes to zero and therefore, processing of nonGaussian signal will be neglected. According to the above discussion, adaptation of weights using MCC filter is more stable when the desired signal has strong outliers or impulsive characteristics. By contrast, whenever a high amplitude outlier is encountered in the desired signal or in the error, $e(n) = d(n) - y(n)$, the LMS weight update rule (8) is forced to make a large increment, which takes the weights away from the optimal values.

By minimizing the zero norm of the filter coefficients vector in cost function of LMS algorithm, the sparsity of parameters has been exploited [6],[7]. In order to apply the zero norm in the MCC algorithm, the negative sign of zero norm should be inserted in the maximum correntropy cost function (12). By combining the correntropy criterion with the zero norm regularization, a new cost function is proposed in this paper,

$$J_{new}(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}} - \gamma\|W(n)\|_0 \quad (15)$$

where γ is a regularization parameter, which represents a trade off between estimation error and sparsity of the parameters. Operator $\|W(n)\|_0$ denotes zero norm, which counts the number of nonzero coefficients of vector $W(n)$. Because solving differentiation of zero norm is not possible, the zero norm is generally approximated by a continuous function [7]:

$$\|W(n)\|_0 \approx \sum_{i=0}^{N-1}\left(1 - e^{-\beta|w_i(n)|}\right) \quad (16)$$

when some elements of vector $W(n)$ are near zero, we have:

$$|w_i(n)| \approx 0 \;\Rightarrow\; \left(1 - e^{-\beta|w_i(n)|}\right) = 0 \quad (17)$$

On the other hand, when some elements of vector $W(n)$ are not zero, and also β is chosen as a large number, then we have:

$$|w_i(n)| \neq 0 \;, \beta \uparrow \;\Rightarrow\; \left(1 - e^{-\beta|w_i(n)|}\right) = 1 \quad (18)$$

According to (17), (18) it can be seen that equation (16) is a general approximation of the zero norm function and the number of nonzero coefficients of vector $W(n)$ is counted. By this general approximation, the gradient of zero norm can be calculated,

$$\nabla\|W(n)\|_0 = \left[\frac{\partial\|W(n)\|_0}{\partial w_0(n)}, \frac{\partial\|W(n)\|_0}{\partial w_1(n)}, \dots, \frac{\partial\|W(n)\|_0}{\partial w_{N-1}(n)}\right]$$
$$= \beta\left[sgn\left(w_0(n)\right)e^{-\beta|w_0(n)|}, \dots, sgn\left(w_{N-1}(n)\right)e^{-\beta|w_{N-1}(n)|}\right]$$
$$= \beta\begin{bmatrix} e^{-\beta|w_0(n)|} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & e^{-\beta|w_{N-1}(n)|} \end{bmatrix} sgn\left(W(n)\right)$$
$$= \beta\, diag\left[e^{-\beta|w_0(n)|}, \dots, e^{-\beta|w_{N-1}(n)|}\right] sgn\left(W(n)\right) \quad (19)$$

in which $diag[.]$ represents a diagonal matrix. By inserting the proposed cost function (15) in a gradient ascent updating, the proposed filter update is obtained as:

$$W(n+1) = W(n) + \mu\nabla J_{new}(n)$$
$$= W(n) + \mu\nabla\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}} - \gamma\|W(n)\|_0\right)$$
$$= W(n) + \mu\nabla\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}}\right) - \mu\gamma\nabla\|W(n)\|_0$$
$$= W(n) + \frac{\mu}{\sigma^3\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}} e(n)X(n) - \mu\gamma\nabla\|W(n)\|_0 \quad (20)$$

By exerting sparse penalty to the MCC cost function, the solution will be sparse and the gradient ascent recursion will improve the performance of near-zero coefficients in the sparse acoustic channel. By inserting (19) in (20), the proposed algorithm can be rewritten as follows:

$$W(n+1) = W(n) + \frac{\mu}{\sigma^3\sqrt{2\pi}} e^{-\frac{e^2(n)}{2\sigma^2}} e(n)X(n) - \mu\gamma\beta\ diag\left[e^{-\beta|w_0(n)|}, \dots, e^{-\beta|w_{N-1}(n)|}\right] sgn(W(n)) \quad (21)$$

In the next section, the proposed algorithm is simulated for sparse acoustic channel along with impulsive and shot noise, like in ANC application. The robustness of the proposed method against channel sparsity and impulsive noise is verified by detailed simulation studies.

4. Simulations

In this section, we have tried to simulate a real life conditions as closely as possible. The speech signal $s(n)$ that has been used is shown in Fig. 2(a). The nonGaussian noise is generated from mixture of multiple Gaussian distributions. After adding the noise to the speech signal, its non stationary characteristics can be seen in Fig. 2(b).

The sparse acoustic channel is that of a typical closed room environment [5], shown in Fig. 3. We use 21 taps to model the sparse acoustic channel path as follows,

$$W^o = [0,0.9,0,0,1,0,0,0,0,0,0.5,0,0,0,0,0,0,0,0,0,0]^T \quad (22)$$

The sparsity ratio of W^o is equal to 3/21 which means vector W^o containing only 3 large coefficients and others are near zero. The impulsive and nonGaussian observation noise is often modeled by a two component Gaussian mixture [11] with the following probability density function (pdf),

$$f_Z(z) = 0.6N(0,0.1) + 0.4N(0,10) \quad (23)$$

in which $N(0,10)$ denotes a Gaussian distribution with mean 0 and variance 10. Clearly, in this pdf, the second Gaussian distribution with variance 10 creates strong outliers as shown in Fig. 4. The kernel width σ for the MCC cost function is set to 2 in this case. According to various experiments and similar to references, other parameters such as step size μ, regularization parameter γ and sparsity parameter β were chosen to be 0.01, 0.001 and 8, respectively.

For comparing the error performance of the algorithms described in the previous section, the Mean Square Deviation (MSD) is defined as,

$$MSD(n) = E[\|W^o - W(n)\|^2] \quad (24)$$

Fig. 5, shows the MSD performances of the presented algorithms in the presence of impulsive noise (23) along with sparse acoustic channel (22). The sudden and high amplitude bursts of samples which occur in speech signals can easily disturb the LMS weight updating. However, MCC algorithm (13) places exponentially decreasing weights on samples that are distant and impulsive. In order to handle the case of channel sparsity, the MCC algorithm was modified further to a novel proposed algorithm (21). In this algorithm, each coefficient is updated with an independent step size that is made proportional to the magnitude of the particular filter coefficient, resulting in better performance for sparse systems. As seen in Fig. 5, the proposed l0-MCC algorithm has a superior performance in a noise cancellation scenario of highly impulsive speech signal.

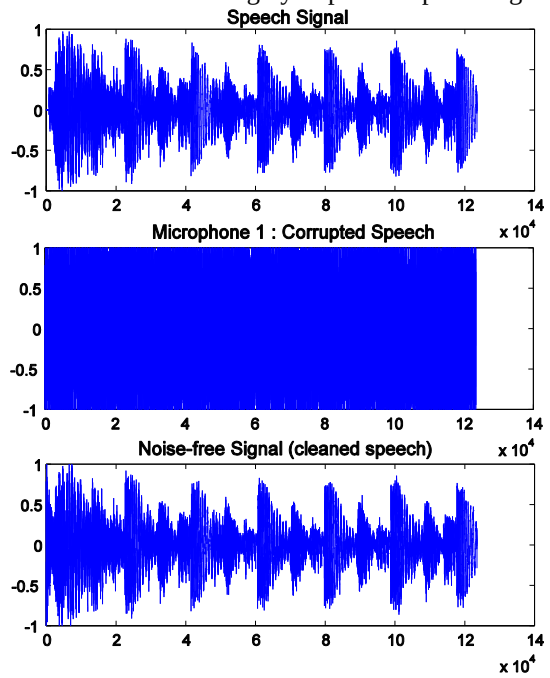


Fig. 2. (a) Original speech signal, (b) corrupted speech signal by implusive noise, (c) cleaned speech signal using proposed algorithm

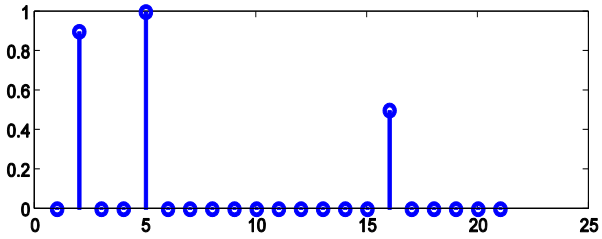


Fig. 3. A typical sparse acoustic channel (22)

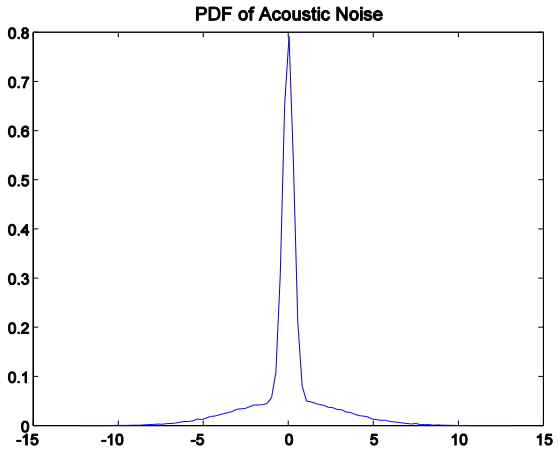


Fig. 4. Impulsive noise pdf (23), containing a two Gaussian components with identical zero means and different variances.

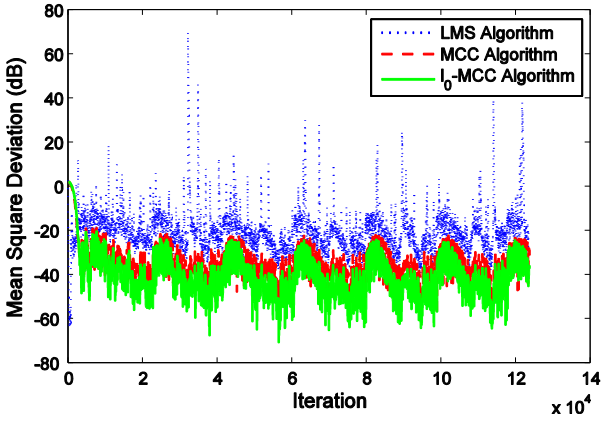


Fig. 5. MSD performance comparison in presence of impulsive noise (23) along with sparse acoustic channel (22) with a real speech signal s(n).

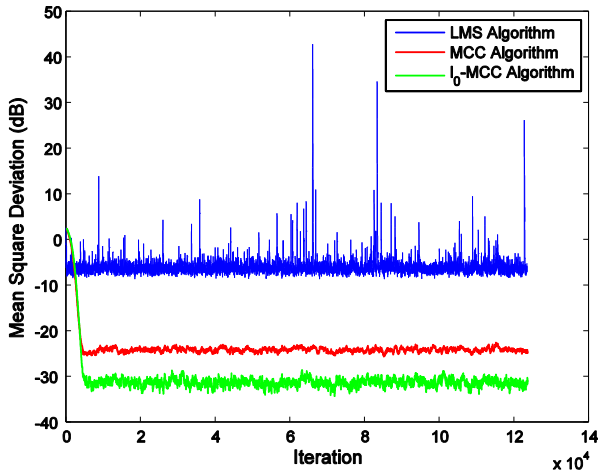


Fig. 6. MSD performance comparison in presence of impulsive noise (23) along with sparse acoustic channel (22) with a white normal signal s(n).

The performance of the described algorithms is also evaluated with a white normal input and is shown in Fig. 6. As seen, the proposed algorithm has superior performance in comparison with LMS (8) and MCC (13) algorithms.

The robustness of the proposed algorithm for diverse conditions of channel sparsity and nonGaussian noises is demonstrated by various simulation studies. Another experiment is utilized to show the superiority of the proposed algorithm in a different noise cancellation scenario. In this experiment, the acoustic channel W^o is assumed to be

$$[0,0.9,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]^T \quad (25)$$

Here, we use an acoustic channel with sparsity ratio of 2/30 as shown in Fig. 7. The nonGaussian noise is modeled by a three component Gaussian mixture with the following pdf,

$$f_Z(z) = 0.2N(-3,0.1) + 0.6N(0,0.1) + 0.2N(3,0.1) \quad (26)$$

as shown in Fig. 8. For comparison purposes, the MSD performances of this experiment are plotted in Fig. 9, by averaging over 20 independent runs. From various simulation studies, it is evident that the proposed filter achieves a 25 dB decrement of steady-state error, when the channel is sparse and the noise is nonGaussian or impulsive.

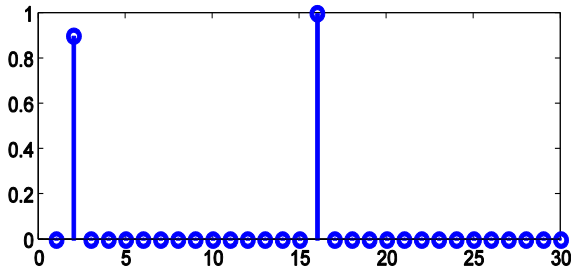


Fig. 7. Another sparse acoustic channel for second experiment (25).

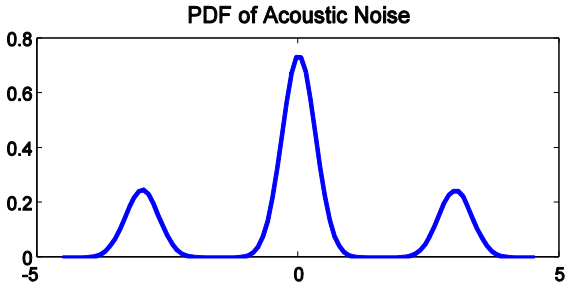


Fig. 8. Gaussian mixture noise pdf (26), containing three Gaussian components with different means and same variances

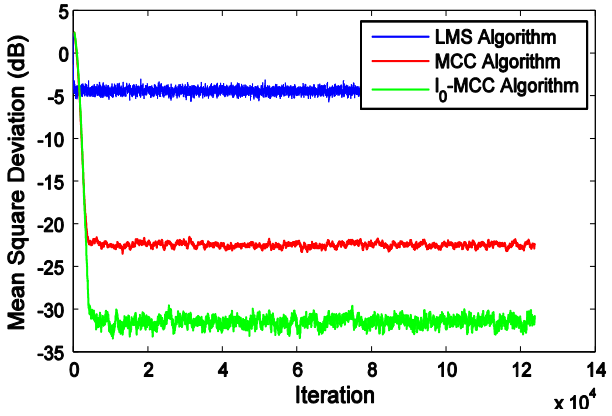


Fig. 9. MSD performance comparison for noise pdf (26) along with sparse acoustic channel (25) and a white normal signal s(n).

In the last experiment, the acoustic channel W^o is assumed to be

$$[0,0.8,0,0.8,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.8,0,0.8,0]^T \quad (27)$$

with sparsity ratio of 4/30 as shown in Fig. 10. The nonGaussian noise is modeled by two component Gaussian mixture with the following pdf,

$$f_Z(z) = 0.5N(-2,0.1) + 0.5N(2,0.1) \quad (28)$$

as shown in Fig. 11. The MSD performances of the last experiment are plotted in Fig. 12. From various simulation studies, it is evident that the proposed filter achieves better performance, when the channel is sparse and the noise is nonGaussian type.

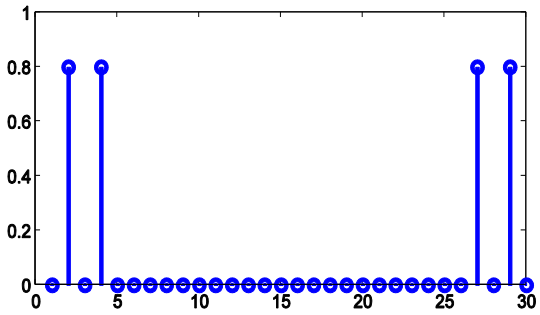


Fig. 10. Typical sparse acoustic channel for third experiment (27).

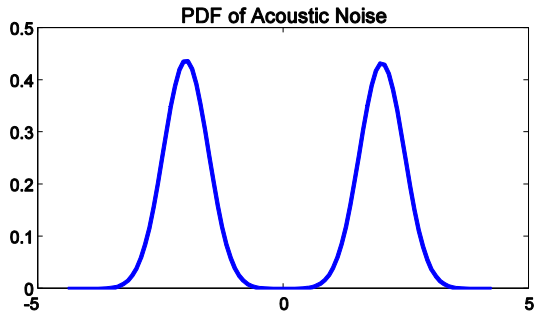


Fig. 11. Gaussian mixture noise pdf (28), containing two Gaussian components with different means and same variances.

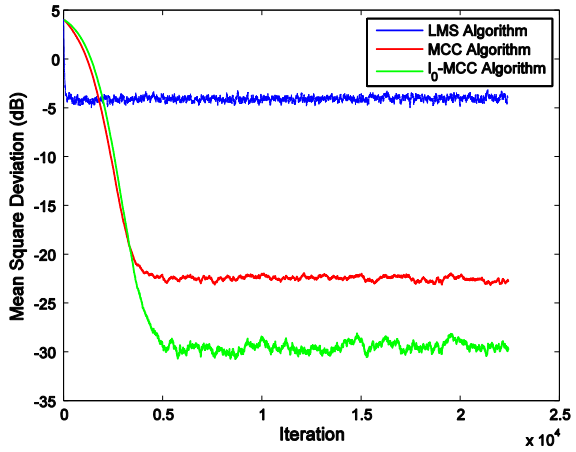


Fig. 12. MSD performance comparison for noise pdf (28) along with sparse acoustic channel (27) and a white normal signal s(n).

5. Conclusions

In a noise cancellation scenario, speech signals usually have high amplitude and sudden variations that are modeled by impulsive disturbances. In this paper, a novel adaptive algorithm has been proposed to improve LMS algorithm performance in impulsive disturbances and sparse acoustic channels. In order to provide robustness against impulsive noise, the cost function is derived by maximizing the correntropy. Additionally, an accurate approximation of zero norm is also utilized to further improve the performance in sparse acoustic channels. Simulation results show that the proposed algorithm achieves a better performance in terms of steady-state error as compared with the LMS and MCC algorithms.

Acknowledgments

The author would like to thank his superviser, Dr. Hossein Zamiri-Jafarian, whose valuable comments and kind evaluations improved the idea and presentation of this paper.

References

[1] B. Farhang-Boroujeny, Adaptive Filters: Theory and Applications, New York: John Wiley, 1998.

[2] A. H. Sayed, Fundamentals of Adaptive Filtering, New York: Wiley, 2003.

[3] J. Arenas-Garcia and A. R. Figueiras-vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 6, pp. 1087-1098, Aug. 2009.

[4] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-Garcia and A. Uncini, "Nonlinear acoustic echo cancellation based on sparse functional link representations," IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 7, pp. 1172-1183, Jul. 2014.

[5] P. A. Naylor, J. Cui and M. Brookes, "Adaptive algorithms for sparse echo cancellation," Signal Processing, vol. 86, p. 1182–1192, 2006.

[6] G. Su, J. Jin, Y. Gu and J. Wang, "Performance analysis of l0-norm constraint least mean square algorithm," IEEE Trans. Signal Process., vol. 60, no. 5, pp. 2223-2235, May 2012.
[7] Y. Gu, J. Jin, and S. Mei, "L0-norm constraint LMS algorithm for sparse system identification," IEEE Signal Process. Lett.,vol. 16, no. 9, pp. 774–777, Sep. 2009.
[8] F. R. Avila and L. W. P. Biscainho, "Bayesian restoration of audio signals degraded by impulsive noise modeled as individual pulses," IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 9, pp. 2470-2481, Nov. 2012.
[9] M. Niedzwiecki and M. Ciolek, "Elimination of impulsive disturbances from archive audio signals using bidirectional processing," IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 5, pp. 1046-1059, May 2013.
[10] I. Kauppinen, "Methods for detecting impulsive noise in speech and audio signals," Int. Conf. Digital Signal Process., vol. 2, pp. 967-970, 2002.
[11] N. J. Bershad, "On error saturation nonlinearities for LMS adaptation in impulsive noise," IEEE Trans. Signal Process., vol. 56, no. 9, pp. 4526-4530, Sep. 2008.
[12] L. Shi and Y. Lin, "Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference," IEEE Signal Process. Lett., vol. 21, no. 11, pp. 1385-1388, Nov. 2014.
[13] W. Lin,P. P. Pokharel and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," IEEE Trans. Signal Process, vol. 55, no. 11, pp. 5286-5298, 2007.
[14] Wentao Ma, et al, "Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under nonGaussian environments," Journal of the Franklin Institute, vol. 352, no. 7, pp. 2708–2727, Jul. 2015.
[15] J. M. Gorriz, J. Ramirez, S. Cruces-Alvarez, C. G. Puntonet, E. W. Lang and D. Erdogmus, "A novel LMS algorithm applied to adaptive noise cancellation," IEEE Signal Process. Lett., vol. 16, no. 1, pp. 34-37, 2009.

Mojtaba Hajiabadi received the degree of diploma in mathematics & physics in NODET) National Organization for Development of Exceptional Talents (school. He received the B.Sc. degree in communication engineering from the University of Birjand in 2012 with honors and the M.Sc. degree in electrical engineering) communication (from Ferdowsi University of Mashhad (FUM) in 2014 with the first rank. He is currently a Ph.D. student in electrical engineering) communication (at FUM. His research interests span several areas including adaptive filters, information theory, statistical signal processing and information theoretic learning

Effects of Wave Polarization on Microwave Imaging Using Linear Sampling Method

Mehdi Salar Kaleji*
Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
mehdi.salar@gmail.com
Mohammad Zoofaghari
Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
zoofaghari@aut.ac.ir
Reza Safian
Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
rsafian@cc.iut.ac.ir
Zaker Hossein Firouzeh
Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
zhfiruozeh@cc.iut.ac.ir

Received: 28/Dec/2013 Revised: 18/Apr/2015 Accepted: 27/May/2015

Abstract

Linear Sampling Method (LSM) is a simple and effective method for the shape reconstruction of unknown objects. It is also a fast and robust method to find the location of an object. This method is based on far field operator which relates the far field radiation to its associated line source in the object. There has been an extensive research on different aspects of the method. But from the experimental point of view there has been little research especially on the effect of polarization on the imaging quality of the method. In this paper, we study the effect of polarization on the quality of shape reconstruction of two dimensional targets. Some examples are illustrated to compare the effect of transverse electric (TE) and transverse magnetic (TM) polarizations, on the reconstruction quality of penetrable and non-penetrable objects.

Keywords: Linear Sampling Method (LSM); Inverse Scattering; Polarization; Singular Value Decomposition (SVD).

1. Introduction

Imaging and identification of targets using electromagnetic waves have a long history, e.g., X-ray imaging goes back to the early twentieth century. Waves in microwave frequency range are more interested than those in higher frequencies for imaging and identification. Microwaves have more penetration depth and less destructive impact on targets than higher frequency waves like X-ray so the great tendency exists towards microwave imaging. The purpose of microwave imaging techniques is to retrieve constitutive parameters and shape of scattering objects using the measurement data collected at a distance from scatterers. The unknown objects are illuminated by electromagnetic waves. This problem has many applications in different subjects such as target identification, geophysics, seismic exploration, remote sensing, atmospheric science, ground penetrating radar (GPR) and medical diagnostics such as cancer and hypothermia detection [1].

Imaging approaches based on the solution of an inverse scattering problem are usually grouped into two classes, weak scattering approximation methods and nonlinear optimization methods. The former exploit low or high frequency approximations of the scattering phenomenon to linearize the data-to-unknown relationship and are typically capable of providing only a rough description of the target's morphology when exploited outside of the range of validity of the underlying approximation (Born and Kirchhof inversion approach). Ill-posedness and non-linearity of the inverse scattering problem are the two main complications in the solution process. One way to avoid these problems is the weak scattering approximation which works for certain category of the inverse problems. Different methods are proposed which approximate the scattering based on certain constraints on the scatterer. For example Born approximation is used when the scatterer has small permittivity and is small compared to the wavelength.

Conversely nonlinear optimization methods tackle the inverse scattering problem in its full nonlinearity to determine both the morphology and the electromagnetic contrast of the targets. Optimization approaches seek for the problem's solution by means of a local iterative optimization scheme, as the large number of unknowns makes global optimization generally unreliable, due to the exponential growth of the computational complexity. On the other hand, local minimization is prone to the occurrence of false solutions, so that these methods have to be equipped with regularization schemes, usually based on the available a priori information. Optimization methods are based on minimizing an error function, starting with an initial guess and this guess is optimized during iteration stages. Because in each iteration stage the inverse scattering problem must be solved, such methods

* Corresponding Author

are slower than the others, however accuracy and quality of imaging is great [2]. Examples of two groups of quantitative methods are the modified gradient method [3], distorted Born approximation [4], the contrast source method [5], the subspace optimization method [6] and the least squares optimization method [7].

Another category is qualitative methods which are aimed to find only the shape and location of the target. Linear sampling method is one of the qualitative methods which is reliable, efficient, and has a noticeable high speed of computational operations to find the object shape and location [2]. Colton and Kirsch [8] regularized sampling methods to operate a reconstruction of the region of support of the scatterer from solution norm of a linear integral equation at each pixel point [9]. Proposing a simple and original "physical" interpretation of the linear sampling methods, which shows its relationship with electromagnetic focusing problems have been done in several articles such as [10,11]. A hybrid approach is proposed to inspect three-dimensional homogeneous dielectric scatterers by using microwaves [12] and unlike optimization methods that are based on an initial guess which can lead to an unsuccessful reconstruction results due to approximate priori information, regularized sampling does not require anything to know about the scatterer shape or composition [13]. The LSM is also very fast in comparison with optimization methods as does not involve iteration. In the last years there were quite number of LSM applications in different area such as Trough Wall Imaging (TWI) [14,15], microwave imaging to detect and characterize nonaccessible target concealed into a wall [16,17]. As a newest one, (the LSM) offers a qualitative image reconstruction approach, which is known as a viable alternative for obstacle support identification to the well-studied filtered back projection (FBP), which depends on a linearized forward scattering model [18].

As the operator used in the LSM is linear it does not need born approximation and can be used for a wide range of the scatterers.

In this paper we analyse the effect of various polarizations on reconstruction of PEC and dielectric shapes using LSM. We first calculate the far filed scattered pattern using the method of moment. Then the target is reconstructed with various polarizations such as TM and TE using LSM and compared. The rest of the paper is organized as follows: In Section 2 and 3 the formulation of the LSM and its physical meaning are briefly summarized. In section 4, we investigate noise effect and number of antennas and in section 5, TM and TE polarization results are compared together. Also the paper includes the results of multiple scatterers and coated objects reconstruction taking account the effect of polarization. Buried objects as one the most important applications in the medicine area has been investigated and the polarization influence in two states is compared for the first time. Finally, the conclusion is given in section 6.

2. Formulation of LSM for TM Polarization

2.1 General Two Dimensional Scatterer

At first, we recall the far-field equation and definition of far-filed operator as the basis of LSM. We consider an infinitely long cylinder located in free space which the axis is parallel to z-axis. For the sake of simplicity, we consider the 2-D scalar electromagnetic scattering problem [19]. Denote that Ω is the region of the test. Let assume some transmitter antenna in the far-filed, radiating field approximated by the TM plane waves and the incidence angle Θ_i varing within $(0,2\pi)$. Let us denote $E_z^s(\Theta,\Theta_i)$ the only z-component of the scattered electric filed which is measured by receiving antenna in observation domain $\underline{r} = (r\cos\Theta, r\sin\Theta)$. Let $E_{z,\infty}^s(\Theta,\Theta_i)$ be the far-field pattern defined by the following equation:

$$E_z{}^s(r,\Theta;\Theta_i) = \frac{e^{ik_0 r}}{\sqrt{r}} E_{z,\infty}^s(\Theta,\Theta_i) + O(r^{-1}) \quad (1)$$

where, $k_0 = \omega\sqrt{\mu_0\epsilon_0}$is the wave number of vacuum,ω is operating angular frequency and μ_0 and ϵ_0 are the free - space permeability and permittivity, respectively.

By introducing of operator F^{TM}:$L^2(0,2\pi) \rightarrow L^2(0,2\pi)$ we define far-field equation as below:

$$F^{TM}g = f(\Theta,\underline{r'}) \quad (2)$$

Where

$$F^{TM}g = \int_0^{2\pi} E_{z,\infty}^s(\Theta,\Theta_i)g(\Theta_i)d\Theta_i \quad (3)$$

which g denotes an indicator function relevant to the amplitude of incident field

and

$$f(\Theta,\underline{r'}) = \sqrt{\frac{2}{\pi k_0}} e^{-i\pi/4} e^{-ik_0 r' \cos(\Theta-\Theta')} \quad (4)$$

which coincides with the far-field pattern of the field radiated by an electric current filament located at $\underline{r'}\epsilon\,\Omega$; $\underline{r'} = (r'\cos\Theta', r'\sin\Theta')$. We solve equation (2) for the all points r'. In this case according to LSM, initial current sources with limited power can radiate so that the scattered field of the object due to these ones is equivalent to the radiated filed of a linear line source which is inside of the object. For the numerical realization of the linear sampling method we assume an approximate measured far field pattern $E_{z,\infty}^s$ that is known for different incident plane waves. According to Nystrom approximation equation (3) can be written as [20]:

$$(Fg)(\Theta_j) \cong \frac{4\pi}{n}\sum_{l=1}^{n} E_{z,\infty}^s(\Theta_j,\Theta_l)\, g(\Theta_l). \quad (5)$$

Now we define a matrix:

$$F = \left(\frac{4\pi}{n} E_{z,\infty}^s(\Theta_j,\Theta_l)\right)_{j,l=1}^{n} \quad (6)$$

And two vectors:

$$g = \left(g(\Theta_1), g(\Theta_2), \dots, g(\Theta_n)\right)^T \quad (7)$$

And

$$f = \left(f(\Theta_1, r'), f(\Theta_2, r'), \dots, g(\Theta_n, r')\right)^T \quad (8)$$

To write matrix form of equation (2) as

$$Fg = f \quad (9)$$

If $\mathcal{G} = \|g\|_{L^2(-\pi,\pi)}$ is the norm of the function g as the solution of the far-filed equation, then $\mathcal{G}$ will be infinite outside the lossless scatterer and finite inside of it. Therefore, the points in the domain where $\|g\|$ is finite are desired to reconstruct the shape of the scatterer. Hence, a regularization scheme must be considered. Using the Tikhonov regularization method [21], $\|g\|$ is given by:

$$\|g\| = \left(\sum_{n=-\infty}^{\infty} \frac{\mu_n^2 (f,\psi_n)}{(\alpha+\mu_n^2)^2}\right)^{1/2} \quad (10)$$

where $\{\mu_n, \psi_n, \psi_n^*\}$ is a singular system of operator F^{TM}, (f, ψ_n) are components of f on the singular function ψ_n s, and α is the regularization parameter which is obtained by Morozov's discrepancy principle and should be considered that it usually chooses one hundredth of the largest of singular values [22].

3. Formulation of LSM for TE Polarization

The same approach can be used to consider TE polarization. In this case a new operator F^{TE} like that of equation (3) is identified. F^{TE} is expressed with the same equation as (3) but $E^s_{z,\infty}$ is replaced by $\left({E^s_{x,\infty}}^2 + {E^s_{y,\infty}}^2\right)^{0.5}$ as the magnitude of electric field in TE polarization. Now F^{TM} can be replaced by F^{TE} in equation (2) to formulate LSM for TE polarization.

4. Noise and Antenna Number Effect

In this section we consider the effect of noise and the number of antennas on the quality of reconstruction. Given the analogy between LSM and focusing problems, it follows that an important role in successful reconstruction of shapes is played by the number of primary sources and measurement points. As a matter of fact, if the number of transmitters is low, nothing could be achieved, even in the full aperture case, the desired focused field, while if the number of receivers is low, one could not properly control the synthesized scattered field to match to the desired one. On the other hand, an arbitrary large number of illuminations and measurements, which could overcome these problems, can be largely redundant and may be unnecessarily increases the measurement costs. According to [23] the selection of N antennas where N is given by$N = 2ka\gamma$ is sufficient. The optimal choice for γ depends on the desired accuracy and$\gamma \approx 2$ is a generally convenient one, a is the largest dimension of the scatterer and k is the wave number. The location of object, configuration of the antennas and how they are placed around the object is considered in Fig.1.

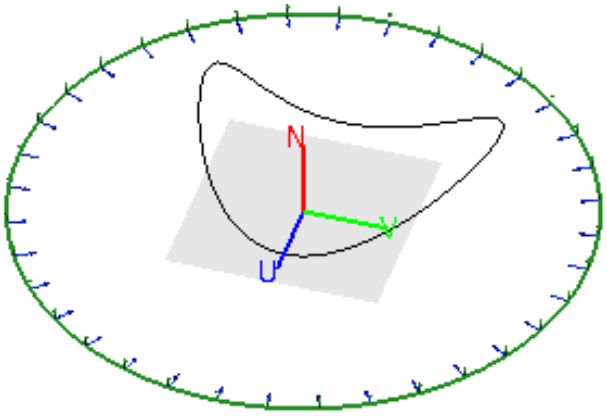


Fig. 1. The Uniform location of antennas around the target

At the end of this section we proceed to the noise issue. We assume that the noise is additive (i.e., $E_s^{\delta}(\theta, \theta_i) = E_s(\theta, \theta_i) + n(\theta, \theta_i)$) and we firstsuppose that, for the sake of simplicity, $n(\theta, \theta_i)$ can be expanded as:

$$n(\theta, \theta_i) = \sum_{-\infty}^{\infty} N_n \, e^{in(\theta-\theta_i)} \quad (11)$$

More details about the effect of noise have been considered in [24]. In the next section we show the effect of noise on reconstruction quality and stability of LSM regarding to noise is considered.

5. Numerical Results and Comparison Discussion

In this section we propose some examples as the numerical results to reconstruct a kite shape object illustrated in Fig.2. We compare the quality of the LSM in TE and TM polarizations for PEC and dielectric targets. The forward scattering data for simulation has been produced by method of moment solver. Also, it is used electric field integral equation (EFIE) for the both of polarizations. Basically, EFIE has been derived by using surface equivalent theorem for PEC objects and volume equivalent theorem (VET) for dielectric ones. That is to say EFIE discretized with pulse basis and delta testing functions. As a test object, a kite plate shape of size about $2.5\lambda \times 3\lambda$ was used and the parametric equation of this object is written below the figure 2. In all the examples discussed throughout this article, we will use units of free-space wavelength for the scatterer geometry, in this examples the wave number k equals 2π. The cell size of surfaces meshes is 0.01λ and for volume equivalent to a cell density of 520 cells per square dielectric wavelength.

The number of receiving and transmitting antennas is n=38. For the kite-shaped cross section shown in Fig 2. $\|g\|$is given by equation (9) where $\alpha = 0.0001$, and $\frac{1}{\|g\|}$ is plotted for two polarizations in Fig. 2 and Fig. 3 for PEC and in Fig.4 and Fig. 5 for dielectric with$\varepsilon_r = 4$.If α is chosen very small so that it couldn't eliminate small singular values; consequently, the reconstruction boundaries of the shape may be oscillated .On the other hand, the selection of large amount of α leads to an unsuccessful

reconstruction. Experimentally selection of $\alpha = 0.0001$ is appropriate. All of the results have been obtained for $k=2\pi$ (the operating frequency is about 300 MHz).

5.1 Polarization Comparison

Comparison between TM and TE polarizations shows that reconstruction of TM is better than TE for PEC objects. On the other hand, TE polarization is more efficient for dielectric objects. It is an expectable result since in forward scattering pattern in dielectric, there are two components of electric field in (i.e. E_x and E_y) TE polarization and just one in TM; thus, the former provides more information for LSM than the latter. However as noted in [8] in resonance frequencies the points where electric field vanishes are better distinguishable in LSM. These points include the boundary of the conducting scatterers in TM polarization and exclude it in TE so the better result is obtained in TM for PEC. It is of great importance to note that both of the results have some problems in the quality of the boundary reconstruction. It refers to the right-hand side of (1). It is proved that the better results are obtained by replacing the right-hand side of (1) by the far field radiated from a dipole instead of monopole [20]. Since the forward pattern has the information of both polarizations, combination of the polarizations leads to a better shape reconstruction which is the future work of the authors.

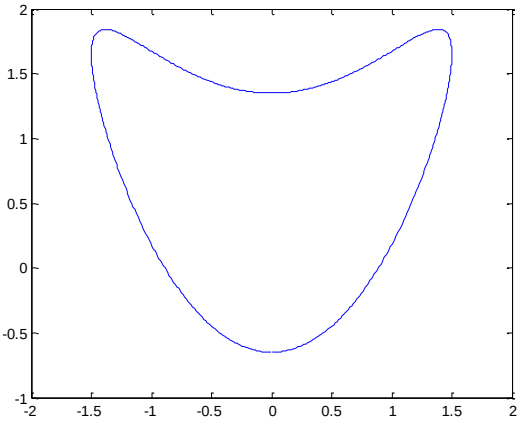


Fig. 2. Kite-Shape scatterer

$$x(t) = (\cos t + 0.65\cos 2t - 0.65, 1.5\sin t), 0 \le t \le 2\pi$$

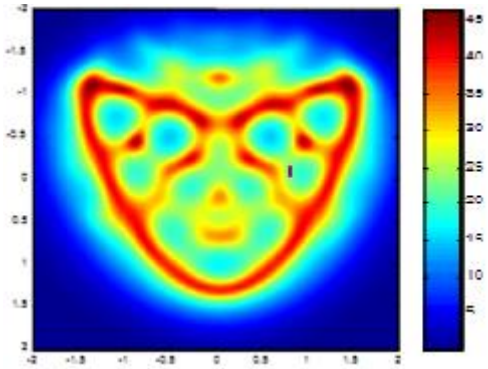


Fig. 3. Reconstruction of PEC object using TM polarization

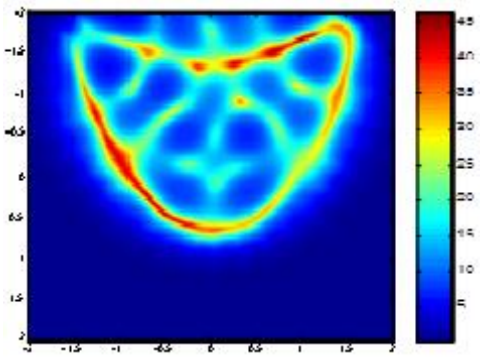


Fig. 4. Reconstruction of PEC object using TE polarization

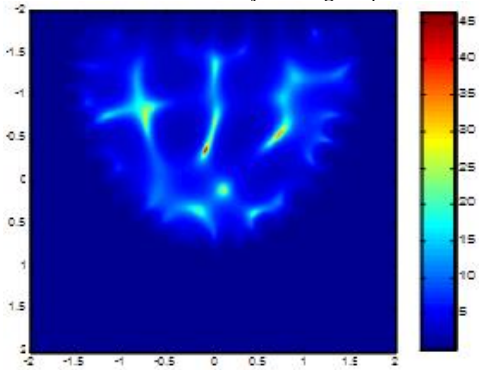


Fig. 5. Reconstruction of dielectric object using TM polarization

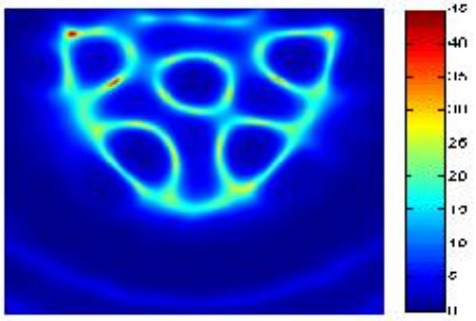


Fig. 6. Reconstruction of dielectric object using TE polarization

5.2 Multi-Scatterers

Here we want to investigate how LSM works for multi scatterers. At first, we study the effect of polarization then pose the number of antenna. Comparison between two polarization TM and TE demonstrated that when we use incident filed with polarization TM, it leads to a better reconstruction toward to TE polarization incident one. It can be observed in Figs. 7 and 8 where we tried for two triangle and circle cross-section shape with 51 transmitter and receiver antennas. On the other hand, if we want to test multi-scatterer with more numbers, we have to increase the number of antennas, as we see in Fig.9, even increasing the number of antennas up to 108 cannot lead to a approvable reconstruction of three shape. However this increase cannot be done arbitrarily. As a result, it can be concluded that we need more transmitting and receiving antennas for LSM to work better as the number of the scatterers increases. It shows the important role of antenna number in the reconstruction of the shape in LSM.

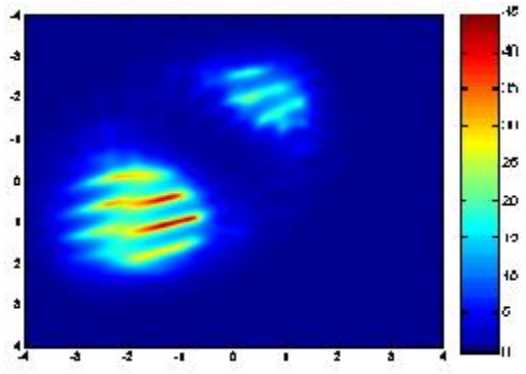


Fig. 7. Reconstruction of multi-scatterers with 51 antennas with Polarization TE

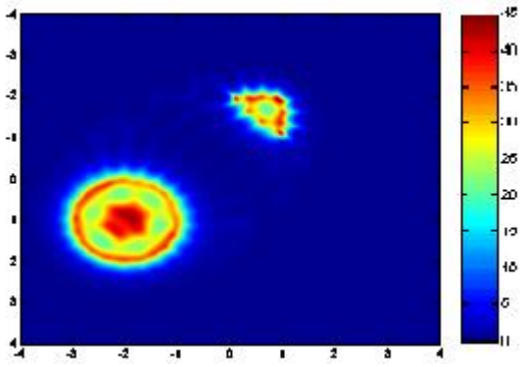


Fig. 8. Reconstruction of multi-scatterer with 51 antennas with Polarization TM

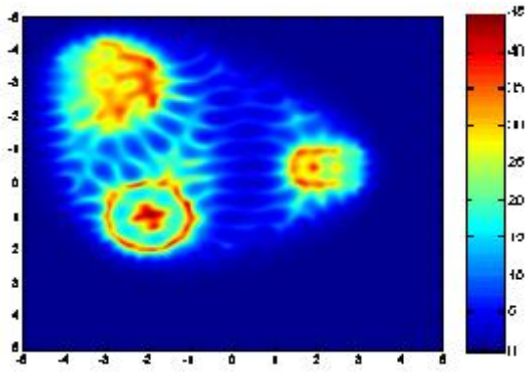


Fig. 9. Reconstruction of three multi-scatterers with 108 antennas with Polarization TM

5.3 Coated Objects

In one part of simulations we considered coated objects. The results for a PEC coated by a dielectric are illustrated in Figs.10 and 11 for two polarizations. We considered a dielectric with thickness of 4 cm.

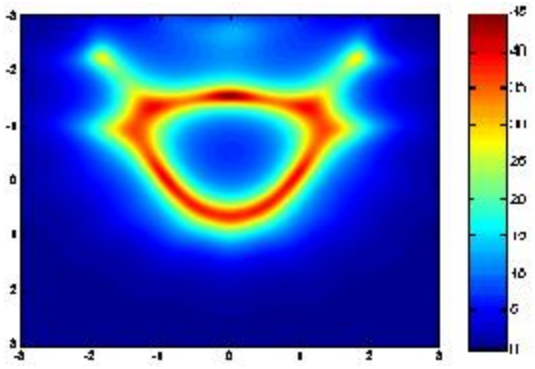


Fig. 10. PEC kite-shape coated with dielectric $\varepsilon_r = 4$ illuminated by TM-polarized waves.

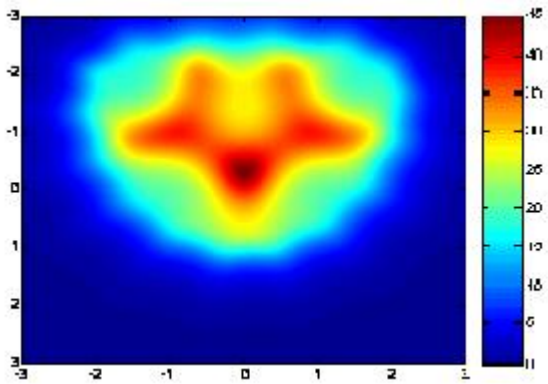


Fig. 11. PEC kite-shape coated with dielectric $\varepsilon_r = 4$ illuminated by TE-polarized waves.

5.4 Buried Objects

One important application of the proposed method is to improve the quality of the shape reconstruction of a hidden object in the other one. It has a numerous applications in medical area such as microwave imaging and detection of cancer cells. Figs 12 and 13 show the location of a PEC object with circular cross section and radius 0.5 lambda, within dielectric object with 2D kite-shape cross section and thickness 4 cm. With comparison between Fifs. 12 and 13, it implies that the TM-polarization result is more clear in the external shape boundary and in contrast the TE one, it has more clarity in the reconstruction of the hidden PEC within the dielectric. To sum up, it seems that, it can be wiser idea to use TE polarization, if the detection of hidden object is more vital.

Despite of this point, the boundary details of internal object in these problems are not so clear and it can be considered as an open problem in LSM to find the shape and material of hidden object more precisely.

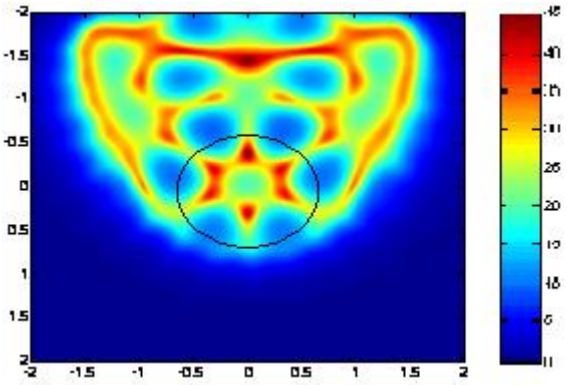


Fig. 12. Dielectric cylinder with $\varepsilon_r = 4$, and thickness 4cm, within a kite- shape cross section, TM polarization.

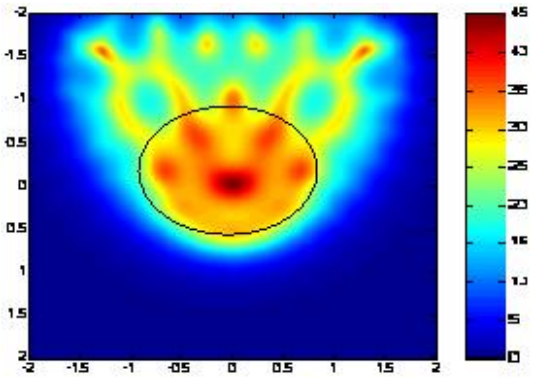


Fig. 13. Dielectric cylinder with $\varepsilon_r = 4$, and thickness 4cm, within a kite-shape cross section, TE polarization.

5.5 Noise Effect

Finally we would like to investigate an important parameter which can affect the reconstruction quality in real areas and that is noise. As a matter of fact, we want to clarify the robustness of LSM against noise for different polarizations. Generally, we know LSM as a method which is not affected by the noise so greatly and here we compare this feature for TE and TM cases. According to Figs. 14 and 16 we can say TM polarization for PEC leads to better results with the same noise condition (SNR=5dB), in contrast in the case of dielectric TE polarization leads to better one on the same condition (Fig. 17 and Fig. 18). This can be implied that it's more efficient to use TE polarization incident field to reconstruction of dielectric objects in the noisy environments.

Generally linear sampling is the most important method which has stability to the noise effect. Note that we can reduce the effect of noise with increasing the number of antennas, but as we mentioned in section 4 the number of antennas cannot be arbitrary due to high cost of computational complexity and simulation time. Actually there is a limitation for the number of antennas that is defined by noise factor.

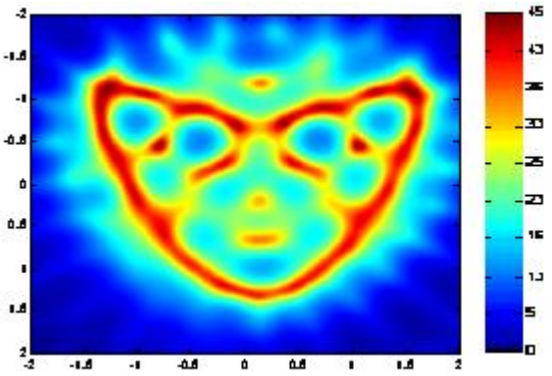


Fig. 14. PEC with noise effect SNR=5dB and TM polarization

The Comparison between Figs. 14 and 15 indicates that as there is more noise at environment, in other words, less SNR; we have a reconstruction with less resolution. The comparison between different polarization and various SNR shows that when we use TM polarization it will be achieved better result toward to TE one in the PEC and on the contrary TE is superior in the case of dielectric scatterer. This issue must be considered that the negative effect of noise on the dielectric is more, related to PEC objects.

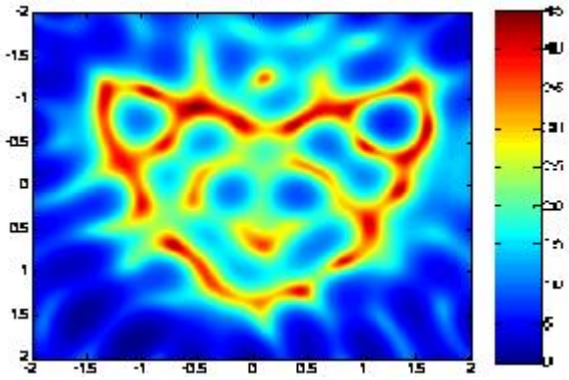


Fig. 15. PEC with noise effect SNR= -5dB and TM polarization

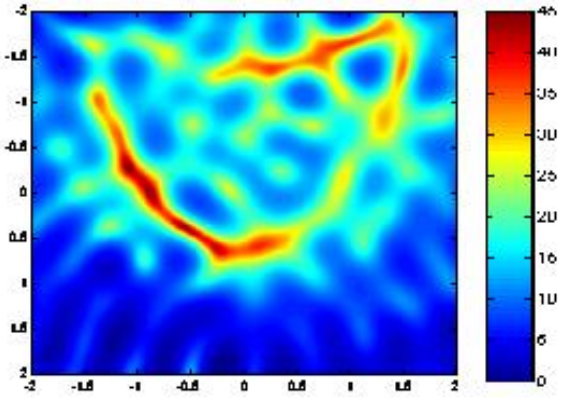


Fig. 16. PEC with noise effect SNR=5dB and TE polarization

For this purpose we tested a dielectric ($\varepsilon_r = 4$) with TM polarization and SNR=-5dB at Fig.15 and shows loss of resolution in the shape reconstruction significantly toward to Fig. 17.

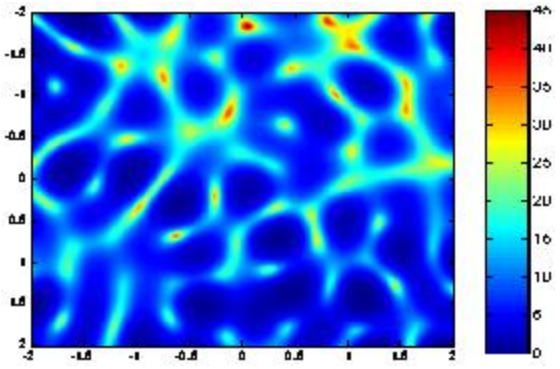


Fig. 17. Dielectric with noise effect SNR=-5dB and TM polarization

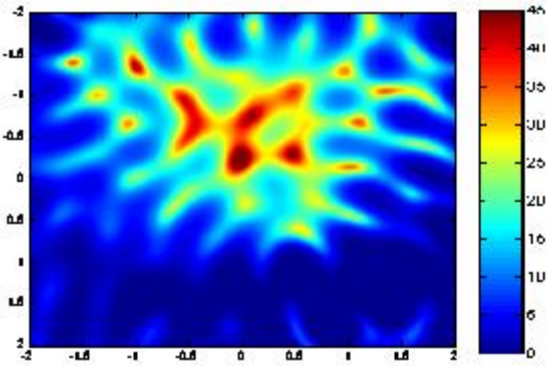


Fig. 18. Dielectric with noise effect SNR= -5dB and TE polarization

6. Conclusions

To put everything in a nutshell, in this paper, we have investigated the performance of linear sampling method (LSM) for TM and TE polarizations based on far-field equation. It was shown by numerical examples that the quality of reconstruction of PEC objects is better than dielectric ones for TM polarization.

However, in the dielectric object, TE polarization operates a little better than TM one. In addition, we have investigated the multiple scattering and coated objects and comparison between various polarizations. Further we posed an open problem about the reconstruction of a hidden shape within other unknown shape that both of them are dielectric especially due to having lots of applications. Moreover we discussed about the stability of LSM toward noise effect and had numerical examples for that.

In the future, we plan to study the combination of TM and TE polarization; because the forward pattern has the information of both polarizations hence it can improve the quality of LSM results. Also we will consider the effect of frequency in various ranges on the shape reconstruction and infer it on the formulation to improve the results.

References

[1] L. Crocco, I. Catapano, L. Di Donato and T. Isernia, "The Linear Sampling Method as a Way to Quantitative," *IEEE Transactions on Antennas and Prpagation*, p. 10, 2012.

[2] D. Colton, H. Haddar and M. Piana, "The linear sampling method in inverse electromagnetic scattering theory," *Inverse Problems 19*, p. S105–S137, 12 November 2003.

[3] L. Souriau, B. Duchene, D. Lesselier and R. Kleunman, "Modified Gradient Approach to Inverse Scattering for Binary Objects In Stratified Media," *Inverse Problems*, vol. 12, pp. 463-481, 1996.

[4] T. Isernia, L. Crocco and M. D'Urso, "New Tools and Series for Forward and Inverse Scattering Problmes in Lossy Media," *IEEE Geosci.Remote Sens.Lett*, vol. 1, pp. 327-331, 2004.

[5] P. VandeBerg and R. Kleiman, "A Contrast Source Inversion Method," *Inverse Problmes*, vol. 13, pp. 1607-1620, 1997.

[6] X. Chen, "Subspace-Based Optimization Method for Solving Inverse Scattering Problmes," *IEEE Trans.Geosci.Remote Sens*, vol. 48, pp. 42-49, 2010.

[7] K. Miller, "Least Squares Methods for Ill-Posed Problems with a Prescribed Bound," *SIAM*, vol. 1, pp. 52-74, 1970.

[8] D. Colton and A. Kirsch, "A simple method for solving inverse scattering problems in the resonance region," *Inverse Problems 12*, p. 383–393, 19 January 1996.

[9] S. N. and K. F. Warnick, "Behavior of The Regularized Sampling Inverse Scattering Method at Integral Resonace Frequencis," *Progress In Electromagnetics Research, PIER 38*, p. 29–45, 2002.

[10] I. Catpano, L. Crocco and T. Isernia, "On Simple Methods for Shape Reconstruction of Unknown Scatterers," *IEEE Trans Antennas Propagation*, vol. 55, p. 1431-1436, 2007.

[11] I. Catapano, L. Crocco, M. D'Urso and T. Isernia, "On the Effect of Support Estimation and of a New Model in 2D Inverse Scattering Problems," *IEEE Trans Antennas Propag*, vol. 55, pp. 1895-1899, 2007.

[12] M. Brignone, G. Bozza, A. Randazzo, M. Piana and M. Pastorino, "A Hybrid Approach to 3d Microwave Imaging by Using Linear Sampling Method and ACO," *IEEE Trans Antennas Propag*, vol. 56, pp. 3224-3232, 2007.

[13] M. Brandfass, A. Lanterman and K. Warnick, "A comparison of the Colton-Kirsch inverse scattering method with linearized tomographic inverse scattering," *Inverse Problems*, vol. 17, p. 1797– 1816, December 2001.

[14] I. Catapano and L. Crocco, "A Qualitative Inverse Scattering Method for Through-The-Wall Imaging," in *IEEE Geoscience and Remote Sensing Lett*, Rome, 2010.

[15] Ö. Özdemir and H. Haddar, "Preprocessing the Reciprocity Gap Sampling Method in Buried-Object Imaging Experiments," *IEEE*, vol. 7, no. Geoscience and Remote Sensing Letters, pp. 756-760, October 2010.

[16] I. Catapano and L. Crocco, "An Imaging Method for Concealed Targets," *IEEE Transaction*, vol. 47, no. Geoscience and Remote Sensing, May 2009.
[17] G. Bozza, M. Brignone, M. Pastorino, M. Piana and A. Randazzo, "A Linear Sampling Approach to Crack Detection in Microwave Imaging," in *In Proceedings of IEEE International Workshop on Imaging Systems and Techniques*, Chni,Crete,Greece, 2008.
[18] H. Alqadah, M. Ferrara, H. Fan and J. Parker, "Single frequency Inverse Obstacle Scattering:A Sparsity Constrained Linear Sampling Method Approach," *IEEE Transaction on Image Processing*, vol. 21, pp. 2062-2075, 2012.
[19] D. Colton and R. Kress, Inverse acoustic and electromagnetic scattering theory, 2nd ed., Berlin: Springer, 1998
[20] I. Catapano, L. Crocco and T. Isernia, "Linear Sampling Mtehod: Physical Interpretation and Guidelines for a Successful Application," in *PIERS proceedings*, Hangzhou,China, 2008.
[21] D. Colton, "Inverse Acoustic and Electromagnetic ScatteringTheory," *Inverse Problems*, vol. 47, 2003.
[22] A. Tikhonov, A. Goncharsky, V. and A. Yagola, Numerical Method for the Solution of Ill-Posed Problmes, Kluwer Academics, 1995.
[23] D. Colton, K. Giebermann and P. Monk, "The linear sampling method for three-dimensional inverse scattering problems," vol. 42, pp. 434-460, 2000.
[24] R. Kress, "Boundary integral equations in time-harmonic acoustic scattering," *Math. Comput. Modelling*, vol. 15, pp. 229-243, 1991.

Mehdi Salarkaleji was born in Babolsar, Iran, in 1986. He received the B.Sc. degree in Electrical Communication Engineering from Shiraz University of Technology (SUT), Shiraz, Iran, in 2008, and the M.Sc. degree in Electrical Communication Engineering (Filed & Wave) from Isfahan University of Technology (IUT), in 2013,Isfahan, Iran. From 2009 to 2011, he was working in HUAWEI company as Microwave Engineer, and from 2013 to 2015 in RCII Lab as a EMC/SAR specialist. Now he is a PhD student In Electrical Engineering at Wayne State University (WSU), Detroit, MI, USA. His research interests include Inverse & Forward Scattering Problems, Microwave Imaging, Computational Electromagnetic and Antenna.

Mohammad Zoofaghari received the B.Sc. and M.Sc. degrees in Electrical Engineering from the Isfahan University of Technology, Isfahan, in 2008 and 2010, respectively, and he is currently working toward the Ph.D. degree in Amirkabir University of Technology, Tehran. His research interests include Scattering of Buried Objects in Multi-layers Media with Rough Interfaces, Through and Intra Wall Imaging and Inverse Scattering.

Reza Safian received the B.Sc. degree in Electrical Engineering from the Isfahan University of Technology, Isfahan, Iran, in 1999, the M.Sc. degree in Electrical Engineering from McMaster University, Hamilton, ON, Canada, in 2003, and the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2008. From 1999 to 2002, he was an RF Design Engineer with the Electrical and Computer Engineering Research Center (ECERC), Isfahan University of Technology. In 2008, he joined the faculty of the Isfahan University of Technology, where he is currently an Assistant Professor with the Electrical and Computer Engineering Department. His technical interests include Basic Electromagnetic Theory, Microwave Circuits, THz and Microwave Imaging.

Zaker Hossein Firouzeh was born in Isfahan, Iran, in 1977. He received the B.Sc. degree in Electrical Engineering from Isfahan University of Technology (IUT), Isfahan, in 1999 and the M.Sc. and Ph.D. degrees in Electrical Engineering from Amirkabir University of Technology (AUT), Tehran, Iran, in 2002, and 2011, respectively. He was a Research Engineer at Information Communication Technology Institute (ICTI) of IUT from 2002 to 2006. He experienced in design and implementation of antenna, Radar, and wireless communication systems. He currently works as an assistant professor in Department of Electrical and Computer Engineering, Isfahan University of Technology. He is the author or coauthor of more than 40 papers in journals and conference proceedings. His current research interests include antenna design and measurements, numerical techniques in electromagnetic, wave propagation and scattering, and EMC/EMI.

A New Approach to the Quantitative Measurement of Software Reliability

Abbas Rasoolzadegan*
Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
rasoolzadegan@um.ac.ir

Received: 02/May/2015 Revised: 07/Jun/2015 Accepted: 04/Jul/2015

Abstract

Nowadays software systems have very important role in a lot of sensitive and critical applications. Sometimes a small error in software could cause financial or even health loss in critical applications. So reliability assurance as a nun-functional requirement, is very vital. One of the key tasks to ensure error-free operation of the software, is to have a quantitative measurement of the software reliability. Software reliability engineering is defined as the quantitative study of the operational behavior of software systems with respect to user requirements concerning reliability. Software Reliability is defined as the probability of failure-free software operation for a specified period of time in a specified environment. Quantifying software reliability is increasingly becoming necessary.

We have recently proposed a new approach (referred to as $SDA_{Flex\&Rel}$) to the development of «reliable yet flexible» software. In this paper, we first present the definitions of a set of key terms that are necessary to communicate with the scope and contributions of this work. Based on the fact that software reliability is directly proportional to the reliability of the development approach used, in this paper, a new approach is proposed to quantitatively measure the reliability of the software developed using $SDA_{Flex\&Rel}$, thereby making precise informal claims on the reliability improvement. The quantitative results confirm the reliability improvement that is informally promised by $SDA_{Flex\&Rel}$.

Keywords: Reliability; Quantitative Measurement; Reliability Assessment; Fault Prevention; Formal Methods.

1. Introduction

The demand for complex software-hardware systems has increased more rapidly than the ability to develop them with highly desired quality [2], [6], [9], [12]. When the requirements for and dependencies on such systems increase, the possibility of crises from software failures also increases. The impact of these failures ranges from inconvenience (e.g., malfunctions of home appliances) to economic damage (e.g., interruptions of banking systems) to loss of life (e.g., failures of flight systems or medical software).

Software reliability engineering (SRE) is defined as the quantitative study of the operational behavior of software-based systems with respect to user requirements concerning reliability. SRE is centered around a very important facet of dependability, i.e., reliability. Software Reliability is defined as the probability of failure-free software operation for a specified period of time in a specified environment [1]. Software reliability has to be a probabilistic measure because the failure process, i.e. the way faults become active and cause failures, depends on the input sequence and operation conditions, and those cannot be predicted with absolute certainty [37-39]. Human behavior introduces uncertainty and hence probability into software reliability, although software usually fails in the same way for same operational conditions and same parameters. An additional reason to claim a probabilistic measure is that it is usually only possible to approximate the number of faults of complex software system.

Many concepts of software reliability can be adapted from the older and successful techniques of hardware reliability [40-41]. However, this must be done with care, since there are some fundamental differences in the nature of hardware and software, and their failure processes. The largest part of hardware failures is considered as result from physical deterioration. Sooner or later, these natural faults will introduce faults into hardware components and hence lead to failures. Experience has shown, that these physical effects are well-described by exponential equations in the relation to time. Usage commonly accelerates the reliability decrease, but even unused hardware deteriorates. Software does not wear outor deteriorate, i.e., its reliability does not decrease with time. Moreover, software generally enjoys reliability growth during testing and operation since software faults can be detected and removed when software failures occur. On the other hand, Software may experience reliability decrease due to abrupt changes of its operational usage or incorrect modifications to the software. Software is also continuously modified throughout its life cycle. The malleability of software makes it inevitable for us to consider variable failure rates.

Design faults are a different source for failures. They result mainly from human error in the development process or maintenance. Design faults will cause a failure under certain circumstances. The probability of the

* Corresponding Author

activation of a design fault is typically only usage dependent and time independent. Unlike hardware faults which are mostly physical faults, software failures are caused by design faults, which are harder to visualize, detect, and correct. In the context of software reliability, the term *design* refers to all software development steps from the requirements to implementation [2-3].

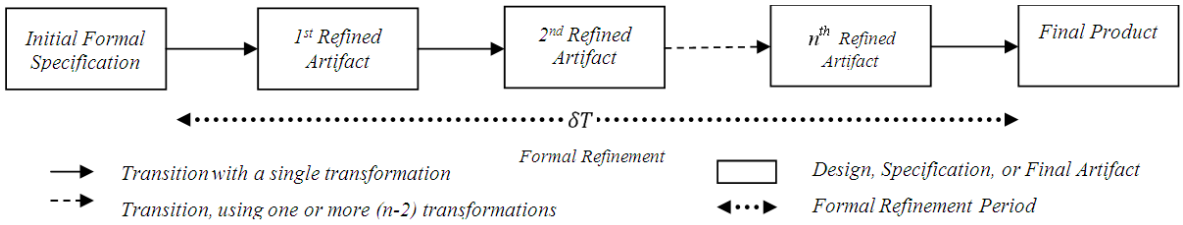


Fig. 1. Formal Software Development Process

In contrast to hardware, software can be perfect (i.e. fault-free). Formal modeling methods (FMMs) are broadly defined as notations with accurate and unambiguous semantics. They are supported by various tools. FMMs mathematically prove the consistency and completeness of activities during software development. Such proofs help detect all faults before they turn into failures. In addition, the correctness insured by proof is more comprehensive and reliable than the correctness guaranteed by test. These advantages facilitate the development of correct and reliable software [3-5].

Fig. 1 illustrates the formal software development process using FMMs. This process starts with an *initial formal specification*, which abstractly states the stakeholders' requirements. Then, the details of design are added to the initial specification through a gradual process (δT), using *formal refinement*. This process contains several intermediate *artifacts* refined by *transformations* and continues until producing the *final product* [6].

FMMs, along with formal refinement and formal verification techniques prove the correctness of software throughout the formal software development process. As a result, the absence of faults is guaranteed. However, lack of knowledge and high cost restrict their use to the development of critical and high integrity software. Critical systems such as spacecraft, aircraft, nuclear power plant and pacemakers require a high level of reliability in their operation. Software failures can lead to fatal consequences in safety-critical systems [4], thereby making it more important than ever to ensure the reliability of such systems. The term 'safety-critical' refers to those software systems whose failure may lead to loss of life or severe injury. In other words, safety-critical systems include software whose failure can lead to a hazardous state.

We have recently proposed a Software Development Approach (SDA). This approach, referred to as $SDA_{Flex\&Rel}$ in this paper, promises to develop reliable yet flexible software [7]. In this approach, Object-Z, as a dominant formal specification language, is used to formally specify and refine requirements – which, in turn, prevent and remove probable faults. Formal modeling and refinement in Object-Z ensure the reliability of software.

So far, many models have been proposed for quantification of the software reliability. Each of these models has its advantages and limitations [11-41]. In [43] we classify different approaches of software reliability modeling and finally, based on the analysis of the advantages and limitations, compare different approaches and mention some challenges and issues.

In this paper, we quantitatively measure the reliability improvement promised by $SDA_{Flex\&Rel}$. Indeed, the contribution of this paper is to measure the reliability of the software developed using $SDA_{Flex\&Rel}$ by measuring the reliability of $SDA_{Flex\&Rel}$ because there is a direct relation between the reliability of software and the reliability of the corresponding development approach. The idea behind this work has been inspired by an existing technique for reliability assessment, i.e., software metric based reliability analysis, as well as a typical type of reliability measurement, i.e., prediction when failure data are not available.

The rest of this paper is organized as follows: Section two presents the definitions of a set of key terms that are necessary to communicate with the scope and contributions of this work. These terms are *dependability*, *failure*, *fault*, and *error*. A brief description of the main approaches to the achievement of reliability, the major classes of reliability assessment, and the main activities of reliability measurement are also presented in section two. The reliability of the software development approach $SDA_{Flex\&Rel}$ is quantitatively measured in section three. Finally, section four discusses the conclusions.

2. Background

2.1 Dependability

Dependability is defined as the trustworthiness of a software-hardware system such that reliance can justifiably be placed on the service it delivers [1-3], [8-9]. The service delivered by a system is its behavior as it is perceptible by its user(s); a user is another system (human or physical) interacting with the former. Depending on the application(s) intended for the system, a different

emphasis may be put on the various facets of dependability, that is, dependability may be viewed according to different, but complementary, properties, which enable the attributes of dependability to be defined:

- The readiness for usage leads to *availability*.
- The continuity of service leads to *reliability*.
- The nonoccurrence of catastrophic consequences on the environment leads to *safety*.
- The nonoccurrence of the unauthorized disclosure of information leads to *confidentiality*.
- The nonoccurrence of improper alterations of information leads to *integrity*.
- The ability to undergo repairs and evolutions leads to *maintainability*.

2.2 Failure

A failure occurs when the user perceives that the system ceases to deliver the expected service [1]. The user may choose to identify several severity levels of failures, such as: catastrophic, major, and minor, depending on their impacts to the system service. The definitions of these severity levels vary from system to system [3].

Failure behavior directly depends on the environment and the number of faults present in the software during execution. Let T denotes a random variable representing the system failure time. Failure density $f(T)$ corresponds to the probability distribution function of T. Failure probability $F(t)$ is the probability that the failure time is less or equal to time t [2], [10]:

$$F(t) = Prob(T \leq t) = \int_0^t f(u).du \qquad (1)$$

Reliability $R(t)$ is the probability that the system delivers the expected services in the time interval:

$$R(t) = 1 - F(t) = Prob(T \geq t) = \int_t^\infty f(u).du \qquad (2)$$

With respect to the type of hardware faults, hardware reliability metrics are usually time dependent. Although the failure behavior of software (design) faults depends on usage and not directly on time, software reliability is usually expressed in relation to time, as well. However, it is possible to define software reliability with respect to other bases such as software runs. A major advantage of time dependent software reliability metrics is that they can be combined with hardware reliability metrics to estimate the system reliability. Only as intermediate results, some reliability models use time-independent metrics.

2.3 Fault

A fault is uncovered when either a failure of the software occurs or an internal error (e.g., an incorrect state) is detected within the software. The cause of the failure or the internal error is said to be a fault. It is also referred as a *bug*. Software faults arise mostly from design issues. The source of software faults include:

- Incorrect requirements, even though the implementation may match them.
- Implementation (software design and coding) deviating from (correct) requirements.
- Uncontrolled or unexpected changes in operational usage or incorrect modifications.

In summary, a software failure is an incorrect result with respect to the specification or an unexpected software behavior perceived by the user at the boundary of the software system, while a software fault is the identified or hypothesized cause of the software failure. When the distinction between fault and failure is not critical, *defect* can be used as a generic term to refer to either a fault (cause) or a failure (effect).

2.4 Error

The term error has two different meanings [3], [10]:

1. A discrepancy between a computed, observed, or measured value, or condition and the true, specified, or theoretically correct value or condition. Errors occur when some part of the software produces an undesired state. Examples include exceptional conditions raised by the activation of existing software faults and an incorrect system status due to an unexpected external interference. This term is especially useful in fault-tolerant computing to describe an intermediate stage in-between faults and failures.
2. A human action that results in software containing a fault. Examples include omission or misinterpretation of user requirements in a software specification, and incorrect translation or omission of a requirement in a software design. However, this is not a preferred usage, and the term *mistake* is used instead to avoid the confusion.

2.5 Approaches to the Achievement of Reliability

The development of a reliable software system calls for the combined utilization of a set of methods and techniques which can be classed into [3], [16], [25-30], [32], [34], [36]:

- *Fault prevention:* how to prevent fault occurrence or introduction. The interactive refinement of the user's system requirement, requirements engineering (RE), the use of sound design methods, and the encouragement of writing clear code are the general approaches to prevent faults in the software. Formal methods develop and refine requirement specifications correctly using languages and tools with sound mathematical bases in order to achieve the following goals: 1) executable specifications for systematic and precise evaluation, 2) proof mechanisms for step-by-step verification using incremental refinement, and 3) every intermediate artifact is a subject to mathematical verification for correctness and appropriateness.
- *Fault removal:* how to reduce the presence (number and seriousness) of faults. Fault removal uses techniques such as testing, inspection, verification, and validation to track and remove faults in software. Formal inspection is a practical fault removal

scheme which is widely implemented in industry. Formal inspection is a rigorous process focused on finding faults, correcting faults, and verifying the corrections.

- *Fault tolerance:* how to ensure a service capable of fulfilling the system's function in the presence of faults. Software fault tolerance is achieved by design diversity in which multiple versions of software are developed. These multiple versions, which are functionally equivalent yet independent, are applied in the system to provide ultimate tolerance to software design faults.
- *Fault forecasting:* how to estimate the present number, future incidence, and consequences of faults. Fault forecasting involves formulation of the fault/failure relationship, an understanding of the operational environment, the establishment of reliability models, the collection of failure data, the application of reliability models by tools, the selection of appropriate models, and the analysis and interpretation of results.

2.6 Reliability assessment

The three major classes of software reliability assessment are [8-9], [14], [24]:

- *Black box reliability analysis:* Estimation of the software reliability based on failure observations from testing or operation. These approaches are called black boxapproaches because internal details of the software are not considered.
- *Software metric based reliability analysis:* Reliability evaluation based on the static analysis of the software (e.g., lines of code, number of statements, complexity) or its development process and conditions (e.g., developer experience, applied testing methods).
- *Architecture-based reliability analysis:* Evaluation of the software system reliability from software component reliabilities and the system architecture (the way the system is composed out of the components). These approaches are sometimes called *component-based reliability estimation* (CBRE), or *grey* or *white box* approaches.

2.7 Reliability measurement

Measurement of software reliability includes two types of activities: reliability *estimation* and reliability *prediction* [11], [13]. Estimation determines current software reliability by applying statistical inference techniques to failure data obtained during system test or during system operation. This is a measure regarding the achieved reliability from the past until the current point. Its main purpose is to assess the current reliability and determine whether a reliability model is a good fit in retrospect. Prediction determines future software reliability based upon available software metrics and measures [15]. Depending on the software development stage, prediction involves different techniques [17-23]:

1. When failure data are available (e.g., software is in system test or operation stage), the estimation techniques can be used to parameterize and verify software reliability models, which can perform future reliability prediction.
2. When failure data are not available (e.g., software is in design or implementation stages), the metrics obtained from the software development process and the characteristics of the resulting product can be used to predict reliability of the software.

Data collected during the test phase is often used to estimate the number of software faults remaining in a system which in turn often is used as input for reliability prediction. This estimation can either be done by looking at the numbers (and the rate) of faults found during testing or just by looking at the effort that was spent on testing. The underlying assumption when looking at testing effort is "more testing leads to higher reliability" [31], [33], [35].

3. Quantifying the reliability of the software developed using $SDA_{Flex\&Rel}$

The software development approach $SDA_{Flex\&Rel}$ has recently been proposed to develop reliable yet flexible software [7]. In $SDA_{Flex\&Rel}$, formal (Object-Z) and semi-formal (UML) models are transformed into each other using a set of bidirectional formal rules. In this approach, Object-Z, as a dominant formal specification language, is used to formally specify, verify, and refine requirements to prevent and remove probable faults. As previously mentioned, fault prevention and fault removal are two main approaches to the development of reliable software systems. Therefore, formal modeling, verification, and refinement in Object-Z ensure the reliability of software. Visual models (UML diagrams) facilitate the interactions among stakeholders who are not familiar enough with the complex mathematical concepts of formal modeling methods. Applying design patterns to visual models improves the flexibility of software. The transformation of formal and visual models into each other through the iterative and evolutionary process, proposed in [7], helps develop the software applications that need to be highly reliable yet flexible. The workflow of $SDA_{Flex\&Rel}$ is illustrated in Fig. 2.

The iterative and evolutionary process illustrated in Fig. 2 continues until a final product with a desired quality (in terms of reliability and flexibility) is achieved. Fig. 3 illustrates the details of an iteration of $SDA_{Flex\&Rel}$ which consists of the following phases:

- Reliability Assurance Phase (RAP) which supports formal specification and refinement in Object-Z.
- Visualization Phase (VP) which transforms Object-Z models into UML ones.
- Flexibility Assurance Phase (FAP) which revises UML models from the viewpoints of design patterns and polymorphism.

- Formalization Phase (FP) which transforms UML models into Object-Z ones.

In order to assess/measure the reliability of the software developed using $SDA_{Flex\&Rel}$, the reliability of $SDA_{Flex\&Rel}$ is evaluated because there is a direct relation between the reliability of software and the reliability of the corresponding development approach [2-3], [6]. In other words, software reliability is directly proportional to the reliability of the development approach used. As previously mentioned, from the view point of assessment, such reliability assessment is categorized as software metric based reliability analysis, and from the viewpoint of measurement, such reliability measurement is categorized as prediction when failure data are not available. According to the details of each iteration in the proposed approach, the total reliability of $SDA_{Flex\&Rel}$ is calculated as:

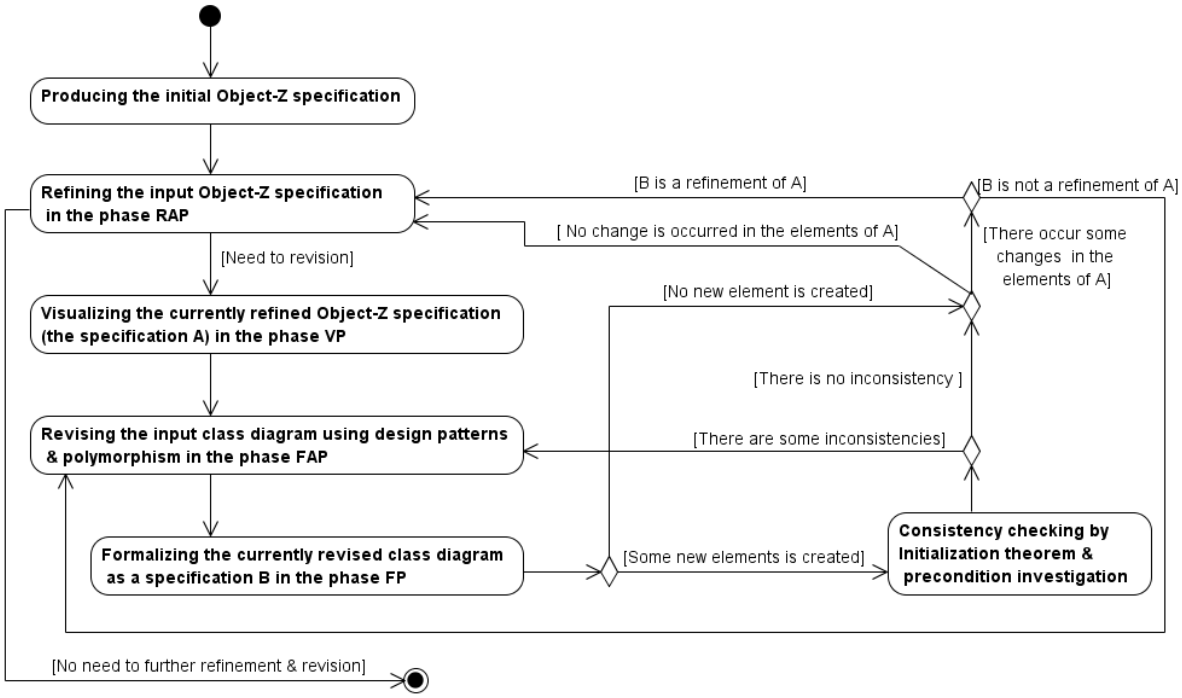


Fig. 2. The workflow of $SDA_{Flex\&Rel}$

$$R_{SDA_{Flex\&Rel}} = \prod_{i=1}^{k}\prod_{j=1}^{n_i} R_{RAP}(i,j) * \prod_{i=1}^{k} R_{VP}(i) * \prod_{i=1}^{k}\prod_{j=1}^{m_i} R_{FAP}(i,j) * \prod_{i=1}^{k} R_{FP}(i) \quad (3)$$

k Number of iterations in the development process proposed by $SDA_{Flex\&Rel}$.

n_i Number of formal refinement steps during RAP in the iteration i of $SDA_{Flex\&Rel}$.

m_i Number of revision steps during FAP in the iteration i of $SDA_{Flex\&Rel}$.

$R_{RAP}(i,j)$ Reliability of the jth formal refinement step in RAP during the ith iteration.

$R_{VP}(i)$ Reliability of the ith formal transformation from Object-Z into UML (formalization) in VP.

$R_{FAP}(i,j)$ The reliability of the jth revision step in FAP during the ith iteration of $SDA_{Flex\&Rel}$.

$R_{FP}(i)$ Reliability of the ith formal transformation from UML into Object-Z (visualization) in FP.

$R_{SDA_{Flex\&Rel}}$ Total reliability of $SDA_{Flex\&Rel}$.

As previously mentioned, contrary to hardware, software does not wear out or deteriorate, i.e., its reliability does not decrease with time due to physical depreciation. However, Software may experience reliability decrease due to abrupt changes of its operational usage or incorrect modifications to the software. Therefore, the reliability of a flexible software or a flexible software development approach (such as $SDA_{Flex\&Rel}$) does not decrease with time because "flexibility" is defined as the ability of a system to respond to potential internal or external changes affecting its value delivery, in a timely and cost-effective manner. In other word, the reliability of software or a software development approach in the presence of flexibility is equivalent to R (reliability in the absence of time) instead of $R(t)$ (general definition of reliability presented in subsection 2.2). According to the fact that the flexibility of $SDA_{Flex\&Rel}$ has been demonstrated in [42], the reliability of $SDA_{Flex\&Rel}$ is calculated regardless of time as $R_{SDA_{Flex\&Rel}}$ instead of $R_{SDA_{Flex\&Rel}}(t)$.

In the current version of $SDA_{Flex\&Rel}$, all activities of the phases RAP, FP, and VP are performed formally with sound mathematical bases. The proposed formal transformation rules make it possible to transform UML class diagrams and Object-Z specifications into each other

without any fault during the phases FP and VP. Moreover, formal refinement, along with formal verification guarantees the correctness of the activities performed during the phase RAP. As previously mentioned, formalism ensures the absence of faults. Therefore, the reliability of each of those activities performed during the phases RAP, FP, and VP equal 1 $(\forall i,j,\ R_{RAP}(i,j) = R_{VP}(i) = R_{FP}(i,j) = 1)$.

However, in the current version of $SDA_{Flex\&Rel}$, during the phase FAP, designers apply the required design patterns to the class diagram of the system being developed without any formal systematic control. This may cause the syntactic or the semantic structure of the class diagram to become inconsistent. Therefore, the reliability of every activity in the phase FAP does not equal 1 $(\forall i,j,\ R_{FAP}(i,j) \neq 1)$. Relation (3) is then simplified as:

$$R_{SDA_{Flex\&Rel}} = (1) * (1) * \prod_{i=1}^{k}\prod_{j=1}^{m_i} R_{FAP}(i,j) * (1) \Rightarrow$$

$$R_{SDA_{Flex\&Rel}} = \prod_{i=1}^{k}\prod_{j=1}^{m_i} R_{FAP}(i,j) \qquad (4)$$

$$\forall i,j,\ 0 < R_{FAP}(i,j) < 1$$

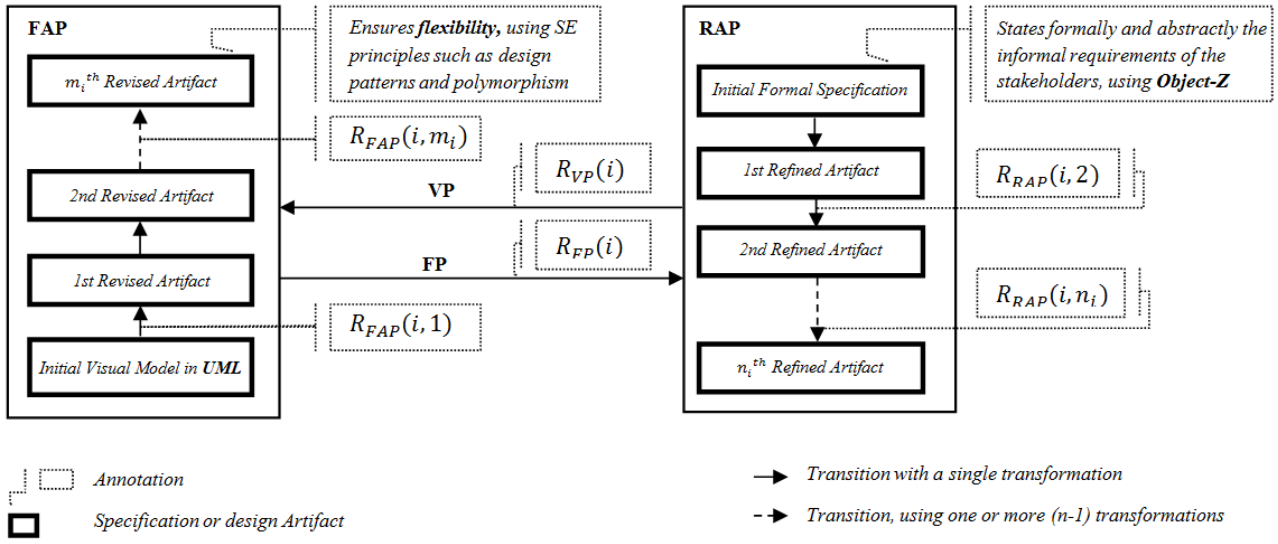


Fig. 3. A schematic view of an iteration i of SDA*Flex&Rel*

A formal mechanism can be proposed to make the class diagram of the software being developed be formally revised when a design pattern is applied to it in FAP. As a result, applying design patterns to the class diagram of software not only improves the flexibility of the software but also preserves the syntactic and the semantic structure of the class diagram – which, in turn, leads to consistency preservation during the revision process of the class diagram in FAP. To do so, a set of formal rules can be defined using model refactoring based on graph transformation at the meta-level of the UML class diagram to make it possible to add/remove/change a modeling element to/from/in a class diagram without making its syntax and semantics become inconsistent. Designers are then allowed to change a class diagram just using the defined rules in order to revise it based on a design pattern. Therefore, the reliability of every activity in the phase FAP will equal 1 $(\forall i,j,\ R_{FAP}(i,j) = 1)$. Relation (3) is then simplified further as:

$$R_{SDA_{Flex\&Rel}} = \prod_{i=1}^{k}\prod_{j=1}^{m_i} 1 = 1 \qquad (5)$$

As previously mentioned, the reliability of software developed using a development approach is directly proportional to the reliability of the development approach. Therefore, the reliability of the software developed using the current version of $SDA_{Flex\&Rel}$ is obtained according to relation (4), but in the future, by proposing a formal mechanism for supporting the revision process of the phase FAP, the reliability of the software increases to 1 according to relation (5).

Generally, a software development process includes several (n) activities [6], [44-45]. Based on the assumption that these activities do not enjoy a sound mathematical (formal) basis, the reliability of each of them dose not equal to 1 $(\forall i, R(i) \neq 1)$. As a result, the process reliability can be formulated as:

$$R_{GenericProcess} = \prod_{i=1}^{n} R(i)\ \forall i,\ 0 < R(i) < 1 \qquad (6)$$

In order to simply compare $R_{GenericProcess}$ with $R_{SDA_{Flex\&Rel}}$, relation (4) is reformulated as follows:

$$R_{SDA_{Flex\&Rel}} = \prod_{i=1}^{k}\prod_{j=1}^{m_i} R_{FAP}(i,j) = \prod_{p=1}^{k'} R_{FAP}(i',j') \qquad (7)$$

such that:

$$\forall\, i', j',\ \ 0 < R_{FAP}(i', j') < 1,$$
$$k' = \sum_{i=1}^{k} m_i,$$
$$\forall p, 0 \le n' < k \cdot \sum_{j=1}^{n'} m_j \le p \le \sum_{j=1}^{n'+1} m_j$$
$$\Rightarrow i' = n' + 1 \ \wedge\ j' = i - \sum_{j=1}^{n'} m_j$$

The following assumptions are made according to relations (6) and (7):

1. In $R_{GenericProcess}, \forall i,\ 0 < R(i) < 1$,
2. In $R_{SDA_{Flex\&Rel}}, \forall\, p, i', j',\ 0 < R_{FAP}(i', j') < 1$,
3. $\forall\, i', j', i,\ R_{FAP}(i', j') \not< R(i)$, because: 1) the input materials of FAP are correct and fault-free and 2) During FAP, the input materials are just revised using design patterns and polymorphism with low possibility of fault occurrence, and 3) the lack of semantic inconsistency between the input and the output of the phase FAP is guaranteed by the existing formal analysis techniques (such as initialization theorem and precondition investigation) and the various formal verification mechanisms that support Object-Z (as illustrated in Fig. 2).
4. $k' \ll n$,

With respect to these assumptions, we can conclude that:

$$\prod_{p=1}^{k'} R_{FAP}(i', j') \prod_{i=1}^{n} R(i) \Rightarrow \tag{8}$$
$$R_{SDA_{Flex\&Rel}} \gg R_{GenericProcess}$$

The above-mentioned analysis shows that the reliability of $SDA_{Flex\&Rel}$ is greater than the reliability of a generic software development process. The main conclusion is that the more widespread the use of formalism including formal specifications, refinement, and verification throughout a software development process, the more reliable the software development process will be. Therefore, supposing that some (m) of the activities of the software development process $GenericProcessare$ performed formally, the reliability of $GenericProcess$, previously formulated as relation (6), is reformulated as relation (9):

$$R_{GenericProcess} = \prod_{i=1}^{n} R(i) = \prod_{i=1}^{m} R(i) * \prod_{i=1}^{n-m} R(i)$$
$$\xRightarrow{\forall i, 1 \le i \le m,\ R(i)=1} R_{GenericProcess} = 1 * \prod_{i=1}^{n-m} R(i) \Rightarrow$$
$$R_{GenericProcess} = \prod_{i=1}^{n-m} R(i)\ \forall i,\ \ 0 < R(i) < 1 \tag{9}$$

Conclusion: $m \to n \ \Rightarrow\ R_{GenericProcess} \to 1$

With respect to the fact that the reliability of a software product is directly proportional to the reliability of the development approach used, the more reliable the software development approach, the more reliable the software product. According to the aforementioned analyses, the reliability of software developed using $SDA_{Flex\&Rel}$ is greater than the reliability of software developed using a generic software development process.

4. Conclusions

In this paper, we quantify the reliability improvement promised by the software development approach $SDA_{Flex\&Rel}$, which has recently been proposed to develop reliable yet flexible software. This approach improves software reliability through preparing the ground for formal modeling, refinement, and verification– which, in turn, prevent and remove probable faults. In order to quantify the reliability of the software developed using $SDA_{Flex\&Rel}$, the reliability of $SDA_{Flex\&Rel}$ is quantitatively measured because there is a direct relation between the reliability of software and the reliability of the corresponding development approach. In other words, software reliability is directly proportional to the reliability of the development approach used. Such reliability assessment is categorized as software metric based reliability analysis. The results confirm the promised reliability improvement

References

[1] ISO/IEC/IEEE, *Systems and software engineering – Vocabulary*, ISO/IEC/IEEE 24765:2010.

[2] H. Pham, *System Software Reliability*, Springer, 1st ed., 2007.

[3] M. R. Lyu, *Handbook of Software Reliability Engineering*, McGraw-Hill, 1996.

[4] A. Pandit, "A Framework-Based Approach for Reliability & Quality Assurance of Safety-Critical Software," *Int. Journal on Computer Science and Eng*., vol. 2 (9), pp. 2874-2879, 2010.

[5] H. B. Christensen, *Flexible, Reliable Software: Using Patterns and Agile Development*, Chapman and Hall/CRC; 1st ed., 2010.

[6] D. Bjørner, *Software Engineering III: Domains, Requirements, and Software Design*, Springer, 2006.

[7] A. Rasoolzadegan, A. Abdollahzadeh, "Reliable yet Flexible Software through Formal Model Transformation (Rule Definition)," *Journal of Knowledge and Information Systems (KAIS)*, vol. 40 (1), 2014.

[8] W. Ecker, W. Müller, Rainer Dömer, *Hardware Dependent Software Principles and Practice*, Springer, 2009.

[9] L. I. Millett, *Software for Dependable Systems: Sufficient Evidence?*, The National Academies Press, 2007.

[10] International Organization for Standardization, *ISO Standard 9126: Software Engineering – Product Quality, parts 1, 2 and 3*, Geneve, Switzerland, 2001 (part 1), 2003 (parts 2 and 3).

[11] M. Rahmani, A. Azadmanesh, "Exploitation of Quantitative Approaches to Software Reliability," *Tech. Rep. cst-2011-002*, Computer Science, University of Nebraska, Omaha, Dec. 2011.

[12] P.H. Seong, *Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems*, Springer, 1st ed., Berlin, Germany, pp. 85–87, 2009.
[13] X. Li, "Software reliability measurement: a survey," *MSc. Thesis*, Dept. Computer Science & Software Engineering, Concordia University, 2002.
[14] P. C. J. P. K. Kapur, H. Pham, A. *Gupta, Software Reliability Assessment with OR Applications*, 1st ed., London, England, Springer, 2011.
[15] A. K. Pandey, N.K. Goyal, *Early Software Reliability Prediction: a Fuzzy Logic Approach*, Springer, 2013.
[16] S. Yamada, *Software reliability modeling Fundamental and Applications*, Japan, Springer, 2014.
[17] Q. P. Hu, Y.-S. Dai, M. Xie, S. H. Ng, "Early software reliability prediction with extended ANN model," *in30th Annual International Conference on Computer Software and Applications*, 2006, vol. 2, pp. 234–239, 2006.
[18] S. Mohanta, G. Vinod, A. K. Ghosh, R. Mall, "An approach for early prediction of software reliability", *ACM SIGSOFT Softw. Eng. Notes*, vol. 35, no. 6, pp. 1–9, 2010.
[19] A. Immonen, E. Niemela, "Survey of reliability and availability prediction methods from the viewpoint of software architecture," *Software & System Modeling*, Springer, vol. 7, no. 1, pp. 49-65, Jan 2007.
[20] S. S. Gokhale, K. S. Trivedi, "Analytical models for architecture-based software reliability prediction: A unification framework," *Reliab. IEEE Trans*., vol. 55, no. 4, pp. 578–590, 2006.
[21] R.H. Reussner, H.W. Schimidt, I.H. Poernomo, "Reliability prediction for component-based software architectures," *J. System Softw*., vol. 66, no. 3, pp. 241-252, 2003.
[22] G.N. Rodrigues, D.S. Rosenblum, S. Uchitel, "Using scenarios to predict the reliability of concurrent component-based software systems," *in Proceedings of the8th international conference on Fundamental Approaches to Software Engineering*, pp. 111-126, 2005.
[23] S. S. Gokhale, K. S. Trivedi, "reliability prediction and sensitivity analysis based on software architecture," *in Proceedings ofthe 3rd international symposiumon Software Reliability Engineering*, pp. 64-75, 2002.
[24] K. Goševa-Popstojanova, K. S. Trivedi, "Architecture-based approach to reliability assessment of software systems," *Journal of Performance Evaluation*, Elsevier, vol. 45, no. 2, pp. 179–204, 2001.
[25] H. A. Stiber, "A family of software reliability growth models," *in Proceeding of 31th Annual International Computer Software and Applications Conference*, IEEE, vol. 2, pp. 217-224, July 2007.
[26] H. Pham, "Software reliability and cost models: Perspectives, comparison, and practice," *Eur. J. Oper. Res*., vol. 149, no. 3, pp. 475–489, 2003.
[27] A. L. Goel, K. Okumoto, "Time-Dependent Error-Detection Rate Model for Software Reliability and Other Performance Measures," *IEEE Transactions on Reliability*, pp. 206 – 211, 2009.
[28] V. Volovoi, "Modeling of System Reliability Using Petri Nets with Aging Tokens," *J. Reliab. Eng. Syst. Saf*., vol. 84, pp. 149–161, 2004.
[29] M. Xie, K.-L. Poh, Y.-S. Dai, *Computing System Reliability: Models and Analysis*, 1st ed., New York, USA: Springer, 2004.
[30] M. Ohba, "Software reliability analysis models," *IBM J. Res. Dev*., vol. 28, no. 4, pp. 428–443, 1984.
[31] L. K. Singh, A. K. Tripathi, G. Vinod, "Software reliability early prediction in architectural design phase: Overview and Limitations," *J. Softw. Eng. Appl*., vol. 4, p. 181, 2011.
[32] W.L. Wang, M.H. Chen, "Heterogeneous software reliability modeling," *in Proceedings of 13th International Symposium on Software Reliability Engineering*, pp. 41 – 52, 2002.
[33] R. Rana, M. Staron, C. Berger, J. Hansson, M. Nilsson, F. Torner, W. Meding, C. Hoglund, "Selecting software reliability growth models and improving their predictive accuracy using historical projects data," *System and Software*, Elsevier, vol. 98, pp. 59–78, 2014.
[34] R. Lai, M. Garg, "A Detailed Study of NHPP Software Reliability Models (Invited Paper)," *J. Softw*., vol. 7, no. 6, pp. 1296–1306, Jun. 2012.
[35] K. Goševa-Popstojanova, K. S. Trivedi, "Architecture-based approaches to software reliability prediction," *International Journal of Computer Mathematics with Applications*, vol. 46, no. 7, pp. 1023–1036, 2003.
[36] S. S. Gokhale, M.-T. Lyu, "A simulation approach to structure-based software reliability analysis," *Softw. Eng. IEEE Trans*., vol. 31, no. 8, pp. 643–656, 2005.
[37] K.C. Chiu, Y.S. Huang, T.Z. Lee, "A study of software reliability growth from the perspective of learning effects," *International Journal of Reliability Engineering & System Safety*, vol. 93, no. 10, pp. 1410–1421, 2008.
[38] K. M. Cheol, J. S. Cheol, J. J. Ha, "Possibilities and Limitations of Applying Software Reliability Growth Models to Safety- Critical Software," *Journal of Nuclear Engineering and Technology*, vol. 39, no. 2, pp. 129–132, 2007.
[39] V. Almering, M. Van Genuchten, G. Cloudt, P. J. M. Sonnemans, "Using software reliability growth models in practice," *Software*, IEEE, vol. 24, no. 6, pp. 82–88, 2007.
[40] B. Cukic, E. Gunel, H. Singh, G. U. O. Lan, "The theory of software reliability corroboration," *IEICE Trans. Inf. Syst*., vol. 86, no. 10, pp. 2121–2129, 2003.
[41] K. Sharma, R. Garg, C. K. Nagpal, R. K. Garg, "Selection of Optimal Software Reliability Growth Models Using a Distance Based Approach," *IEEE Transactions on Reliability*, pp. 266–276, 2010.
[42] A. Rasoolzadegan, "A New Approach to the Quantitative Measurement of Software Flexibility," *Journal of Soft Computing and Information Technology*, submitted, to be evaluated.
[43] M. Hashemi, Z. Ghavidel, A. Rasoolzadegan, "A Systematic Literature Review on Software Reliability Modeling," *Journal of Modeling in Engineering*, submitted, to be evaluated.
[44] R. S. Pressman, *Software Engineering-A Practitioner's Approach-Required*, 7th ed. McGraw Hill, 2009.
[45] I. Sommerville, *Software Engineering*, 9th ed. Addison Wesley, 2011.

Abbas Rasoolzadegan has received his B.Sc. degree in Software Engineering from Air Force University in 2004, Tehran, Iran. He has also received M.Sc. and Ph.D. degrees in Software Engineering from Amirkabir University of Technology, Tehran, Iran, respectively in 2007 and 2013. During his Ph.D., he has worked on formal software engineering and model transformation. He is currently an assistant professor in the Computer Engineering Department of Ferdowsi University of Mashhad. His main research focus is on software quality engineering, model transformation, testing, and design patterns.

Fusion Infrared and Visible Images Using Optimal Weights

Mehrnoosh Gholampour
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
mehrnosh.gholampor@birjand.ac.ir
Hassan Farsi*
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
hfarsi@birjand.ac.ir
Sajjad Mohamadzadeh
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
s.mohamadzadeh@birjand.ac.ir

Received: 20/Jan/2015 Revised: 17/Mar/2015 Accepted: 12/Apr/2015

Abstract

Image fusion is a process in which different images recorded by several sensors from one scene are combined to provide a final image with higher quality compared to each individual input image. In fact, combination of different images recorded by different sensors is one of image fusion methods. The fusion is performed based on maintaining useful features and reducing or removing useless features. The aim of fusion has to be clearly specified. In this paper we propose a new method which combines vision and infrared images by weighting average to provide better image quality. The weighting average is performed in gradient domain. The weight of each image depends on its useful features. Since these images are recorded in night vision, the useful features are related to clear scene details. For this reason, object detection is applied on the infrared image and considered as its weight. The vision image is also considered as a complementary of infrared image weight. The averaging is performed in gradient of input images, and final composed image is obtained by Gauss-Seidel method. The quality of resulted image by the proposed algorithm is compared to the obtained images by state-of-the-art algorithms using quantitative and qualitative measures. The obtained results show that the proposed algorithm provides better image quality.

Keywords: Image Fusion; Useful Features; Infrared Image; Vision Image; Weighted Averaging.

1. Introduction

Image fusion is a process of combining different images recorded from one scene to provide a final image with higher quality compared to the individual input images [1]. Image fusion algorithm has three important stages as: input images that are classified in four categories [1,2,3], image registration [4,5] and combining algorithm [1]. The image fusion is performed in three levels as: pixel-level [1], feature level [1] and decision level [1]. In this paper, the image fusion is performed in pixel level. According to the aim of fusion, definition of the quality is specified. One of image fusion methods is combination of different images recorded by different sensors. The fusion is based on maintaining useful features and reducing or removing useless features. The aim of fusion must be clearly specified. This depends on the type of images and its application. The useful and useless features depend on the aim of fusion [1]. In this paper input images have been recorded by visible and infrared sensors from same scene and in sunset time. Since visible imaging sensors are dependent to scene light image, clearness also depends on scene light. However, due to imaging time (sunset time) there is not enough light for some points of scene and therefore there is not possibility to present all scene details.

The problem can be resolved by using infrared imaging sensor. Because this performs imaging based on heat of objects and radiant heat difference is displayed as a visible image. But cold object is not clear in infrared image [1]. Therefore, it seems that combining visible and infrared images using appropriate method can resolve the shortcoming of both sensors and provide a final image with quality of day time [1]. Thus, in this case, the aim of fusion is to obtain an image with better quality such that all scene details are clear [2]. The useful features of input images are clear objects of the scene which have to retain in final image [12]. In this paper the performance of the proposed method is compared against general methods including averaging by principle component analysis [14], averaging, laplacian pyramid [5] and wavelet transform [6].

* Corresponding Author

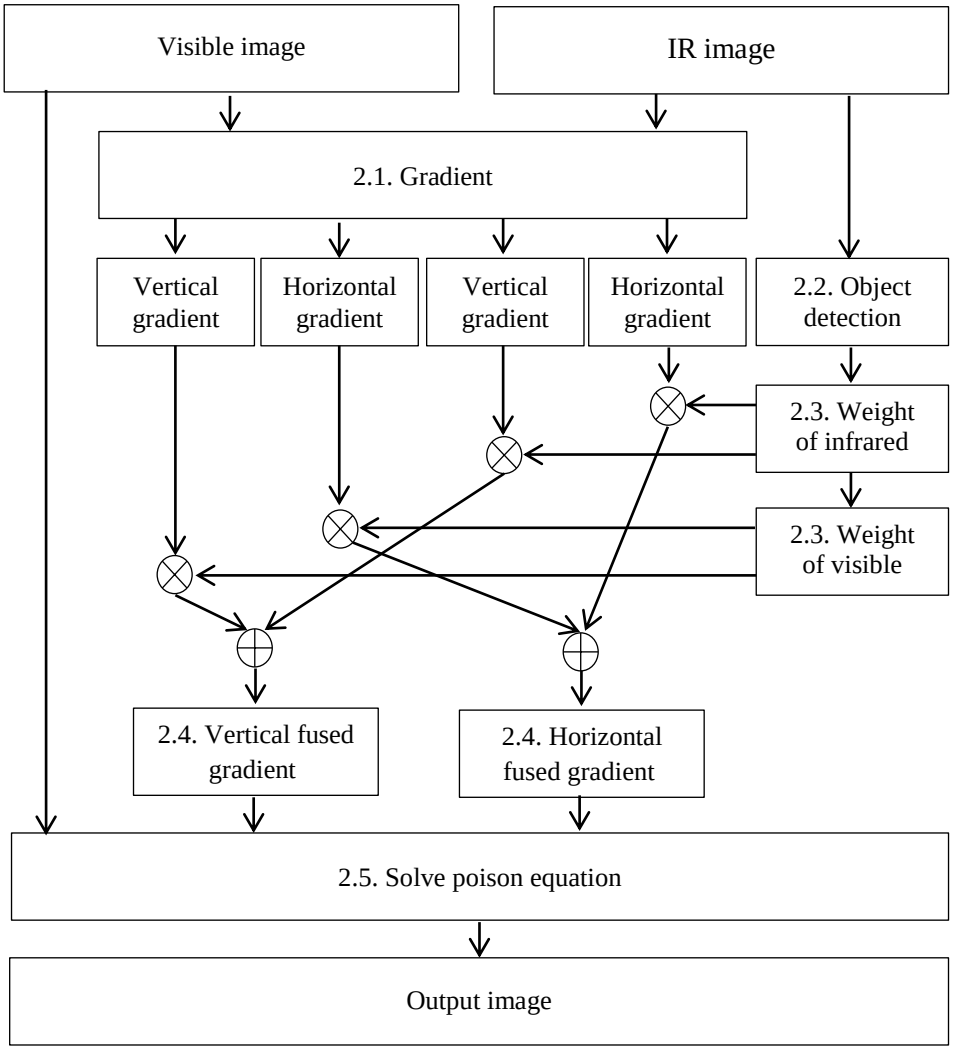


Fig. 1. Block diagram of the proposed algorithm

In averaging by principle component analysis method, the weight of each image is the normalized weight proportional to its principle component [14]. In averaging method, pixels of the output image are obtained by averaging on pixels of input image. In Laplacian method, each input image is decomposed into its Laplacian pyramid surfaces, then fusion is performed by combination of pyramid surfaces and final image is obtained using inverse pyramid transform [15]. In wavelet method, each input image is decomposed by wavelet transform and then the wavelet coefficients of the final image is calculated by maximization of high frequency coefficients and averaging on low frequency coefficients. The final image is obtained by using inverse wavelet transform [16]. In these methods, there are two problems: reduction in final image contrast and reduction in clarity of final image background, compared to visible input image. These problems occur because firstly low frequency components are blurred by averaging in whole image and so the contrast of final image reduces. Secondly, the weight of each image is not proportional to useful features. This causes that the useful information to be useless. In following section, we propose a new method which provides optimum weights and prevents contrast reduction in the final image.

2. The Proposed Method

An effective method to prevent contrast reduction is to apply averaging on gradient of each image. Because the gradient represents sudden changes or in other words high frequency image components and so low frequency image components remain without any changes. This results in the contrast reduction to be removed and low frequency information remains without any changes. To retain the useful information, we find the optimal weight of each image. The optimum weight includes only useful image features.

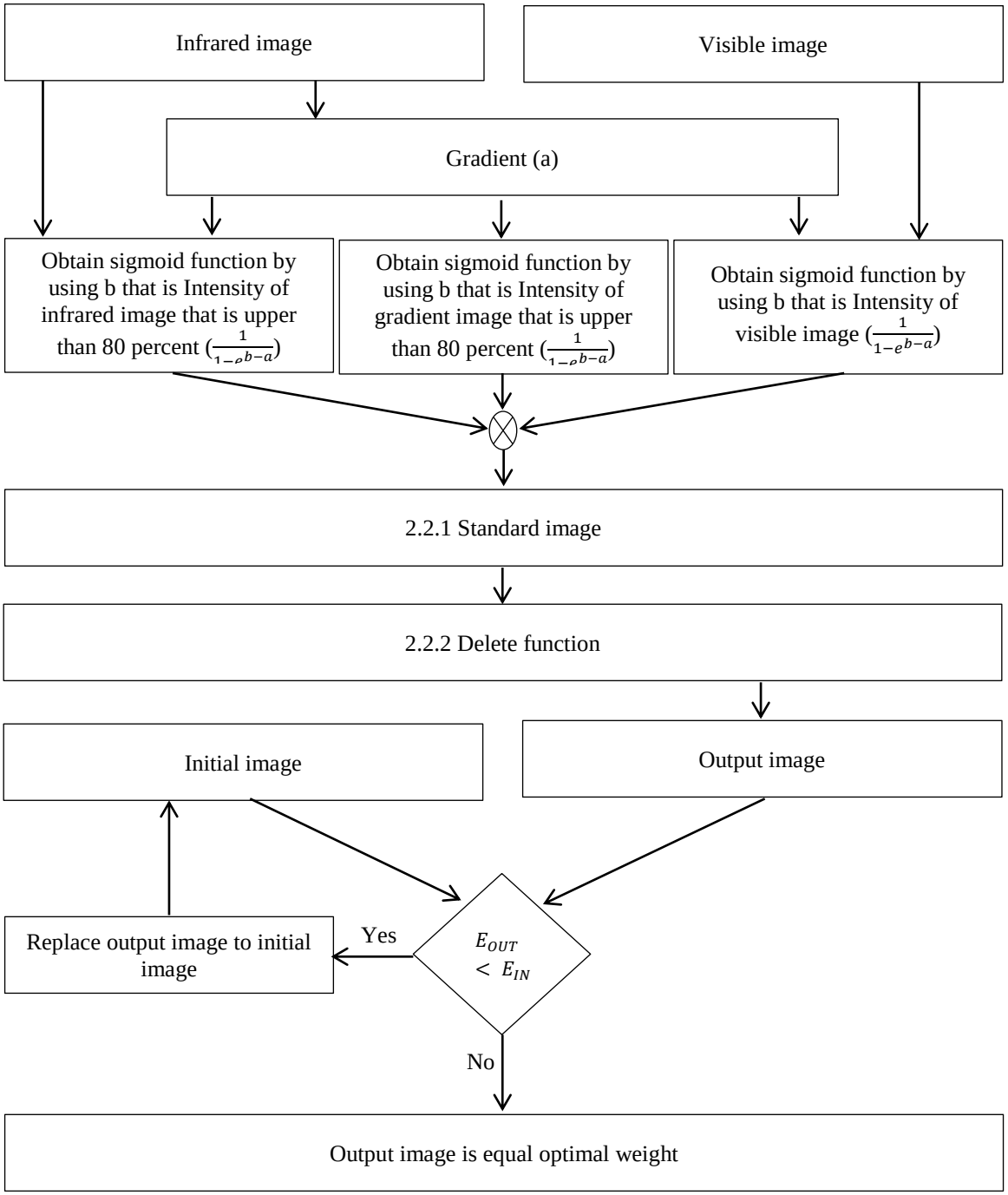


Fig. 2. Block diagram of the proposed object detection

The useful features are background information and hidden objects for the visible image and infrared image, respectively. Note that we use the term of "hidden objects" for infrared images because these objects cannot be observed in visible images. Therefore, the weight of infrared image is complementary of the weight of visible image. The aim of fusion is to detect the hidden objects existed in infrared image and to display them in a same background as existed in visible image. The block diagram of the proposed method is shown in Figure 1 which is detailed on following subsections.

2.1 Gradient Block

The inputs to the gradient block are the input images and the outputs are horizontal and vertical gradient of each image. The image gradient is obtained by convolution of input image with a vector of [-1/2, 0, 1/2] in row and column directions [17]. The gradient is used to extract the required information from images for instance, the edges and the changes in intensity. After computation of the gradient images, the pixels with large gradient values will be considered as possible edge pixels. The

pixels with the largest gradient values in the direction of the gradient are considered as edge pixels, and the edges may be traced in the direction perpendicular to the gradient direction. One example of edge detection algorithm based on gradient is the Canny edge detector.

2.2 Object Detection

In order to find the optimum weight of the infrared image, or in other words, hidden objects, it is applied as an input to the object detection block. For more classification, the process of object detection is shown as a block diagram in Figure 2. The main criterion for object detection is to exist higher intensity of object area compared to the background image. During the detection of high density areas in the image, undesired and additional areas will be also detected. Thus, it requires to use an effective method to remove undesired areas and maintain target area. This can be achieved by detection of target area borders, detailed in following section.

2.2.1 To Obtain Standard Image

Standard image includes the borders of the target area. This image is a multiplication of three threshold images. The threshold image is obtained by sigmoid function which is applied on gradient image. The gradient image is compared to three threshold levels. These thresholds are: (a) visible image intensities, (b) intensity upper than 80 percent of infrared image and (c) intensity upper than 80 percent of infrared gradient image.

2.2.2 To Execute Delete Function

Gamut of input and output of delete function is zero and one. For detection of target area completely, equality of neighbor pixels is used as a criterion. The delete function is given by:

$$F_{OUT}(i,j) = |f_{in}(i,j)) - IM(i,j)_{base}| + 0.7*(\left|\sum_{(p,q)\in N(x,y)} L - f_{in}(i,j))\right| \quad (1)$$

Where N is quartette neighborhood of current pixel located in position of (i, j) from input image. This function is executed twice for each pixel (for L=0, and L=1). The value of L corresponding to lower function value is selected and the lower function value is considered for output pixel. In first step, this function is executed on initial image. In next step, this function is applied on output image of the previous step and it is repeated until output image energy is less than input image energy. The energy of each image is obtained by:

$$E(i,j) = |f_{OUT}(i,j)) - IM(i,j)_{base}| + 0.7*(\left|\sum_{(p,q)\in N(x,y)} N - f_{OUT}(i,j))\right| \quad (2)$$

Where N is quartette neighborhood of current pixel located in position of (i, j) from input image. Note that the coefficient of 0.7 has been obtained in simulation by manual changing to get the best performance in the proposed method. By repeating the execution of delete function, we obtain an image which satisfies the criterion as mentioned in block diagram of Figure 2 which is $E_{out}<E_{in}$. Therefore the target area is detected and weight of infrared image is calculated. Since useful information of two images are complementary, we consider the weight of infrared and visible images complementary as well.

2.3 To Obtain Weight of Input Images

The weight of infrared and visible images are multiplied in gradient infrared and visible images, respectively.

2.4 To Obtain Fused Image Gradients

The horizontal and vertical gradients of two images are independently added together. The obtained horizontal and vertical gradients are considered as fused gradients in same direction.

$$I_{F'_x} = (1\text{-w})*I_{A_x} + (w)*I_{B_x} \quad (3)$$

$$I_{F'_y} = (1\text{-w})*I_{A_y} + (w)*I_{B_y} \quad (4)$$

Where W is the optimum fused weight.

2.5 Final Image Construction Using Fused Gradient

In order to construct fused image using fused gradients, we use solving Poisson equation [8]. Suppose final fused image and the gradient fused image are presented by ∇F and $\nabla F'$, respectively. It requires minimizing following equation:

$$\int abs\,(\nabla F - \nabla F')d\emptyset \quad (5)$$

Where $\emptyset$ is image area and equation (5) should be zero. To solve final fused problem, it requires solving Poisson equation:

$$\Delta F = div\,(\nabla F') \quad (6)$$

This implies that following equation is solved in terms of F(i,j):

$$F(i,j+1) + F(i,j-1) - 2*F(i, j) + F(i-1,j)+F(i+1,j) - 2*F(i, j) = div(\nabla F') \quad (7)$$

Thus, it requires to find the image pixel located in (i,j). Equation (7) can be rewritten as:

$$F(i,j) = 0.25*(F(i,j+1) + F(i,j-1) + F(i-1,j) + F(i+1,j) - div(\nabla F') \quad (8)$$

This equation is used for all image pixels. Thus, the number of unknown quantities is equal to all fused image pixels. Iterative method is used to solve Eq. (8). We use Gauss-seidel method due to having high speed convergence. According to the aim of fusion, visible image pixels are considered as initial values.

3. Output of the Proposed Object Detection Method

As an example, we apply the proposed object detection algorithm on three groups of input images

named: Nato Camp, Duine and Trees. The obtained results are shown in Figures 3, 4 and 5. In these Figures, (a) shows infrared image that is the input of the object detection, (b) shows the infrared image that is the output of first step of the object detection, (c) shows initial image that is the input of the delete function and (d) shows weight of infrared image that is the output of the object detection. As observed, the proposed method detects the hidden object (which is a person) and therefore it considers the hidden object as a weight of the infrared image. Figures 3(b) through 5(b) and 3(d) through 5(d) contain difference features. For instance, Figures 3(b) through 5(b) extract only the edges of the hidden object, but Figures 3(d) through 5(d) detect the hidden object. Meanwhile, Figures 3(b) through 5(b) consists of a lot of noise, and the proposed detection method removes the noise of Figures 3(b) through 5(b) and the obtained results are shown in the Figures 3(d) through 5(d).

4. Evaluation of the Proposed Algorithm

The proposed algorithm is evaluated using qualitative and quantitative measures. Visual comparison and numerical comparison are used as quantitative and qualitative measures, respectively.

4.1 Visual Comparison

Figures 6, 7 and 8 show the obtained images by different fusion algorithms. Visual and infrared images are shown in figures (a) and (b), respectively. The resulted image by using PCA [18,19], averaging [6], Laplacian pyramid [20], Wavelet [21] and the proposed algorithms are presented in Figures (c), (d), (e), (f) and (g), respectively.

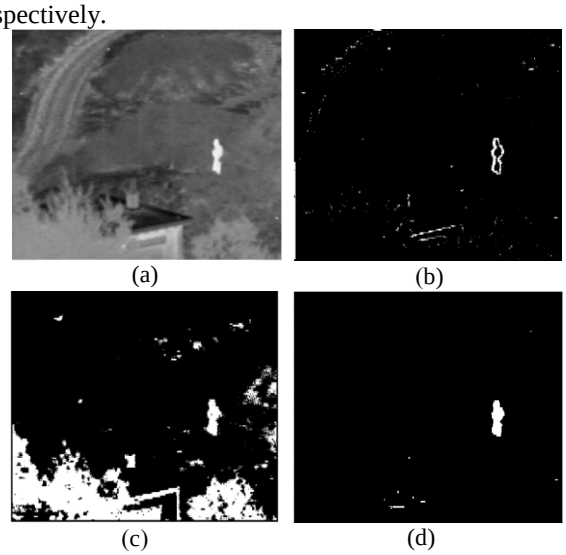


Fig. 3. Object detection on Nato Camp: (a) Infrared image (b) Standard image (c) Initial image (d) Weight of infrared image.

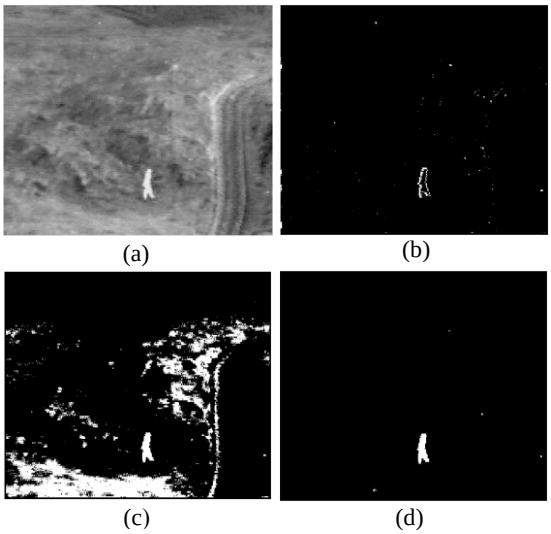


Fig. 4. Object detection on Duine: (a) Infrared image (b) Standard image (c) Initial image (d) Weight of infrared image.

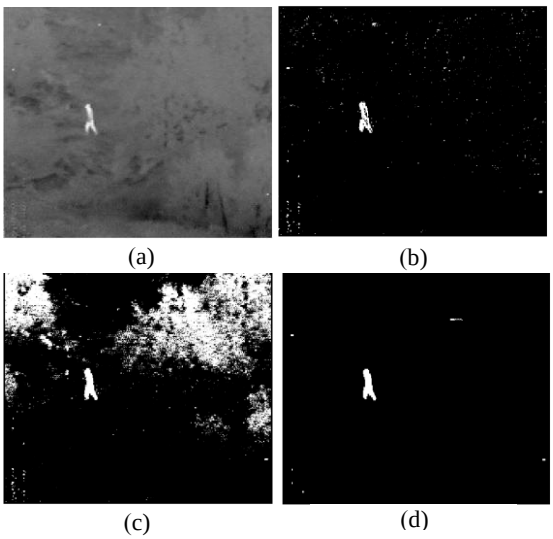


Fig. 5. Object detection Trees: (a) Infrared image (b) Standard image (c) Initial image (d) Weight of infrared image.

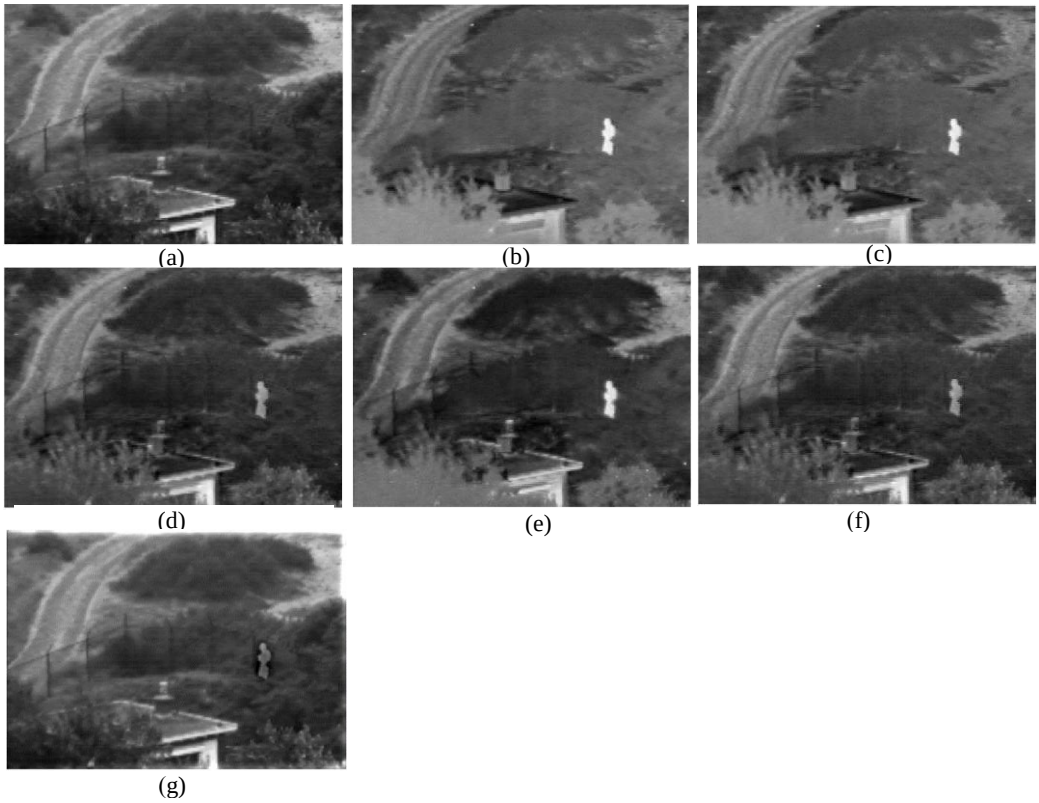


Fig. 6. Input and output of different methods on Nato Camp: (a) visible image (b) infrared image (c) output of PCA (d) Output of averaging method (e) Output of laplacian pyramid (f) output of wavelet (g) Output of the proposed algorithm.

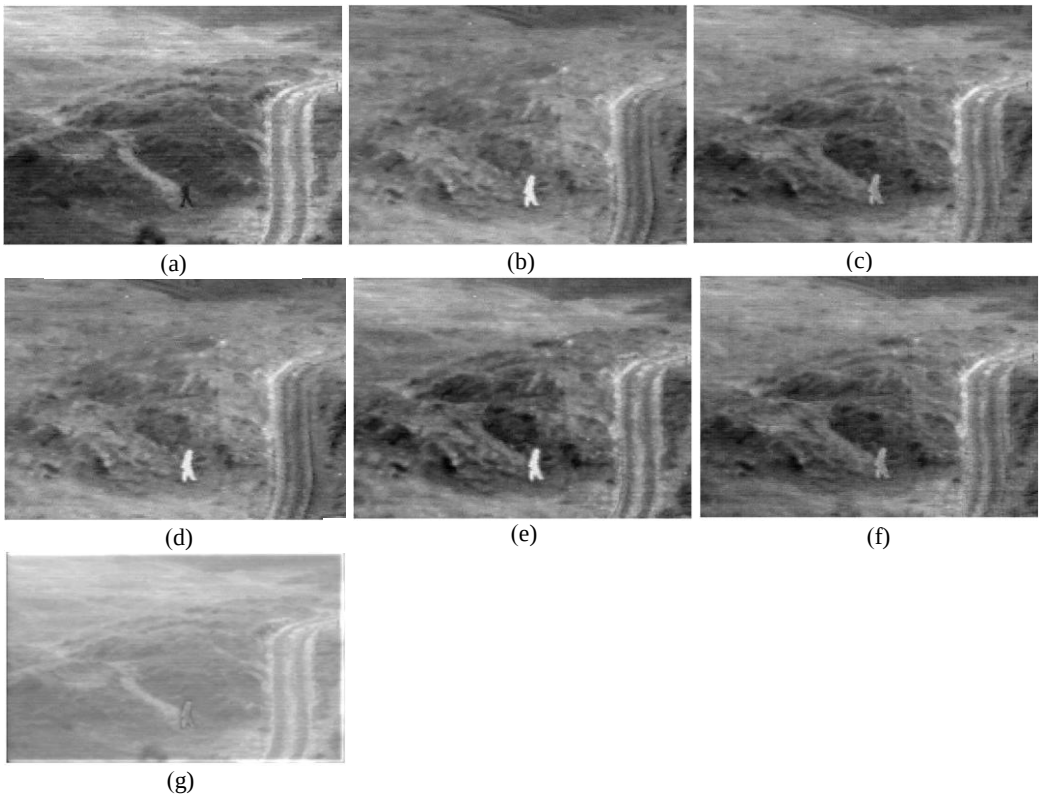


Fig. 7. Input and output of different methods on Duin: (a) visible image (b) infrared image (c) output of PCA (d) Output of averaging method (e) Output of laplacian pyramid (f) output of wavelet (g) Output of the proposed algorithm.

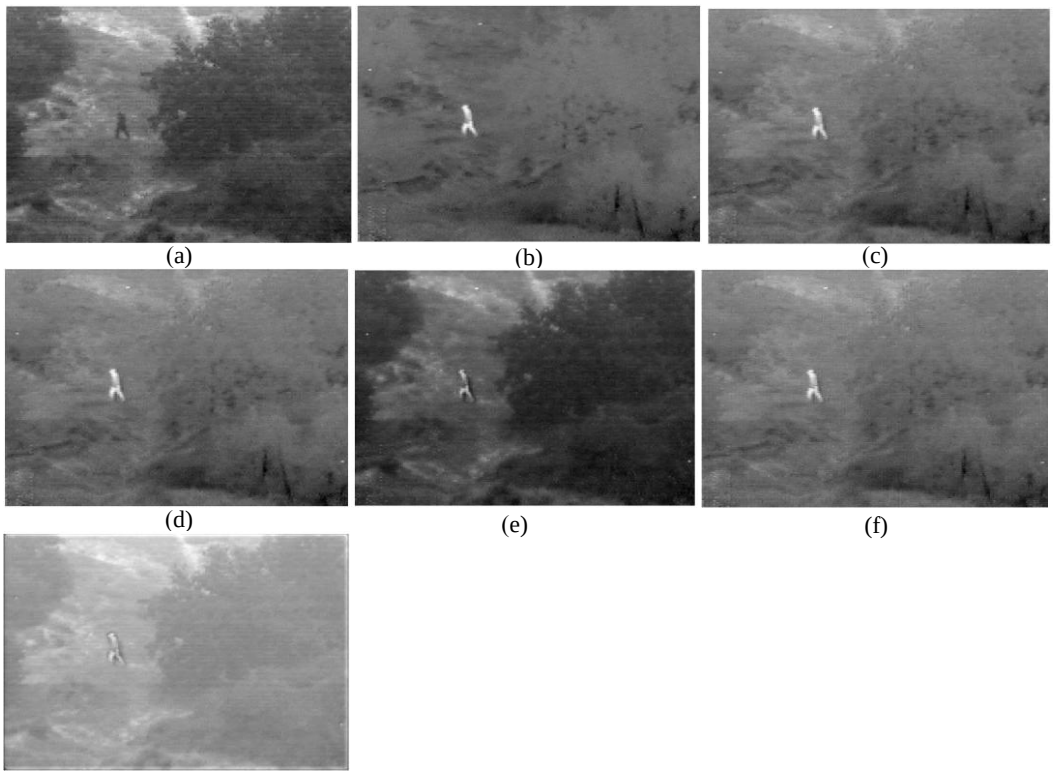


Fig. 8. Input and output of different methods on Trees: (a) visible image (b) infrared image (c) output of PCA (d) Output of averaging method (e) Output of Laplacian pyramid (f) output of wavelet (g) Output of the proposed algorithm.

As observed, the output image provided by the proposed method maintains useful features of the infrared image which is hidden person and also maintains the background of visible image better than other fusion algorithms. Therefore, the proposed algorithm provides more effective fusion compared to the other algorithms.

4.2 Quantitative Comparison

Quantitative comparisons are performed by using two criteria named structural similarity [22] and maintain edge [23,24].

4.2.1 Structural Similarity

Structural similarity is given by:

$$SSIM\ (A,B) = [I(A,B)][C(A,B)][S(A,B)] \quad (9)$$

Where I(A,B), C(A,B) and S(A,B) are intensity, contrast and intensity correlation of A and B images given by:

$$I(A,B) = \frac{2\mu_A\mu_B}{\mu_A^2+\mu_B^2}\ ;\ C(A,B) = \frac{2\sigma_A\sigma_B}{\sigma_A^2+\sigma_B^2};\ S(A,B) = \frac{\sigma_{AB}}{\sigma_A\sigma_B} \quad (10)$$

Where σ is standard deviation and μ is mean of image.

This criterion must have low value for infrared images and high value for visible images because in infrared images the only useful information is the hidden object (the hidden person) whereas all background information related to visible images is reflected to fusion image and therefore the structural similarity between the visible image and final image will be high. The obtained structural similarity of PCA, Averaging, Laplacian pyramid, Wavelet and the proposed method are given in Table 1, Table 2 and Table 3 for Nato Camp, Duine and Trees images, respectively. As observed, the proposed method provides lower value for infrared images and higher values for visible images. This means that the proposed method results in higher performance compared to other methods.

4.2.2 Maintain Edge

Maintain edge is a measure for evaluation of fused algorithms and its value is between zeros to one. The value of zero means that the edge information has been lost whereas the value of one indicates all edge information has completely maintained. Therefore, higher value of maintain edge indicates the fusion algorithm is more appropriate and edge information is maintained.

Table 1. Comparison of similarity structural that is related to Nato Camp input

Fusion algorithms	Sum of structural similarity	structural similarity with visible image	structural similarity with infrared image
PCA	12.8484	1.7531	11.954
Averaging	16.5413	10.4168	6.1245
Laplacian pyramid	16.6348	7.9466	8.6881
Wavelet	16.5742	10.4180	6.1562
The proposed method	**17.3109**	**15.7363**	**1.5742**

Table 2. Comparison structural similarity that is related to Duine input images

Fusion algorithms	Sum of structural similarity	structural similarity with visible image	structural similarity with infrared image
PCA	31.9668	14.4797	17.4871
Averaging	47.9530	17.1848	30.7692
Laplacian pyramid	36.3190	14.8922	21.2468
Wavelet	48.1597	17.2093	30.9504
The proposed method	**17.9357**	**17.2586**	**0.6771**

Table 3. Comparison structural similarity that is related to Trees input images

Fusion algorithms	Sum of structural similarity	structural similarity with visible image	structural similarity with infrared image
PCA	31.9668	14.4797	17.4871
Averaging	31.9884	15.2024	16.7860
Laplacian pyramid	19.7236	15.4598	4.2638
Wavelet	31.9886	15.2176	16.7711
The proposed method	**19.7086**	**16.210**	**3.8675**

As an example, we calculate the maintain edge resulted by averaging with PCA, averaging, Laplacian pyramid, wavelet and the proposed algorithms for three different images: Nato Camp, Duin and Trees. The obtained results are presented in Table 4. As observed, the proposed algorithm provides higher value compared to other methods. This means that the proposed method is more capable to maintain the information of input image edges.

Table 4. Maintain edge of Nato Camp, Duin, Trees

Fusion algorithms	Nato Camp	Duin	Trees
Averaging with PCA	0.3626	0.3137	0.3137
Averaging	0.4153	0.3500	0.3183
Laplacian pyramid	0.2958	0.2706	0.2703
Wavelet	0.4144	0.3500	0.3176
The proposed method	**0.6806**	**0.6955**	**0.5757**

4.2.3 Running Time

A running time of the averaging PCA, averaging, Laplacian pyramid, wavelet and the proposed method has been computed by using MATLAB on the computer (Intel, CPU 2.0, 2GB memory) and has been represented in Table 5. The size of three images is 1280*960. As observed, the running time of the proposed method is lower than other methods.

Table 5. Time duration of Nato Camp, Duin, Trees

Fusion algorithms	Nato Camp	Duin	Trees
Averaging with PCA	287.345 s	279.344 s	291.478 s
Averaging	250.073 s	246.691 s	251.679 s
Laplacian pyramid	257.867 s	249.675 s	259.213 s
Wavelet	256.355 s	247.543 s	258.467 s
The proposed method	**203.578 s**	**245.468 s**	**232.741 s**

5. Conclusions

In this paper, we proposed a fusion algorithm based on infrared and visible images by averaging on gradients of input images. The gradient of image includes high frequency features of the image and therefore low frequency components maintain and prevent blending in output image. By obtaining the optimum weights of the infrared and visible images for fusion, we ensure that useful features are used in final image. The useful features are hidden objects and background for the infrared and visible images, respectively. In order to evaluate the proposed method, the obtained results are compared to averaging by PCA, averaging, Laplacian pyramid and wavelet methods for three different images named: Nato Camp, Duin and Trees. The evaluation is performed by using qualitative and quantitative measures. Visual comparison is used as qualitative measure and numerical comparison including structural similarity and maintain edge is used as quantitative measure.

In visual comparison, the obtain results show that the proposed method maintains the background information (useful information related to visible image) and hidden objects (useful information related to infrared image) more effectively compared to other methods.

In numerical comparison, the resulted structural similarity has lower value for infrared and higher value for visible image by applying the proposed method compared to other methods. This means that the proposed method maintains the useful information for both infrared and visible images in fused image. Also, the obtained results for maintain edge using the proposed method have higher values compared to other methods. This indicates that the proposed method maintains edge information belonging to both infrared and visible images better than other methods. Thus, the obtaining results show that the proposed algorithm results in higher quality in the fused image compared to other methods.

References

[1] H. B. Mitchel, *Image Fusion: Theories, Techniques and Applications*, Springer, 2010.

[2] R.S. Blum, Z. Liu, *Multi- Sensor Image Fusion and Its Applications*, Taylor and Francis Group, New York, 2005.

[3] S. Li, B. Yang, H. Jianwen, "Performance Comparison of Different Multi-Resolution Transforms for Image Fusion", *International Journal on Multi-Sensor, Multi-Source Information Fusion*, Vol. 3, No. 2, pp. 741-746, 2011.

[4] M. Gong, Z. Zhou, J. Ma, "Change Detection in Synthetic Aperture Radar based on Image fusion and Fuzzy Clustering", *IEEE transaction on Image Processing*, Vol.: 21, No. 4, pp. 2141-2151, 2012.

[5] J. Black, "Multi View Image Surveillance and Tracking", *Motion and Video Computing*, Vol. 32, pp. 169-174, 2002.

[6] B. Zitova, J. Flusser, "Image Registration Methods: A Survey ", *Image and Vision Computing*, Vol. 21, pp. 977–1000, 2003.

[7] L. Zagorchev, A. Goshtasby, "A Comparative Study of Transformation Functions for Non-Rigid Image Registration", *IEEE Transactions on Image Processing*, Vol. 15, pp. 529-538, 2006.

[8] R. G. Radke, S. Andra, O. Al-Kofahi, B. Roysam, "Image Change Detection Algorithms: A Systematic Survey", IEEE Trans. Image Process., Vol. 14, No. 3, pp. 294 -307 2005

[9] S. Zheng, "Pixel-Level Image Fusion Algorithms for Multi Camera Imaging System", *M.Sc. thesis, University of Tennessee, Knoxville, United States*, 2010.

[10] R. Wong, F. Bu, H. Jin, L. Li, "A Feature Level Image Fusion Algorithm Based on Neural Network", *International Conference on IEEE Bioinformatics and Biomedical Engineering*, pp. 821-824, 2007.

[11] Y. Zhao, Y. Yin, D. Fu, "Decision Level Fusion of Infrared and Visible Images for Face Detection", *IEEE Conference on Control and Decision*, pp. 2411-2414, 2008.

[12] J. B. Priestly, "Night Vision Devices", *Defense Scientific Information & Documentation Centre*, Vol. 18, No. 1, 1985.

[13] T. Neo, "Fusion of Night Vision and Thermal Images", *M.Sc. Thesis, Science, Naval Postgraduate School Montary, California, United Stats*, 2006.

[14] V. Nadio, and J. Raol, "Pixel-Level Image Fusion Using Wavelet and Principle Component Analysis", *Defense science journal*, Vol. 58, pp. 338-352, 2008.

[15] S. M. Mukane, Y. S. Ghodake, P. S. Khandagle, "Image Enhancement Using Fusion by Wavelet and Laplacian Pyramid", *IJCSI International Journal of Computer Science*, Vol. 10, pp. 1694-0814, 2013.

[16] G. Pajares, and J. Cruz, "A Wavelet-Based Image Fusion Tutorial", *Pattern Recognition*, Vol. 37, pp. 1855 – 1872, 2004.

[17] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing*, Second Edition, Prentice Hall, U.S.A, 1987.

[18] V. Nadio and J. Raol, "Pixel-Level Image Fusion Using Wavelet and Principle Component Analysis", *Defense science journal*, Vol. 8, pp. 338-352, 2008.

[19] S. K. Sadhasivam and et al, "Implementation of Max Principle with PCA in Image Fusion for Surveillance and Navigation", *Electronic letters on Computer Vision and Image Analysis*, Vol. 10, No. 1, pp. 1-10, 2011.

[20] R. Chandrakantha, J. Saibabaa, Geeta Varadana & P. Ananth Raj, "A Novel Image Fusion System for Multi-sensor and Multiband Remote Sensing Data", *IETE Journal of Research*, Vol. 60, No. 2, pp. 168-182, 2014.

[21] G.Pajares and J.M.Cruz, "A Wavelet-based Image Fusion tutorial", *Pattern Recognition*, Vol. 37, pp. 1855-1872, 2004.

[22] Z. Liu, E. Blasch, Z. Xue, J. Zhao, "Objective Assessment of Multi-resolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 1, pp. 94-109, 2012.

[23] V. Petrovic, and C. Xydeas, "Objective Evaluation of Signal-Level Image Fusion performance", *Optical engineering*, Vol. 44, pp. 1-8, 2005.

[24] V. Tsagaris, and V. Anastassopoulos, "Global Measure for Assessing Image Fusion Methods", *Optical Engineering*, Vol. 45, pp. 1-8, 2006.

Mehrnoosh Gholampour received his B.Sc. and M.Sc. degrees from University of Birjand, Birjand, Iran in 2012 and 2014 respectively, all in Communication Engineering. Her research interest and contributions are in image processing, image fusion and signal processing with a focus on image and speech coding. She is also interested in applications signal processing Biomedical with a focus on image fusion.

Hassan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D. in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D. degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as associate professor in communication engineering in department of Electrical and Computer Eng., university of Birjand, Birjand, IRAN.

Sajjad Mohammadzadeh received the B.Sc. degree in electrical engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. degree in communication engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012. He is currently Ph.D. student in Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. His area research interests include Image Processing and retrieval, Pattern recognition, Digital Signal Processing and Sparse representation.

A Persian Fuzzy Plagiarism Detection Approach

Shima Rakian*
Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
shima.rakian@yahoo.com
Faramarz Safi Esfahani
Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
fsafi@iaun.ac.ir
Hamid Rastegari
Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
rastegari@iaun.ac.ir

Received: 05/Dec/2014 Revised: 06/Feb/2015 Accepted: 18/Feb/2015

Abstract

Plagiarism is one of the common problems that is present in all organizations that deal with electronic content. At present, plagiarism detection tools, only detect word by word or exact copy phrases and paraphrasing is often mixed. One of the successful and applicable methods in paraphrasing detection is fuzzy method. In this study, a new fuzzy approach has been proposed to detect external plagiarism in Persian texts which is called Persian Fuzzy Plagiarism Detection (PFPD). The proposed approach compares paraphrased texts with the aim to recognize text similarities. External plagiarism detection, evaluates through a comparison between query document and a document collection. To avoid un-necessary comparisons this tool employs intelligent technology for comparing, suspicious documents, in different levels hierarchically. This method intends to conformed Fuzzy model to Persian language and improves previous methods to evaluate similarity degree between two sentences. Experiments on three corpora TMC, Irandoc and extracted corpus from prozhe.com, are performed to get confidence on proposed method performance. The obtained results showed that using proposed method in candidate documents retrieval, and in evaluating text similarity, increases the precision, recall and F measurement in comparing with one of the best previous fuzzy methods, respectively 22.41, 17.61, and 18.54 percent on the average.

Keywords: Text Retrieval; Plagiarism Detection; External Plagiarism Detection; Text Similarity; Fuzzy Similarity Detection.

1. Introduction

In this article, different types of plagiarism and their detection methods are studied. Also, a method based on fuzzy information retrieval is proposed to detect plagiarism.

Precision and recall are significant performance factors in plagiarism detection system. In this paper, we present an approach to external plagiarism detection in Persian texts, PFPD (Persian Fuzzy Plagiarism Detection). The aim of this tool is to make compatible fuzzy method in Persian language, thus the procedures put forward by [1], would be improved and the precision and recall increased.

The fuzzy statement here, is extracted from previous researches done in this field, and denotes the calculation of similarity between zero to one range.

The precision and recall in the tools for plagiarism detection is very important. The problem with previous systems in intelligent plagiarism detection is the embedding of plagiarized parts in varied sentence structures and synonym replacement [2].

The Language-independent tools may be inefficient for particular languages such as Arabic and Farsi [4].

The main problem is to present an approach to verify plagiarism through an efficient algorithm in order to recognize similarities and improve the precision and recall in obtained results in a timely manner. The other problem is prevention or minimizing the unnecessary repetitious operations [5]. The existing solutions are time-consuming [6].

In the phase of candidate document selections and plagiarism analysis, presented methods do not encompass adequate precision and recall to detect paraphrasing [2].

Therefore, the problem of this research is:

Increasing precision and recall in candidate documents retrieval in Persian language, through hierarchical methods, and in measuring the similarity of a Persian text by using fuzzy-methods.

We believe that the use of semantic methods, based on fuzzy IR for paraphrasing detection establishes more precision and recall, in comparison with other methods [2]. Using the retrieved candidate's documents of hierarchical methods at any phase, the program checks the documents with an increased possibility of plagiarism detection and prevents checking of unimportant documents [4].

For the design of this system, Apache nutch and Apache solr are applied. Nutch used for crawling while solr is for indexing and searching data of candidate documents retrieval. Experiments on three corpora TMC, Irandoc and extracted corpus from prozhe. com, are performed to get confidence on proposed method performance. The experiments denote that using proposed method in candidate documents retrieval, and in evaluating text similarity, increases the precision, recall and F measurement in comparing with one of the best previous fuzzy methods, respectively 22.41, 17.61, and 18.54 percent on the average.

* Corresponding Author

The organization of the paper is as follows. Section 2 describes related works to the paper and their limitations. The approach proposed here is described in Section 3. Section 4 states a description of test collection which is used for evaluation and presents the results of this evaluation. Finally, Conclusions are presented in Section 5.

2. Related Work

Depending on the language of compared texts, plagiarism detection, is classified into two groups as; mono-lingual or cross-lingual (Figure 1).

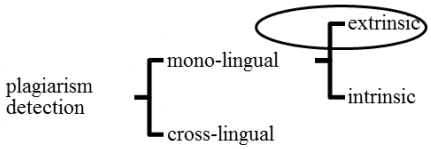


Fig. 1. Plagiarism Detection Techniques

For external plagiarism detection, a suspicious document is compared with one or more other documents. The operational platform for that is briefly outlined below [3]: (Figure 2)

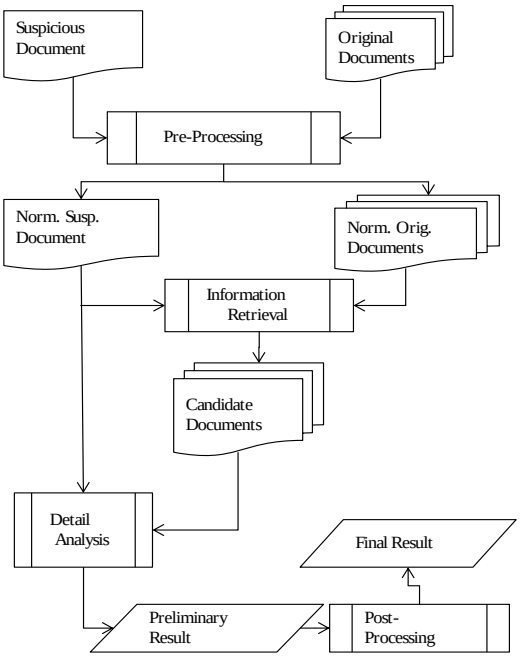


Fig. 2. Methodology of extrinsic plagiarism detection [2,3]

1. The suspicious document and original documents; including the sources that may contain plagiarism which are regarded as input.
2. Three main steps are required:
   a. The retrieval of a list of candidate documents, using the models of information retrieval.
   b. Comprehensive analysis to compare suspicious document with candidate documents.
   c. Post-processing to mix small detected units to be presented to the viewer. At this stage, it should be decided whether plagiarism exists or not.

Meyer zu Eissen, S et. al. [7] discuss the different types of plagiarism. This represented classification in [2], is a more comprehensive, where plagiarism is divided into two groups: i) Literal ii) Intelligent. Fig. 3 illustrates the different types of plagiarism. Hence, the focus of this paper is detecting paraphrased sentences.

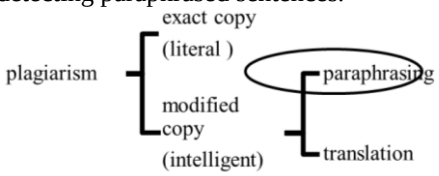


Fig. 3. Types of plagiarism

The two main stages in plagiarism detection are candidate document retrieval and detail analysis, to compare suspicious document with candidate documents. In candidate documents retrieval there is global similarity, but in detail analysis step, local similarity is considered. In local similarity, two documents are checked from a semantic point of view which forms the basis for plagiarism detection [8].

In this section only the most related works are mentioned. In research [9], to obtain semantic similarity, the depth and length of the shortest path to the word, in sequence of Synset in Wordnet, is applied. The structure of Wordnet changes this into a useful tool for a natural language processing.

Table 1. Most Related Work.

	Date	Specification	Advantage	Disadvantage
[5]	2005	Using the term-term correlation matrix in fuzzy information retrieval approach	Able to distinguish synonym replacement and structure change	High expenses to make and keep matrix
[6]	2006	Using the depth and length of the shortest path to the word, in sequence of Synset in Wordnet	Able to distinguish synonym replacement and structure change	Unable to be used in Persian language, because Farsnet is unable to calculate depth and length of the path to the word
[1]	2010	Fuzzy analysis for plagiarism detection	Because of its nature, Fuzzy analysis for paraphrasing detection is more effective than other approaches	Using shingling and Jaccard coefficient in the candidate selection step reduces the precision. In cases where sentences are long and complex, this procedure is of low precision and may not detect plagiarism.
[4]	2012	In candidate documents, documents	Performs the analysis on several levels,	Low precision, due to synonym replacement.

	Date	Specification	Advantage	Disadvantage
		selection, research in sequences of the documents, paragraphs and sentences through intersection rates, was carried out.	to avoid un-necessary comparisons.	To obtain intersection rates in three levels is costly. Using fingerprint in paraphrasing detection, is not suitable
Proposed Method (PFPD)	2014-2015	Fuzzy analysis for plagiarism detection and in candidate documents retrieval	Increasing precision and recall in candidate documents retrieval and in measuring the similarity. Avoid un-necessary comparisons	

In [1], proposed Web based plagiarism detection using fuzzy information retrieval, a mixture of fuzzy similarity model [10] and semantic similarity, obtained from lexical database [9] were employed. Similar to model [9], instead of using term-term correlation matrix, in model [10], the extracted Synset from Wordnet was employed. Thus Semantic observation of the text in which synonym of the word; was extracted using Wordnet database [11] was strongly increased. After that the similarity degree of two sentences was calculated. This system was proposed to detect external plagiarism, capable of paraphrasing detection in English language. Pre-processing of this system was done using different procedures including: tokenization, stemming, Stopword removal and candidate selection with Shingling and Jaccard coefficient. In the analysis phase of plagiarism detection, makes use of fuzzy plagiarism analysis. This system was tested using datasets PAN'09, PAN'10. The advantage of this system is that fuzzy analysis for paraphrasing detection is more effective than other approaches such as Shingling. The disadvantage of this system may be caused by imprecision of shingling and Jaccard coefficient in candidate selection. The goal in this fuzzy analysis for plagiarism detection was to detect paraphrasing, and in the case of long sentences, the already proposed fuzzy analysis proved to be imprecise due to addition of words and other sentences. The proposed tool aims to improve the plagiarism detection method using fuzzy IR.

In [4], APLAG was proposed which was able to detect paraphrasing in the Arabic language. In this system, during the pre-processing phase, tokenization, Stopword removal, stemming and synonym replacement were used and in candidate documents selection, research in hierarchy of the documents, paragraphs and sentences through intersection rates, was done. In the analysis phase of plagiarism detection, this system made use of fingerprint techniques which was tested using three datasets provided by the writer. The advantage of this system is performing the analysis on several levels, to avoid un-necessary comparisons. The disadvantage of this system is its low precision, due to synonym replacement. To obtain intersection rates in three levels is costly. Additionally, the use of fingerprint in paraphrasing detection is not suitable. The idea of intersection blocks in this paper is the result of the aforementioned hierarchal procedure. In Table these approaches and their advantages and disadvantages are highlighted.

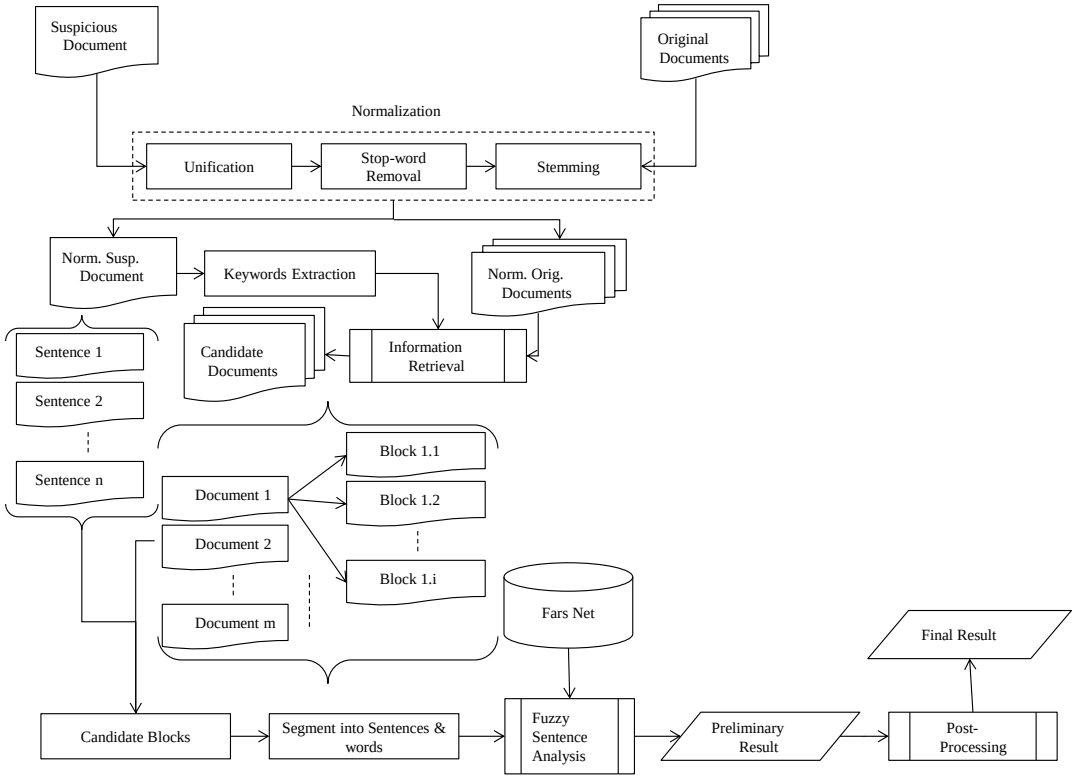


Fig. 4. PFPD Plagiarism Detection Steps

3. PFPD-Persian Fuzzy Plagiarism Detection

In this paper, a tool is proposed to detect external plagiarism in Persian texts, namely Persian Fuzzy Plagiarism Detection (PFPD). This tool is to compare Persian documents, using a fuzzy approach, to recognize similarities.

Compared with other methods, this tool makes use of intelligent technology for comparison of suspicious documents in different levels sequentially, in order to avoid un-necessary comparisons.

In the article, sentences-level representation is used at the analysis phase of plagiarism. Our aim is to adopt fuzzy model in Persian language and improve the methods offered in [1], to calculate the degree of similarity between two sentences. In this method, the degree of similarities between two sentences is calculated within 0-1. If this figure is larger than threshold, the two sentences are considered similar. The threshold is regarded 0.65, the same as [1].

To investigate the effectiveness of this method, three datasets were used. The performance of the system is measured with precision, recall, and F-measure metric.

3.1 PFPD Approach

We use an algorithm for mono-lingual external-paraphrasing detection. This method is particularly useful for recognition of different paraphrasing levels, by using semantic similarity based on fuzzy IR.

Fig. 4 illustrates the operational platform for this method. Input encompasses the suspicious document and source collection; including sources which may have plagiarism. In order to avoid the unnecessary steps in recognition of the plagiarized sentences, keywords from the suspicious texts are used as verifiers of the original documents. Afterwards, certain blocks are chosen in the original document based on the similarity of the sentences of the suspicious document and the blocks in original text.

After pre-processing operations, a list of candidate documents, related to suspicious documents are retrieved and the sentences are checked for plagiarism in details. The method of semantic similarity based on fuzzy IR was used. The fuzzy IR method investigates the suspicious blocks in details. To detect exact copy, 100% similarity should be reached. However, since the system is intended to detect paraphrasing, if the compatibility is more than a certain level, we regard the texts as similar. The detail of the illustrated steps are described below.

3.2 Phase I: Pre-processing

The process begins with eliminating excess data in the original and suspicious documents. The text should be pre-processed, in advance. Pre-processing includes: text unification, Stopword removal, and stemming. Pre-processing is a key stage in obtaining satisfactory results when facing the difficulties in natural language processing. Stopword removal leads to non-sense words. Stemming algorithm is also used to eliminate prefix and postfix to establish word roots. To do that, we produced Aria Package[1], which performs the required operations for Persian language processing.

3.3 Phase II: Candidate Document Retrieval

In this phase keywords in query document and their Synsets are extracted and among original documents, those which include these words, are retrieved. The retrieved and query document are inspected and for each sentence of query document, related blocks with higher intersection are retrieved.

Operations are done before sending a paragraph to the plagiarism detection system to retrieve candidate documents.

Note that using this method, the number of candidate for each suspicious document is dynamic and may be small which could reduce calculation time, while having a fixed number of candidate.

3.4 Phase III: Plagiarism Analysis

After above stage is complete, main operations of analysis are started. In this stage, Candidate blocks with suspicious document are analyzed, sentence by sentence to have a precise study.

At first, blocks are broken by punctuation. At this stage, the values below, for each sentence of the suspicious document and sentence of the blocks of original documents are calculated. The operation unit at this stage is "word". The reason not to use word n-gram, is that in the text paraphrasing, there is the possibility of word order changes. At first similarity of the words in the two sentences (δ) is obtained (1). Synonym words are obtained using Farsnet [12], that is the first Persian Wordnet.

$$\delta = \#\text{Synonym words} * 0.5 + \#\text{exact words} * 1 \quad (1)$$

The intersection ratio of two sentences to the first sentence (α) and the intersection ratio of two sentences to the second sentence are calculated with (2) and (3).

$$\alpha = \frac{\delta}{|S_1|} \quad (2)$$

$$\beta = \frac{\delta}{|S_2|} \quad (3)$$

Intersection to union of two sentence ratio (γ) is calculated with (4).

$$\gamma = \frac{\delta}{(|S_1| + |S_2|) - \delta} \quad (4)$$

Afterwards, the similarity ratio (sim(s1,s2)) is calculated with (5).

$$sim(S_1, S_2) = \frac{A * \min(\alpha, \beta) + B * \max(\alpha, \beta) + C\gamma}{A + B + C} \quad (5)$$

In this similarity metrics order of the words does not affect the degree of final similarity between them. The

[1] www.aria-ware.com

values of coefficients A, B, and C are considered 20, 8, and 3 according to tests.

Finally if $sim(S_1,S_2)$ is greater than the threshold value (T), the sentence is marked as plagiarism, otherwise it is considered to be not plagiarized (6). Following tests, appropriate value for T is obtained to be 0.65, similar to [1].

$$\mathrm{EQ}(S_q, S_x) = \begin{cases} 1 & \text{If } \mathrm{sim}(S_1, S_2) \geq T \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

3.5 Phase IV: Post-processing

Finally, the output of this algorithm is a list of sentences, indicated as similar/plagiarized. Since the sentence is a comparison unit, they are combined in paragraphs.

4. Experimental Evaluation

In text paraphrasing, two sentences may have the same meaning, but different structures; e.g. replacing synonym and adding or shortening the sentences.

For the design of this system, Apache nutch and Apache solr are applied. Nutch used for crawling while solr is for indexing and searching data.

Considering that this method is to improve the method proposed in [1], a comparison between the two was made.

4.1 Test Collections

During the evaluation of the method, we used three different corpora:

1) The Tehran Monolingual Corpus (TMC)[1]: TMC, established by Tehran University, was employed, and after pre-processing, the proposed method was performed. The dataset included the news that was extracted from Hamshahri Corpus and ISNA (Iran students' broadcasting) news agency website. TMC is a large-scale Persian monolingual corpus. From TMC, 1000 documents were achieved as source documents, from which 400 suspicious documents were produced.
2) The IRANDOC Test Collection: Iranian Research Institute for Information Science and Technology (IRANDOC) [2] previously known as Iranian Research Institute for Scientific Information and Documentation is an Iranian research center with a national mission to meet the country's needs in the field of information science and technology. IRANDOC provided 230 documents from which 220 suspicious documents were produced.
3) The prozhe.com Test Collection[3]: The prozhe.com is a website presents student research documents and articles, from which 440 documents are extracted as source documents, and 160 suspicious documents were produced from these.

[1] Available from http://ece.ut.ac.ir/nlp/resources.htm
[2] http://www.irandoc.ac.ir
[3] http://www.prozhe.com

4.2 Query Collections

Two query collections for each corpus, is established. Each collection includes four types of query that is produced artificially.

The four types include the cases below:

1) Synonym replacement of 50% of the words in each sentence.
2) Sentence structure changes with an increase in 45% in the number of words in each sentence.
3) A combination of points 1 and 2, in a sentence.
4) A combination of points 1 and 2, in different sentences of document.

In order to have a query, 50% of the sentences, depending on the type of query (which maybe of the four cases above) are replaced. The difference between first and second series is in the replacement of the second 50% of the document.

In the queries of the first set, 25% of the sentences with exact copy, and 25% as non-copied sentences are replaced (the rest set with ratio of 1 to 1).

In the queries of the second set, 12.5% of the sentences with exact copy, and 37.5% as non-copied sentences are replaced (the rest set with ratio of 1 to 3).

For IRANDOC corpus, one new query was added, which includes 100 paraphrased sentences that are created manually.

The reason for establishing two query collection is to highlight the faults in [1], and ratification of these problems with the present method. Two main fault in [1] are:

1) Lack of precision in the recognition of plagiarism and inability to distinguish paraphrased cases, in which the length of the sentences are increased by adding the words. This phenomena is due to poor of fuzzy formula in [1].
2) Inability to detect plagiarism, due to the approach of weak candidate retrieval. In [1], the Jaccard coefficient is used which is operation heavy. In addition, when the plagiarized text of less volume in comparison with original documents, the Jaccard coefficient is less than 0.1, that document is not selected as candidate to follow other stages. For example, assume of a thesis with 5 chapters, just one is copied, considering the above mentioned problem, research [1] is not able to detect that copy.

The first query collection, was used to highlight the fault of the fuzzy method applied in [1] and a solution was in proposed method. The second query collection was used to represent the fault of second case.

The description of applied corpora and their test results are presented in the following section.

4.3 Evaluation

Each sentence of the suspicious document is compared with original documents. Assuming S_1 to have a bigger sentence length and S_2 as the second sentence length, then $1 \leq S_2 \leq S_1$. In Fig. 5 and Table similarity values are presented for $9 \leq S_1 \leq 12$.

Obtained results indicated that, when lengths S_1 and S2 are close to each other, results of PFPD and method [1] are closely in compatibility. But when S_1 and S_2 are far from each other and small sentence is extracted from the bigger sentence (i.e. more than 85% of bigger sentence), PFPD is able to detect plagiarism, while the method outlined in [1] is not able to detect these cases.

As you see in Table , there is only one row in $9 \leq S_1 \leq 12$, which is detected by the method [1], but the present method is not able to detect that. In which $\gamma \leq 0.5$, indicating ratio of intersection on union is less than 0.5, and the method [1] detected that wrongly.

Fig. 6 illustrates the percent of precision, recall and overlap calculated correctly, For $1 \leq S_1 \leq 30$, in PFPD, and the method [1].

Studying the results indicated that the average amount of precision and recall in PFPD respectively are 100%, 99.97%, but in [1] are 99.10%, 85.05% respectively. Additionally, the PFPD is able to detect 99.97% of the correct cases detected in [1] on average.

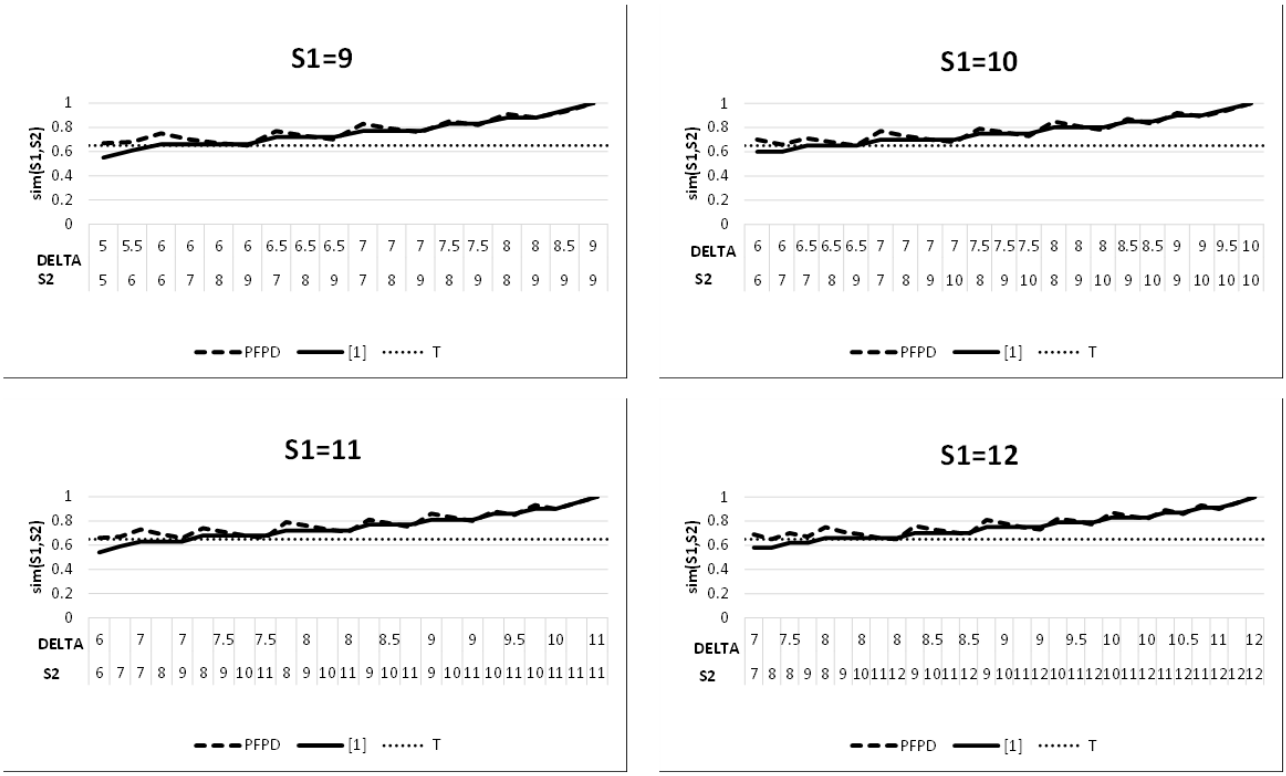


Fig. 5. Similarity ratio calculated for $9 \leq S1 \leq 12$ and $PFPDSim(S1,S2) \geq T$ when T=0.65

Table 2. similarity ratio calculated in $9 \leq S1 \leq 12$ for the cases where ($PFPDSim(S1,S2) \geq T$ or $[1]Sim(S1,S2) < T$) and ($PFPDSim(S1,S2) < T$ or $[1]Sim(S1,S2) \geq T$)

S_1	S_2	□	□	□	□	PFPD	[1]
9	5	5	0.55	1	0.55	0.67	0.55
	6	5.5	0.61	0.91	0.57	0.68	0.61
10	6	6	0.6	1	0.6	0.7	0.6
	7	6	0.6	0.85	0.54	0.66	0.6
	10	6.5	0.65	0.65	0.48	0.63	0.65
11	6	6	0.54	1	0.54	0.66	0.54
	7	6.5	0.59	0.92	0.56	0.67	0.59
	7	7	0.63	1	0.63	0.73	0.63
	8	7	0.63	0.87	0.58	0.69	0.63
	9	7	0.63	0.77	0.53	0.66	0.63
12	7	7	0.58	1	0.58	0.69	0.58
	8	7	0.58	0.87	0.53	0.65	0.58
	8	7.5	0.62	0.93	0.6	0.7	0.62
	9	7.5	0.62	0.83	0.55	0.67	0.62

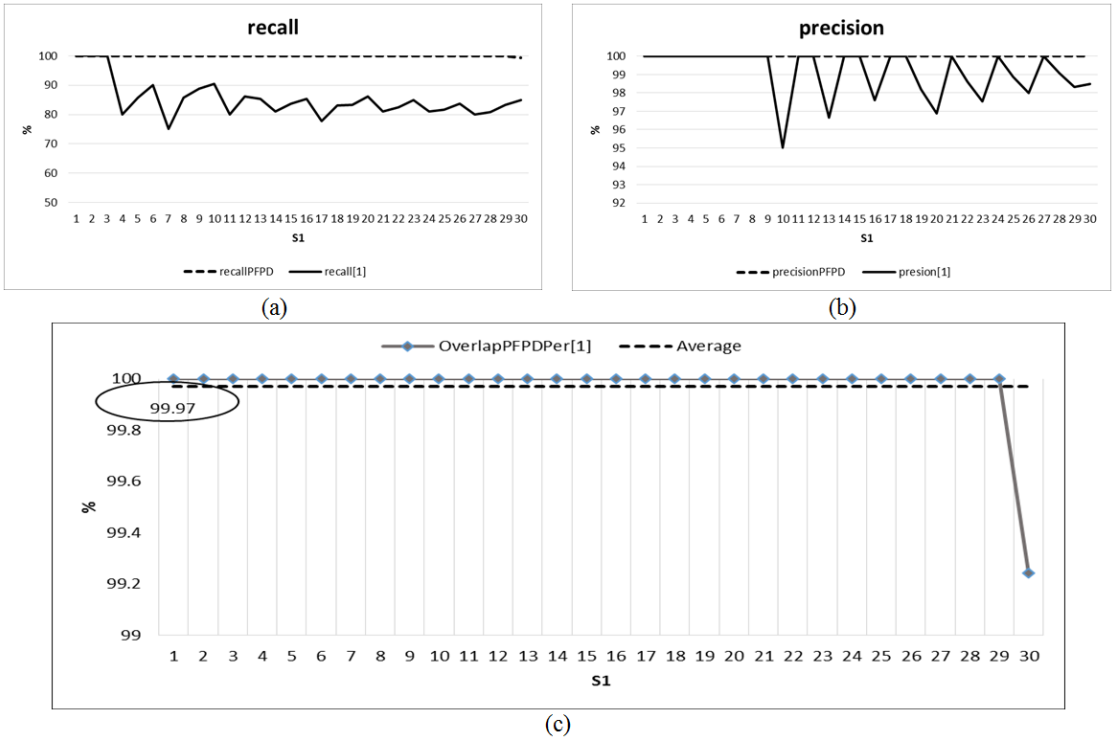


Fig. 6. (a) Recall (b) Precision (c) overlap PFPD per [1] for 1≤S1≤30

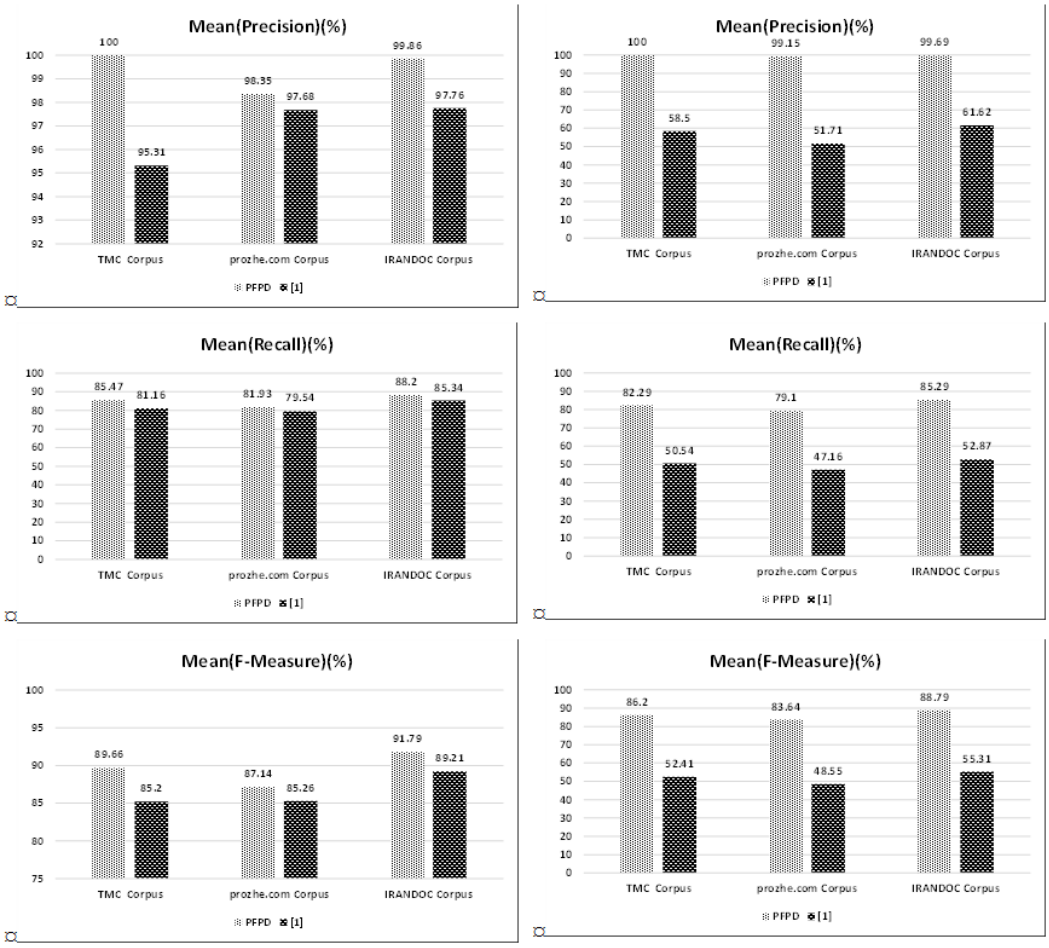


Fig. 7. Precision, Recall and F-Measurement (a) first experiment series (b) second experiment series.

The three datasets described in Section 4.1, and their queries were input in to the system. The results are presented in Fig. 7.

Based on these results, all cases picked up by method [1], could be also picked up by PFPD.

In the first experiment series used for test fuzzy formula, in four types of query increase of precision, recall and F-measurement are obtained respectively 2.49, 3.19 and 2.97. In the second experiment series which were used for test candidate document retrieval stage, it was found that precision and recall in method [1] is strongly dependent on volume of the copied document. If it is of low value in the candidate retrieval stage, this document would not be selected, and this method would not progress to the next stage and the plagiarized text would not be detected. PFPD method could obtained a high degree of precision and recall by improvement to the candidate retrieval stage. Increase of precision, recall and F-measurement are obtained respectively 42.34, 32.04 and 34.12.

Regarding the first and second experiments, increase of the precision, recall and F measurement are obtained respectively 22.41, 17.61, and 18.54 percent on the average.

4.4 Time Complexity

According to the investigations done, the number of source documents have no effect on precision and recall. Because according to the suggested algorithm, all the source documents including one of the keywords in suspected document, are selected and checked as candidate for next stage. Therefore the number of these documents is of no effect on the precision and recall. But influencing on algorithm speed which are examined in the following.

The experiments were done on a HP-Pavilion dv4 laptop. In these experiments, the volume of the input document was 3 KB. At first the source documents were 50, and in each experiment, 50 documents were added. The time of each experiment was measured. Fig. 8 demonstrates the comparison between PFPD and [1] in terms of time required for Plagiarism Detection. The results from the comparison show that the proposed method achieved better results in terms of time required for Detection.

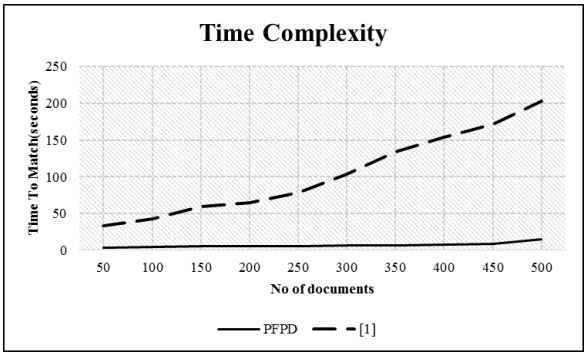


Fig. 8. Time Complexity

The method [1] investigates the whole content of the candidate documents, to obtain Jaccard coefficient. While the suggested method compares suspicious documents, in different levels hierarchically. This leads to reduction of the operations and increase of speed.

5. Conclusions

To identify paraphrasing based on sentences, fuzzy method is effective, as it has the capacity to distinguish similar sentences, based on the similarity among a set of synonym words. In this article, PFPD to detect external mono-lingual plagiarism was performed. This semantic fuzzy method is designed to detect different degrees of paraphrasing.

The obtained results showed that using proposed method in candidate documents retrieval, and in evaluating text similarity, increases the precision, recall and F measurement in comparing with one of the best previous fuzzy methods, respectively 22.41, 17.61, and 18.54 percent on the average. Also the results from the comparison show that the proposed method achieved better results in terms of time required for Detection.

Previous presented fuzzy methods are unable to distinguish paraphrased cases, in which the length of the sentences are increased by adding the words. While PFPD is able to detect such cases, it is also capable of picking up cases where the length of the plagiarized sentences are close to each other.

References

[1] S. M. Alzahrani and N. Salim, "Fuzzy Semantic-Based String Similarity for Extrinsic Plagiarism Detection," Braschler and Harman, 2010.

[2] S. M. Alzahrani, et al., "Understanding Plagiarism linguistic patterns, textual features, and detection methods," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 42, pp. 133-149, 2012.

[3] B. Stein, et al., "Strategies for retrieving plagiarized documents," in Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, 2007, pp. 825-826.

[4] M. E. B. Menai, "Detection of Plagiarism in Arabic Documents," International Journal of Information Technology, vol. 4, 2012.

[5] A. H. Osman, et al., "CONCEPTUAL SIMILARITY AND GRAPH-BASED METHOD FOR PLAGIARISM

DETECTION," Journal of Theoretical and Applied Information Technology, vol. 32, pp. 135-145, 2011.
[6] D. Ceglarek and K. Haniewicz, "Fast Plagiarism Detection by Sentence Hashing," in Artificial Intelligence and Soft Computing, 2012, pp. 30-37.
[7] S. Meyer zu Eissen, et al., "Plagiarism detection without reference collections," Advances in data analysis, pp. 359-366, 2007.
[8] S. M. Alzahrani and N. Salim, "Plagiarism detection techniques," 2008.
[9] Y. Li, et al., "Sentence similarity based on semantic nets and corpus statistics," Knowledge and Data Engineering, IEEE Transactions on, vol. 18, pp. 1138-1150, 2006.
[10] R. Yerra and Y. K. Ng, "A sentence-based copy detection approach for web documents," Fuzzy systems and knowledge discovery, pp. 481-482, 2005.
[11] G. A. Miller, "WordNet: a lexical database for English," Communications of the ACM, vol. 38, pp. 39-41, 1995.
[12] M. Shamsfard, et al., "Semi automatic development of farsnet; the persian wordnet," in Proceedings of 5th Global WordNet Conference, 2010.

Shima Rakian was born in Iran. She has recently received her master's degree in software engineering from Islamic Azad University, Najafabad Branch, Iran. Her current research interests include Data Mining and Text Retrieval.

Faramarz Safi Esfahani was born in Iran. He got his Ph.D. in Intelligent Computing from University of Putra Malaysia in 2011. He is currently on faculty at Department of Computer Engineering, Islamic Azad University, Najafabad branch, Iran, since 2002. His research interests include intelligent computing, Cloud Computing, Autonomic Computing, and Bio-inspired Computing.

Hamid Rastegari was born in Iran. He got his Ph.D. in Computer Science – Soft Computing from University of UTM Malaysia in 2011. He is currently Assistant Professor on faculty at Department of Computer Engineering, Islamic Azad University, Najafabad branch, Iran. Experiences 1-Title: Head of Computer Department- Organization: University of Applied Science and Technology from 2002 to 2007 2-Title: Head of Postgraduate Department- Organization: IAUN from 2013 to Pres. 3-Title: Coordinator of 1st National Conference on Computer Science ament- Organization: IAUN from 2013 to 2013.

Simultaneous Methods of Image Registration and Super-Resolution Using Analytical Combinational Jacobian Matrix

Hossein Rezayi*
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
hrezayi2001@yahoo.com
Seyed Alireza Seyedin
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
seyedin@um.ac.ir

Received: 13/Mar/2015 Revised: 05/Aug/2015 Accepted: 09/Aug/2015

Abstract

In this paper we propose two simultaneous image registration (IR) and super-resolution (SR) methods using a novel approach in calculating the Jacobian matrix. SR is the process of fusing several low resolution (LR) images to reconstruct a high resolution (HR) image; however, as an inverse problem, it consists of three principal operations of warping, blurring and down-sampling that should be applied sequentially to the desired HR image to produce the existing LR images. Unlike the previous methods, we neither calculate the Jacobian matrix numerically nor derive it by treating the three principal operations separately. We develop a new approach to derive the Jacobian matrix analytically from the combination of the three principal operations. In this approach, a Gaussian kernel (as it is more realistic in a wide range of applications) is considered for blurring, which can be adaptively resized for each LR image. The main intended method is established by applying the aforementioned ideas to the joint methods, a class of simultaneous iterative methods in which the incremental values for both registration parameters and HR image are obtained by solving one linear system of equations per iteration. Our second proposed method is formed by applying these ideas to the alternating minimization (AM) methods, a class of simultaneous iterative methods in which the incremental values of registration parameters are obtained after calculating the HR image at each iteration. The results show that our proposed joint and AM methods are superior to the recently proposed methods such as Tian's joint and Hardie's AM methods respectively.

Keywords: Super-Resolution; Image Registration; Jacobian Matrix; Combinational Coefficient Matrix; Joint Methods.

1. Introduction

Super-resolution (SR) refers to a series of techniques that integrate the information of a low resolution (LR) image sequence captured from a scene to produce a high resolution (HR) image with higher quality. Each LR image frame is required to have partial unique information or it will not have any positive effect on the final HR image. This partial unique information can be obtained in a number of ways such as camera/scene movement or zooming. There are various SR techniques which have been developed and reviewed in the literature including [1]-[4]. Generally SR techniques include three phases either implicitly or explicitly: 1) image registration (IR), 2) image interpolation, and 3) image deblurring and denoising [1]. In a small category of the techniques such as interpolation-based methods, these phases are performed separately [5]-[9]. To overcome the intensive presence of error propagation in these methods, a majority of other SR techniques attempt to perform the last two phases in an integrative phase called image reconstruction. However, an important source of error is the inaccuracy of registration parameters. Therefore, to further prevent the error propagation, a large group of SR techniques have been developed recently that deal with the inaccuracy of registration parameters. Some of these techniques utilize median estimator to reduce the artifacts caused by errors and outliers of registration parameters [10], [11]. Some others use Bayesian methods in which the unknowns (including registration parameters) are treated as stochastic variables [12]-[14]. In Tipping's method [12], marginalization is applied to HR image but in Pickup's method [13], it is applied to registration and blurring parameters. Nevertheless, the latter provides a wide range of various priors (regularizations) to select, but in both of them, IR and image reconstruction are implemented in relatively separate steps without persistent interaction. It seems that such values are not reliable enough. Babacan [14] extends the Pickup's method to consider hyperparameters (such as the regularization parameter) as stochastic variables, and as an alternating minimization (AM) method, establishes persistent interaction between the estimate of the reconstructed HR image, registration parameters and hyperparameters. AM methods are a class of iterative SR techniques in which the HR image and registration parameters are improved in two consecutive steps at each iteration [15]. A group of AM methods use expectation-maximization (EM) to estimate the HR image (in the expectation phase) and registration parameters (in the maximization phase) iteratively

* Corresponding Author

[16],[17]. The AM methods, nevertheless, may lead to suboptimal solutions [18].

There is another category of SR techniques which are also iterative similar to AM methods, but in this category, HR image and registration parameters are not calculated separately at each iteration [19]-[22]. A nonlinear cost function was used by Chung et al. [19] to estimate HR image and registration parameters. Using Euler–Lagrange necessary conditions for the cost function, they derived a nonlinear system of equations, proposing three methods for its solution. Their first method (called decoupled) resembled an AM method, but their second method (called partially coupled) was a kind of variable projection (VP) method [18]. A similar method was also proposed by Robinson et al. [20] where they used a similar nonlinear cost function to derive the maximum likelihood (ML)/maximum a posteriori (MAP) solution for the HR image. After substituting this solution of HR image into the cost function, the reduced cost function [19] was minimized with respect to the reminder of unknowns, i.e. registration parameters. Finally, these registration parameters were used to obtain the final HR image. Chung et al. [19] attempted to solve the nonlinear system of equations through Gauss-Newton algorithm in their third method. This led to the development of a new class of methods called fully coupled. In these techniques, which are referred to as joint methods in this paper, the incremental values of the HR image and registration parameters are jointly calculated in only one system of equations. He et al. [21] used a similar cost function by linearizing it at existing current values for HR image and registration parameters using Taylor series approximation. In this linearization, they obtained the Jacobian matrix analytically (in contrast to some methods like [12], [13] where it is calculated numerically). Finally, this linear system of equations (in terms of incremental values of HR image and registration parameters) has been solved through a conjugate gradient (CG) optimization algorithm. They have used Euclidean motion model and it has been extended to the similarity motion model by Tian et al. [22].

In all of these joint methods [19]-[22], the blurring is assumed to be the same for all LR images, which is impractical in many applications. In reality, when the motion model is more complex than Euclidean motion model, the size of blurring function is no longer the same for all LR images [23]. This has not been considered in the [19] and [22] methods, which have not restricted their motion model to the Euclidean motion model. However, in the methods proposed in [21] and [22], the convergence to the global solution is more probable (especially when the initial values are close enough to optimal values) but the derivation of Jacobian matrix is based on bilinear interpolation of warped pixels [19], [21], [22]. This may introduce the restriction in which only four neighboring pixels are effective in determining the values of warped pixels [21].

As an inverse problem, SR dictates three principal operations that should be applied sequentially to the original HR image to produce the corresponding LR images. These operations are image warping, image blurring and image down-sampling (for brevity they are referred to as principal operations). In many approaches, the principal operations are treated separately and in others they are combined into a unit operation [2], [12], [13], [23]. The advantages of this combination include the possibility of incorporating the pixels in any arbitrary neighboring radius to derive the warped pixels without changing the framework of the problem and employing a new interpolation method. Another advantage of combining the principal operations is error propagation reduction because all three operations are performed in one stage. Moreover, it allows having an adaptive kernel for blurring (blurring is treated as a function of zooming, which can be different for each LR image).

This paper focuses on the joint methods [21], [22] and treats the three principal operations in the inverse problem in a combinational form as [12] and [13]. Then, a new joint method based on this combinational form is proposed. In contrast to [12] and [13], the Jacobian matrix is not calculated numerically and unlike common joint methods [21], [22], the Jacobian matrix is not derived by treating the three principal operations separately. In the proposed method, the bilinear interpolation has not been used in the warping operation and its derivative. Moreover, the same blurring for all LR images [19]-[22] has not been considered. We developed a new approach to derive the Jacobian matrix analytically based on the combinational form of the three principal operations. In this regard, a Gaussian kernel blur (as is more realistic) was adopted the radius of which is adaptive to each LR image. We also used a bilateral total variation (BTV) regularization [11], which incorporates the eight directions of each pixel in the cost function. Although the main goal of this paper is to develop a new joint method of IR and SR using the combinational coefficient matrix and analytical combinational Jacobian matrix, the application of these concepts to the framework of AM methods gives rise to new method which will be discussed later. In this paper, the similarity motion model is used (which consists of translation, rotation and zooming) similar to [22].

The rest of this paper is organized as follows. In Section 2, the problem formulation including notation of SR problem, Gaussian kernel blur and combinational coefficient matrix are introduced. The proposed iterative joint and AM methods are developed in Section 3. In Section 4, experimental results on simulated and real images are presented. Conclusion and future works are discussed in Section 5.

2. Problem Formulation

2.1 Super Resolution Notations

Let us consider a series of K discrete LR images g_k of size $M_g \times N_g$ where $1 \leq k \leq K$. The lexicographically

ordered LR images are denoted by column vectors g_k and all these vectors are stacked in one column vector, i.e. $g = [g_1^T, \dots, g_K^T]^T$. The purpose of the SR technique is to reconstruct the original HR image f of the size $M_f \times N_f$ using existing LR images as well as some prior information about the original HR image. The lexicographically ordered original HR image is presented by column vector f. Here, it is assumed that the decimation factor is the same in both vertical and horizontal directions ($\rho = M_f/M_g = N_f/N_g$). Each LR image g_k is obtained by applying the three principal operations to the original HR image f as follows (generative model):

$$g_k = DH_kB(a_k)f + n_k \quad (1)$$

Where n_k is the column vector of additive white Gaussian noise, D is the down-sampling operator (which is realized as a $M_gN_g \times M_fN_f$ matrix), H_k is the blurring operator (which is realized as a $M_fN_f \times M_fN_f$ matrix) and $B(a_k)$ is the warping operator (which is realized as a $M_fN_f \times M_fN_f$ matrix) [14], [21], [22]. a_k is the vector of unknown registration parameters used for warping the grid of original HR image f (called reference grid) onto the up-scaled grid of *k* th LR image. Practically one of the existing LR images (the first LR image in this paper) is selected as the reference image and the up-scaled grid of the reference image is considered as the reference grid. Generally, in all simultaneous IR and SR methods like AM and joint methods, it is assumed that initial and imprecision values of registration parameters can be provided by some IR techniques. In this paper, enhanced correlation coefficient (ECC) method [24] is used for the IR.

The combination of the three principal operations can be considered as a unit combinational operation, which is realized by a matrix $W_k(a_k)$ of the size $M_gN_g \times M_fN_f$ as follows:

$$g_k = W_k(a_k)f + n_k \quad (2)$$

This paper adopts the similarity motion model with four degrees of freedom (zooming, rotation, vertical and horizontal translation). Hence, $a_k = [h_{1k}, \dots, h_{4k}]$ where $h_{1k}, \dots, h_{4k}$ are the variable elements of the kth 3x3 homogenous matrix [25] with the current motion model, as follows:

$$\mathbf{M}_k = \mathbf{M}(\boldsymbol{\alpha}_k) = \begin{bmatrix} h_{1k} & -h_{2k} & h_{3k} \\ h_{2k} & h_{1k} & h_{4k} \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

It is worth noting that the above homogenous matrix can be decomposed into pure translation, rotation and zooming matrices respectively as $M_k = M^T(u_k)M^R(\theta_k)M^Z(z_k)$ [25] in which $u_k = [u_{xk} \dots u_{yk}]^T = [h_{3k} \dots h_{4k}]^T$, $\theta_k = tan^{-1}(h_{2k}/h_{1k})$ and $z_k = (h_{1k}^2 + h_{2k}^2)^{0.5}$.

Equation (2) shows a series of relations between the original HR image and each LR image. These relations can be written as one equation as follows:

$$g = Wf + n \quad (4)$$

where $W = [W_1(a_1)^T, \dots, W_k(a_k)^T]^T$ is called combinational coefficient matrix and $n = [n_1^T, \dots, n_K^T]^T$.

2.2 Blur Considerations

Consistent with the literature, in this paper Gaussian kernel, which is more realistic, has been used to model the blur caused by the atmosphere turbulence and camera lens, and the motion blur has been excluded. Usually, the blur is assumed to be isotropic in the imaging plane [23]. When the motion model is similarity, the kernel of back-projected blur into the scene plane will be isotropic too. However, the greater the distance of a scene plane from the image plane (or less zoom is applied) the more extensive is the area encompassed in the scene to contribute to the blurring. Therefore, when the LR images are registered to the reference image, they have isotropic Gaussian blur, but possibly with different radiuses in the reference image [23].

2.3 Combinational Coefficient Matrix

Three methods of computing the combinational coefficient matrix has been addressed by Capel [23]. The simplest one is the directly computing the warping, blurring and down-sampling matrices separately and then calculating their multiplication (separately applying the principal operations) as (2) has been derived from (1). However this method involves large amount of memory usage. Additionally, since the operations are implemented separately the error propagation can occur. Although in the Capel's second and third methods the principal operations are not implemented separately, they have complex implementations. Another method has been proposed in [12] which is theoretically very similar to the Capel's third method but it has simpler implementation. This method which has also been used in some other papers (e.g. [13]) will be described in the reminder of this section and applied in this paper.

The elements of *i*th row of the $W_k(a_k)$ are the coefficients of linear combination of f required to generate the gray scale value of *i*th pixel of g_k. Thus, the sum of these coefficients should be equal to one. As such, the elements of $W_k(a_k)$ are calculated as:

$$W_{ijk}(\boldsymbol{\alpha}_k) = \frac{\widetilde{W}_{ijk}(\boldsymbol{\alpha}_k)}{\sum_{j'=1}^{M_fN_f} \widetilde{W}_{ij'k}(\boldsymbol{\alpha}_k)} \quad (5)$$

where the elements $\widetilde{W}_{ijk}(a_k)$ are obtained as [12]:

$$\widetilde{W}_{ijk}(a_k) = \exp\left\{-\frac{1}{2\sigma^2|M_k|}(v_j - s_k(u_i)^T(v_j - s_k(u_i))\right\} \quad (6)$$

Here $v_j = [v_{xj} \;\; v_{yj}]^T$ is the position of *j*th pixel of the original HR image in the reference grid, $u_i = [u_{xi} \;\; u_{yi}]^T$ is the position of *i*th pixel of *k*th LR image and $s_k(u_i) = [s_{xk}(u_i) \;\; s_{yk}(u_i)]^T$ is its transformation through the motion model, which is characterized by a_k with respect to the reference grid:

$$s_{xk}(\mathbf{u}_i) = h_{1k}u_{xi} - h_{2k}u_{yi} + h_{k3} \quad (7)$$

$$s_{yk}(u_i) = h_{2k}u_{xi} + h_{1k}u_{yi} + h_{4k} \quad (8)$$

Actually, $s_k(u_i)$ is the center of isotropic Gaussian kernel after projection to the reference grid. σ is the standard deviation of isotropic Gaussian kernel of the blur and $|\bullet|$ is the determinant operator.

It is worth noting that in the combinational generative models used by Capel [23] and Tipping et.al [12], Gaussian interpolation method [26] has been used implicitly instead of bilinear interpolation method, which is explicitly common in several SR techniques, e.g. [14], [19], [21] and [22]. This is shown in Fig.1 in which both bilinear and Gaussian interpolation methods have been displayed from a new aspect. As shown in this schematic view, the reverse mapping is adopted in the image warping implementation. Hence, instead of warping the HR pixels, the LR pixels are warped in the reference grid. Then, the neighboring HR pixels/samples around a warped LR pixel (the center of kernel in Fig.1) are assigned some weights equal to the height of the point in which the kernel and the neighboring sample intersect. As shown in Fig.1 (a), in the bilinear interpolation, at most four samples can be included in the domain of the square pyramid kernel, but there is not such limitation in the Gaussian interpolation (Fig.1 (b)).

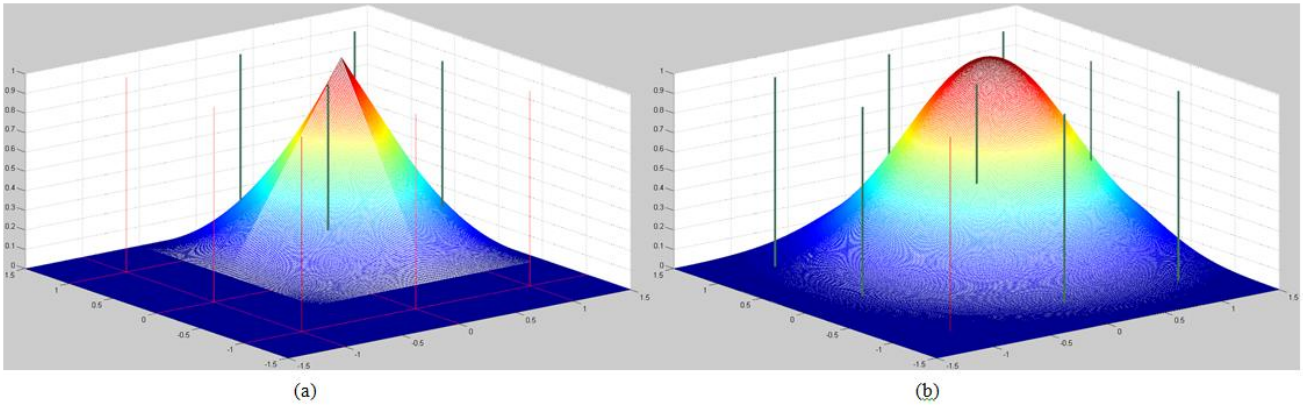


Fig. 1. Two kernels for interpolation; thick vertical lines show the samples located in the domain of the kernel. (a) Square pyramid kernel used in the bilinear interpolation, (b) Gaussian kernel used in the Gaussian interpolation.

3. Development of New Iterative Simultaneous Methods

3.1 Cost Function of the New Joint Method

As super resolution is an ill-posed problem, there are infinite or instable solutions which can satisfy Equation (4). Therefore, to make the solution unique and stable, prior information is necessary. In the joint methods, both original HR image and registration parameters are unknown. TV of the HR image is an important regularization which is often used in SR techniques as the prior information. For the registration parameters, a simple Tikhonov regularization, i.e. the minimum energy, has been used. Given the generative model of LR images expressed in (4) and considering the mentioned regularizations, the framework of the cost function for the joint method is [19], [21], [22]

$$E(f,a) = \|g - Wf\|^2 + \lambda T_V(f) + \beta R(a) \quad (9)$$

where $a = [a_1^T, \ldots, a_K^T]^T$, λ and β are regularization parameters, $T_V(f)$ is TV of the HR image. Similar to Farsiu's [11] and Tian's [22] methods, BTV is used in our cost function, but here it encompasses all eight neighboring pixels to reduce edge penalization in any directions

$$T_V(f) = \sum_{\substack{m=-1 \\ |m|+|n|\neq 0}}^{1} \sum_{n=-1}^{1} \sum_{ij} \frac{|f(i+m, j+n) - f(i,j)|}{\sqrt{m^2+n^2}} \quad (10)$$

Because $T_V(f)$ is a nonlinear function of f using half-quadratic scheme [27] and fixed-point techniques [28], it can be written as:

$$T_V(f) = \sum_{\substack{m=-1 \\ |m|+|n|\neq 0}}^{1} \sum_{n=-1}^{1} \sum_{ij} \frac{|f(i+m, j+n) - f(i,j)|^2}{\mu_{m,n}} \quad (11)$$

Where $\mu_{m,n}$ is calculated in the previous iteration as follows:

$$\mu_{m,n} = \sqrt{(m^2+n^2)\{(f(i+m, j+n) - f(i,j))^2 + \gamma} \quad (12)$$

and γ is the small positive value to ensure $\mu_{m,n}$ is nonzero. $T_V(f)$ can be expressed in a matrix-vector form as [22]:

$$T_V(f) = f^{\mathrm{T}} L^{\mathrm{T}} L f = f^{\mathrm{T}} T f = L f^2 \quad (13)$$

$R(a) = \|a - \bar{a}\|^2$ is the Tikhonov regularization for registration parameters and $\bar{a}$ is a vector containing the average values of registration parameters during all previous iterations. This method of $\bar{a}$ selection can help reduce

resonance around the final values of registration parameters. Hence, the cost function (9) can be expressed as:

$$E(f,a)=\left\|[r(a,f)^{\mathrm{T}} \quad \sqrt{\lambda}(Lf)^{\mathrm{T}} \quad \sqrt{\beta}(a-\bar{a})^{\mathrm{T}}]^{\mathrm{T}}\right\|^2 \qquad (14)$$

where $r(a,f)=g-Wf$ is called residual vector.

To estimate the unknown HR image and registration parameters, the cost function should be minimized. Although this optimization problem is convex with respect to f, it is nonconvex in terms of a because the term $r(a,f)$ is nonlinear with respect to a. To alleviate this problem, linear approximation has been used for $r(a,f)$. Linear approximation requires initial values for unknowns. As mentioned earlier, initial values for registration parameters may be obtained using registration techniques such as ECC algorithm [24]. Given these initial values for registration parameters, the initial value for HR image can be obtained using the simple interpolation-based method [5], the approach used by [21] or a simple up-scaling of the reference image. If Δf and Δa are incremental values for f and a then $r(a+\Delta a, f+\Delta f)$ can be approximately linearized with respect to Δf and Δa as follows:

$$r(a+\Delta a,\mathrm{f}+\Delta f)\approx r(a,\mathrm{f})+\begin{bmatrix}\frac{\partial r(a,\mathrm{f})}{\partial a} & \frac{\partial r(a,\mathrm{f})}{\partial f}\end{bmatrix}\begin{bmatrix}\Delta a\\ \Delta f\end{bmatrix} \qquad (15)$$

The two derivative terms in the above equation are calculated in the two following relations:

$$\frac{\partial r(a,\mathrm{f})}{\partial a}=\frac{\partial(g-Wf)}{\partial a}=-\frac{\partial Wf}{\partial a}=-J(a,f) \qquad (16)$$

$$\frac{\partial r(a,\mathrm{f})}{\partial f}=\frac{\partial(g-Wf)}{\partial f}=-W \qquad (17)$$

Substituting (16) and (17) in (15) yields

$$r(a+\Delta a,\mathrm{f}+\Delta f)\approx r(a,\mathrm{f})-J(a,\mathrm{f})\Delta a-W\Delta f \qquad (18)$$

Where $J(a,f)$ is the Jacobian matrix, the calculation of which presents a challenging issue, as will be discussed in the next section (3.2.).

Substituting (18) in (14) and rewriting the content of the norm as a linear combination of incremental values of unknowns leads to

$$E(f+\Delta f,a+\Delta a)=\left\|\begin{matrix}-r(a,f)+J(a,f)\Delta a+W\Delta f\\ \sqrt{\lambda}Lf+\sqrt{\lambda}L\Delta f\\ \sqrt{\beta}(a-\bar{a})+\sqrt{\beta}(\Delta a)\end{matrix}\right\|^2 \qquad (19)$$

$$=\left\|\begin{pmatrix}J(a,f) & W\\ 0 & \sqrt{\lambda}L\\ \sqrt{\beta}I & 0\end{pmatrix}\begin{pmatrix}\Delta a\\ \Delta f\end{pmatrix}+\begin{pmatrix}-r(a,f)\\ \sqrt{\lambda}Lf\\ \sqrt{\beta}(a-\bar{a})\end{pmatrix}\right\|^2$$

Now, instead of minimizing the nonlinear cost function in (14) with respect to the main unknowns, i.e. f and a, the linear cost function in (19) is minimized with respect to their incremental values, i.e. Δf and Δa respectively [19],[21]. The incremental values are used to update the main unknowns. Obtaining the optimal solution for the cost function in (14) is not guaranteed, but starting from initial values close to optimal values and continuing the optimization of the linear cost function in (19) may yield the optimal solution [21].

3.2 The proposed Method for Derivation of Jacobian Matrix $J(a,f)$

Jacobian matrix in the existing joint methods is derived from the three principal operations separately [19], [21], [22]. In this paper, however, we propose the derivation of Jacobian matrix using the combinational operation (5), and in contrast to [13], it is calculated analytically rather than numerically. As mentioned in (16) it can be written as follows:

$$J(a,f)=\frac{\partial Wf}{\partial a}=\frac{\partial Wf}{\partial[h_{11},...,h_{41},h_{12},...,h_{4K}]} \qquad (20)$$

$J(a,f)$ is a $KM_gN_g\times 4K$ block diagonal matrix [21] of matrices $J(a,f)$ where

$$J(a_k,f)=\frac{\partial W_k(a_k)f}{\partial a_k}=\frac{\partial W_k(a_k)f}{\partial[h_{1k},...,h_{4k}]} \qquad (21)$$

Naturally $J(a_k,f)$ is a $M_gN_g\times 4$ matrix and the nth column of which can be obtained as follows:

$$J_n(a_k,f)=\frac{\partial W_k(a_k)f}{\partial h_{nk}}=\frac{\partial W_k(a_k)}{\partial h_{nk}}f \qquad (22)$$

where $1\le n\le 4$. Using equations (5) and (6), each entry of matrix $\partial W_k(a_k)/\partial h_{nk}$ is calculated as follows:

$$\frac{\partial W_{ijk}(a_k)}{\partial h_{nk}}=\frac{\sum_{j'=1}^{M_fN_f}(\nabla_n Q_{ijk}-\nabla_n Q_{ij'k})\tilde{W}_{ij'k}(a_k)}{\sum_{j'=1}^{M_fN_f}\tilde{W}_{ij'k}(a_k)}W_{ijk}(a_k) \qquad (23)$$

where the operator ∇_n is derivative with respect to h_{nk} and

$$Q_{ijk}=-\frac{1}{2\sigma^2|M_k|}(v_j-S_k(u_i))^{\mathrm{T}}(v_j-S_k(u_i)) \qquad (24)$$

therefore

$$\nabla_n Q_{ijk}=\frac{1}{\sigma^2|M_k|}\nabla_n S_k(u_i)^{\mathrm{T}}(v_j-S_k(u_i))-\frac{\nabla_n|M_k|^{-1}}{2\sigma^2}(v_j-S_k(u_i))^{\mathrm{T}}(v_j-S_k(u_i)) \qquad (25)$$

Where $\nabla_n|M_k|^{-1}$ and $\nabla_n S_k(u_i)$ can be calculated simply using Equations (3), (7) and (8).

The proposed method can be implemented in the framework of an algorithm as described in the next subsection:

3.3 Algorithmic Steps of the Proposed Joint Method

Step 1: Calculating initial values for registration parameters ($\hat{a}$) using ECC algorithm.

Step 2: Calculating the initial value of HR image using Delaunay triangulation-based interpolation method of SR, given the LR images and the estimated $\hat{a}$.

Step 3: (At the iteration i) Calculating the combinational coefficient matrix $W(a)$, Jacobian Matrix $J(a,f)$, residual vector $r(a,f)$ and BTV matrix T as discussed earlier.

Step 4: Solving the linear system of equations (26) using a CG algorithm:

$$\begin{bmatrix} J^{\mathrm{T}}J+\beta I & J^{\mathrm{T}}W \\ W^{\mathrm{T}}J & W^{\mathrm{T}}W+\lambda T \end{bmatrix}\begin{pmatrix} \Delta a \\ \Delta f \end{pmatrix}=\begin{bmatrix} J^{\mathrm{T}}r-\beta(a-\bar{a}) \\ W^{\mathrm{T}}r-\lambda Tf \end{bmatrix} \tag{26}$$

This linear system of equations is the result of minimizing the cost function expressed in (19), which is calculated by taking derivative of the cost function with respect to Δf and Δa and then equating the result to zero.

Step 5: Updating the unknown variables using the estimated incremental values:

$$\begin{pmatrix} a \\ f \end{pmatrix}^{i+1}=\begin{pmatrix} \Delta a \\ \Delta f \end{pmatrix}+\begin{pmatrix} a \\ f \end{pmatrix}^{i} \tag{27}$$

Where the superscript shows the iteration number.

Step 6: Updating $\bar{a}$ according to the following relation:

$$\bar{a}^{i+1}=\frac{i\bar{a}^{i}+a^{i+1}}{i+1} \tag{28}$$

Step 7: If the following condition of HR image, (29), is satisfied for a specified threshold (Thr), which is assumed 10-6 here, or a maximum number of iteration is reached, Then stop Else go to step 3.

$$\frac{\left\|f^{i}-f^{i-1}\right\|^{2}}{\left\|f^{i-1}\right\|^{2}}<\mathrm{Thr} \tag{29}$$

3.4 Cost Function of the New AM Method

In the AM methods, the cost function is minimized in two separate phases. In the first phase, it is minimized with respect to f. Fortunately, the cost function (14) is quadratic with respect to f and its MAP estimation is obtained as follows [15]:

$$f(a)=\left(W^{\mathrm{T}}W+\lambda T\right)^{-1}W^{\mathrm{T}}g \tag{30}$$

In the second phase, the cost function, which is not quadratic with respect to a, is minimized as a function of a. For this reason, the cost function is linearized only with respect to incremental values of registration parameters Δa using (19)

$$E(f,a+\Delta a)=\left\|\begin{pmatrix} J(a,f) \\ 0 \\ \sqrt{\beta}I \end{pmatrix}\Delta a+\begin{pmatrix} -r(a,f) \\ \sqrt{\lambda}Lf \\ \sqrt{\beta}(a-\bar{a}) \end{pmatrix}\right\|^{2} \tag{31}$$

Where $J(a,f)$ is obtained as discussed in Section 3.2, using proposed method. Finally, the vector Δa is calculated by solving the following linear equation:

$$\left(J^{\mathrm{T}}J+\beta I\right)\Delta a=J^{\mathrm{T}}r-\beta(a-\bar{a}) \tag{32}$$

Other details, the same as the proposed iterative joint method in section 3.3, have not been included here.

This proposed AM method, like other AM methods, is more sensitive to the initial values compared to the joint method.

3.5 Algorithmic Steps of the Proposed AM Method

Step 1: Calculating initial values for registration parameters ($\hat{a}$) using ECC algorithm and then calculating the combinational coefficient matrix $W(a)$.

Step 2: Calculating the initial value of HR image using Delaunay triangulation-based interpolation method of SR based on the LR images and the estimated $\hat{a}$.

Step 3: (At the iteration i) Calculating the Jacobian Matrix $J(a,f)$ and the residual vector $r(a,f)$.

Step 4: Solving the linear system of equations (32) using a CG algorithm.

Step 5: Updating the registration parameters using the estimated incremental values:

$$(a)^{i+1}=(\Delta a)+(a)^{i} \tag{33}$$

Step 6: Updating $\bar{a}$ using Equation (27).

Step 7: Calculating the combinational coefficient matrix $W(a)$ and BTV matrix T.

Step 8: Solving the linear system of equations (30) to obtain a new update of the HR image f.

Step 9: If the condition (29) is satisfied or a maximum number of iteration is reached, then stop Else go to step 3.

3.6 Implementation Details

To realize the proposed methods and compare them with other methods, MATLAB R2012b is used. The ECC library provided by Evangelidis [29] is used to obtain initial values for registration parameters. The preconditioned conjugate gradient (PCG) routine "pcg" is used where a linear system of equations should be solved. We partially used (just subroutine "makeW.c") the codes provided by Pickup [30], making some modifications to produce the desired combinational coefficient matrix. The Gaussian kernel used in this subroutine is truncated after three standard deviations in terms of LR pixels. This has a significant impact on reducing the computational cost of Equations (5) and (23), as will be discussed in the next subsection 3.7. The above-mentioned truncation is also considered in the blur mask produced by the routine "fspecial", which is used to perform the blurring operation in Tian's and Hardie's methods. Other parameters like λ and β are taken as [22].

The original Hardie's method proposed for translational motion model and direct search in neighboring pixels was used in the optimization of registration parameters. To draw a fair comparison, we extended this method to include similarity motion model, and the nonlinear least square optimization method used in the proposed AM method was also applied to the Hardie's method because a motion model more complex than translational was not applicable to original Hardie's method.

In the methods that perform the warping and blurring operations separately [21], [22], after warping the HR image, some blank areas like corners usually appear. These blank regions may be filled by black pixels or nearest pixels [21]. When the motion is considerable, selecting each of these tricks may negatively affect the final results because the blur mask can significantly change the gray level of the pixels located in the borders of these blank regions. To overcome this problem, we used another trick. When using the reverse mapping model, the pixels that leave the frame are marked during the warping operation. This process does not require considerable additional computation since it is a part of warping operation. Then, the columns of blurring matrix H_k associated to these pixels are set to zero and the rows of the matrix are renormalized. These operations are all conducted through matrix calculations. This trick has been used in our implementations in the next section to increase the performance of Tian's and Hardie's methods.

3.7 Computational Complexity

In this subsection, the computational complexity of the proposed methods is calculated and compared with two other simultaneous methods (Tian's [22] and Hardie's [15] methods). The overall computational cost of the proposed methods (as well as Tian's and Hardie's methods) is affected by several factors including the dimensions of LR images, the number LR images, the increasing factor, the size of blur mask and the termination criteria of algorithm (the number of iterations allowed in the algorithm and the pcg routine). Similar to [31] and [22] it is assumed that the addition and multiplication operations are the same. In all methods, the sparsity of matrices is considered in the computational complexity. There are three processes that their complexity dominates others: the calculation of combinational coefficient matrix $W(a)$, the calculation of Jacobian Matrix $J(a,f)$ and solving the linear system of equations (26) or (30). Since the frameworks of the methods are the same, the complexity is calculated only for one iteratio n. In Tian's and Hardie's methods $W(a)$ is calculated in accordance with (1) [22] by separately calculating its component matrices, i.e. D, H_k and b(a_k). Because the motion model is similarity and the bilinear interpolation is used, the computational cost of $B = blkdiag\{B(a_1), \dots, B(a_K)\}$ is $O(8kM_fN_f)$. The blur matrix $H = blkdiag\{H_1, \dots, H_K\}$ has M_hN_h non-zero elements in each row and the multiplication of HB can be considered as the convolution of M_hN_h blur mask (since the blur mask is truncated) with a 2×2 interpolation mask. Hence, the computational cost of this multiplication is O(4($M_h + 1)(N_h + 1)KM_fN_f)$. Finally since $O(8KM_fN_f) < O(4(M_h + 1)(N_h + 1)KM_fN_f)$ the computational cost of $W(a)$ is $O\big(4KM_hN_hM_fN_f\big) = O(K\rho^2M_hN_hM_gN_g)$. In these two methods $J(a,f) = blkdiag\{J(a_1,f), \dots, J(a_{K-1},f)\}$ is derived by separate calculation of its component matrices, i.e. $J(a_k,f) = \mathrm{DH}_k(\partial B(a_k)/\partial a_k) = \mathrm{DH}_k\mathrm{E}_kC$ where C is a $2M_fN_f \times 4$ matrix that requires no considerable calculation and E_k is a $M_fN_f \times 2M_fN_f$ matrix that consists of two adjacent diagonal matrices (see [21] or [22] for further details). The computational cost of E_k is $O(4M_fN_f)$ and its multiplication with C yields the computational cost of $O(8M_fN_f)$. Since C is not sparse, the new matrix E_KC will not be sparse too, hence the computational cost of H_KE_KC will be $O(4M_hN_hM_fN_f)$. Finally, the computational cost of $J(a,f)$ will be $O(K\rho^2M_hN_hM_gN_g)$. For Hardie's AM method, the HR image is obtained in each iteration by solving the linear system of equation (30) using the iterative PCG algorithm. The computational cost of the PCG for one iteration is equal to the number non-zero elements of the coefficient matrix of the linear system of equations [32]. Although the dimensions and the number of non-zero elements of the sparse matrix $W(a)$ are dependent on the number of LR images and the increasing factor, this is not the case for the coefficient matrix of the linear system of equations (30), i.e. $W^TW + \lambda T$. The number of non-zero elements of this sparse matrix is approximately equal to $(M_f(2M_h + 1) - M_h^2 - M_h)(N_f(2N_h + 1) - N_h^2 - N_h)$. Hence, if the iteration number allowed for PCG is L, the computational cost of PCG algorithm will be $O(L\rho^2M_hN_hM_gN_g)$. For the Tian's joint method, the larger linear system of equation (26) should be solved. As mentioned earlier, since $J(a_k,f)$ is not sparse, W^TJ will not be sparse too. Hence, the computational cost of PCG algorithm for solving linear system of equations (26) is $O((M_hN_h + 2K)L\rho^2M_gN_g)$. In the proposed methods, the combinational coefficient matrix $W(a)$ is calculated in accordance with relations (5)-(8). Each row of this matrix, has $(M_h + 1) \times (N_h + 1)$ elements (as before) and requires $3(M_h + 1)(N_h + 1) + 7$ multiplications and $(M_h + 1)(N_h + 1)$ exponentiations. Since, at most three standard deviations of a Gaussian is preserved, its argument will have a maximum absolute value of 3^2/2=4.5. Experimentally, it was verified that such an exponent has a computational cost which is at most 15 times greater than the computational cost of a multiplication. Hence, the computational cost of $W(a)$ is $O\Big(\big(3(M_h + 1)(N_h + 1) + 7 + 15(M_h + 1)(N_h + 1)\big)KM_gN_g\Big) = O(KM_hN_hM_gN_g)$. To calculate the computational complexity of the combinational Jacobian matrix $J(a,f)$ in the proposed methods, we found that the computational cost of $J(a_k,f)$ was 4 times greater than $J_n(a_k,f)$ because the motion model is similarity. According to (22) $J_n(a_k,f)$ is the multiplication of matrix $\partial W_k(a_k)/\partial h_{nk}$ with vector f. Each row of $\partial W_k(a_k)/\partial h_{nk}$ has the same number of non-zero elements as $W_k(a_k)$, i.e. $(M_h + 1)(N_h + 1)$, hence this matrix-vector product will require $(M_h + 1)(N_h + 1)M_gN_g$ multiplications. According to (23)-(25), each element of $\partial W_k(a_k)/\partial h_{nk}$ requires $(M_h + 1)(N_h + 1) + 8$ multiplications. Finally, the computational cost of $J(a,f)$ is $O\big(4(K-1)(M_h + 1)(N_h + 1)M_gN_g + 4(K-1)(M_h + 1)(N_h + 1) + \big((M_h + 1)(N_h + 1) + 8\big)M_gN_g\big) = O(KM_h^2N_h^2M_gN_{g)}$.

Now, the summery of computational complexity of the four methods are as follows:

A. Extended Hardie's AM method:

$Max\ (O(K\rho^2 M_h N_h M_g N_g)\ ,\ O(L\rho^2 M_h N_h M_g N_g)) = O(Max(K,L)\rho^2 M_h N_h M_g N_g)$

B. The proposed AM method:

$Max\ (O(KM_h^2 N_h^2 M_g N_g)\ ,\ O(L\rho^2 M_h N_h M_g N_g)) = O(Max(KM_h N_h, L\rho^2) M_h N_h M_g N_g)$

C. Tian's joint method:

$Max\ (O(K\rho^2 M_h N_h M_g N_g)\ , O((M_h N_h + 2K)L\rho^2 M_g N_g)) = O(Max(KM_h N_h, LM_h N_h + 2KL)\rho^2 M_g N_g)$

D. The proposed joint method:

Max $(O(KM_h^2 N_h^2 M_g N_g)\ , O((M_h N_h + 2K)L\rho^2 M_g N_g)) = O(Max(KM_h^2 N_h^2, (M_h N_h + 2K)L\rho^2) M_g N_g).$

Generally, it can be concluded that joint methods have higher computational cost than AM methods. Practically, when a higher increasing factor is selected, the larger blur mask is desirable so that $M_h N_h$ is approximately proportional to ρ^2 and then $O(KM_h^2 N_h^2 M_g N_g) \approx O(K\rho^2 M_h N_h M_g N_g)$. Hence, the proposed AM and joint methods will have the same computational cost as Hardie's AM and Tian's joint methods respectively. It should be noted that the computational complexity derived here is more detailed than the one used in [22].

4. Experimental Results

In this section, the performance of the proposed Joint and AM methods are discussed and compared with a recently joint method (Tian's method [22]) and the famous AM method (Hardie's method [15]). We have provided five experiments including three synthetic image sequences and two real-life images sequences. The synthetic sequences, produced by warping some test images, are used to evaluate the performance of methods according to the following metrics: normalized mean square error (NMSE) for the estimated registration parameters vector and the peak signal to noise ratio (PSNR) for the reconstructed HR image, which are defined as follows [21],[22]:

$$\mathrm{NMSE}(\tilde{a}) = 100\frac{\|a-\tilde{a}\|^2}{\|a\|^2} \tag{34}$$

$$\mathrm{PSNR}(\tilde{f}) = 10\log_{10}\left(\frac{M_f N_f}{\|f-\tilde{f}\|^2}\right) \tag{35}$$

where "~" denotes the currently estimated values of the unknowns. Lower values of NMSE and higher values of PSNR are preferred in a method. Apart from the two objective measures, the final HR images are used to compare the performance of the methods subjectively.

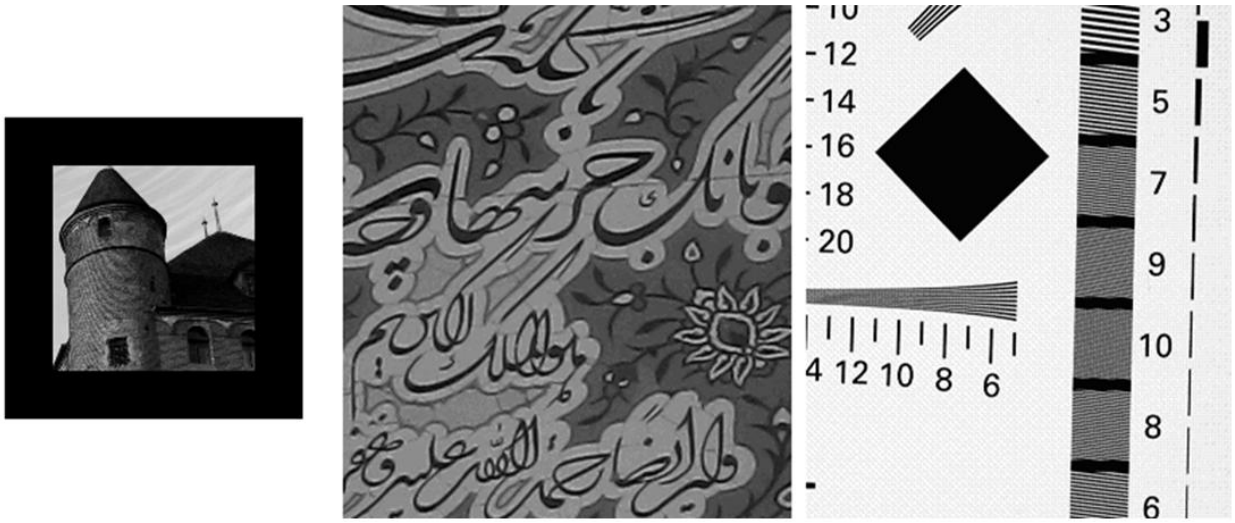


Fig. 2. Three test images, 'Castle', 'Khayam' and 'Chart' which are considered as the ground truth in the reconstruction.

Fig. 3. Four LR images of the sequence were created from the first test image.

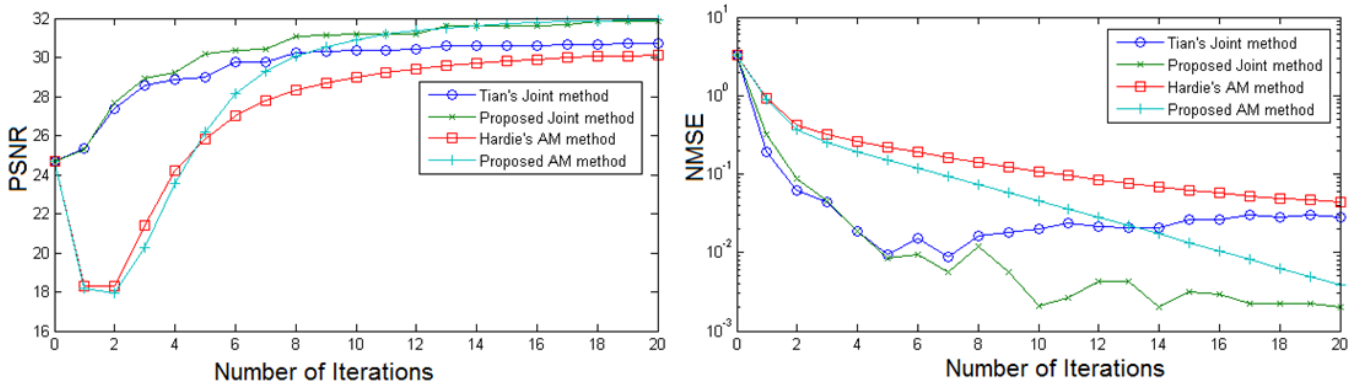


Fig. 4. PSNR of the reconstructed HR image and NMSE of the estimated registration parameters using the four methods for the 'Castle' image.

4.1 Experimental Results of Degraded Test Images

Three test images are used for the experiments in this subsection. The first one (188x186 pixels) is 'Caslte' image [33], which has been placed at the center of a frame surrounded by a thick strip of 30 black pixels (zero gray value). By doing so, the main part of the image is not even pushed out of the frame partially after warping. The second and third test images, namely 'Khayam' and 'Chart' images [33], are relatively large (580x640 and 512x640 pixels respectively). After warping these test images, the central part of images (320x280 pixels), which are still in the frame, are selected as warped HR images. Therefore, the original HR images (the ground truth) be reconstructed are shown in Fig. 2. In the first experiment, a sequence of six images were created by warping one of test images through the use of different homogenous matrices (M_k) each of which containing a zooming factor (z_k), a rotation angle (θ_k) and a translation in both vertical and horizontal directions (v_{xk}, v_{yk}), randomly chosen from the ranges $[0.95,1.05]$, $[-5°, 5°]$ and $[-3,3]$ respectively. Then, they were blurred by an isotropic Gaussian kernel with a variance of 0.333 LR pixel. These images, degraded by Additive White Gaussian Noise (AWGN) to have 30 dB SNR, were down-sampled by a decimation factor of 2. Four degraded LR images for the first test image are shown in Fig.3. Using four methods, i.e. Tian's method, Hardie's method and our two proposed methods, the HR image was reconstructed. To obtain initial values for registration parameters, no explicit IR techniques were used in the experiments of this subsection; instead, the inverse of each original homogeneous matrix M_k was multiplied by a homogeneous error matrix. This homogenous error matrix, similar to M_k, contains a zooming factor, a rotation angle and a translation in both vertical and horizontal direction randomly chosen from the ranges $[0.995,1.005]$, $[-0.5°, 0.5°]$ and $[-0.5,0.5]$ respectively. The PSNR of the reconstructed HR images and NMSE of estimated registration parameters using these four methods for the 'Castle' image are shown in Fig.4. Also, the HR images reconstructed by these methods for the test image are shown in Fig.5. As can be seen, the proposed methods have improved NMSE and PSNR more than Tian's and Hardie's methods. Additionally, the superior performance of the proposed methods is evident in the reconstructed HR images highlighted in Fig.5 (e-h). As can be seen, the PSNR is reduced in the AM methods at the first iteration, while the reconstructed HR image deviates from the desired solution, though it returns to the desired solution in the later iterations. We have also examined other initial

values of HR image such as up-scaled version of reference image and MAP estimation of the HR image using the initial values of registration parameters [21]. However, it had no significant impact on the performance of joint methods, but it could affect the performance of the AM methods. This experiment was repeated with a homogeneous error matrix that contained twice the values of zooming factor, rotation angle and translation in both vertical and horizontal directions. Nevertheless, it was observed that the joint methods converged during 20 iterations, with most of the attempts to perform the AM methods leading to divergence. In the second experiment, in which the second and third test images were used, we focused on the zooming part of the motion model. A sequence of six images with zooming factors $z_k = 1 + k/6$ for $k = 0,2,3,4,5,6$ and vertical translations $v_{xk} = -26, -39, -53, -66, -78$ and horizontal translations $v_{yk} = -23, -35, -46, -57, -67$ were created. Vertical and horizontal translations were selected in such a way that the center of the test image remained unchanged after warping. Other steps were the same as the first experiment. Four degraded LR images for the 'Chart' image are shown in Fig.6. The PSNR of the reconstructed HR images and the NMSE of the estimated registration parameters for the test image are shown in Fig.7 and HR images reconstructed by the four methods are illustrated in Fig.8. The results show that our proposed methods can reconstruct the edge regions with greater precision, especially when the relative zooming is considerable.

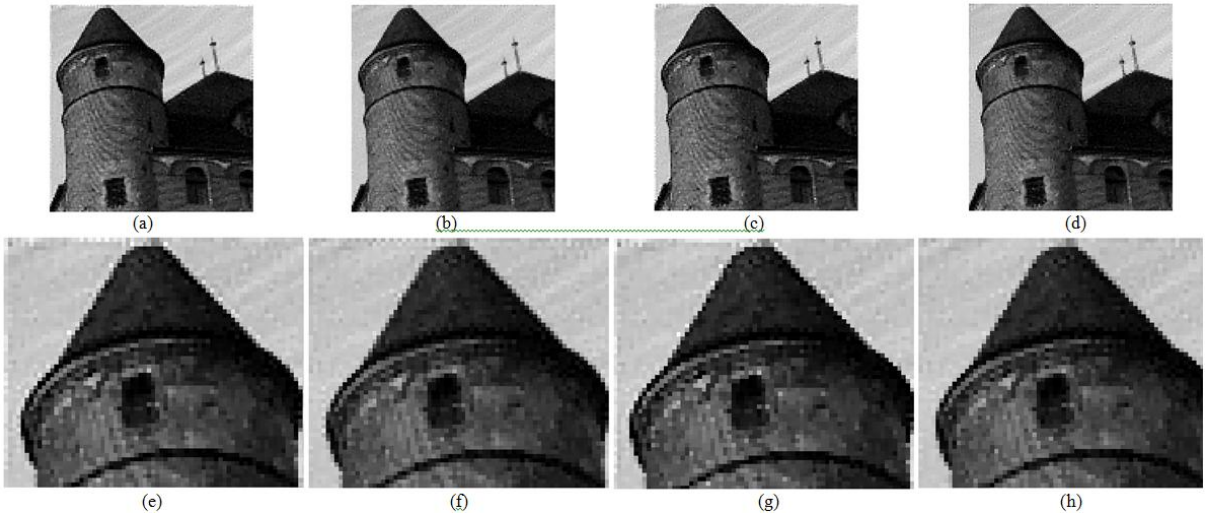


Fig. 5. the reconstructed HR images using Tian's Joint method (a), proposed Joint method (b), Hardie's AM method (c) and proposed AM method (d). A zoomed part of these images have also been shown in (e-h) respectively.

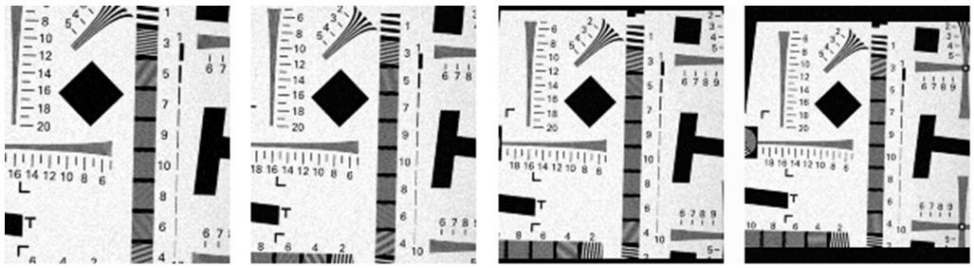


Fig. 6. A sequence of four degraded LR images was created from the third test image by emphasizing the zooming.

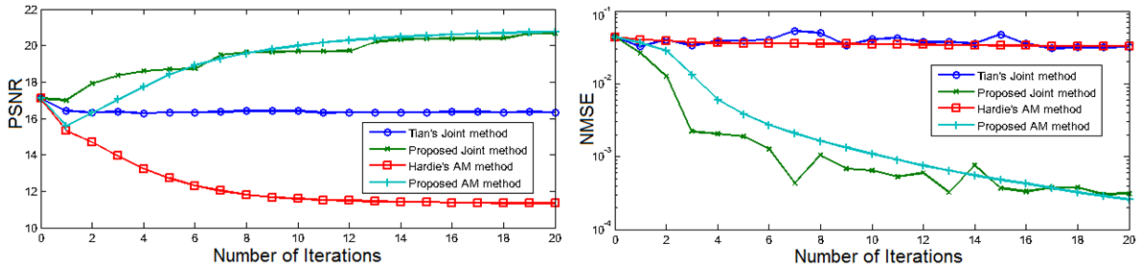


Fig. 7. PSNR of the reconstructed HR images and the NMSE of the estimated registration parameters using all four methods for the 'Chart' image.

4.2 Experimental Results of Real-life Images

In this subsection, two sequences were extracted from two videos of 'text' and 'car', which are accessible in [34]. Four frames of each video are shown in Fig. 9. In the first experiment, a sequence of 10 frames with a size of 57x49 pixels was extracted from the 'text' video. Also, σ was set at 0.55 LR pixel and the increasing factor was set at 4. Initial values for registration parameters were obtained by setting the homogeneous matrix M_k equal to the identity matrix for all frames, because the relative motion between successive frames was small in this experiment. This can be considered as a special application of these SR methods. The HR images reconstructed by the four methods are displayed in Fig.10. The results of this experiment also confirm the superior performance of our proposed methods in the edges. Also, it is observed that letters in reconstructed images are clearly distinguishable in the proposed method. Selecting a larger standard deviation increases the shadows around the letters, especially in Tian's and Hardie's methods. In the second experiment, a sequence of 16 frames with a size of 121x72 pixels was extracted from 'car' video. σ was set at 0.45 LR pixel and the increasing factor was set at 3. Initial values for registration parameters were obtained by adopting the ECC IR technique [24]. The HR images reconstructed by the four methods are shown in Fig. 11.

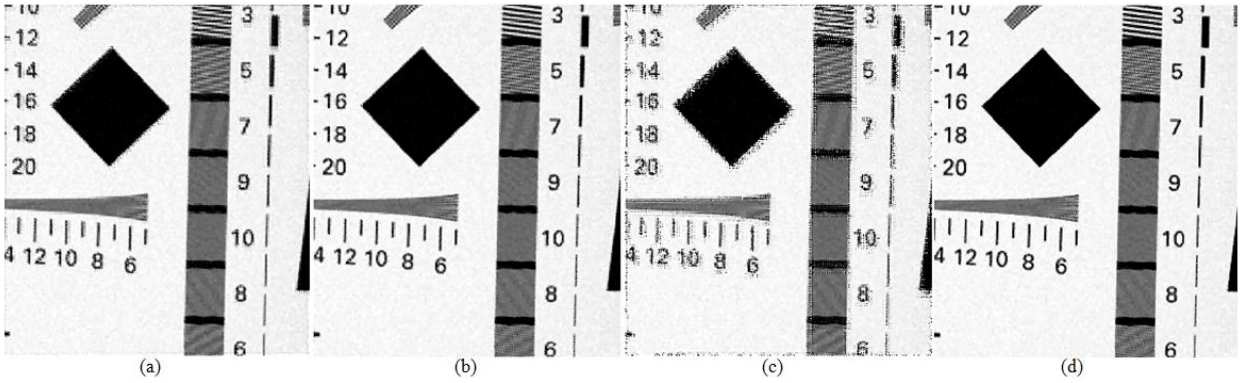


Fig. 8. Reconstructed HR images using Tian's joint method (a), our proposed joint method (b), Hardie's AM method (c) and our proposed AM method (d).

Fig. 9. Four LR images of the sequences were extracted from videos 'text' and 'car'

Our proposed joint method produced better results in this experiment. Although both license plates, namely '3PLK273' and the name of vehicle manufacturer 'SUBARU', are recognizable, the letters of license plate are more distinctive in the proposed method. Our proposed AM method detected the license plate letters precisely, though some shadows in edge regions could be observed in other parts of image.

We used a DELL/Vostro notebook with 4 GB RAM and a 2.5 GHz dual core processor in our experiments. The maximum iteration number for both pcg routine and the four discussed simultaneous methods was set at 20. At each iteration, however, the proposed AM and joint methods had the same computational complexity as Hardie's and Tian's methods respectively, the run time of the proposed joint and AM methods were less than Tian's and Hardie's methods respectively. For example, the run times of the first experiment for 6 LR images with a size of 94x93 pixels were 64s, 40.5s, 52.5s and 24s for Tian's method, the proposed joint, Hardie's method and the proposed AM methods respectively. This is due to three reasons. First of all, the proposed methods require fewer iterations for the convergence the two others, as shown in Fig.4. Also, the three principal operations are merged into one operation and hence fewer loops are employed in implementations. Although the use of loops instead of matrix calculation has no significant impact on complexity, it affects the run time. Finally, additional

operations are required in Tian's and Hardie's methods to avoid the blank regions discussed in subsection 3.6.

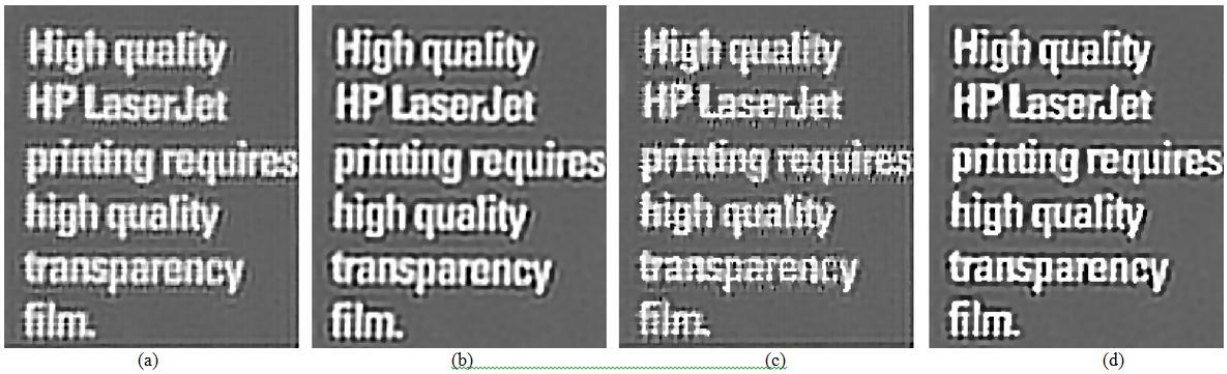


Fig. 10. Reconstructed HR images using Tian's Joint method (a), proposed Joint method (b), Hardie's AM method (c) and proposed AM method (d).

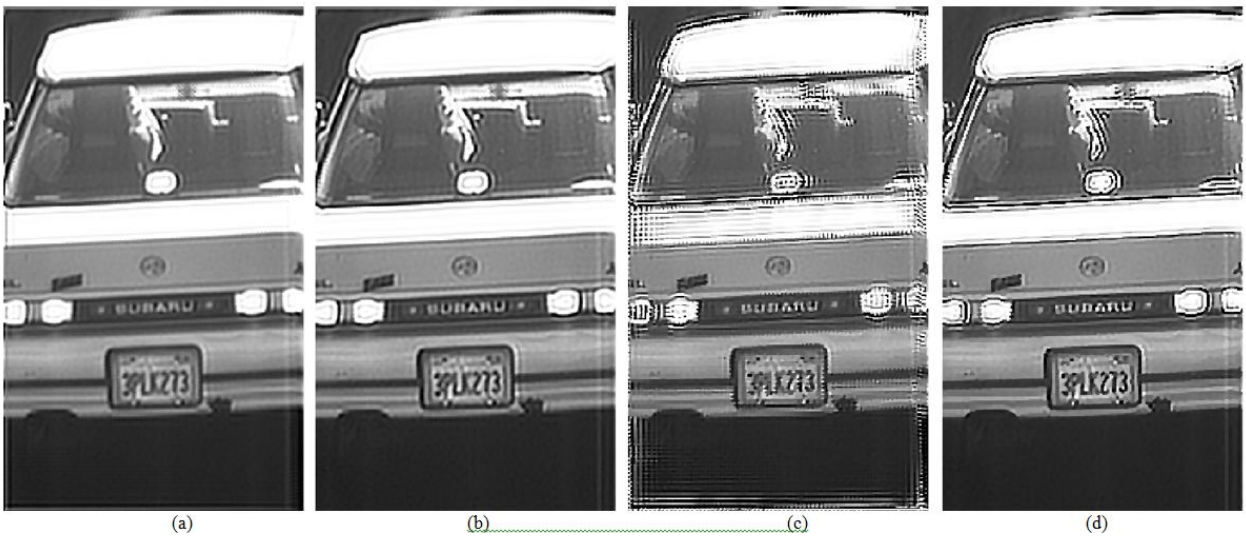


Fig. 11. Reconstructed HR images using Tian's Joint method (a), proposed Joint method (b), Hardie's AM method (c) and proposed AM method (d).

5. Conclusion and Future Works

In this paper we proposed a new joint method that combined the three principal operations in one operation. The application of this combinational operation in the calculation of Jacobian matrix is one of the most important contributions of this paper. The proposed joint method reduced error propagation, was less likely to be trapped into suboptimal solutions, especially when the relative zooming between frames was considerable, and finally increased the quality of the reconstructed HR images. Applying this combinational operation to the framework of AM methods presented a new AM method. The proposed AM method was not as stable as the proposed joint method, but it was more reliable than existing AM methods such as Hardie's method. Similar to other AM methods, the convergence of the proposed AM method was highly dependent on initial values of HR image and the registration parameters but if the initial values were close to optimal values, this method provided fast convergence. In future works, we will extend the motion model to the affine and finally to the homography, and the size of blur kernel in the image plane will be refined during the iterations.

References

[1] S. Park, M. Park and M. Kang. "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp.21– 36, 2003.

[2] S. Borman and R.L. Stevenson. "Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research," Laboratory for Image and Signal Analysis (LISA), University of Notre Dame, Notre Dame, Ind, USA, Tech. Rep., 1998.

[3] K. Katsaggelos, R. Molina and J. Mateos. *Super Resolution of Images and Video*, Morgan and Claypool Publishers, San Rafael, 2007.

[4] P. Milanfar. *Super-Resolution Imaging*, CRC Press, 2011.

[5] S. Lertrattanapanich and N. Bose. "High Resolution Image Formation from Low Resolution Frames Using Delaunay Triangulation," *IEEE Transaction on Image Processing*, vol. 11, no. 12, pp. 1427–1441, 2002.

[6] F. Xu, H. Wang, L. Xu and C. Huang. "A new framework of normalized convolution for superresolution using robust certainty," *IEEE International Conference on Computer and Automation Engineering*, vol. 2, pp. 144-148, 2010.

[7] T.Q. Pham, L.J.V. Vliet and K. Schutte. "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1-12, 2006.

[8] K. Zhang, G. Mu, Y. Yuan, X. Gao and D. Tao, "Video super-resolution with 3D adaptive normalized convolution", *Elsevier Journal of Neurocomputing*, vol. 94, pp. 140-151, 2012.

[9] A. Sánchez-Beato and G. Pajares, "Noniterative Interpolation-Based Super-Resolution Minimizing Aliasing in the Reconstructed Image", *IEEE Transaction on Image Processing*, VOL. 17, NO. 10, pp. 1817-1826, 2008.

[10] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," *IEEE International Conference on Computer Vision and Pattern Recognition* , pp. 645–650, 2001.

[11] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[12] M.E. Tipping and C.M. Bishop, "Bayesian image super-resolution," *Conference on Advances in Neural Information Processing Systems 15 (NIPS)*. Cambridge: MIT Press, 2003.

[13] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. "Bayesian Methods for Image Super-Resolution," *Computer Journal*, 2007

[14] S. D. Babacan, R. Molina and A. K. Katsaggelos. "Variational Bayesian Super Resolution," *IEEE Transaction on Image Processing*, vol. 20, no. 4, 2011.

[15] R. C. Hardie, K.J. Barnard and E.E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images, " *IEEE Transaction on Image Processing*, vol. 6, no. 12, pp. 1621-1633, 1997.

[16] N. A. Woods, N. P. Galatsanos and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Transaction on Image Processing*, vol. 15, no. 1, 201–213. 2006.

[17] B. C. Tom, A. K. Katsaggelos and N. P. Galatsanos, "Reconstruction of a high resolution image from registration and restoration of low resolution images," *IEEE International Conference on Image Processing*, vol. 3, pp. 553–557, 1994.

[18] G. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Journal of Inverse. Problems*, vol. 19, no. 2, pp. 1–26, 2003.

[19] J. Chung, E. Haber, and J. Nagy, "Numerical methods for coupled super-resolution," *Journal of Inverse Problems*, vol. 22, no. 4, pp. 1261–1272, 2006.

[20] D. Robinson, S. Farsiu, and P. Milanfar, "Optimal registration of aliased images using variable projection with applications to super-resolution," *Computer. Journal*., vol. 52, no. 1, pp. 31–42, 2007.

[21] Y. He, K.-H. Yap, L. Chen and L.-P. Chau. "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution," *IEEE Transaction on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.

[22] Y. Tian and K.H. Yap. "Joint Image Registration and Super-Resolution from Low-Resolution Images with Zooming Motion," *IEEE Transaction on Circuits and System for Video Technology*, vol. 27, no. 3, pp. 1224-1234, 2013.

[23] Capel, D. Image Mosaicing and Super-resolution (Distinguished Dissertations). Springer, ISBN: 1852337710, 2004.

[24] G.D. Evangelidis, E.Z. Psarakis, "Parametric Image Alignment using Enhanced Correlation Coefficient", *IEEE Transaction on Pattern Analysis and Machine Inteligence*, vol. 30, no. 10, 2008.

[25] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, 2nd ed., Ed. Cambridge, United Kingdom: Cambridge University Press, 2003.

[26] T. Lehmann, C. Gonner, and K. Spetzer, "Survey: Interpolation Methods in Medical Image Processing, " *IEEE Transaction on Medical Imaging*, vol. 18, no. 11, pp. 1049-1067, 1999.

[27] F. Sroubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Transaction on Image Processing*, vol. 12, no. 9, pp. 1094–1106, 2003.

[28] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," *IEEE Transaction on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.

[29] [On-line], Available: http://xanthippi.ceid.upatras.gr/people /evangelidis/ecc/ [2014-09-29].

[30] [On-line], Available: http://www.robots.ox.ac.uk/~elle/SRcode/ [2014-09-29].

[31] M. V. W. Zibetti and J. Mayer, "A robust and computationally efficient simultaneous super-resolution scheme for image sequences," *IEEE Trans. Circuits Syst. Video Technol*., vol. 17, no. 10, pp. 1288-1300, 2007.

[32] L. Olikery, X. Liy, P. Husbandsy and R. Biswasz, "Effects of Ordering Strategies and Programming Paradigms on Sparse Matrix Computations." *Journal of SIAM Review*, Vol. 44, No. 3, pp. 373-393, 2002.

[33] [On-line], Available: http://lcav.epfl.ch/software/superresolution [2014-09-29].

[34] [On-line], Available: http://users.soe.ucsc.edu/~milanfar/ software /sr-datasets.html [2014-09-29].

Hossein Rezayi received his B.Sc. degree in Electrical Engineering in 2001 and his M.Sc. degree in Communication Engineering in 2004 from the Ferdowsi University of Mashhad, Iran, where he is currently a Ph.D. candidate in Communication Engineering. His current research interests include machine vision, image processing, pattern recognition and artificial intelligence algorithms.

Seyed Alireza Seyedin received his B.Sc. degree in Electronics Engineering from Isfahan University of Technology, Isfahan, Iran in 1986, his M.Sc. degree in Control and Guidance Engineering from Roorkee University, Roorkee, India in1992, and his Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia in, and 1996.In 1996, he joined the Faculty of Engineering of Ferdowsi University of Mashhad as an assistant professor, where he has been teaching as an associate professor since 2007. He was the Head of the Department of the Electrical Engineering from 1998 to 2000. His current research interests include machine vision, DSP-based signal processing algorithms, and digital image processing techniques. Dr. Seyedin is a member of the Technical Committee of the Iranian Conference on Machine Vision and Image Processing.