

Editor in Chief Professor Hu, Yu-Chen

International Journal of Image Processing (IJIP)

Book: 2009 Volume 3, Issue 5

Publishing Date: 30-11-2009

Proceedings

ISSN (Online): 1985-2304

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

IJIP Journal is a part of CSC Publishers

<http://www.cscjournals.org>

©IJIP Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversion by CSC Publishing Services – CSC Journals, Malaysia

Table of Contents

Volume 3, Issue 5, November 2009.

Pages

- 184 – 194 Offline Signature Verification Using Local Radon Transform and Support Vector Machines
Vahid Kiani, Reza Pourreza, Hamid Reza Pourreza
- 195 - 202 Filter for Removal of Impulse Noise By Using Fuzzy Logic
Er. Harish Kundra, Er. Monika Verma, Er. Aashima
- 203 - 217 Multiple Ant Colony Optimizations for Stereo Matching
Lili Nurliyana Abdullah, Fatimah Khalid
- 218- 228 Image Registration for Recovering Affine Transformation Using Nelder Mead Simplex Method for Optimization
Mehfuza Suleman Holia, V.K.Thakar
- 229 - 245 A Texture Based Methodology for Text Region Extraction from Low Resolution Natural Scene Images
S A Angadi, M. M. Kodabagi
- 246- 251 Lossless Grey-scale Image Compression Using Source Symbols Reduction and Huffman Coding
Saravanan C, Ponalagusamy R

252 –264 An Intelligent Control System Using an Efficient License Plate
Location and Recognition Approach
**Saeed Rastegar, Reza Ghaderi, Gholamreza Ardeshipr, Nima
Asadi**

Offline Signature Verification Using Local Radon Transform and Support Vector Machines

Vahid Kiani

*Computer Engineering Department
Ferdowsi University of Mashhad
Mashhad, Iran*

vahid_keyany@yahoo.com

Reza Pourreza

*Computer Engineering Department
Ferdowsi University of Mashhad
Mashhad, Iran*

reza_pourreza@yahoo.com

Hamid Reza Pourreza

*Associate Professor
Computer Engineering Department
Ferdowsi University of Mashhad
Mashhad, Iran*

hpourreza@um.ac.ir

Abstract

In this paper, we propose a new method for signature verification using local Radon Transform. The proposed method uses Radon Transform locally as feature extractor and Support Vector Machine (SVM) as classifier. The main idea of our method is using Radon Transform locally for line segments detection and feature extraction, against using it globally. The advantages of the proposed method are robustness to noise, size invariance and shift invariance. Having used a dataset of 600 signatures from 20 Persian writers, and another dataset of 924 signatures from 22 English writers, our system achieves good results. The experimental results of our method are compared with two other methods. This comparison shows that our method has good performance for signature identification and verification in different cultures.

Keywords: Offline Signature Verification, Radon Transform, Support Vector Machine.

1. INTRODUCTION

Signatures are most legal and common means for individual's identity verification. People are familiar with the use of signatures in their daily life. Automatic signature recognition has many applications including credit card validation, security systems, cheques, contracts, etc [1], [2].

There are two types of systems in this field, signature verification systems and signature identification systems. A signature verification system just decides whether a given signature belongs to a claimed writer or not. A signature identification system, on the other hand, has to decide a given signature belongs to which one of a certain number of writers [3].

Major methods of signature recognition can be divided into two classes, on-line methods and off-line methods. On-line methods measure the sequential data such as coordinates of writing points, pen pressure, angle and direction of the pen. While off-line methods use an optical scanner to obtain signature image [4], [5]. Offline systems are of interest in scenarios where only hard copies of signatures are available. Since online signatures also contain dynamic information, they are difficult to forge. Therefore, offline signature verification methods are less reliable than online methods [3].

In signature verification systems, two common classes of forgeries are considered: casual and skilled. A casual forgery is produced by only knowing the name of the writer, and without access to a sample of the genuine signature. When forger uses his own signature or genuine signature of another writer as a casual forgery, it is called a substitution forgery. So, stylistic differences are common in casual forgeries. In skilled forgeries, the forger has access to a sample of genuine signature and knows the signature very well. Since skilled forgeries are very similar to genuine signatures, some appropriate features for detection of casual forgeries are ineffective in detection of skilled forgeries [2], [4].

The precision of signature verification systems can be expressed by two types of error: the percentage of genuine signatures rejected as forgery which is called False Rejection Rate (FRR); and the percentage of forgery signatures accepted as genuine which is called False Acceptance Rate (FAR) [4].

The signature verification is performed in two steps, feature extraction and classification. During the feature extraction phase, personal features of each training signature are extracted and trained to the classifier. In the classification phase, personal features extracted from a given signature are fed into classifier in order to judge its validity.

Offline signature verification generally involves extraction of global or local features. Global features describe the characteristics of the whole signature and include the discrete Wavelet transform, the Hough transform, horizontal and vertical projections, edge points of signature, signature area, and smoothness features [2], [7]. Local features describe only a small part of signature and extract more detailed information from image. These features include unballistic motion and tremor information in stroke segments, stroke elements, local shape descriptors, and pressure and slant features [3], [7]. This paper presents a new offline signature verification method based on local Radon Transform.

The rest of paper is organized as follows. After this introduction, Section 2 presents some related works done in the field. Our proposed method is described in section 3. Experimental results of the proposed method on two signature sets are discussed in Section 4. Finally, Section 5 draws the conclusions and further work.

2. RELATED WORK

The problem of automatic signature verification has received big attention in past years because of its potential applications in banking transactions and security systems. Cavalcanti et al [8] investigates the feature selection for signature identification. He used structural features, pseudo-dynamic features and five moments in his study. Ozgunduz et al [9] has presented an off-line signature verification and recognition method using the global, directional and grid features. He has showed that SVM classifier has better performance than MLP for his proposed method. Mohamadi [10] has presented a Persian offline signature identification system using Principal Component Analysis (PCA) and Multilayer Perceptron (MLP) neural network.

Sigari and Pourshahabi [11], [12] proposed a method for signature identification based on Gabor Wavelet Transform (GWT) as feature extractor and Support Vector Machine (SVM) as classifier. In their study after size normalization and noise removal, a virtual grid is placed on signature

image and Gabor coefficients are computed on each point of grid. Next, all Gabor coefficients are fed into a layer of SVM classifiers as feature vector. The number of SVM classifiers is equal to the number of classes. Each SVM classifier determines whether the input image belongs to corresponding class or not (one against all method). In their study two experiments on two signature sets were done. They have achieved identification rate of 96% on Persian signature set and more than 93% on Turkish signature set. Their Persian signature set was the same as signature set that has been used in [10].

Coetzer [3], have used Discrete Radon Transform as global feature extractor and a Hidden Markov Model in a new signature verification algorithm. In their proposed method, The Discrete Radon Transform is calculated at angles that range from 0° to 360° and each observation sequence is then modeled by an HMM of which the states are organized in a ring. To model and verify signatures of each writer one HMM is considered. Their system is rotation invariant and robust with respect to moderate levels of noise. Using a dataset of 924 signatures from 22 writers, their system achieves an equal error rate (EER) of 18% when only high-quality forgeries (skilled forgeries) are considered and an EER of 4.5% in the case of only casual forgeries. These signatures were originally captured offline. Using another dataset of 4800 signatures from 51 writers, their system achieves an EER of 12.2% when only skilled forgeries are considered. These signatures were originally captured online and then digitally converted into static signature images.

3. THE PROPOSED METHOD

As shown in Figure 1, our proposed method consists of two major modules: (i) Learning genuine signatures, (ii) Verification or recognition of given signature. These modules share two common prior steps: preprocessing and feature extraction.

Preprocessing phase makes signature image ready for feature extraction. When system is in learning mode, extracted features resulting from feature extraction step are used by learning module and fed into SVM classifiers to learn signature. But, when system is in testing mode, extracted features resulting from feature extraction step are used by classification module and fed into SVM classifiers to classify given signature.

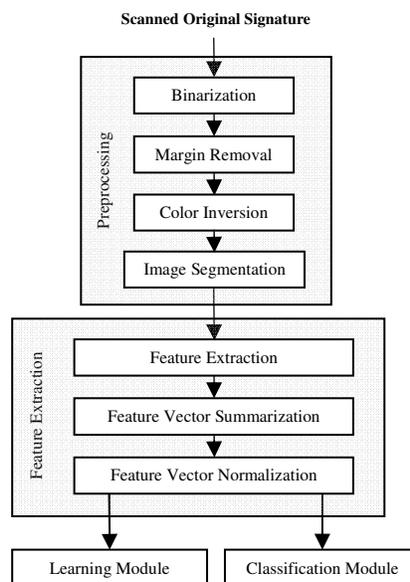


FIGURE 1: Block diagram of proposed system.

Removing margins of signature image in the preprocessing step leads to shift invariance property in our algorithm. Also, our method is scale invariant due to feature vector normalization in the

feature extraction phase. The proposed method can tolerate small rotations in signature image, however in the case of big rotations its performance will reduce.

The next sections describe these modules in more details. The data collection and preprocessing is described in Section 3.1. Feature extraction and idea of the system are presented in Section 3.2. In Section 3.3 some collected samples are used for training and finally the coming samples are tested to obtain the percentage of FRR and FAR.

3.1 Preprocessing

The purpose of preprocessing phase is to make signatures ready for feature extraction. The preprocessing stage includes three steps: Binarization and margin removal, Color inversion, Image segmentation.

Binarization and margin removal

In the first step, signature image is converted to binary image using Otsu binarization algorithm [13]. The next step is finding the outer rectangle of the signature and removing signature image margins. This gives us shift invariance property in our algorithm. We found the outer rectangle using horizontal and vertical projections of binary image. Figure 2 shows a sample original signature before preprocessing. Figure 3 shows horizontal and vertical projections of binary image.

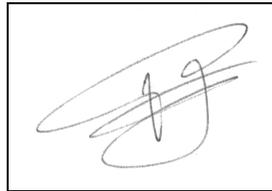


FIGURE 2: An original sample signature.

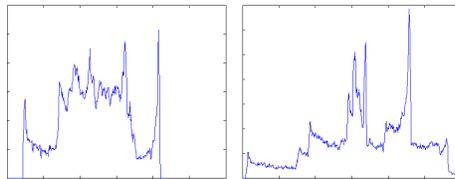


FIGURE 3: Horizontal (left) and vertical (right) projections of binary image.

Color inversion

The Radon Transform counts pixels with nonzero value in desired direction to produce image projection. In our binary images foreground is black with zero value and background is white with nonzero value. Hence, in this step we invert image before giving it to Radon Transform. Thereafter, when we give inverted image to Radon Transform, the biggest peak in the result of Radon Transform will be corresponding to line segment direction. Figure 4 shows inverted signature image.



FIGURE 4: Inverted signature image

Image Segmentation

Our proposed method works locally, so signature image must be segmented to some local windows and then orientation and width of line segment must be detected in each local window. The size of local window (n) has a direct effect on precision of line detection process. A small/big window will result in narrow/wide line segments detection. By choosing an appropriate window size, all line segments can be detected.

Another parameter that has great effect on our algorithm precision is overlay rate of neighboring windows. Non-overlaid windows will reduce algorithm precision substantially. Therefore, we define a new parameter "step" that its combination with window size will determine overlay rate of neighboring windows. Figure 5 shows overlaying windows.

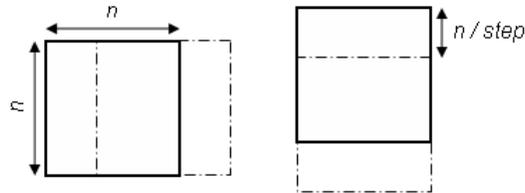


FIGURE 5: Overlaying windows

3.2. Feature extraction

The feature extraction stage includes four steps: line segment detection, line segment existence validation, feature vector extraction and summarization, and feature vector normalization.

Radon Transform and line segment detection

“The Radon Transform computes projection sum of the image intensity along a radial line oriented at a specific angle” [14]. For each specific angle θ , the radon transform produce a vector R containing the projection sum of image intensity at angle θ . A mathematical definition of Radon Transform is given in [14], [15], [16], [17]. The radon transform of a function $g(x,y)$ in 2-D Euclidean space is defined by

$$R(p, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(p - x \cos \theta - y \sin \theta) dx dy \quad (1)$$

Where the $\delta(r)$ is Dirac function. Computation of Radon Transform of a two dimensional image intensity function $g(x,y)$, results in its projections across the image at arbitrary orientations θ and offsets p [17].

The local image is squared shape; therefore occurrence of peak value in diagonal directions is more probable than other orientations. We solve this problem, by applying a circular mask to the local image before giving it to Radon Transform. Figure 6 shows this circular mask and its application.

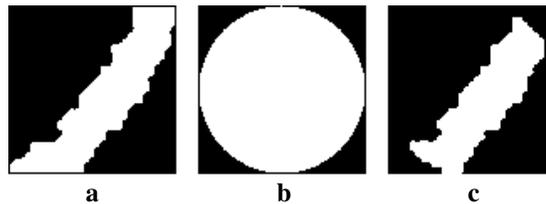


FIGURE 6: Filtering local window using a circular mask
a) local window b) mask c) masked window

The biggest peak value in the result of Radon Transform is the projection of probable line segment in its orientation angle. We call this orientation angle α and the projection peak value along it P_α . In the next step, P_α is processed to specify existence of the line segment in the local window.

To determine line width, we go left and right from peak offset in projection along angle α until reach a value smaller than $level \cdot P_{\alpha}$. These two places are line segment start and end positions. Level is another parameter of our algorithm and must be in (0, 1].

Line segment existence validation

In the next step, we detect existence of the line segment in the local window by comparing projection peak value P_{α} with a predetermined threshold value. To use the same threshold value for windows with different size, the Projection peak value P_{α} must be normalized before comparing with the line validation threshold. To do this we compute line validity value by dividing P_{α} to window size (n) as below:

$$line\ validity = \frac{P_{\alpha}}{n} \tag{2}$$

If the line validity value is greater than the line validation threshold then a line segment is detected in the current local window.

Feature vector extraction and summarization

Based on detected line segments with different orientations and widths, for each line width a feature vector containing histogram of detected line segments in orientations of 0° to 179° is produced. This feature vector is computed for line widths of 1, 2, 3, 4, 5 and 6 pixels. All line segments with line width greater than 6 counted in sixths histogram. This approach for feature extraction gives us a long feature vector with 1080 elements that is inappropriate for classification purpose. To solve this problem, we summarize this feature vector by combining some line widths together and also considering a degree resolution. To do this, we first combine some line widths together by summing corresponding feature vector elements of these line widths. Then based on selected degree resolution dr , we extend number of angle values spanned by each bin from 1 to dr angles. To do this, we combine each dr neighboring elements in the resulted feature vector from prior step by summing them into a new corresponding bin in final feature vector. This approach gives us a good flexibility in feature vector summarization. The optimum number of bins in the histogram is a function of the desired accuracy and the amount of data to be examined. Figure 7 shows this summarization process for {1,2,3} {4,5} {6} width combination and degree resolution of 3. This sample combination results in a feature vector with 180 elements.

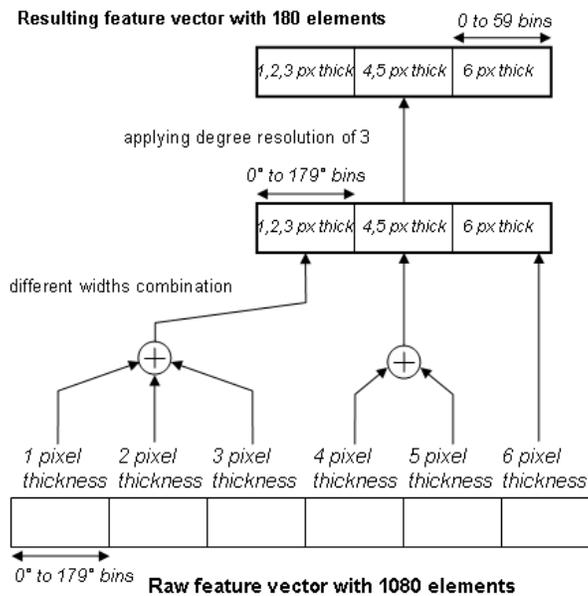


FIGURE 6: Feature vector summarization process

Feature vector normalization

In the last step of feature vector generation, we normalize feature vector by dividing its elements to its maximum element value. This leads to scale invariance property in our algorithm.

3.3. Classification

In the classification step, signature images are classified using a layer of Support vector machine (SVM) classifiers. The number of SVM classifiers in classification layer is equal to number of signature classes.

The main characteristic of a learning machine is its generalization property [18]. This is the ability of the classifier to correctly classify unseen data which were not present in the training set. Recent advances in the field of statistical learning have led to SVM classifier which has been applied with success in many applications like face and speaker recognition [19], [21]. The concept of SVM classifier was introduced by Vapnik in late of 1970's. The main idea of SVM is to construct a hyperplane as the decision surface with maximal margin of separation between positive and negative examples [20]. This leads to high generalization ability in SVM classifier with respect to other statistical classifiers. When samples are nonlinearly separable in input space, SVM must used in feature space using appropriate kernel function. In our experiment, we have used SVM classifier with Radial Basis Functions (RBF) kernel to achieve the best results.

Since SVM is a binary classifier (can categorize two classes) for classification of N classes, N SVM classifiers are needed. So in our application, number of SVM classifiers is equal with the number of writers. Each SVM classifier is used for identification of one writer signatures against all other writers (one against all strategy).

We have used two rules in our method for signature identification and verification. If all classifiers except only one generate negative result, the corresponding class of the classifier that generates positive result is considered as the class of input signature. For identification purpose this class is notified as the signer identity. For verification purpose, input signature is genuine if this class is equal to claimed signer class. In the case of skilled or casual forgeries, output of all classifiers can be negative or two or more classifier outputs can be positive. In this case the input signature will not belong to known classes, and is detected as a forgery signature.

4. EXPERIMENTAL RESULTS

Two experiments were done to evaluate our method and compare it with other methods. The first experiment was on a Persian signature set and the second experiment was on an English signature set.

4.1. Persian signature dataset

This signature set is same as the signature set that is used in [10-12]. It contains 20 classes and 30 signatures per class. For each class, 10 signatures for training and 10 genuine signatures and 10 skilled forgery signatures for test were used.

The results of our algorithm are compared to the results of algorithm developed by M.H. Sigari on this dataset [12]. Sigari developed SVM-based algorithm only for identification of offline signatures. He used Grid Gabor Wavelet coefficients to identify signature images, and achieved 96% identification rate.

Table 1 shows the performance of our method on this dataset for different width combinations and degree of resolutions. Comparing our results with the Sigari results on this dataset, our algorithm in the best case achieved the same 96% identification rate (FRR of 4%) and FAR of 17%. The main advantage of our algorithm is that it can also give us good results for verification purposes. We achieved these results using local window size of $n=31$, line validation threshold of $L=0.7$, level=0.95 and step=3.

Width combination	Degree resolution	FRR	FAR
{1,2,3,4,5,6}	3	0.09	0.23
{1,2,3,4,5,6}	5	0.09	0.25
{1,2,3,4,5,6}	10	0.09	0.29
{1,2,3} {4,5,6}	3	0.12	0.16
{1,2,3} {4,5,6}	5	0.06	0.21
{1,2,3} {4,5,6}	10	0.08	0.27
{1,2,3,4} {5,6}	3	0.07	0.17
{1,2,3,4} {5,6}	5	0.04	0.22
{1,2,3,4} {5,6}	10	0.08	0.26
{1,2} {3,4} {5,6}	3	0.11	0.14
{1,2} {3,4} {5,6}	5	0.09	0.19
{1,2} {3,4} {5,6}	10	0.06	0.22
{1,2} {3,4,5} {6}	3	0.07	0.20
{1,2} {3,4,5} {6}	5	0.08	0.23
{1,2} {3,4,5} {6}	10	0.07	0.26
{1,2,3} {4,5} {6}	3	0.09	0.13
{1,2,3} {4,5} {6}	5	0.04	0.17
{1,2,3} {4,5} {6}	10	0.08	0.22
{1} {2} {3} {4} {5} {6}	3	0.24	0.08
{1} {2} {3} {4} {5} {6}	5	0.13	0.11
{1} {2} {3} {4} {5} {6}	10	0.06	0.19

TABLE 1: Performance on the Persian signature dataset.

4.2. The Stellenbosch dataset

This signature set is same as the signature set that is used in [3]. It contains 22 classes and 30 genuine signatures, 6 skilled forgeries, and 6 casual forgeries in each class. For each writer, 10 genuine signatures are used for training and 20 genuine signatures for testing.

The results of our algorithm are compared to the results of a HMM-based algorithm developed by J. Coetzer on this dataset [3]. Coetzer algorithm is flexible and can achieve different FRR and FAR pairs. He achieved equal error rate (ERR) of 4.5% when only casual forgeries are considered and ERR of 18% when only skilled forgeries are considered.

Width combination	Deg. Res.	FRR	FAR casual	FAR skilled
{1,2,3,4,5,6}	3	0.49	0	0.07
{1,2,3,4,5,6}	5	0.33	0	0.14
{1,2,3,4,5,6}	10	0.24	0.01	0.19
{1,2,3,4,5,6}	15	0.24	0.01	0.20
{1,2,3,4,5,6}	20	0.19	0.02	0.22
{1,2,3} {4,5,6}	3	0.61	0	0.05
{1,2,3} {4,5,6}	5	0.43	0	0.11
{1,2,3} {4,5,6}	10	0.28	0	0.18
{1,2,3} {4,5,6}	15	0.27	0.01	0.19
{1,2,3} {4,5,6}	20	0.23	0.04	0.18

{1,2,3,4} {5,6}	3	0.70	0	0.03
{1,2,3,4} {5,6}	5	0.54	0	0.05
{1,2,3,4} {5,6}	10	0.34	0	0.14
{1,2,3,4} {5,6}	15	0.29	0	0.17
{1,2,3,4} {5,6}	20	0.28	0.01	0.18
{1,2} {3,4} {5,6}	3	0.7	0	0.03
{1,2} {3,4} {5,6}	5	0.54	0	0.05
{1,2} {3,4} {5,6}	10	0.34	0	0.14
{1,2} {3,4} {5,6}	15	0.29	0	0.17
{1,2} {3,4} {5,6}	20	0.28	0.01	0.18
{1,2} {3,4,5} {6}	3	0.66	0	0.05
{1,2} {3,4,5} {6}	5	0.49	0	0.10
{1,2} {3,4,5} {6}	10	0.33	0.01	0.18
{1,2} {3,4,5} {6}	15	0.28	0.01	0.19
{1,2} {3,4,5} {6}	20	0.24	0.02	0.22
{1,2,3} {4,5} {6}	3	0.75	0	0.01
{1,2,3} {4,5} {6}	5	0.65	0	0.04
{1,2,3} {4,5} {6}	10	0.42	0.01	0.10
{1,2,3} {4,5} {6}	15	0.33	0	0.18
{1,2,3} {4,5} {6}	20	0.29	0.01	0.19
{1} {2} {3} {4} {5} {6}	3	0.77	0	0.02
{1} {2} {3} {4} {5} {6}	5	0.74	0	0.01
{1} {2} {3} {4} {5} {6}	10	0.57	0	0.06
{1} {2} {3} {4} {5} {6}	15	0.46	0	0.09
{1} {2} {3} {4} {5} {6}	20	0.36	0	0.13

TABLE 2: Performance on the English signature dataset.

Table 2 shows the performance of our method on this dataset. Comparing our results to Coetzer results on this dataset, our algorithm in the best case achieved the FRR of 19% and FAR of 2% when only casual forgeries are considered and FAR of 22% in the case of only skilled forgeries. The best case for our algorithm is when no line width information is used on this dataset. Due to using a HMM classifier with ring topology, when only casual forgeries are considered Coetzer result in identification are better than our results on this dataset. However, the results of our algorithm are satisfying on this dataset. We achieved these results using local window size of $n=31$, line validation threshold of $L=0.7$, level=0.95 and step=3.

5. CONCLUSION

In this work we presented an approach to offline signature identification and verification problems based on local Radon Transform and SVM classifier. Using Radon Transform as a local feature extraction method gives us fine information and more detailed features. The main advantage of our algorithm with respect to identification method in [12] is its ability to produce good results for verification purposes beside identification purposes. Also, it exhibits a good performance for signature identification and verification purposes in different cultures.

6. ACKNOWLEDGMENT

The authors would like to thank M.H. Sigari and J. Coetzer for providing the signature datasets.

7. REFERENCES

1. D. Impedovo, G. Pirlo. "Automatic Signature Verification - The State of the Art". IEEE Transactions on Systems Man and Cybernetics, 38(5):609-635, 2008
2. A.C. Ramachandra, J.S. Rao, K.B. Raja, K.R. Venugopla, L.M. Patnaik. "Robust Offline Signature Verification Based On Global Features". In Proceedings of the IEEE International Advance Computing Conference. Patiala, India, 2009
3. J. Coetzer, B.M. Herbst, J.A. du Preez. "Offline Signature Verification Using the Discrete Radon Transform and a Hidden Markov Model". EURASIP Journal on Applied Signal Processing, 2004(4):559-571, 2004
4. W. Hou, X. Ye, K. Wang. "A Survey of Off-line Signature Verification". Proceedings of the 2004 International Conference on intelligent Mechatronics and Automation. Chengdu, China, 2004
5. S. Sayeed, N.S. Kamel, R. Besar. "A Sensor-Based Approach for Dynamic Signature Verification using Data Glove". Signal Processing: An International Journal, 2(1):1-10, 2008
6. D.S. Guru, H.N. Prakash. "Online Signature Verification and Recognition: An Approach Based on Symbolic Representation". IEEE Transaction on Pattern Analysis and Machine Intelligence. 31(6):1059-1073, 2009
7. D.S. Guru, H.N. Prakash and S. Manjunath. "On-line Signature Verification: An Approach Based on Cluster Representations of Global Features". Seventh International Conference on Advances in Pattern Recognition. Kolkata, India, 2009
8. G.D.D.C. Cavalcanti, R.C. Doria, E.Cde.B.C. Filho. "Feature Selection for Off-line Recognition of Different Size Signatures". Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing. Martigny, Switzerland, 2002
9. E. Ozgunduz, T. Senturk, E. Karsligil. "Off-Line Signature Verification and Identification by Support Vector Machine". 11th International Conference on Computer Analysis of Images and Patterns. Versailles, France, 2005
10. Z. Mohamadi. "Persian Static Signature Recognition". Thesis of Bachelor of Engineering, Ferdowsi University of Mashhad, March 2006
11. M.H. Sigari, M.R. Pourshahabi. "Static Handwritten Signature Identification and Verification". Thesis of Bachelor of Engineering, Ferdowsi University of Mashhad, July 2006
12. M.H. Sigari, M.R. Pourshahabi, H.R. Pourreza. "Offline Handwritten Signature Identification using Grid Gabor Features and Support Vector Machine". 16th Iranian Conference on Electrical Engineering. Tehran, Iran, 2008
13. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms". IEEE Transaction on Systems Man and Cybernetics, 9(1):62-66, 1979
14. T.O. Gulum, P.E. Pace, R. Cristi. "Extraction of polyphase radar modulation parameters using a wigner-ville distribution - radon transform". IEEE International Conference on Acoustics, Speech and Signal Processing, 2008
15. F. Hjouj, D.W. Kammler. "Identification of Reflected, Scaled, Translated, and Rotated Objects From Their Radon Projections". IEEE Transactions on Image Processing, 17(3):301-310, 2008

16. M. R. Hejazi, G. Shevlyakov, Y-S Ho. *"Modified Discrete Radon Transforms and Their Application to Rotation-Invariant Image Analysis"*. IEEE 8th Workshop on Multimedia Signal Processing, 2006
17. Q. Zhang, I. Couloigner. *"Accurate Centerline Detection and Line Width Estimation of Thick Lines Using the Radon Transform"*. IEEE Transactions on Image Processing, 16(2):310–316, 2007
18. C. Burgess. *"A tutorial on support vector machines for pattern recognition"*. Data Mining and Knowledge Discovery, 2(2):121–167, 1998
19. E.J.R. Justino , F. Bortolozzi, R. Sabourin. *"A comparison of SVM and HMM classifiers in the off-line signature verification"*. Pattern Recognition Letters, 26(9):1377-1385, 2005
20. S. Haykin. *"Neural Networks: A Comprehensive Foundation"*, Englewood Cliffs, pp. 340-341 (1999).
21. D. Mohammad. *"Multi Local Feature Selection Using Genetic Algorithm for Face Identification"*. International Journal of Image Processing, 1(2):1-10, 2007

Filter for Removal of Impulse Noise by Using Fuzzy Logic

Er. Harish Kundra

*Asistant Professor
R.I.E.I.T., Railmajra
Distt. Ropar, Punjab, India.*

hodcseit@rayatbahra.com

Er. Monika Verma

*Asistant Professor
S.V.I.E.I.T, Banur
Distt. Patiala, Punjab, India.*

monikaverma007@gmail.com

Er. Aashima

*Lecturer
R.B.I.E.B.T., Sahauran
Distt. Kharar, Punjab, India.*

er.aashima@yahoo.co.in

Abstract

Digital image processing is a subset of the electronic domain wherein the image is converted to an array of small integers, called pixels, representing a physical quantity such as scene radiance, stored in a digital memory, and processed by computer or other digital hardware. Fuzzy logic represents a good mathematical framework to deal with uncertainty of information. Fuzzy image processing [4] is the collection of all approaches that understand, represent and process the images, their segments and features as fuzzy sets. The representation and processing depend on the selected fuzzy technique and on the problem to be solved. This paper combines the features of Image Enhancement and fuzzy logic. This research problem deals with Fuzzy inference system (FIS) which help to take the decision about the pixels of the image under consideration. This paper focuses on the removal of the impulse noise with the preservation of edge sharpness and image details along with improving the contrast of the images which is considered as the one of the most difficult tasks in image processing.

Keywords: Digital Image Processing (DIP), Image Enhancement (IE), Fuzzy Logic (FL), Peak-signal-to-noise-ratio (PSNR).

1. INTRODUCTION

1.1 Image Processing

An image is digitized to convert it to a form which can be stored in a computer memory or on some form of storage media such as hard disk or CD-ROM. This digitization procedure can be done by scanner, or by video camera connected to frame grabber board in computer. Once the image has been digitized, it can be operated upon by various image processing operations.

Image processing operations [1] can be roughly divided into three major categories, Image Compression, Image Enhancement and Restoration and Measurement Extraction. Image

compression involves in reducing the amount of memory needed to store a digital image. Image restoration is the process of taking an image with some known, or estimated, degradation, and restoring it to its original appearance. Image restoration is often used in the field of photography or publication where an image was somehow degraded, but need to be improved before it can be printed. Image enhancement is improving an image visually.

The main advantage of IE is in the removal of noise in the images. Removing or reducing noise in the images is very active research area in the field of DIP.

1.2 Noise in Images

Image noise is the random variation of brightness or color information in images produced by the sensor and circuitry of a scanner or digital camera. Image noise can also originate in film grain and in the unavoidable shot noise of an ideal photon detector. Image noise is generally regarded as an undesirable by-product of image capture. Although these unwanted fluctuations became known as "noise" by analogy with unwanted sound, they are inaudible and actually beneficial in some applications, such as dithering.

The impulse noise (or salt and pepper noise) is caused by sharp, sudden disturbances in the image signal; its appearance is randomly scattered white or black (or both) pixels over the image. Fig. 1.1 shows an original image and the image which is corrupted with salt and pepper noise.

Noise filtering can be viewed as removing the noise from the corrupted image and smoothen it so that the original image can be viewed. Noise filtering can be viewed as replacing every pixel in the image with a new value depending on the fuzzy based rules. Ideally, the filtering algorithm should vary from pixel to pixel based on the local context.

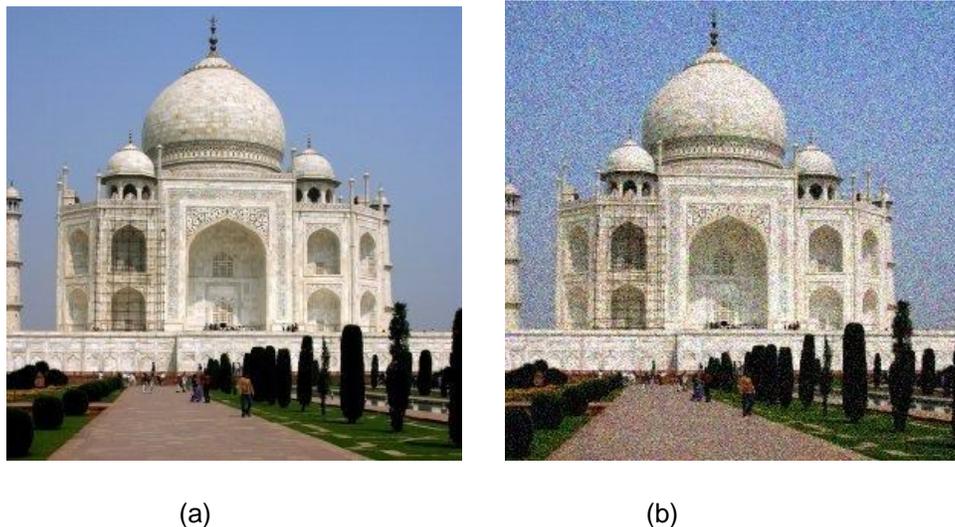


Figure. 1.1: Noise in Images (a) Original Image (b) Image with noise.

1.3 Objectives

The objective of the paper is to give a new better, faster and efficient solution for removing the noise from the corrupted images. The main point under consideration is that the noise-free pixels must remain unchanged. The main focus will be on:

1. Removal of the noise from the test image.
2. Noise free pixels must remain unchanged.
3. Edges must be preserved.
4. Improve the contrast

2. PROPOSED WORK

In literature several (fuzzy and non-fuzzy) filters have been studied [2] [3] [5] [6] for impulse noise reduction. These techniques are often complementary to existing techniques and can contribute to the development of better and robust methods. Impulse noise is caused by errors in the data transmission generated in noisy sensors or communication channels, or by errors during the data capture from digital cameras. Noise is usually quantified by the percentage of pixels which are corrupted. Removing impulsive noise while preserving the edges and image details is the difficult issue.

Traditionally, IE techniques such as mean and median filtering have been employed in various applications in the past and are still being used. Although these techniques remove the impulsive noise but they were unable to preserve the sharpness of the edges. They smooth the noise as well as the edge sharpness. They were unable to improve the contrast of the image. A fuzzy theory based IE avoids these problems and is a better method than the traditional methods. The proposed filter provides an alternative approach in which the noise of colored image is removed and the contrast is improved.

To achieve a good performance, a noise reduction algorithm should adapt itself to the spatial context. Noise smoothing and edge enhancement are inherently conflicting processes, since smoothing a region might destroy an edge, while sharpening edges might lead to unnecessary noise. Many techniques to overcome these problems have been proposed in literature. In this thesis a new filter, based on the concepts of IE and FL have been introduced that not only smooth the noise but also preserves the edges and improve its contrast. The test images taken into consideration have impulse noise or salt and pepper noise.

The work is done in two phases. In the first phase, the noise in the images is removed and in the second phase, contrast is improved. The output image generated is noise-free high-contrast image.

The noise intensity in the same test image varies as 10%, 20%, 30%, 40% and 50%. For each case the PSNR and Execution time is calculated.

2.1 Phase 1: Removal of Impulsive Noise

For each pixel (i, j) of the image (that isn't a border pixel) we use a 3x3 neighborhood window. For each pixel position we have the gradient values. The two related gradient values for the pixel in each direction are given by the following table:

TABLE 1. Basic and two related gradient values for each direction.

<i>R</i>	basic gradient	related gradients
NW	$\nabla_{NW} A(i, j)$	$\nabla_{NW} A(i + 1, j - 1), \nabla_{NW} A(i - 1, j + 1)$
N	$\nabla_N A(i, j)$	$\nabla_N A(i, j - 1), \nabla_N A(i, j + 1)$
NE	$\nabla_{NE} A(i, j)$	$\nabla_{NE} A(i - 1, j - 1), \nabla_{NE} A(i + 1, j + 1)$
E	$\nabla_E A(i, j)$	$\nabla_E A(i - 1, j), \nabla_E A(i + 1, j)$
SE	$\nabla_{SE} A(i, j)$	$\nabla_{SE} A(i - 1, j + 1), \nabla_{SE} A(i + 1, j - 1)$
S	$\nabla_S A(i, j)$	$\nabla_S A(i, j - 1), \nabla_S A(i, j + 1)$
SW	$\nabla_{SW} A(i, j)$	$\nabla_{SW} A(i - 1, j - 1), \nabla_{SW} A(i + 1, j + 1)$
W	$\nabla_W A(i, j)$	$\nabla_W A(i - 1, j), \nabla_W A(i + 1, j)$

These values indicate in which degree the central pixel can be seen as an impulse noise pixel.

The fuzzy gradient value $\nabla_R^F A(i, j)$ for direction R (R ∈ {NW, N, NE, E, SE, S, SW, W}), is calculated by the following fuzzy rule:

If $|\nabla_R A(i, j)|$ is large AND $|\nabla_R' A(i, j)|$ is small
 OR
 $|\nabla_R A(i, j)|$ is large AND $|\nabla_R'' A(i, j)|$ is small
 OR

$\nabla_R A(i, j)$ is big positive AND ($\nabla'_R A(i, j)$ AND $\nabla''_R A(i, j)$) are big negative

OR

$\nabla_R A(i, j)$ is big negative AND ($\nabla'_R A(i, j)$ AND $\nabla''_R A(i, j)$) are big positive

Then $\nabla^F_R A(i, j)$ is large.

Where $\nabla_R A(i, j)$ is the basic gradient and $\nabla'_R A(i, j)$ and $\nabla''_R A(i, j)$ are the two related gradient values for the direction R. Because “large”, “small”, “big negative” and “big positive” are non-deterministic features, these terms can be represented as fuzzy sets. Fuzzy sets can be represented by a membership function. Examples of the membership functions LARGE (for the fuzzy set *large*), SMALL (for the fuzzy set *small*), BIG POSITIVE (for the fuzzy set *big positive*) and BIG NEGATIVE (for the fuzzy set *big negative*)

When we get the gradient values we apply the similarity function. The similarity function is $\mu: [0; \infty) \rightarrow R$. We will need the following assumptions for μ :

1. μ is decreasing in $[0; \infty)$,
2. μ is convex in $[0; \infty)$,
3. $\mu(0) = 1, \mu(\infty) = 0$.

In the construction, the central pixel in the window W is replaced by that one, which maximizes the sum of similarities between all its neighbors. Basic assumption is that a new pixel must be taken from the window W . Each of the neighbors of the central pixel is moved to the center of the filtering window and the central pixel is rejected from W . For each pixel of the neighborhood, which is being placed in the center of W , the total sum of similarities is calculated and then compared with maximum sum. The total sum of similarities is calculated without taking into account the original central pixel, which is rejected from the filter window. In this way, the central pixel is replaced by that pixel from the neighborhood, for which the total similarity function, which is a sum of all values of similarities between the central pixel and its neighbors, reaches its maximum. The filter tends to replace the original pixel only when it is really noisy and preserves in this way the image structures.

2.2 Improving the Contrast of the Image

In the first phase we remove the noise from the image. Now the test image generated after removing the noise is operated upon again to improve its contrast, which is the second phase of the algorithm.

For improving the contrast of the image following steps are done:

1. Setting the shape of membership function (regarding to the actual image)
2. Setting the value of fuzzifier Beta
3. Calculation of membership values
4. Modification of the membership values by linguistic hedge
5. Generation of new gray-levels

3. RESULTS

The test images are operated on different intensities of noise as 10%, 20%, 30%, 40% and 50%. Different PSNR and evaluation time are calculated for each image with different noise intensities.

The results are shown:

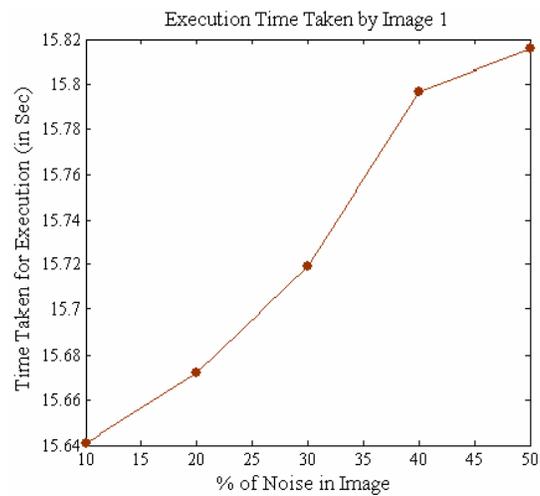
10% Noise in Image 1.



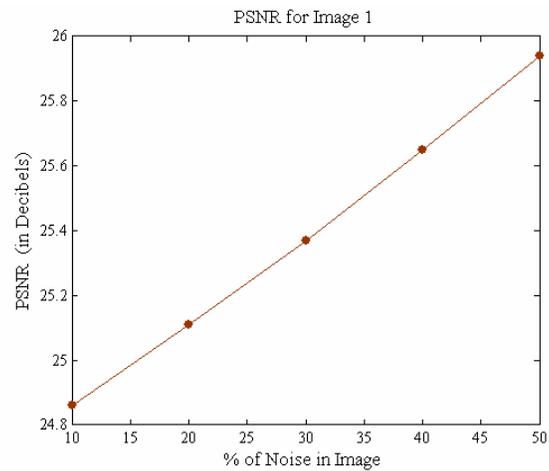
Figure 3.1: Image 1

TABLE 2: Performance Evaluation of Image 1

% of Noise	Parameters	PSNR (in Decibels)	TIME (in Seconds)
10		24.86	15.641000
20		25.11	15.672000
30		25.37	15.719000
40		25.65	15.797000
50		25.94	15.816000



GRAPH 1: Execution Time Taken by Image 1.



GRAPH 2: PSNR for Image 1.

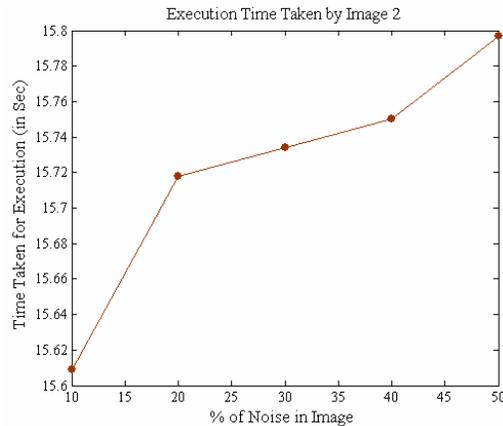
10% Noise in Image 2:



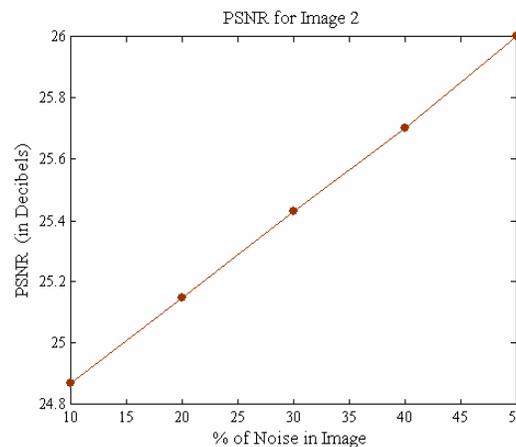
Figure 3.2 Image 2

TABLE 3: Performance Evaluation of Image 2

Parameters		
% of Noise	PSNR (in Decibels)	TIME (in Seconds)
10	24.87	15.609000
20	25.15	15.718000
30	25.43	15.734000
40	25.70	15.750000
50	26.00	15.797000



GRAPH 3: Execution Time Taken by Image 2.



GRAPH 4: PSNR for Image 2.

4. CONSLUSION & FUTURE WORK

Various test images of different extensions are fed to the system. The images are corrupted with salt and pepper noise as well as are of low contrast. The filter is seen to preserve intricate features of the image while removing heavy impulse noise where as the conventional mean and median filters fail in this context even at low corruption levels. The learning of fuzzy rules in a fuzzy image filter with a true hierarchical fuzzy logic structure where the output of the first layer is fed in to the second layer to obtain an 'improved' final output. The evaluation parameters PSNR and Evaluation time taken are evaluated. The program generates positive PSNR and is above 20dB which is considered to be the best ratio. The overall execution time which the program takes is approximately 15 seconds.

In future, modification of fuzzy rules can produce better result. Other techniques such as PSO can also be used for image enhancement.

5. REFERENCES

- [1] Gonzalez, R.C., Woods, R.E., Book on "Digital Image Processing", 2nd Ed, Prentice-Hall of India Pvt. Ltd.
- [2] Carl Steven Rapp, "Image Processing and Image Enhancement", Texas, 1996.
- [3] R. Vorobel, "Contrast Enhancement of Remotely-Sensed Images," in *6th Int. Conf. Math. Methods in Electromagnetic Theory*, Lviv, Ukraine, Sept 1996, pp. 472-475.

- [4] Tizhoosh, "Fuzzy Image Processing", © Copyright Springer, 1997.
- [5] Farzam Farbiz, Mohammad Bager Menhaj, Seyed A. Motamedi, and Martin T. Hagan, "A new Fuzzy Logic Filter for image Enhancement" IEEE Transactions on Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 30, No. 1, February 2000
- [6] P. Fridman, "Radio Astronomy Image Enhancement in the Presence of Phase Errors using Genetic Algorithms," *in Int. Conf. on Image Process.*, Thessaloniki, Greece, Oct 2001, pp. 612-615.

Multiple Ant Colony Optimizations for Stereo Matching

XiaoNian Wang

*The School of Electronics and Information Engineering
Tong Ji University
ShangHai, 201804, China*

Dawnyear@Tongji.edu.cn

Ping Jiang

*The School of Electronics and Information Engineering
Tong Ji University
ShangHai, 201804, China*

P.Jiang@Bradford.ac.uk

Abstract

The stereo matching problem, which obtains the correspondence between left and right images, can be cast as a search problem. The matching of all candidates in the same line forms a 2D optimization task, which is an NP-hard problem in nature. Two characteristics are often utilized to enhance the performance of stereo matching, i.e. concurrent optimization of several scan-lines and correlations among adjacent scan-lines. Such correlations are considered to be posterior, which require several trails for their discovery. In this paper, a Multiple Ant Colony based approach is proposed for stereo matching because of the Ant Colony optimization's inherent capability of relation discovery through parallel searching. The Multiple Ant Colony Optimization (MACO) is efficient to solve large scale problems. For stereo matching, it evaluates sub-solutions and propagates the discovered information by pheromone, taking into account the ordering and uniqueness constraints of candidates in images. The proposed algorithm is proved to be able to find the optimal matched pairs theoretically and verified by experiments.

Keywords: Multiple Ant Colony Optimizations, Stereo Matching, Iteration, Constraints.

1. INTRODUCTION

The purpose of computer stereo vision is to obtain depth information of objects with the help of two or more cameras. Generally speaking, there are four steps to accomplish it, which are image pre-processing, matching primitive defining & extracting, feature matching and disparity refining. The image pre-processing includes image enhancement and epipolar rectification; the second step includes the definition of feature and its extraction; disparity refining is to get a smooth depth map in which sub-pixel interpolation is involved for example. Feature matching has been one of the most challenging research topics in computer vision.

The stereo matching problem [1-7], that is to obtain a correspondence between right and left images, can be cast as a search problem. When a pair of stereo images is rectified, corresponding points can be searched within the same scanline, this is a two dimensional (2D) optimization, which can be shown as a NP-hard problem [3]. An optimization method, such as Dynamic Programming (DP)[2,8-16], Simulated Annealing(SA) [17], Genetic Algorithm(GA)[18],

max-flow[19], graph-cut [3,20], etc., can be used to find the optimal or sub-optimal solutions with different efficiency.

Baker [1] describes a fast, robust, and parallel implementable edge-based line-by-line stereo correlation scheme. Based on the fact that a connected sequence of edges in one image should be a connected sequence of edges in the other, a cooperative procedure to deal with edge correspondences is proposed. The dynamic programming algorithm performs a local optimization for the correlation of individual lines in the image, and the edge connectivity is used to remove miscorrelations.

Ohta [8] defines two different searches, intra-scanline and intera-scanline search. The intra-scanline search can be treated as finding a matching path on 2D search plane whose axes are the right and left scanlines. Vertically connected edges in the images provide consistency constraints across the 2D search planes. Inter-scanline search in a 3D search space, with a stack of the 2D search planes, needs to utilize the vertically connected edge information. Dynamic programming is used in both searches.

Birchfield [2] proposes a new algorithm based on three heuristic functions. During the matching the occluded pixels are allowed to remain unmatched, the information between scanlines is propagated by a postprocessor. The global post-process propagates reliable disparities to the regions with unreliable disparities.

Bobick [4] develops a stereo algorithm that integrates matching and occlusion analysis into a single process. After highly-reliable matches, the ground control points (GCPs) are introduced. The matching sensitivity to occlusion-cost and algorithmic complexity can be significantly reduced. The use of ground control points eliminates both the need for biasing the process towards a smooth solution and the task of selecting critical prior probabilities describing image formation.

Raymond [10] proposes the use of a multi-level dynamic programming method to solve the matching problem of stereo vision. At level 1, the line segment pairs that have a very high local similarity measure are selected for the matching process. By considering the geometric properties between the matched and the unmatched line segments, a global similarity measure is calculated for each unmatched line segments pair, and then the second level starts.

In [Kim 13], first, a new generalized ground control points (GGCPs) scheme is introduced, where one or more disparity candidates for the true disparity of each pixel are assigned by local matching using the oriented spatial filters. Second, it performs optimization both along and across the scanlines by employing a two-pass dynamic programming technique. Combined with the GGCPs, the stability and efficiency of the optimization are improved significantly.

[Sorgi 15] presents a symmetric stereo matching algorithm, based on the bidirectional dynamic programming scanline optimization. The Sum of the Squared Differences (SSD) map is treated as a decoding trellis and inspected twice: the forward optimization produces the optimal path from the upper left to the lower right corner, and the backward optimization produces the optimal path from the lower right back to the upper left corner. The final operation, a consistency check between the selected forward and backward optimal paths, can produce an occlusion-robust matcher without defining an empirical occlusion cost.

[Sung 16] proposes a stereo matching algorithm which employs an adaptive multi-directional dynamic programming scheme using edge orientations. Chain codes are introduced to find the accurate edge orientations which provide the DP scheme with optimal multidirectional paths. The proposed algorithm eliminates the streaking problem of conventional DP based algorithms, and estimates more accurate disparity information in boundary areas.

On the assumption that the neighboring elements have consistent match values, in [Zitnick 5], local support area that determines which and to what extent neighboring elements should contribute to averaging is introduced. An iterative algorithm updating the match values by diffusing support among neighboring values and inhibiting others along similar lines of sight is proposed. After the match values have converged, occluded areas are explicitly identified and the final results are obtained.

Marr and Poggio [9] present two basic assumptions for a stereo vision algorithm. The first assumption, uniqueness assumption, states that at most a single unique match exists for each pixel if surfaces are opaque; The second one, continuity assumption, states that disparity values are generally continuous, i.e., smooth within a local neighborhood.

[Scharstein 6, Brown 7] reviews the development in decades. In [Leung 12, Selzer 14], the process is speeded up with special data structure. All the articles mentioned above have the following common characteristics. 1) There are two kinds or levels of optimization, the local one which accomplishes the optimization in the corresponding scanline and the global one which finds the best solution among all scanlines. 2) Based on the Marr's assumptions, many constraints must be obeyed, for example, ordering constraint, uniqueness constraint and bi-directional monotonicity constraint. 3) To get the global optimal solution, the reliability should be propagated. How to propagate the reliability is skillful. In [8], the optimization function is update by the inter-scanline information; in [4, 13], the preprocessed GCPs are introduced; in [5, 10], the iteration is used to remove the wrong matches or enhance the correct matches; in [2], a postprocessor is employed to remove wrong matches after the optimization.

In this paper, the expected merits of a good algorithm for stereo matching are analyzed. Then a new multiple ant colony optimization (MACO) method is proposed to solve the stereo matching problem, and the convergence of the proposed algorithm is also discussed. In the last part the experiments show the results of the algorithm.

2. PRELIMINARIES

Stereo matching is still an open task to be investigated. The following two questions are argued firstly in this paper.

- Marr's assumptions is correct, but dose it need to be obeyed during the optimizing process?
- The reliability propagation is necessary, but how to get more reliable one and how to do?

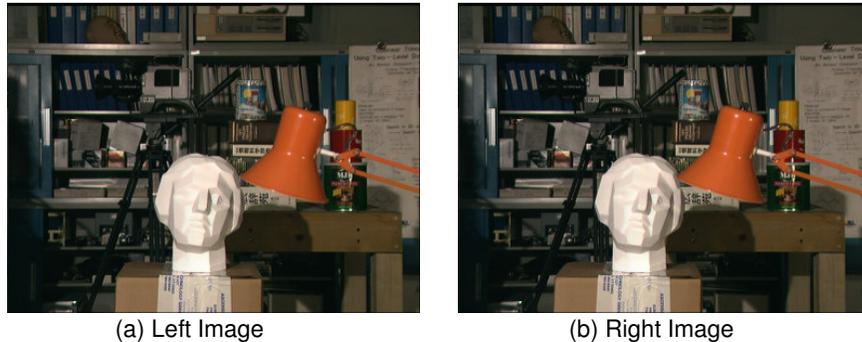


FIGURE 1: Tsukuba Pairs

One pair of standard test images for stereo algorithm is shown in Fig.1. Supposed the size of epipolar rectified image is K by L , there are N features (the features can be point, line, curve and area) on the k^{th} ($k=1\dots K$) scanline of left image and M features on the k^{th} scanline of right image. A matrix $L \times L$, named as similarity matrix, stores all possible matches on the k^{th} scan line. The element at (n^{th}, m^{th}) is the Sum of Absolute Differences (SAD) similarity of the feature on (k^{th}, n^{th}) in left image and the feature on (k^{th}, m^{th}) in right image. In the similarity matrix, only N by M elements are meaningful, the others are zero. If minimum & maximum parallax restraint is considered, define v_n as the search space of the n^{th} feature in the left image, all v_n ($n=1\dots L$) form a banded region (actually, n should be from 1 to N , to simplify statement and not loss of the generality, in this paper $n=1\dots L$), as the white banded region shown in Fig.2, marked with notation Ω (including Ω' , in which every element is zero, marked as the black area) in this paper. The d_{min} and d_{max} represent the minimum and maximum parallax respectively. The parallax on the dot black line is zero. One similarity matrix corresponds to a specific scan line, and then K similarity matrixes are available, That is to say there are K tasks to be optimized in the stereo matching (every one is a sub-task).

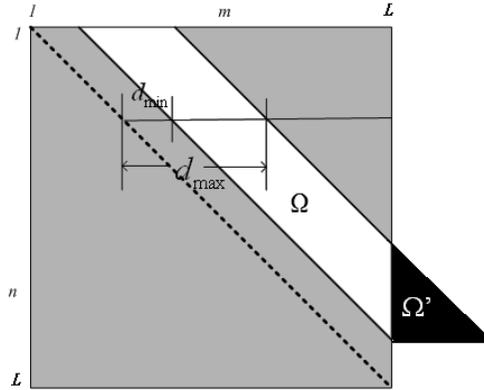


FIGURE 2: Similarity Region

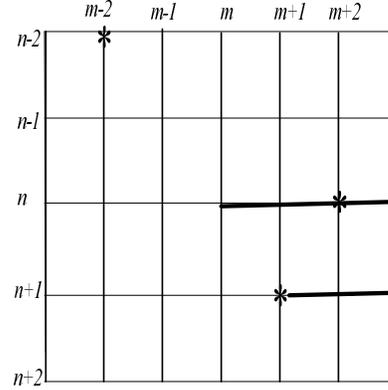


FIGURE 3: Constraints in Matching Process

2.1 Constraints during the Optimization

The dynamic programming algorithm requires the uniqueness and ordering constraints [11]. The current state is undoubtedly decided by the state of the previous result, when the matching error occurs in the previous stage, it will directly affect the current stage of the match. There is no opportunity to rectify this error in the later match if the simple DP is used.

In the matching process what's happened if the ordering and uniqueness constraints are exerted? In the k^{th} scanline, suppose there are 3 features at $n-2^{th}$, n^{th} , $n+1^{th}$ columns of left image and 5 features from $m-2^{th}$ to $m+2^{th}$ columns of right image. The matching process is illustrated by Fig.3, in which $d_{min} \leq m-n \leq d_{max}$ is satisfied. Set the features at $n-2^{th}$ is the best match with one at $m-2^{th}$, then the possible matchers with n^{th} can be found from m^{th} to $m+d_{max}^{th}$ as marked with a dark black line. If there is no match with $n-1^{th}$ features and the n^{th} is matched with $m+2^{th}$, considering the ordering constraint, the match with the $n+1$ can only be found from $m+3$ to $m+1+d_{max}$. If there is a wrong math between the n^{th} and $m+2^{th}$, the abuse of such constraints will miss the correct math between the $n+1^{th}$ and $m+1^{th}$! Clearly we cannot trust the ordering constraint based on maybe false assumptions, the same as the uniqueness constraint.

- In this paper, during the procedure of matching, the only constraint is the minimum & maximum parallax constraint.

Such strategy is good for finding more real matches, simplifying the computation and promoting the parallelism greatly. That is to say the sub-optimization can start from random position instead of the rigid left-top or right-bottom corner.

2.2 Reliability Propagation

In literature[4,13], GCPs are used to increase the real match probability, GCPs not only generate a sufficient number of seed pixels for guiding the subsequent matching process, but also remarkably reduce the risk of false match. It is known that the false matches in GCPs could severely degrade the final matching results. In practice, the reliability of GCPs/GGCPs is far away from expectation. The ordering and uniqueness constraints may propagate the error and make the result worse, additionally, the GCPs must be identified before the DP optimization process, but how to automatically get more reliable GCPs?

The vertical edges information among scanlines may be the most frequently considered one. Vertical edges, especially the edges with high threshold, are robust features, that is to say the possibility of an edge which finds the correct match is high. There is no any priori-knowledge about which edge of the right image is matched with one of the left image. Every vertical edge are discretized by sacnlines, so many features are fomed. According to the large probability hypothesis, most of the features on the vertical edge in the left image will be matched with the features in the same edge in the right image. That is to say after the optimization of features sharing the same vertical edge, if the matching results are voted, the matched edge in the right image can be identified. **This information is posterior, means that this information only can be obtained after every sub-optimization.** Such automatically obtained knowledge is relative

reliable and should be propagated to the sub-matching process, so the feedback or iterative idea should be introduced into the whole optimizing procedure. The confirmed matching edges after voting can serve as the GCPs for the next optimization. **The first contribution of this paper is that such voted GCPs are obtained automatically during the optimization procedure.**

3. PARALLEL ACO Based STEREO MATCHING

An ideal stereo matching algorithm should have three merits, first the ordering and uniqueness constraints are ignored during the optimizing process but the result must satisfy such constraints; second the optimizing process of every line is relatively independent, so every process can be done concurrently; third if two scanlines share the same vertical edge, the hint of vertical edge should be exploited to enhance the certainty of every line's optimization next time, That is to say the reliability should be propagated. To sum up, a parallel, iterative and feedback algorithm is proposed in this paper.

Ant-based system is recently developed for the solution of combinatorial optimization problems [21]. After this the Ant Colony Optimization (ACO) emerged. In ACO, an ant builds a solution to a combinatorial optimization problem by exploiting pheromone trails and heuristic information. The main characteristics of the ACO are inherent parallelism, stochastic nature, adaptability, and the use of positive feedback. Paper [22] shows that ACO is always better than Genetic Algorithm (GA) and Simulated Annealing (SA), if the parameters are selected properly.

When dealing with complex and large-scale issues, a single-group ant colony optimization algorithm is prone to be slow and premature. The parallel multiple Ant Colony Optimization (MACO) algorithms can be exploited through the acceleration of the construction procedure. Various parallel approaches [23-28] are proposed to promote the efficiency with the help of communication and parallelism. Most parallelization models can be classified into fine-grained, in which the population of ants is separated into a large number of very small sub-populations, and coarse-grained models, in which the population of ants is divided into few sub-populations. [Bullnheimer 23] introduces two parallel implementations called the Synchronous Parallel Implementation (SPI) and the Partially Asynchronous Parallel Implementation (PAPI). SPI is based on a master-slave paradigm in which every ant finds a solution in the slaves and sends the result to the master. When the solutions are available the master updates the pheromone information and sends the updated information back to all slaves. PAPI is based on the coarse-grained model in which information is exchanged among colonies every fixed number of iterations. The simulation indicates that PAPI performs better than SPI in terms of running time and speedup. [Talbi 24] presents a parallel model for ant colonies to solve the Quadratic Assignment Problem. The programming style used is a synchronous master/workers paradigm. During every iteration, the master broadcasts the pheromone matrix to all the workers. Each worker receives the pheromone matrix, constructs a complete solution, and sends the found solution to the master. When the master receives all the solutions, it updates the pheromone, and then the process is iterated. In [Rahoual 25], the Set Covering Problem is solved by master/slaver colonies. Each ant process is set on an independent processor. The master process sends the necessary information (pheromone) to each ant. In [Randall 26], several parallel decomposition strategies, Parallel Independent Ant Colonies, Parallel Interacting Ant Colonies, Parallel Ants, Parallel Evaluation of Solution Elements, Parallel Combination of Ants and Evaluation of Solution Elements, are examined. These techniques are applied to Traveling Salesman Problem (TSP). In [Chu SC 27], the artificial ants are partitioned into several groups. Seven communication methods for updating the pheromone between groups are proposed. In [Ellabib 28], a performance study is carried out to evaluate the effectiveness of the exchange strategies and demonstrate the potential of applying MACO to solve the Vehicle Routing Problem with Time Windows. Experiments using the proposed assessment technique demonstrate that the exchange strategies have a considerable influence on the search diversity. The results indicate that the multiple colony system approach outperforms the single colony.

As mentioned above, the total stereo matching consists of many sub-optimization problems, and every sub-task can be optimized at the same time. So in this paper single ACO is employed to solve a sub-task. As mentioned above there are some relationship among sub-optimizations if the scanlines share the same vertical edge. Based on the large probability assumption voting can be

used to decide the correctness of all sub-solutions after the optimization of all sub-tasks are finished. This posterior knowledge requires a master to gather the results of slavers and get the best one by evaluation. When a good result is obtained by voting, it is propagated iteratively to enhance reliability. In this paper, a parallel MACO is employed. It is based on the master-slaver mode that the parallel slaver optimizes every sub-problems and the master broadcasts the pheromone formed according to the results from slavers.

3.1 Construction of MACO

The constructed MACO for stereo matching is introduced in this section by means of the following definitions.

- Definition of Pheromone

It is a 2D optimization for every sub-task to find the best matchers. In literature [33], the 2D optimization can be cast as a path finding problem. All possible paths in a search space should be stored at a pheromone matrix. The search space of sub-optimization task is Ω , which consists of $v_n (n=1, \dots, K)$. Suppose that in some iteration, the best matched pair is the n^{th} and m^{th} features.

All possible match feature with the $n+1^{th}$ feature is in v_{n+1} if there is. It means there are $d_{max} - d_{min}$ possible choices, so as every element in v_n . Actually without the ordering and uniqueness constraint, the number of all possible choices is independent with the position in v_n . The pheromone matrix τ_{ijp} is defined, Where $i=1, \dots, L-1, j, p=1 \dots d_{max} - d_{min}$. The size of pheromone field matrix is with a dimension of $(L-1) \times (d_{max} - d_{min}) \times (d_{max} - d_{min})$.

- Definition of Heuristic Information

Heuristic information is used for an ant to choose the correct match, the bigger value of similarity the higher possibility of a match. If the best match pair is the n^{th} and m^{th} features, all possible match with the $n+1^{th}$ feature is all elements in v_{n+1} . Which one is the most likely match with the $n+1^{th}$ feature depends on whose SAD is the smallest one from $m+1+d_{min}$ to $m+1+d_{max}$. In this paper, the heuristic Information is defined as the SAD if all SAD values in v_{n+1} are not equal to zero, clearly heuristic information satisfies $0 < \eta_{min}$.

- Probability Decision

In search domain Ω , set the ant at the position $x_i^j \in v_i$. According to the following probability, it will select $x_{i+1}^p \in v_{i+1}$ according to Equation (1).

$$p_{jp} = \begin{cases} \frac{\tau(i, x_i^j, x_{i+1}^p)^\alpha \eta(x_{i+1}^p)^\beta}{\sum_{l \in x_{i+1}} \tau(i, x_i^j, x_{i+1}^l)^\alpha \cdot \eta(x_{i+1}^l)^\beta}, & \max(x_{i+1}^l) > \theta \\ 0, & \text{others} \end{cases} \quad (1)$$

where $i=1, \dots, L-1; j=i+d_{min}, \dots, i+d_{max}; l, p=i+1+d_{min}, \dots, i+1+d_{max}$. $\alpha, \beta > 0$ are the weight parameters for pheromone and heuristic information; $\tau(i, x_i^j, x_{i+1}^p)$, simply note as τ_{ijp} , is the pheromone value between $x_i^j \in v_i$ and $x_{i+1}^p \in v_{i+1}$; $\eta(x_{i+1}^l)$ is the heuristic information of x_{i+1}^l ; $\max(x_{i+1}^l)$ is the maximum value in v_{i+1} ; θ is the threshold for occlusion, refer to section 5.1.

- Definition of Exchanged Information

The images are shown in Fig.4 with Canny operator, the 26th, 30th, 37th row are marked as white solid line, dotted line, stippling line respectively. There are two edges named B, E as marked in Fig.4 (a), and β, ϵ in Fig. 4 (b). The 26th, 30th scanlines pass through the B, E edge and the 37th doesn't. In the similarity matrix corresponding to the 26th scan line, if the B feature in left image matches with β in the right image, and so as the left E with the right ϵ feature, then in the 30th similarity matrix, the same match should exist, while the phenomena do not exist

between the 30th and 37th similarity matrix. Clearly the relationship among sub-optimization relies on whether the scan lines share the same vertical edge. It means that there are some uncertain relationships among every sub task.

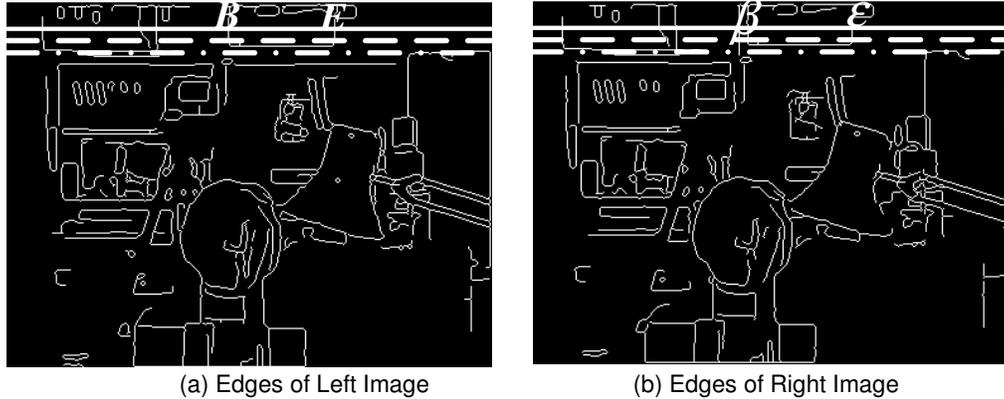


FIGURE 4: Edges Images

The master collects all solutions of every sub-task and votes to get the most probably matched vertical edge. In order to broadcast this possibly existed information to impact the process of the next iteration, the pheromone, corresponding to these voted edges, is selected as the exchange information between master and slavers. We know that the behavior of an ant is indirectly and possibly influenced by pheromone. The exchanged information (constructed pheromone) acts as a kind of soft GCPs indeed, avoiding the shortcomings of the hard one. **The second contribution of this paper is that the soft GCPs (pheromone) are used.**

In order that the optimization of the next generation can be induced by the master, the pheromone corresponding to the same edge decided by voting should be increased. Take the edge named B of left image for example, assume it crosses with the 11th to 40th scanlines, that is to say there are 30 features, among which there are 26 features matched with the same edge β of the right image. According to the big probability assumption, β edge of the right image should be the real matched edge with left B . In order to propagate this information, the pheromone matrix T_{\max}^k ($k = 1, \dots, K$) is constructed in which the values corresponding to matched edge are τ_{\max} and the others are 0. Take the 11th, 12th scan line for example, the 99th, 100th column crossed with B edge, the 119th, 120th column with β , That is to say the best matched pairs are (99th, 119th) and (100th, 120th). So in T_{\max}^{11} , all pheromone recording the connection with (99th, 119th) are τ_{\max} , and the others are zero, as well as (100th, 120th) in T_{\max}^{12} . **The exchanged pheromone is re-constructed instead of directly getting from one of the best solutions of sub-task, which is the third discrimination with the past articles.**

- Pheromone Updating

Every isolated ant colony fulfilling every sub-task has its own pheromone updating strategy. While in MACO, the pheromone updating process consists of two parts, local and global updating. The local updating, the same as the single ACO, is done after every generation is finished, shown as Equation (2). The global updating, which reflects the influence of the master, is done after all colonies are finished, shown as Equation (3). That is to say that the pheromone of every colony must be influenced not only by itself but also the master when each iteration is finished.

The Local pheromone updating procedure can be described as follows after a generation is finished.

$$\forall (i, j, p) : \tau_{ijp}(n) \leftarrow (1 - \rho) \cdot \tau_{ijp}(n) \tag{2}$$

if $J(R_a) < J(\hat{R})$ then $\hat{R} \leftarrow R_a$

$$\forall (i, j, p) \in \hat{R} \tau_{ijp}(n+1) \leftarrow \tau_{ijp}(n) + \rho \cdot g(\hat{R})$$

where $0 < \rho < 1$ is the evaporation rate; $J(R_a)$ is the energy of path belonging to the a^{th} ant; \hat{R} is the best matched pairs at present; the boundary function $g(\hat{R})$ returns the pheromone matrix corresponding to \hat{R} . Every element after update must obey $0 < \tau_{\min} \leq \tau_{ijp}(n+1)$.

The global pheromone updating procedure is executed in two phases, an evaporation phase where a fraction of the all sub-task's pheromone evaporates, and a reinforcement phase where the pheromone is increased corresponding to a better solution which the master thinks. Take the k^{th} colony for example, the update rule is:

$$\forall(i, j, p) : \tau_{ijp}(n) \leftarrow (1 - \mu) \cdot \tau_{ijp}(n) \tag{3}$$

$$\tau_{ijp}(n+1) = \tau_{ijp}(n) + \mu \cdot T_{\max}^k$$

Where $k = 1, \dots, K$; $0 < \mu < 1$ is the evaporation rate, $\mu = 0$ means the new pheromone totally comes form the master, and $\mu = 1$ means the new pheromone ignored the information form master; Every element must obey $\tau_{\min} \leq \tau_{ijp}(n+1)$. The new global updated pheromone works as the initial value of every sub-task, means one iteration is finished; T_{\max}^k is the reconstructed pheromone for the k^{th} colony by master.

3.2 Flow Chart

The flow chart of MACO is illustrated as Fig.5, the slavers and the master are marked as different shading respectively, where $\tau_{jk}^0 = \tau_{\min}$ is the initial pheromone matrix to all sub-tasks. At the beginning, set $T_{\max}^k = 0$, firstly every sub-task gets initial pheromone matrix of this iteration according to Equation (3), and then the isolated colonies start its own optimization (inner loop), in which the local pheromone updating is done according to Equation (2) after every generation is finished. After all sub-tasks are finished, the master evaluates all solutions and constructs every T_{\max}^k , the outer loop starts again. This process is iterated until a stopping criterion is met.

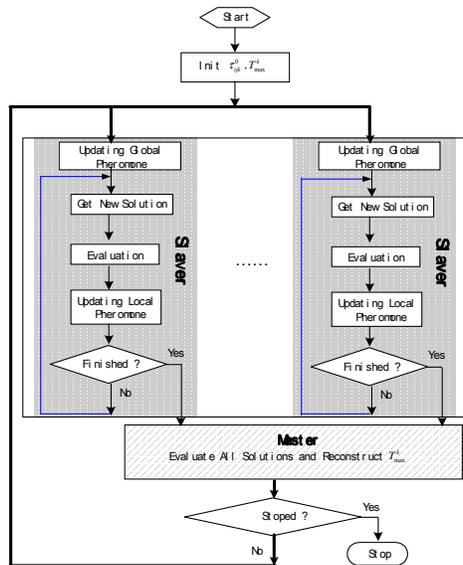


FIGURE 5: Flow Chart of MACO

4. CONVERGENCE PROOF

The convergence proof of single colony can refer to literature [29, 30, 31]. The MACO is based on the single colony, and with the help of mutual communication. The convergence has dealings with not only how every colony is defined but also how the information is interacted. Based on literature [33], the convergence proof of this paper is the following.

Set $\tau_{ijp}(n)$ represent the pheromone after iterated n times, $\hat{R}(n)$ represent the optimal path iterated n times later, and $\hat{J}(n)$ be the energy function value. A stochastic process $X(n) = \{\tau_{ijp}(n), \hat{R}(n), \hat{J}(n)\} (n = 1, \dots)$ is defined. It can be deduced that $X(n)$ depends on new status and current status merely. Therefore, $X(n)$ is an inhomogeneous Markov chain in discrete time.

Lemma 1: let T_0 be the initial pheromone value of every sub-task. Supposing that pheromone evaporation coefficient obeys $0 < \rho < 1$, for arbitrary pheromone τ_{ijp} , it holds $\tau_{\min} \leq \lim_{x \rightarrow \infty} \tau_{ijp}(n) \leq (\rho+1) \cdot g(R^*)$, where τ_{\min} and $(\rho+1) \cdot g(R^*)$ is the lower and upper bound of pheromone.

Proof :

Let R^* be the best matched pairs, after n generations the arbitrary τ_{ijp} meets

$$\tau_{ijp}(n) \leq \left(\prod_{i=1}^n (1-\rho) \right) \cdot T_0 + \sum_{i=1}^{n-1} \left(\prod_{j=i+1}^n (1-\rho) \right) \cdot g(R^*)$$

When $n \rightarrow \infty$, we get

$$\lim_{n \rightarrow \infty} \tau_{ijp}(n) \leq \lim_{n \rightarrow \infty} \left[\prod_{i=1}^n (1-\rho) \right] \cdot T_0 + \lim_{n \rightarrow \infty} \left[\sum_{i=1}^{n-1} \left(\rho \prod_{j=i+1}^n (1-\rho) \right) \right] \cdot g(R^*)$$

The first factor is $\lim_{n \rightarrow \infty} \left[\prod_{i=1}^n (1-\rho) \right] = \lim_{n \rightarrow \infty} (1-\rho)^n = 0$

The second factor is $\lim_{n \rightarrow \infty} \left[\sum_{i=1}^{n-1} \left(\rho \prod_{j=i+1}^n (1-\rho) \right) + \rho \right] = \rho + \lim_{n \rightarrow \infty} \left[\rho \sum_{i=1}^{n-1} (1-\rho)^{n-i} \right] = \rho + 1$, then we have

$$\lim_{n \rightarrow \infty} \tau_{ijp}(n) \leq \tau_{\max} = (\rho+1) \cdot g(R^*) \cdot$$

Function $g(x)$ has bound, so τ_{ijp} is limited by $(\rho+1) \cdot g(R^*)$ after n interactions.

Set the initial value of some elements τ_{ijp} to be τ_{\min} , and its value is not increased, after a generation although we have $\tau_{ijp} = (1-\rho) \tau_{\min} < \tau_{\min}$, actually it is obliged to lower bound τ_{\min} .

Finally, we get $\tau_{\min} \leq \tau_{ijp}(n) \leq (\rho+1) \cdot g(R^*)$. □

Lemma 2: Set the initial value of every sub-task as the exchanged pheromone according to the master, for arbitrary pheromone τ_{ijp} , it also holds $\tau_{\min} \leq \lim_{x \rightarrow \infty} \tau_{ijp}(n) \leq (\rho+1) \cdot g(R^*)$

Proof :

At the end of generation optimization of sub-task, the arbitrary τ_{ijp} meets $\tau_{\min} \leq \lim_{x \rightarrow \infty} \tau_{ijp}(n) \leq (\rho+1) \cdot g(R^*)$. Set the max value of pheromone from master is $(\rho+1) \cdot g(R^*)$, the initial value of next generation is the result according to Equation (3), after evaporation we get

$$\tau_{ijp}^{new}$$

$\tau_{\min} \leq \tau_{ijp}^{new} \leq (1-\mu) \cdot (\rho+1) \cdot g(R^*)$, if the pheromone τ_{\min} is enhanced, then it meets

$\tau_{\min} + \mu \cdot (\rho+1) \cdot g(R^*) \leq \tau_{ijp}^{new} \leq (1-\mu) \cdot (\rho+1) \cdot g(R^*) + \mu \cdot (\rho+1) \cdot g(R^*)$, if the pheromone which is τ_{\min} is not enhanced, then it meets

$$\tau_{\min} \leq \tau_{ijp}^{new} \leq (\rho+1) \cdot g(R^*)$$

Set τ_{ijp}^{new} as initial value, according to **Lemma 1**, after n steps, arbitrary τ_{ijp} holds $\tau_{\min} \leq \lim_{x \rightarrow \infty} \tau_{ijp}(n) \leq (\rho+1) \cdot g(R^*)$. That is to say the new initial value has the lower and upper bound. □

Lemma 3: Heuristic information η has a bound, that is, $\eta_{\min} \leq \eta \leq \eta_{\max}$.

Proof : According to the definition of heuristic information, firstly, the minimum value of η is above zero , and set the search windows to be R by T , then the maximum value of SAD is $R \times T \times 255$. \square

Theorem 1: Set $W \in Z^+$, for an arbitrary $n \geq W$, if there exists $\tau_{\min}(n) > 0$ to guarantee $\tau_{ijp}(n) \geq \tau_{\min}(n) > 0$, the inhomogeneous Markov process in discrete time $X(n) = \{\tau_{ijp}(n), \hat{R}(n), \hat{J}(n)\}$ will be convergent at the optimal status $(\tau_{ijp}[R^*], R^*, J^*)$ with probability one when $n \rightarrow +\infty$, where R^* represents the optimal path; J^* the minimal energy function value, and $\tau_{ijp}[R^*]$ is defined as follows:

$$\tau_{ijp}|_{R^*} = \begin{cases} \tau_{\max} & (i, j, p) \in R^* \\ 0 & \text{others} \end{cases}$$

Proof : According to Equation (1), there exists $p_{jp} = \frac{\tau(i, x_i^j, x_{i+1}^p)^\alpha \eta(x_{i+1}^p)^\beta}{\sum_{l \in x_{i+1}} \tau(i, x_i^j, x_{i+1}^l)^\alpha \cdot \eta(x_{i+1}^l)^\beta}$. Set $N = d_{\max} - d_{\min}$.

According to lemma 1, lemma2, lemma3 and $\tau_{ijp}(n) \geq \tau_{\min}(n) > 0$, the following holds:

$$p_{jp}(n) \geq \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^\alpha \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^\beta$$

Then, the probability of an artificial ant producing a solution (including the optimal solution R^*) after n steps iterations is $\hat{p} \geq \hat{p}_{\min}^M = \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta} > 0$, where $M < +\infty$ is the maximal length of the sequence. The minimal probability of Markov chain X_n being convergent at the optimal solution X_n^* after n steps iterations can be given by [11]:

$$\hat{P}^*(n) = 1 - (1 - \hat{p})^n \geq 1 - \left(1 - \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta}\right)^n$$

When $n \rightarrow +\infty$, considering the second term of $\hat{P}^*(n)$, $\left(1 - \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta}\right)^n$ and taking the

logarithm and limit of this product we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} \log \left(1 - \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta}\right)^n \\ &= \sum_{n=W}^{\infty} \log \left(1 - \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta}\right) \\ &\leq - \sum_{n=W}^{\infty} \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta} = -\infty \end{aligned}$$

Therefore, $\lim_{n \rightarrow \infty} \left(1 - \left(\frac{\tau_{\min}(n)}{N \cdot \tau_{\max}}\right)^{M \cdot \alpha} \cdot \left(\frac{\eta_{\min}}{\eta_{\max}}\right)^{M \cdot \beta}\right)^n = 0$. Then $\lim_{n \rightarrow \infty} \hat{P}^*(n) = 1$, that is, when $n \rightarrow +\infty$, X_n will be

convergent at the optimal status $(\tau_{ijp}[R^*], R^*, J^*)$ with probability one. \square

Reasoning 1. For every colony of multi colonies, after information exchange, in the worst case, the probability of finding the best solution is bigger than that of the single colony.

Proof : Due to the pheromone and heuristic information limits τ_{\min} , τ_{\max} , η_{\min} , η_{\max} , and set $N = d_{\max} - d_{\min}$. A trivial lower bound^[29] can be given as

$$p_{\min} \geq p_0 = \frac{\tau_{\min}^\alpha \eta_{\min}^\beta}{(N-1)\tau_{\max}^\alpha \eta_{\max}^\beta + \tau_{\min}^\alpha \eta_{\min}^\beta}$$

For the derivation of this bound we consider the following “worst case” situation: the pheromone trail associated with the desired decision is τ_{\min}, η_{\min} , while all the other feasible choices (there are at most $N - 1$) have an associated pheromone trail of τ_{\max}, η_{\max} .

When the pheromone is updated by the master, in “worst case”, the elements whose value is τ_{\max} is increased, according to the Equation (1), a new lower bound can be given as

$$p'_{\min} \geq p'_0 = \frac{\tau_{\min}^{\alpha} \eta_{\min}^{\beta}}{(1-\mu)(N-2)(\tau_{\max})^{\alpha} \eta_{\max}^{\beta} + ((1-\mu)\tau_{\max} + u\tau_{\max})^{\alpha} \eta_{\max}^{\beta} + \tau_{\min}^{\alpha} \eta_{\min}^{\beta}}$$

The denominator of p'_0 is

$$\begin{aligned} & (1-\mu)(N-2)(\tau_{\max})^{\alpha} \eta_{\max}^{\beta} + ((1-\mu)\tau_{\max} + u\tau_{\max})^{\alpha} \eta_{\max}^{\beta} + \tau_{\min}^{\alpha} \eta_{\min}^{\beta} \\ & = (1-\mu)(N-2)(\tau_{\max})^{\alpha} \eta_{\max}^{\beta} + \tau_{\max}^{\alpha} \eta_{\max}^{\beta} + \tau_{\min}^{\alpha} \eta_{\min}^{\beta} \end{aligned}$$

Clearly, it is smaller than the denominator of p_0 . That is to say the probability of finding best solution is bigger than the single colony. □

Reasoning 2. The colony in MACO can find better solutions after information exchanged.

Proof: Heuristic information is unchanged with iteration. In order to simplify the description, and don not lose generality, set $\eta(x_{i+1}^p) = 1$, in the i^{th} colony, the pheromone of the j^{th} and p^{th} feature is $\tau(i, x_i^j, x_{i+1}^p) = \tau_m$, and they are true match pair. In current generation, the actual matched feature is the j^{th} and s^{th} feature, its corresponding pheromone is $\tau(i, x_i^j, x_{i+1}^s) = t'_{\max} > \tau_m, v \neq p$, then the probability of selecting the p^{th} feature is

$$p_{jp} = \frac{\tau_m^{\alpha}}{sum + t'_{\max}{}^{\alpha} + \tau_m^{\alpha}}, \text{ where } sum \text{ is the summary of pheromone of the others } d_{\max} - d_{\min} - 2.$$

After voting, suppose the master thinks the j^{th} feature should be matched with s^{th} feature, then its will increase the pheromone being $\tau(i, x_i^j, x_{i+1}^p) = (1-\mu) \cdot \tau_m + \mu \cdot \tau_{\max}$, then the probability of selecting p^{th} feature is

$$\begin{aligned} p'_{jp} & = \frac{((1-\mu)\tau_m + \mu t_{\max})^{\alpha}}{(1-\mu)^{\alpha}(sum) + (1-\mu)^{\alpha} t'_{\max}{}^{\alpha} + ((1-\mu)\tau_m + \mu t_{\max})^{\alpha}}, \text{ simplify the equation, we get} \\ p'_{jp} & = \frac{(\tau_m + (1/(1-\mu) - 1)t_{\max})^{\alpha}}{sum + t'_{\max}{}^{\alpha} + (\tau_m + (1/(1-\mu) - 1)t_{\max})^{\alpha}}, \text{ because } (\tau_m + (1/(1-\mu) - 1)t_{\max})^{\alpha} > \tau_m^{\alpha} \text{ and} \\ & sum + t'_{\max}{}^{\alpha} > 0, \text{ so } p'_{jp} > p_{jp}. \end{aligned}$$

The probability is bigger after master influenced, that is to say the colony in MACO can find better solutions after information exchanged. □

Every sub-optimization is a not-convex and the total optimization target is consistent, the master broadcasts the better pheromone to induce the optimizing process of every sub-task. Every sub-task is convergent, so the algorithm proposed in this paper can find the global best solution with probability one.

5. SIMULATION

5.1 Optimization Target

Given energy function, the proposed algorithm can be used to find the best match in every Ω space. Just as what's said above, every sub-task finishes optimization according to its own energy function and master's guidance, and the master gathers all solutions from sub-tasks and evaluates them to get exchanged information. The local and global energy functions are defined in this section.

The function must have the following two traits. Firstly, because of the change of viewpoint, some features in left image cannot match with any one in right image, so the occlusion should also be considered. If the maximum similarity with n^{th} feature is θ' and θ' is smaller than threshold θ , then we can conclude that the counter feature with n^{th} is occluded. Secondly, the ordering and

uniqueness constraint are all ignored during the optimization procedure. So there is a special phenomenon that one feature of left image may be matched with many features of right image, called collision in this paper and it must be forbidden.

Suppose there are N features on k^{th} ($k=1...K$) line of left image and M features on k^{th} line of right image. Because of occlusion and collision, there are L real matches. Refer to literature [2], local energy function for the k^{th} sub-task is defined as Equation (4).

$$\min J_k = - \left(\sum_{i=1}^L D(x_i^j) - k_{occ} N_{occ} - k_{coll} N_{coll} \right) \tag{4}$$

where $x_i^j \in R^*$, R^* represent the best path; N_{occ} is the number of occlusion; k_{occ} is the penalty coefficient of occlusion; N_{coll} is the number of collision; k_{coll} is the penalty coefficient of collision; function D is the similarity of SAD. It is clear such definition is to encourage more match and punish the unmatched and the collided case.

The vertical edge is discretized by every sub-task. Based on large probability, most of features matched with discretized point of left edge will lie on the same edge of right image. After voting the corresponding edges are decided, the pheromone is reconstructed according to the voting results. The target of master is to get the most consistent voting results. Set $L_i > R_i$, the energy function for the master is:

$$\min J_{global} = \frac{1}{N} \sum_{i=1}^N \left(\frac{L_i - R_i}{L_i} \right) \tag{5}$$

where, N is the number of vertical edges of left image; L_i is the length of i^{th} edge in left image; R_i is the quantity of matched features in right edge which corresponds to the i^{th} left edge after voting.

5.2 Depth Restoration

The MACO is implemented by the multi-thread technology on PC. The resources are limited, so the quantity of sub-groups is limited. Set 10 colonies in a group (the maximum length of L_i is 10), 5 ants in every colony, information exchange every 10 generation. When J_{global} is smaller than 0.2 the whole optimization procedure is stopped. The matching primitive is the intensity and window size of the SAD is 9×9 .

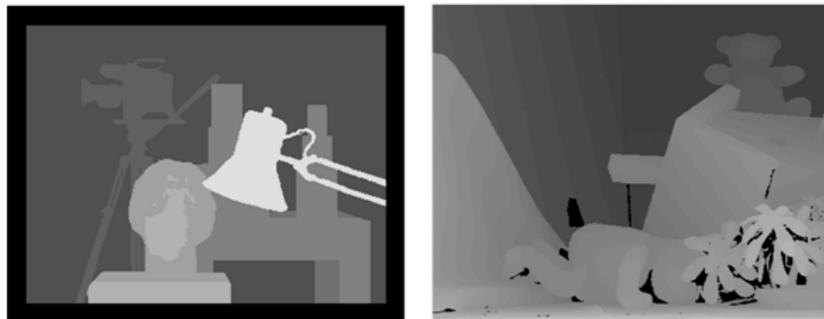
α	β	ρ	k_{occ}	k_{coll}	μ	θ
2	0.5	0.6	1	1	0.6	0.2

TABLE 1: Parameters for depth Restoration

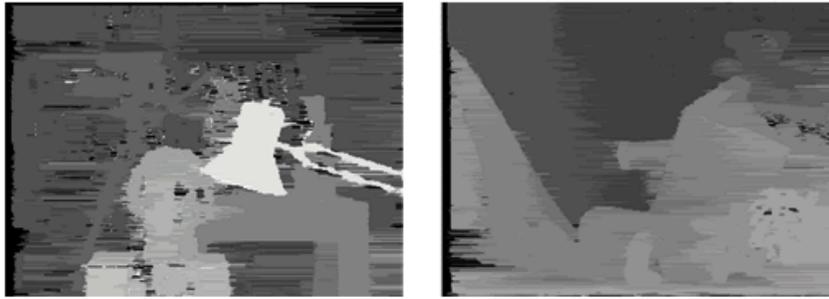
To obey the ordering and uniqueness constraint, after the whole procedure is finished, such constraints is obliged to the final results.

Because of the exhaustive parallelism (there are no any constraints among sub task optimization but minimum & maximum parallax restraint), the start point can be at arbitrary n^{th} feature (if the ordering or uniqueness constraints are exerted, such merits will no longer exist). So during the sub-task, the start position can be variable at every generation, this manner can eliminate the collision.

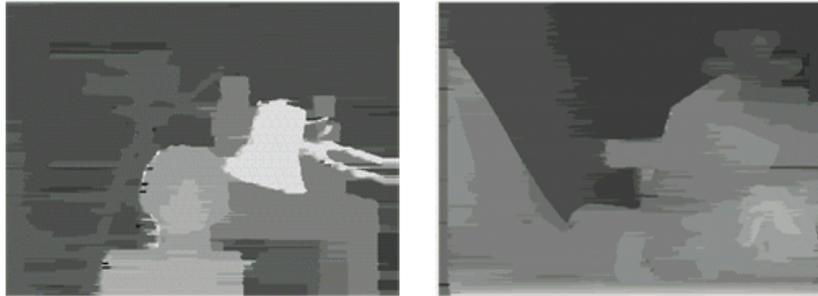
In Fig.6, the tests and comparisons are shown. Clearly, the result is better than DP, but there are big step to improve.



a) True Depth Map of Tsukuba, Teddy



b) Results of DP



c) Results of this paper

FIGURE 6: the Results of Depth Map

6. CONCLUSIONS

This paper presented a parallel, iterative and feedback MACO method for stereo matching. In this method the slaver optimizes every sub-task, which is the target to find the best matchers along every scan-line. During the iterative process, the master gathers and analyzes the results from sub-groups the results, decides the matched edge based on voting, and then reconstructs the pheromone corresponding to the matched edges, feeds the pheromone field back to the sub-task. Each sub-optimization problem starts a new matching process under the reconstructed pheromones until the stopping criterion is met. The proposed method has two outstanding merits. Firstly, this method makes full use of the matching problem of parallelism, that each relatively independent sub-task can be solved in parallel. Secondly, also makes full use of a posteriori information. Besides, the reconstructed pheromones which reflects the result of voting plays the role of soft GCPs, avoiding the misleading of hard GCPs, theory and experiments show that this idea is better than the dynamic programming algorithm with hard GCPs. The convergence proof of the proposed method gives the strong support for its application. Finally, there are problems, such as the efficiency, parameter tuning and other issues should be resolved.

7. REFERENCES

1. H. Baker and T. Binford. "Depth from edge and intensity based stereo". In IJCAI81, pages 631–636, 1981.
2. S. Birchfield and C. Tomasi. "Depth discontinuities by pixel-to-pixel stereo". In ICCV, pages 1073–1080, 1998.
3. O. Veksler. "Efficient Graph-based Energy Minimization Methods in Computer Vision". PhD thesis, Cornell University, 1999.
4. A. F. Bobick and S. S. Intille. "Large occlusion stereo". IJCV, 33(3):181–200, 1999.
5. C. L. Zitnick and T. Kanade. "A cooperative algorithm for stereo matching and occlusion detection". IEEE TPAMI, 22(7):675–684, 2000.
6. Scharstein D and Szeliski R. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", Int'l J. Computer Vision", 2002,47(1):7-42,.

7. Brown MZ, Burschka D, Hager GD. "Advances in Computational Stereo". Transactions on Pattern Analysis and Machine Intelligence, August 2003, 25(8):993-1008.
8. Y. Ohta and T. Kanade, "Stereo by two-level dynamic programming". IEEE TPAMI, 7(2):139-154, 1985.
9. D. Marr and T. Poggio, "A Computational Theory of Human Stereo Vision", Proc. Royal Soc. London B, vol. 204, pp. 301-328, 1979.
10. Yip, Raymond K.K., Ho, W.P., "A multi-level dynamic programming method for stereo line matching", PRL(19), No. 9, 31 July 1998, pp. 839-855.
11. A. L. Yuille and T. Poggio. "A generalized ordering constraint for stereo correspondence". A.I. Memo 777, AI Lab, MIT, 1984.
12. C. Leung, B. Appleton and C. Sun, "Fast stereo matching by iterated dynamic programming and quadtree subregioning". British Machine Vision Conference vol. 1, Kingston University, London 2004: 97-106.
13. Kim J, Lee KM, Choi BT, et al. "A dense stereo matching using two-pass dynamic programming with generalized ground control points", IEEE CVPR, 2005, 2:1075-1082.
14. C. Lei, J. Selzer, and Y.H. Yang, "Region-Tree based Stereo using Dynamic Programming Optimization", Proc. IEEE Conf. on Computer Vision and Pattern Recognition, New York, NY: June 17-22, 2006, pp. 2378-2385.
15. Lorenzo Sorigi, Alessandro Neri: "Bidirectional Dynamic Programming for Stereo Matching". ICIP 2006: 1013-1016.
16. Min Chul Sung, Sang Hwa Lee, Nam Ik Cho: "Stereo Matching using Multi-Directional Dynamic Programming and Edge Orientations". ICIP (1) 2007: 233-236.
17. Babu Thomas, B. Yegnanarayana, S. Das: "Stereo-correspondence using Gabor logons and neural networks". ICIP 1995: 2386-2389.
18. M. Gong and Y. H. Yang, "Multi-resolution stereo matching using genetic algorithm", IEEE Workshop on Stereo and Multi-Baseline Vision, Dec. 2001.
19. S. Roy and I. J. Cox. "A maximum-flow formulation of the N-camera stereo correspondence problem". In ICCV, pages 492-499, 1998.
20. Y. Boykov, O. Veksler, and R. Zabih. "Fast approximate energy minimization via graph cuts". IEEE TPAMI, 23(11):1222-1239, 2001.
21. Dorigo M, Manjezzo V, Colorni A. "The ant system: Optimization by a colony of cooperating agents". IEEE Transaction on Systems, Man & Cybernetics B, 1996, 26(2): 29-41.
22. M. Dorigo, V. Maniezzo & A. Colorni, 1996. "Ant System: Optimization by a Colony of Cooperating Agents", IEEE Transactions on Systems, Man, and Cybernetics-Part B, 26 (1): 29-41.
23. Bullnheimer B, Kotsis G, Steauss C. "Parallelization strategies for the ant system. High Performance and Algorithms and Software in Nonlinear Optimization", Applied Optimization, 1998, 24:87-100.
24. Talbi EG, Roux O, Fonlupt C, et al. "Parallel ant colonies for the quadratic assignment problem". Future Generation Computer Systems, 2001, 17:441-449.
25. M. Rahoual, R. Hadji and V. Bachelet, "Parallel ant system for the set covering problem". Third International Workshop on Ant Algorithms, Lecture Notes in Computer Science vol. 2463, Springer-Verlag, Heidelberg, Germany 2002: 262-267.
26. M. Randall, A. Lewis, "A Parallel Implementation of Ant Colony Optimization", Journal of Parallel and Distributed Computing, Volume 62, Number 9, 1421-1432, September 2002
27. Chu SC, Roddick JF, Pan JS. "Ant colony system with communication strategies". Information Science, 2004, 167:63-76.
28. Ellabib I, Calamai P, Basir O. "Exchange strategies for multiple Ant Colony System". Information Sciences, 2007, 177(5): 1248-1264.
29. Stutzle T, Dorigo M. "A short convergence proof for a class of ant colony optimization algorithm", IEEE Transactions on evolutionary computation, 2002, 6(4): 358-365.
30. Gutjahr WJ. "ACO Algorithms with Guaranteed Convergence to the Optional Solution". Info. Processing Lett. 2002, 82(3):145-153.
31. M. Dorigo and C. Blum. "Ant colony optimization theory: A survey". Theoretical Computer Science, 344(2-3):243-278, 2005.

32. Jun O, Yan GR. "*A Multi-Group Ant Colony System*". International Conference on Machine Learning and Cybernetics. New York: IEEE Press, 2004: 117-121.
33. Xiao-Nian Wang, Yuan-jing Feng, Zu-Ren Feng. "*Ant Colony Optimization for Image Segmentation*". Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005, 9(1): 5355-5360.

Image registration for recovering affine transformation using Nelder Mead Simplex method for optimization.

Mehfuza Holia

*Senior Lecturer, Dept. of Electronics
B.V.M engineering College, Vallabh Vidyanagar
Anand-388120, Gujarat ,India.*

mehfuza_1@yahoo.com

Prof. (Dr.) V.K.Thakar

*Professor & Head,
Dept. of Electronics & Communication
A.D.Patel Institute of Technology
New Vallabh Vidyanagar, Anand-388120. Gujarat ,India*

ec.vishvjit.thakar @adit.ac.in

Abstract

In computer vision system sets of data acquired by sampling of the same scene or object at different times or from different perspectives, will be in different coordinate systems. Image registration is the process of transforming the different sets of data into one coordinate system. Registration is necessary in order to be able to compare or integrate the data obtained from different measurements such as different view points, different times, different sensors etc. Image Registration is an important problem and a fundamental task in image processing technique. This paper presents an algorithm for recovering translation parameter from two images that differ by Rotation, Scaling, Transformation and Rotation-scale-Translation (RST) also known as similarity transformation. It is a transformation expressed as a pixel mapping function that maps a reference image into a pattern image. The images having rotational, scaling, translation differences are registered using correlation with Nelder-mead method for function minimization. The algorithm finds the correlation between original image and sensed images. It applies the transformation parameters on sensed images so that maximum correlation between original image and sensed images are achieved. Simulation results (Using Matlab) on images show the Performances of the method.

Keywords: Image registration, Optimization, Correlation, Affine transformation

1. INTRODUCTION

Image registration is the process of transforming the different sets of data into one coordinate system. Registration is necessary in order to be able to compare or integrate the data obtained from different measurements such as different view points, different times, different sensors etc. Image Registration is a crucial step in all image analysis tasks in which the final information is

gained from the combination of various data sources like in image fusion, change detection, and multichannel image restoration .It geometrically aligns two images—the reference and sensed images. The present differences between images are introduced due to different imaging conditions, detection, and multichannel image restoration. Registration is required in remote sensing (multispectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images, integrating information into geographic information systems (GIS), in medicine (combining computer tomography (CT) and NMR data to obtain the complete information about the patient, monitoring tumor growth, treatment verification, comparison of the patient's data with anatomical atlases), and in computer vision (target localization, automatic quality control).

Image registration can be divided into main four groups according to the manner of image acquisition [2].

Different view points: Images of the same scene are acquired from different viewpoints. Examples: Remote sensing, shape recovery

Different times: Images of the same scene are acquired at different times, often on regular basis, and possibly under different conditions.

Examples: automatic change detection for security monitoring, maps from satellite images, remote sensing, healing therapy in medicine, monitoring of the tumor growth

Different sensors: Images of the same scene are acquired by different sensors. The aim is to Integrate of information obtained from different to gain detailed scene presentation.

Examples: offering better spatial resolution in Medical imaging for combination of sensors recording the anatomical body structure like magnetic resonance image (MRI), CCD image sensors, CMOS image sensors, Bayer sensor are different image sensors for better resolution.

Scene to model registration: Images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance maps or digital elevation models (DEM), another scene with similar content (another patient). The aim is to localize the acquired image in the scene/model and/or to compare them. Examples: automatic quality inspection, specimen classification.

There are two main methods for image registration.

1. Area Based methods: Area based methods are correlation like methods or template matching. These methods deal with the images without attempting to detect salient objects. W.K.Pratt [13], has given correlation techniques of image registration. Windows of predefined size or even entire images are used for the correspondence estimation. Available area based methods are: correlation-like methods, Fourier methods, Mutual information methods and optimization methods.

2. Feature Based methods: : Significant regions (forests and fields), lines (region boundaries, coastlines (road, lakes, mountains, rivers) or points (region corners, line intersections, points or curves with high curvature) are features here. They should be distinct, spread all over the image and efficiently detectable. They are expected to be stable in time to stay at fixed positions during the whole experiment. Two sets of features in the reference and sensed images represented by the Control points (points themselves, end points or centers of line features and centers of gravity of regions) have been detected. The aim is to find the pair wise correspondence between them using their spatial relations or various descriptors of features.

In this paper algorithm for recovering translation parameter from two images that differ by RST(Rotation-Scale-Translation).An RST transformation may be expressed as a combination of single translation, single rotation and single scale factor, all operating in the plane of the image. This in fact is a transformation expressed as a pixel mapping function that maps a reference image into a pattern image. RST is also known as geometric spatial Transformation.

2. GEOMETRIC SPATIAL TRANSFORMATION (RST TRANSFORMATION):

Let us consider an image function f defined over a (w, z) coordinate system, undergoes geometric distortion to produced an image g defined over an (x,y) coordinate system. This transformation may be expressed as

$$(x,y) = T \{(w,z)\} \dots \dots \dots (1)$$

Rotation

The new coordinates of a point in the $x-y$ plane rotated by an angle θ around the z -axis can be directly derived through elementary trigonometry [7]. Here $(x,y) = T\{(w,z)\}$, where T is the rotation transformation applied such as

$$\begin{aligned} x &= w \cos \theta - z \sin \theta \\ y &= w \sin \theta + z \cos \theta \end{aligned} \dots \dots \dots (2)$$

which can be represented in matrix form as follows.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix} \dots \dots \dots (3)$$

Scaling

If the x -coordinate of each point in the plane is multiplied by a positive constant S_x , then the effect of this transformation is to expand or compress each plane figure in the x -direction. If $0 < S_x < 1$, the result is a compression; and if $S_x > 1$, the result is an expansion. The same can also be done along the y -axis. This class of transformations is called scaling [7]. Here $(x,y) = T\{(w,z)\}$, where T is the Scaling transformation applied such as

$$\begin{aligned} x &= S_x w \\ y &= S_y z \end{aligned} \dots \dots \dots (4)$$

The above transformation can also be written in matrix form

$$S_x = \begin{bmatrix} S_x & 0 \\ 0 & 1 \end{bmatrix}, \quad S_y = \begin{bmatrix} 1 & 0 \\ 0 & S_y \end{bmatrix} \dots \dots \dots (5)$$

Translation

A translation is defined by a vector $T=(dx,dy)$ and the transformation of the Coordinates is given simply by

$$\begin{aligned} x &= w + dx \\ y &= z + dy \end{aligned} \dots \dots \dots (6)$$

The above transformation can also be written in matrix form

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} w \\ z \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} \dots \dots \dots (7)$$

3. ALGORITHM FOR RECOVERING ROTATION, SCALING AND TRANSLATION

The images are having rotational, scaling, translation differences are registered are registered using correlation with Nelder-mead method [9] for function minimization. The minimum value of correlation between original and sensed images are found by optimization method. The Nelder-Mead algorithm is most widely used methods for nonlinear unconstrained optimization.[9]. The Nelder-Mead method attempts to minimize a scalar valued nonlinear function of n real variables using only function values, without any derivative information. The method approximately finds a locally optimum solution to problem with n variables when the objective function varies smoothly. The Nelder-Mead algorithm [8] was proposed as a method for minimizing a real-valued function $f(x)$ for $x \in R^n$. Four scalar parameters must be specified to define a complete Nelder-Mead method: coefficients of reflection (ρ), expansion (χ), contraction (Υ) and shrinkage(σ).

The following algorithm finds the correlation between original image and sensed images. It applies the transformation parameters on sensed images so that maximum correlation between original image and sensed images are achieved.

If $A(m,n)$ is reference image and $B(m,n)$ is sensed image and then correlation coefficient between A and B is found by

$$r = \frac{\sum_m \sum_n (A_{mn} - A_o)(B_{mn} - B_o)}{\sqrt{\sum_m \sum_n (A_{mn} - A_o)^2 \sum_m \sum_n (B_{mn} - B_o)^2}} \dots\dots\dots (8)$$

Where r is correlation coefficient which value should be between 0 to 1. Minimum value of r shows the dissimilarity of image and for the same images it will have value 1.

A_0 and B_0 represent mean of Image A and B respectively.

Algorithm:

- (1) Acquire the sensed image, It is assumed that the acquire image differs by rotation, scaling and translation as compared to reference image. As a preprocessing process ,the spatial filter is applied to a sensed image.
- (2) Apply rotational, scaling and translational transformation on sense image by placing the initial value for θ, dx, dy, S_x and S_y in equation (3) and (7). Find the correlation coefficient r between reference and sensed image using equation (8).
- (3) Apply NELDER-MEAD simplex method for optimization in step 2, which gives RST transformation parameters when the maximum correlation occur between sensed image and reference image.
- (4) Repeat the process by changing the value of θ, dx, dy, S_x and S_y till optimum value of r is obtained.
- (5) Step 4 gives the value of θ, dx, dy, S_x and S_y for the maximum correlation. This obtained transformation parameters are applied to sensed image using equation (3) and (7) to get the aligned image.
- (6) If $r_{max} < r_{threshold}$, obtained θ, dx, dy, S_x and S_y in step 4 will be taken as initial transformation parameters and obtained aligned image in step 5 will be taken as sensed image and repeat steps 2 to 5.

If $r_{max} \geq r_{threshold}$, the aligned image obtained in step 5 is the registered image.

4. RESULTS

In figure 1, the original images of different sizes are shown .



Image 1



Image 2



Image 3



Image 4

FIGURE 1 Original Images

Table 1 Correlation of sensed images differ by only rotation from original image

Sr. No	Original Images	Sensed image(rotated by θ angle from original)	Maximum Correlation r
1.	Image1.jpg (576 x 720)	5°	0.9996
		10°	0.9996
		15°	0.9996
		20°	0.9996
		25°	0.9997
		30°	0.9997
		35°	0.9997
		40°	0.9997
2.	Image2.jpg (113x 150)	5°	0.9811
		10°	0.9869
		15°	0.9879
		20°	0.9883
		25°	0.9890
		30°	0.9891
		35°	0.9893
3.	Image3.jpg (245 x326)	5°	0.9967
		10°	0.9970
		15°	0.9973
		20°	0.9974
		25°	0.9976
		30°	0.9976
		35°	0.9978
		40°	0.9978
	45°	0.9989	

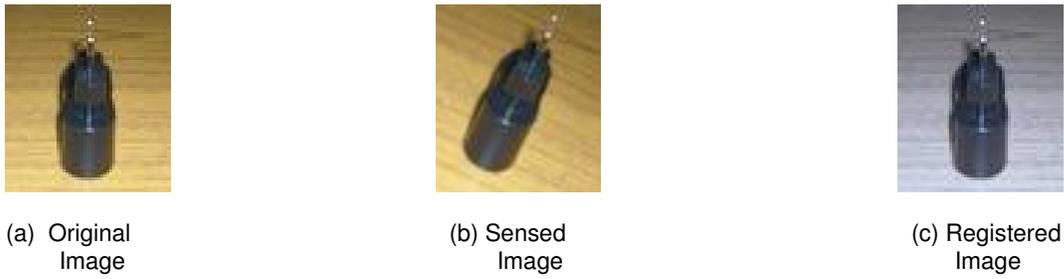


FIGURE 2 Registration of Sensed image rotated by 25 degree from original.

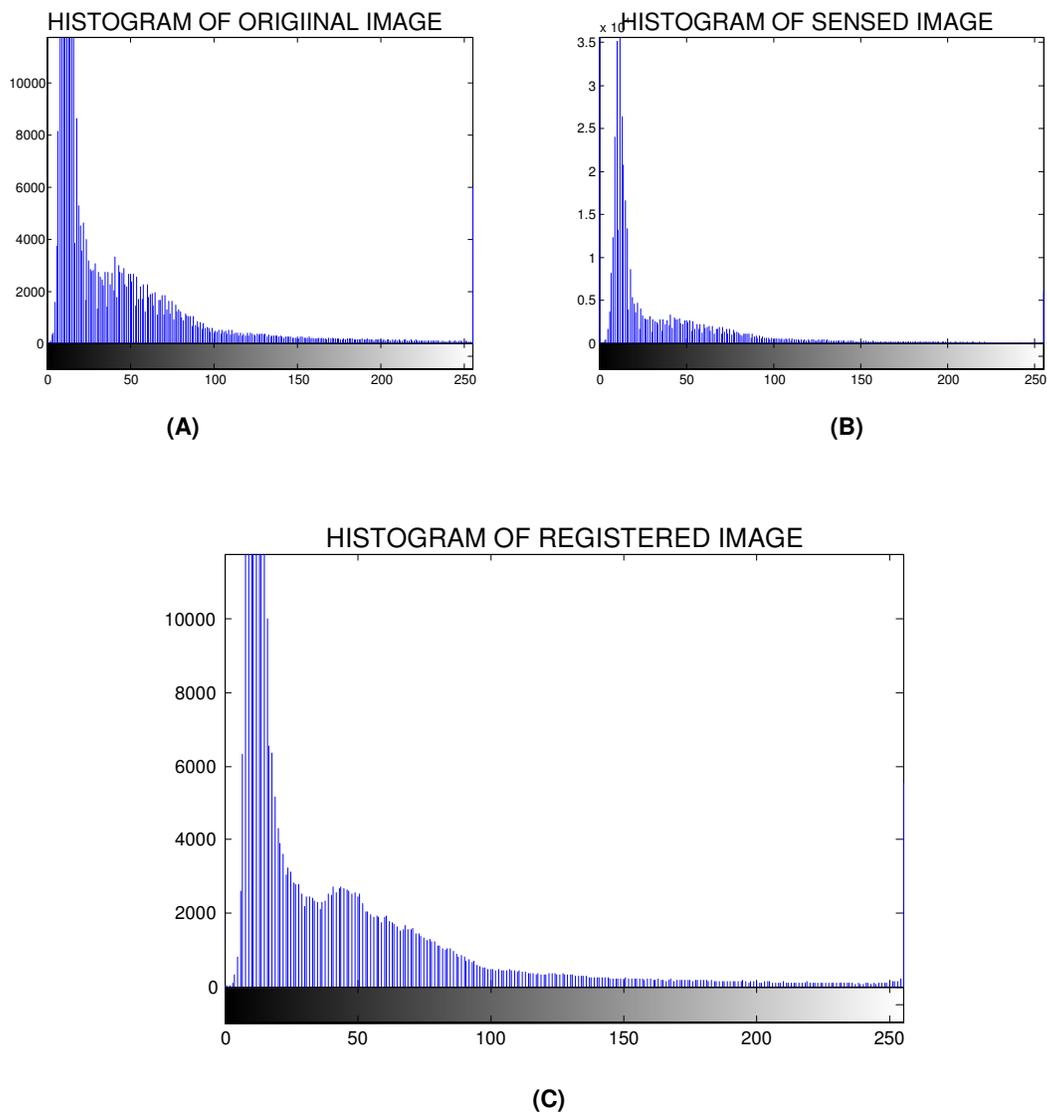
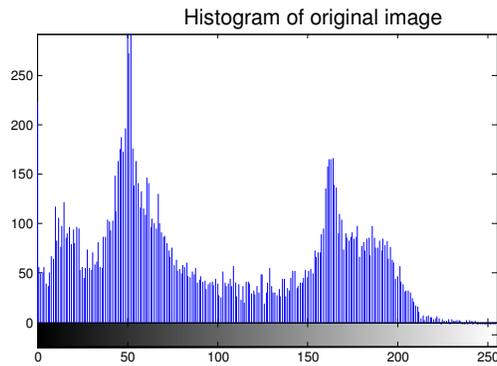


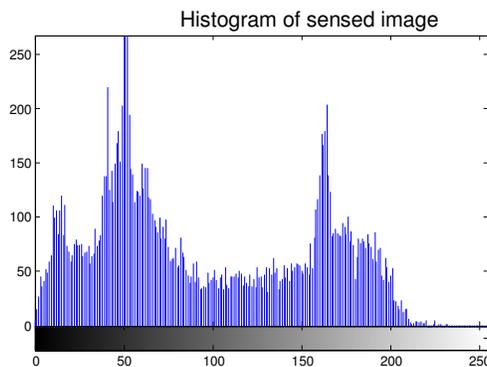
FIGURE 3 (A) Histogram of Original Image (image1.jpg)
(B) Histogram of Sensed Image (differ by rotation of 15 degree from original)
(C) Histogram of registered image

TABLE 2. Correlation of sensed images differ by only scaling from original image

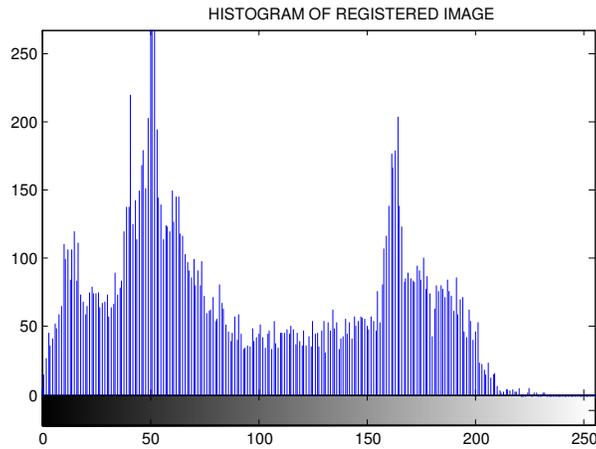
Sr. No	Original Images	Sensed image (scaled by <i>S</i> factor from original)	Size of sensed Image	Maximum Correlation
1.	Image1. jpg (576x720)	0.6	345 x 432	0.9972
		0.7	403 x 503	0.9971
		0.8	460 x 576	0.9989
		0.9	518 x 648	0.9968
		1.1	633 x 792	0.9966
		1.2	691 x 864	0.9967
		1.3	748 x 936	0.9968
		1.4	806 x 1007	0.9955
2.	Image2. jpg (113x150)	0.6	67 x 90	0.9044
		0.7	79 x 105	0.9676
		0.8	90x120	0.9733
		0.9	101x135	0.9843
		1.1	124x165	0.9893
		1.2	135x180	0.9764
		1.3	146x195	0.9949



(A)



(B)



(C)
FIGURE 4 (A) Histogram of original image (Image2.jpg 113 x150)
 (B) Histogram of sensed image (Image2. jpg 90x120) scaled by 0.8 from original.
 (C) Histogram of Registered image (image2.jpg 113 x150)

TABLE 3 Correlations of sensed images differ by only translation from original image.

Sr. No	Original Images	Sensed image (transformed by dx and dy from original)		Maximum Correlation
		dx	dy	
1.	Image1. jpg (576x720)	5	5	1
		10	25	1
		-3	-4	1
		-20	15	1
		50	-50	1
		120	70	1
2.	Image2. jpg (113x150)	0	100	1
		34	-7	1
		-200	12	1
		-130	-78	1
		56	90	1
		198	-198	1
3.	Image3. jpg (245x326)	34	-34	1
		12	200	1
		-8	21	1
		-65	-90	1
		-75	180	1
		75	-180	1
4.	Image4. jpg (113x150)	0	100	1
		34	7	1
		-200	-12	1
		-130	78	1
		56	90	1
		198	-198	1
		-8	21	1
		75	-180	1



FIGURE 5: Sensed image is translated by $dx= 34$, $dy=-34$

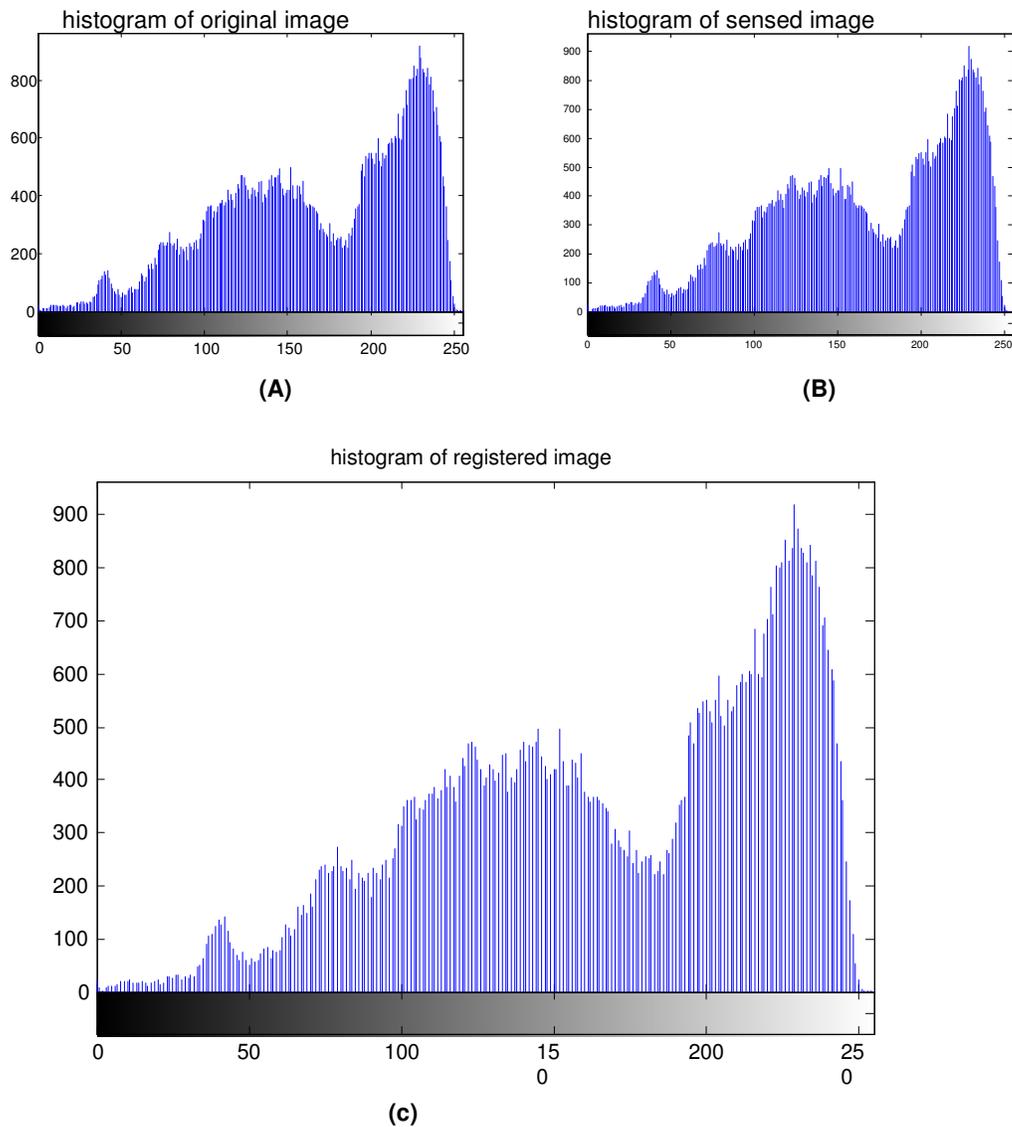


FIGURE 6 (A)Histogram of original image(Image3.jpg)
(B)Histogram of translated sensed image($dx=-15$, $dy=20$)
(C)Histogram of registered image (image3.jpg)



FIGURE 7 Registration of sensed images differ by rotation, scaling and translation from original image.

For FIGURE 7 Size of original image: 427 x 317
Size of sensed image: 423 x 319
Sensed image rotated by original : 3
Correlation : 0.9909

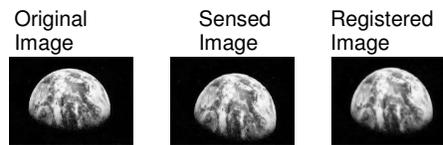


FIGURE 8 Registration of sensed images differ by rotation, scaling and translation from original image.

For FIGURE 8 Size of image: 384 x 276
Size of sensed image: 382 x 255
Sensed image rotated by original : 7
Correlation : 0.9602

In the results by observing the histogram of Image1.jpg we can see that it is having narrow range of gray level so in the NM simplex methods it also goes to the expansion along with reflection. So it will take more time to find the value at which the maximum correlation achieved. In histogram of Image2.jpg and Image3.jpg we can observe that pixels are having all range of gray level(0-255), so in the NM simplex method within reflection only maximum correlation is achieved. It will not go to expansion so it takes less time to compare to Image1.jpg. As shown in TABLE 1, this method gives the registration of rotated sensed image for maximum 45 degree. From TABLE 2 it can be seen that registration can be achieved if the sensed image is scaled between 0.5 to 1.5 times original image. TABLE 3 shows the result in which the sensed image is translated and using this method we can get the registration for any translation parameters. Figures 3,4 and 6 show the histograms of original, sensed and registered images. Histograms of registered images are same as original images.

5. CONCLUSIONS

In this paper authors have presented a technique for registering the images differ by rotation, scaling and translation. The techniques presented can be applied a wide class of problems involving determination of correspondence between two images related by a similarity and affine transformation. This algorithm is useful for images taken from same sensor and which are misaligned by small transformation such as scaling, rotation or translation. But limitation of this algorithm is inability to register the dissimilar images, having different information. Here different optimization algorithm is implemented for maximization of correlation among which NM simplex method gives the best result.

6. REFERENCES

- [1] A. Goshtasby, G.C. Stockman, *A region-based approach to digital image registration with sub pixel accuracy*, IEEE Transactions on Geoscience and Remote Sensing 24 (1986) 390-399.
- [2] Barbara Zitova, Jan Flusser, *Image registration methods: a survey*. Academy of Sciences of the Czech Republic, Image and vision computing 21(2003) 977-1000.
- [3] R.N. Bracewell, *the Fourier Transform and Its Applications*, McGraw-Hill, New York, 1965.
- [4] E.D. Castro, C. Morandi, *Registration of translated and rotated images using finite Fourier transform*, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (1987) 700–703.
- [5] J. B. Antoine Maintz_ and Max A. Viergever, *A Survey of Medical Image Registration*. Image Sciences Institute, Utrecht University Hospital, Utrecht, the Netherlands
- [6] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, *Digital Image Processing Using MATLAB*, Pearson Education.
- [7] Jamal T. Manassah, *Elementary Mathematical and Computational Tools for electrical and Computer engineers*, CRC Press, Boca Raton London New York Washington
- [8] J. A. Nelder and R. Mead, *A simplex method for function minimization*, Computer Journal 7 (1965), 308-313.
- [9] Jeffrey C. Lagarias, James A. Reedsz, Margaret H. Wright, And Paul E. Wright, *Convergence properties of the NELDER-MEAD Simplex method in low dimensions*. SIAM J Optimization, 1999, Vol. 9, pp 112-1247.
- [10] P.Viola, W.M.Wells, *Alignment by maximization of mutual information*, International Journal of Computer Vision 24 (1997) 137-154.
- [11] P.The´venaz, M. Unser, “An efficient mutual information optimizer for multi resolution image registration”, Proceedings of the IEEE International Conference on Image Processing ICIP’98, Chicago,IL, 2000 833-837.
- [12] A. Roche, G. Malandain, N. Ayache, *Unifying maximum likelihood approaches in medical image registration*, International Journal of Imaging Systems and Technology 11(2000) 71-80.
- [13] W.K. Pratt, *Correlation techniques of image registration*, IEEE Transactions on Aerospace and Electronic Systems 10 (1974) 353-358.
- [14] Raman Maini, Himanshu Aggarwal, *Study and Comparison of Various Image Edge Detection Techniques* International Journal of Image Processing (IJIP), 3(1):1-12, 2009

A Texture Based Methodology for Text Region Extraction from Low Resolution Natural Scene Images

S. A. Angadi

*Department of Computer Science & Engineering
Basaveshwar Engineering College
Bagalkot, 587102, Karnataka, India*

vinay_angadi@yahoo.com

M. M. Kodabagi

*Department of Computer Science & Engineering
Basaveshwar Engineering College
Bagalkot, 587102, Karnataka, India*

malik123_mk@rediffmail.com

Abstract

Automated systems for understanding display boards are finding many applications useful in guiding tourists, assisting visually challenged and also in providing location aware information. Such systems require an automated method to detect and extract text prior to further image analysis. In this paper, a methodology to detect and extract text regions from low resolution natural scene images is presented. The proposed work is texture based and uses DCT based high pass filter to remove constant background. The texture features are then obtained on every 50x50 block of the processed image and potential text blocks are identified using newly defined discriminant functions. Further, the detected text blocks are merged and refined to extract text regions. The proposed method is robust and achieves a detection rate of 96.6% on a variety of 100 low resolution natural scene images each of size 240x320.

Keywords: Text Region Extraction, Texture Features, Low Resolution Natural scene image.

1. INTRODUCTION

As the people move across world for business, field works and/or pleasure, they find it difficult to understand the text written on display boards in foreign environment. In such a scenario, people either look for guides or intelligent devices that can help them in providing translated information to their native language. As most of the individuals carry camera embedded, hand held devices such as mobile phones and PDA's, there is a possibility to integrate technological solutions into such systems in order to provide facilities for automatically understanding display boards in foreign environment. These facilities may be provided as an integral solution through web service as necessary computing function, which are not available in hand held systems. Such web based hand held systems must be enabled to capture natural scene images containing display boards and query the web service to retrieve translated localized information of the text written on display boards.

The written matter on display/name boards provides information necessary for the needs and safety of people, and may be written in languages unknown. And the written matter can be street names, restaurant names, building names, company names, traffic directions, warning signs etc. Hence, lot of commercial and academic interest is veered towards development of techniques for web service based hand held systems useful in understanding written text in display boards. There is a spurt of activity in development of web based intelligent hand held tour guide systems, blind assistants to read written text and Location aware computing systems and many more in recent years. A few such works are presented in the following and a more elaborate survey of related works is given in the next section. A *point by photograph paradigm* where users can specify an object by simply taking picture to retrieve matching images from the web is found in [1]. The *comMotion* is a location aware hand held system that links personal information to locations. It reminds users about shopping list when he/she nears a shopping mall [2]. At Hewlett Packard (HP), mobile Optical Character Reading (OCR) applications were developed to retrieve information related to the text image captured through a pen-size camera [3]. Mobile phone image matching and retrieval has been used by insurance and trading firms for remote item appraisal and verification with a central database [4].

The image matching and retrieval applications cannot be embedded in hand held devices such as mobile phones due to limited availability of computing resources, hence such services are being developed as web services. The researchers have also worked towards development of web based intelligent hand held tour guide systems. The *cyberGuide* [5] is an intelligent hand held tour guide system, which provides the information based on user's location. The *cyberGuide* continuously monitors the users location using Global Positioning System (GPS) and provides new information at the right time. Museums could provide these tour guides to visitors allowing them to take personalized tours observing any displayed object. As the visitors move across museum floors, the information about the location is pushed to hand held tour guides. The research prototypes used to search information about an object image captured by cameras embedded in mobile phones are described in [6-7].

The state of art hand held systems available across the world are not automated for understanding written text on display boards in foreign environment. Scope exists for exploring such possibilities through automation of hand held systems. One of the very important processing steps for development of such systems is automatic detection and extraction of text regions from low resolution natural scene images prior to further analysis. The written text provides important information and it is not an easy problem to reliably detect and localize text embedded in natural scene images [8]. The size of the characters can vary from very small to very big. The font of the text can be different. Text present in the image may have multiple colors. The text may appear in different orientation. Text can occur in a complex background. And also the textual and other information captured is affected by significant degradations such as perspective distortion, blur, shadow and uneven lighting. Hence, the automatic detection and segmentation of text is a difficult and challenging problem. Reported works have identified a number of approaches for text localization from natural scene images. The existing approaches are categorized as connected component based, edge based and texture based methods. Connected component based methods use bottom up approach to group smaller components into larger components until all regions are identified in the image. A geometrical analysis is later needed to identify text components and group them to localize text regions. Edge based methods focus on the high contrast between the background and text and the edges of the text boundary are identified and merged. Later several heuristics are required to filter out nontext regions. But, the presence of noise, complex background, and significant degradation in the low resolution natural scene image can affect the extraction of connected components and identification of boundary lines, thus making both the approaches inefficient. Texture analysis techniques are good choice for solving such a problem as they give global measure of properties of a region.

In this paper, a new texture based text detection and segmentation method is proposed. The proposed method uses high pass filtering in the DCT domain to suppress most of the background. Later texture features [33] such as *homogeneity* and *contrast* are computed on image blocks to identify and segment text regions in the image. Each unit block is classified as

either text or nontext based on newly defined discriminant functions. In addition, merging algorithms are used to merge text blocks to obtain text regions. The regions are further refined using post processing. The proposed method is robust enough to detect text regions from low resolution natural scene images, and achieves a detection rate of 96.6%. The system is developed in MATLAB and evaluated for 100 low resolution natural scene images on Intel Celeron (1.4GHz) computer. It was observed that the processing time lies in the range of 6 to 10 seconds due to varying background. The proposed methodology is described in the following sections of the paper.

The rest of the paper is organized as follows; the detailed survey related to text extraction from natural scene images is described in section 2. The proposed method is presented in Section 3. The experimental results and analysis are given in Section 4. Section 5 concludes the work and lists future directions.

2. RELATED WORK

The web based hand held systems useful in understanding display boards requires analysis of natural scene images to extract text regions for further processing. A number of methods for text localization have been published in recent years and are categorized into connected component based, edge based and texture based methods. The performance of the methods belonging to first two categories is found to be inefficient and computationally expensive for low resolution natural scene images due to the presence of noise, complex background and significant degradation. Hence, the techniques based on texture analysis have become a good choice for image analysis, and texture analysis is further investigated in the proposed work.

A few state of the art approaches that use texture features for text localization have been summarized here; the use of horizontal window of size 1×21 (Mask size) to compute the spatial variance for identification of edges in an image, which are further used to locate the boundaries of a text line is proposed in [9]. However, the approach will only detect horizontal components with a large variation compared to the background and a processing time of 6.6 seconds with 256×256 images on SPARC station 20 is reported. The Vehicle license plate localization method that uses similar criteria is presented in [10]. It uses time delay neural networks (TDNNs) as a texture discriminator in the HSI color space to decide whether the window of an image contains a license plate number. The detected windows are later merged for extracting license plates. A multi-scale texture segmentation schemes are presented in [11-12]. The methods detect potential text regions based on nine second-order Gaussian derivatives and is evaluated for different images including video frames, newspapers, magazines, envelopes etc. The approach is insensitive to the image resolution and tends to miss very small text and gives a localization rate of 90%.

A methodology that uses frequency features such as the number of edge pixels in horizontal and vertical directions and Fourier spectrum to detect text regions in real scene images is discussed in [13]. The texture-based text localization method using Wavelet transform is proposed in [14]. The techniques for text extraction in complex color images, where a neural network is employed to train a set of texture discrimination masks that minimize the classification error for the two texture classes: text regions and non-text regions are reported in [16-17].

Learning-based methods for localizing text in documents and video are proposed in [15] and [18]. The method [18] is evaluated for various video images and the text localization procedure required about 1 second to process a 352×240 image on Sun workstation. And for text detection a precision rate of 91% and a recall rate of 92.8% is reported. This method was subsequently enhanced for skewed text in [19]. A work similar to the proposed method which uses DCT coefficients to capture textural properties for caption localization is presented in [8]. The authors claim that the method is very fast and gives a detection rate of 99.17%. However, the precise localization results are not reported. In recent years, several approaches for sign detection, text detection and segmentation from natural images are also reported [20-32].

Out of many works cited in the literature it is generally agreed that the robustness of texture based methods depends on texture features extracted from the window/block or the region of interest that are used by the discriminant functions for classification decisions. The probability of misclassification is directly related to the number and quality of details available in texture features. Hence, the extracted texture features must give sufficient information to distinguish text from the background. And also suppression/removal of background information is an essential preprocessing step needed before extracting distinguishable texture features to reduce the probability of misclassification. But most of the works cited in the literature directly operate on the image without suppressing the background. Hence, there is a scope to explore such possibilities. The proposed method performs preprocessing on the image for suppressing the uniform background in the DCT domain and further uses texture features for text localization. The detailed description of the proposed methodology is given in the next section.

3. TEXTURE BASED METHODOLOGY FOR TEXT EXTRACTION

The proposed methodology is texture based, and operates on low resolution natural scene images captured by cameras embedded in mobile phones to detect and segment text regions. The methodology uses high pass filter in the DCT domain to suppress the background, and texture features such as *homogeneity* and *contrast* to detect and segment text regions. The processing is carried out on 8x8 sized image blocks during background suppression phase and the remaining phases use 50x50 sized image blocks. There are several benefits of using larger sized image blocks for extracting texture features. One such benefit is, the larger size image blocks cover more details and hence extracted features give sufficient information for correct classification of blocks into text and nontext categories. The other benefits include; robustness and insensitiveness to variation in size, font and alignment.

The proposed method comprises of 5 phases; Background removal/suppression in the DCT domain, texture features computation on every 50x50 block and obtaining a feature matrix D , Classification of blocks, merging of text blocks to detect text regions, and refinement of text regions. The block schematic diagram of the proposed model is given in figure 1. The detailed description of each phase is presented in the following subsections;

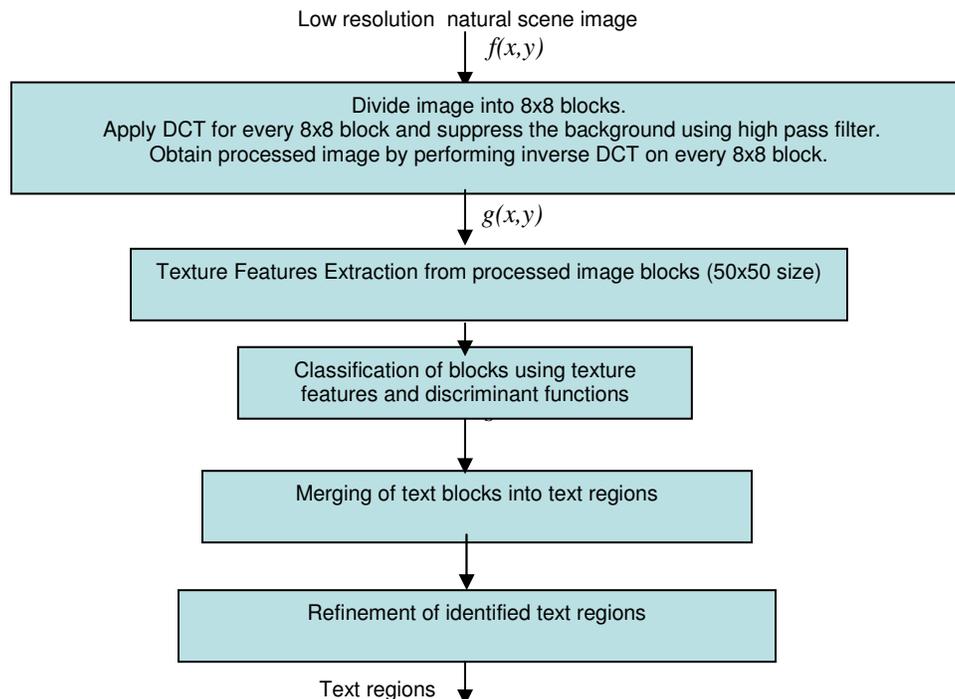


FIGURE 1: Block diagram of proposed method

3.1 Background removal/suppression in the DCT domain

The removal of constant background from the low resolution natural scene image resulting from sources such as building walls, windows, trees, and others is an essential preprocessing step, required to reduce the effort needed for further image analysis. The proposed method uses DCT coefficients to remove constant background. The DCT coefficient values are computed on every 8x8 block of the image. In the corresponding DCT block, the values from top to bottom indicate horizontal variations, with increasing frequencies. The value at the top left corner (first entry in the DCT matrix/block) corresponds to DC component. And the values from left to right indicate vertical variations, with increasing frequencies. Therefore the top left values which represent low frequency components contain most of the energy, while the high frequency components that are located towards the bottom right corner are mostly blank (contain zero values). Hence, the constant background is removed successfully by applying a high pass filter that attenuates the DC component of every 8x8 DCT block of the image $f(x, y)$ of size $L \times M$, where x and y are spatial coordinates. The transform function of high pass filter that operates on every 8x8 DCT block is given in equation 1. Later the processed/background suppressed image $g(x,y)$ is obtained by applying inverse DCT on every 8x8 DCT block, which will be used in subsequent phases. The steps of background suppression are depicted in equations 2, 3 and 4. The block diagram of background suppression using DCT is given in figure 2.

$$H(u,v) = \begin{cases} 0 & (u,v)=(1,1), \text{ where } u=1\dots8, v=1\dots8 \\ 1 & \text{Otherwise} \end{cases} \quad (1)$$

The high pass filter attenuates the DC component by storing value zero in every top left coordinate of 8x8 DCT block. This process is also called removing low frequency component from top left corner from every DCT block.

$$G(u,v) = DCT[f(x,y)] \text{ where } 1 \geq x, u \leq L, \text{ and } 1 \geq y, v \leq M \quad (2)$$

$$P(u, v) = H(u, v) G(u, v). \quad (3)$$

$$g(x, y) = DCT^{-1}[P(u, v)]. \quad (4)$$

Where,
 $G(u,v)$ is DCT matrix of input image $f(x,y)$.
 $P(u,v)$ is Processed DCT matrix.
 $g(x,y)$ is background suppressed image.

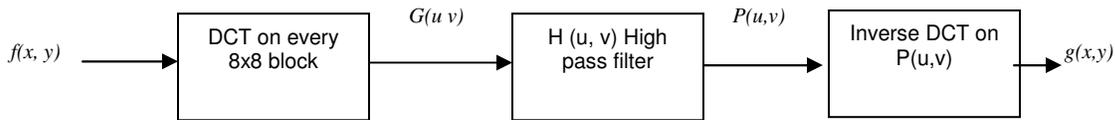


FIGURE 2: High Pass Filter for Background Removal using DCT

3.2 Features Extraction

In this phase, the *texture* features such as *homogeneity* and *contrast* are obtained from every 50 x 50 block of the processed image $g(x,y)$ at $0^\circ, 45^\circ, 90^\circ$, and 135° orientations. Totally 8 features are extracted from every block and are stored into a feature vector X_i (Subscript “i” corresponds to i^{th} block). The feature vector X_i also records block coordinates which corresponds to minimum and maximum row and column numbers of the block. Feature vectors of all N blocks are combined to form a feature matrix D as depicted in equation 5. The feature vector is described in equation 6.

$$D = [X_1, X_2, X_3, \dots, X_N]^T \quad (5)$$

$$X_i = [r_{min}, r_{max}, c_{min}, c_{max}, f_j, j=1,8]; \quad (6)$$

Where;

$r_{min}, r_{max}, c_{min}, c_{max}$ corresponds to coordinates of i^{th} block in terms of minimum and maximum row and column numbers.
 f_1 and f_2 corresponds to homogeneity and contrast at 0 degree orientation.
 f_3 and f_4 corresponds to homogeneity and contrast at 45 degree orientation.
 f_5 and f_6 corresponds to homogeneity and contrast at 90 degree orientation.
 f_7 and f_8 corresponds to homogeneity and contrast at 135 degree orientation.

The features *homogeneity* and *contrast* are calculated as in equations 7 and 8.

$$\text{Homogeneity} = \sum_{i=1}^Q \sum_{j=1}^Q (P(i, j) / R)^2 \tag{7}$$

$$\text{Contrast} = \sum_{n=1}^{Q-1} n^2 \sum_{|i-j|=n}^Q (P(i, j) / R) \tag{8}$$

Where R is given in equation 9.

$$R = \sum_{i=1}^Q \sum_{j=1}^Q P(i, j) \tag{9}$$

P corresponds to cooccurrence matrix at a given degree.
 R is normalized value of cooccurrence matrix P .
 N is total number of blocks.
 $Q \times Q$ is dimension of block size which is chosen as 50×50 .

3.3 Classification

The classification phase of the proposed model uses discriminant functions to classify every block into two classes' w_1 and w_2 based on feature vector X_i . Where, w_1 corresponds to text blocks and w_2 corresponds to nontext blocks category. The discriminant functions uses two thresholds T_1 initialized to 0.4 and T_2 to 50 corresponding to *homogeneity* and *contrast* values respectively. The values 0.4 for T_1 and 50 for T_2 are heuristics chosen based on experiments conducted on several different images and are used by classifiers to produce correct classification results. The discriminant functions d_1 and d_2 together decides a block as text block if the *homogeneity* and *contrast* features values at $0^\circ, 45^\circ, 90^\circ$, and 135° orientations are less than thresholds T_1 and T_2 . The classification rules using discriminant functions are stated in equations 10 and 11.

Given a feature matrix D of features X_i , assign the corresponding image block to;

Class w_1 if $d_1(X_i)$ is satisfied, and $d_2(X_i)$ is satisfied
 Class w_2 otherwise

Where,

$d_1(X_i)$ is a discriminant function which defines/specifies constraint on homogeneity value.
 $d_2(X_i)$ is a discriminant function which defines/specifies constraint on Contrast value.

$$d_1(X_i) = \begin{cases} \text{satisfied} & \text{if } X_i(f_j) \leq T_1, \forall i=1, N \text{ and } j = 1,3,5,7 \\ \text{Not satisfied} & \text{otherwise} \end{cases} \tag{10}$$

$$d_2(X_i) = \begin{cases} \text{satisfied} & \text{if } X_i(f_j) \geq T_2 \quad \forall i=1, N \text{ and } j = 2,4,6,8 \\ \text{Not satisfied} & \text{otherwise} \end{cases} \tag{11}$$

Where,

T_1 corresponds to threshold on homogeneity ($T_1 = 0.4$), chosen empirically.
 T_2 corresponds to threshold on contrast $T_2 = (50)$, chosen empirically.

The coordinates $r_{min}, r_{max}, C_{min}, C_{max}$ which corresponds to minimum and maximum row and column numbers of every classified text block C_i will be stored into a new vector B as in equation 12, which are later used during merging process to obtain contiguous text regions.

$$B = [C_i, i = 1, N1] \tag{12}$$

$$C_i = [r_{min}, r_{max}, C_{min}, C_{max}] \tag{13}$$

Where,

C_i corresponds to i^{th} text block.
 $r_{min}, r_{max}, C_{min}, C_{max}$ corresponds to the coordinates of i^{th} block in terms of minimum and maximum row and column numbers.
 $N1$ is Number of text blocks.

The thresholds $T1$ and $T2$ are highly dependent on the size of the block. For 50 X 50 block size the values 0.4 for $T1$ and 50 for $T2$ have given correct classification results and have been chosen after exhaustive experimentation. It is also found during the experiments that, the probability of misclassification decreases by careful selection of values for $T1$ and $T2$. After this phase, the classified text blocks are subjected to merging process as given in the next section.

3.4 Merging of text blocks to detect text regions

The merging process combines the potential text blocks C_i , connected in rows and columns, to obtain new text regions r_i , whose coordinates are recorded into vector R as depicted in equation 14.

$$R = [r_i, i = 1, W] \tag{14}$$

$$r_i = [r_{min}, r_{max}, C_{min}, C_{max}] \tag{15}$$

r_i corresponds to i^{th} text region
 $r_{min}, r_{max}, C_{min}, C_{max}$ corresponds to the coordinates of i^{th} text region.
 W is number of text regions.

The merging procedure is described in algorithm 1;

Algorithm1

Input: Vector B which contains coordinates of identified text blocks

Output: Vector R which records text regions

Begin

1. Choose the first block C_s from vector B .
2. Initialize coordinates of a new text region r_i to coordinates of **block** C_s .
3. **Select** next block C_p from the vector B .
4. **if** (the block C_p is connected to r_i in row or column) **then**
 begin
 Merge and update coordinates $r_{min}, r_{max}, C_{min}, C_{max}$ of block r_i .
 $r_{min} = \min\{r_i[r_{min}], C_p[r_{min}]\}$ $r_{max} = \max\{r_i[r_{max}], C_p[r_{max}]\}$
 $C_{min} = \min\{r_i[C_{min}], C_p[C_{min}]\}$ $C_{max} = \max\{r_i[C_{max}], C_p[C_{max}]\}$
 else
 Store text region r_i into vector R .
 Initialize coordinates of a new text region r_i to coordinates of current **block** C_p .
 end
5. Repeat steps 2 -5 until $p=N1$.

End

3.5 Refinement of text regions

The refinement phase is a post processing step used to improve the detection accuracy. This phase is concerned with refining the size of the detected text regions to cover small portions of missed text present in adjacent undetected blocks and unprocessed regions. The refinement process is carried out in two steps; Adjacent undetected blocks processing and Combining unprocessed region. The functionality of each step is described below;

3.5.1 Adjacent undetected blocks processing

In this step, every detected text region r_i is refined either by combining the entire size (50 rows and 50 columns) or selected rows and columns of adjacent undetected blocks (in rows and columns) which contain/cover small portions of missed text. The *average contrast* value $(f_2+f_4+f_6+f_8) / 4$ is computed for every adjacent undetected block (in row or column), If the computed value is greater than or equal to 35 then the entire block size is combined with region r_i for refinement assuming that the missed adjacent block may contain significant text information. The heuristic value 35 is chosen empirically based on experiments. Similarly, if the *average contrast* value is between 5-9, 10-19, and 20-34, then 5, 10, and 20 respective adjacent rows/columns are added to region r_i for refinement. Again the heuristic values 5, 10 and 20 are chosen empirically based on experiments and results were encouraging. The procedure for combining adjacent undetected blocks is described in algorithm 2.

Algorithm 2

Input: - Vector R which contains coordinates of extracted text regions.
 - Vector D which contains texture features and coordinates of all 24 blocks of preprocessed image.
 - Unprocessed 40 rows and 20 columns of the preprocessed image

Output: Vector R which records refined text regions.

Begin

1. Start with first text region r_i .
2. Select the next feature vector X_p from the feature matrix D .
3. **if** (the feature vector X_p is connected to text region r_i in row or column) **then**

begin

 - 3.1 Find average *contrast* value = $(f_2+f_4+f_6+f_8) / 4$.
 - 3.2 **if** (average *contrast* value is > 35) **then**

begin

Merge and update coordinates r_{min} , r_{max} , C_{min} , C_{max} of text region r_i by adding adjacent block.

$$r_{min} = \min\{r_i[r_{min}], C_p[r_{min}]\} \quad r_{max} = \max\{r_i[r_{max}], C_p[r_{max}]\}$$

$$C_{min} = \min\{r_i[C_{min}], C_p[C_{min}]\} \quad C_{max} = \max\{r_i[C_{max}], C_p[C_{max}]\}$$

else

if (average *contrast* value is between 5 to 9) **then**

begin

$$C_{min} = C_{min} - 5; \text{ // add 5 columns if feature vector } X_p \text{ left connected}$$

OR

$$C_{max} = C_{max} + 5; \text{ // add 5 columns if feature vector } X_p \text{ right connected}$$

OR

$$r_{min} = r_{min} - 5; \text{ // add 5 rows if feature vector } X_p \text{ top connected}$$

OR

$$r_{max} = r_{max} + 5; \text{ // add 5 rows if feature vector } X_p \text{ bottom connected}$$

end

if (average *contrast* value is between 10 to 19) **then**

begin

$$C_{min} = C_{min} - 10; \text{ // add 5 columns if feature vector } X_p \text{ left connected}$$

OR

$$C_{max} = C_{max} + 10; \text{ // add 5 columns if feature vector } X_p \text{ right connected}$$

OR

$$r_{min} = r_{min} - 10; \text{ // add 5 rows if feature vector } X_p \text{ top connected}$$

OR

$$r_{max} = r_{max} + 10; \text{ // add 5 rows if feature vector } X_p \text{ bottom connected}$$

end

if (average *contrast* value is between 20 to 34) **then**

begin

$$C_{min} = C_{min} - 20; \text{ // add 5 columns if feature vector } X_p \text{ left connected}$$

OR

$$C_{max} = C_{max} + 20; \text{ // add 5 columns if feature vector } X_p \text{ right connected}$$

OR

$$r_{min} = r_{min} - 20; \text{ // add 5 rows if feature vector } X_p \text{ top connected}$$

OR

$$r_{max} = r_{max} + 20; \text{ // add 5 rows if feature vector } X_p \text{ bottom connected}$$

end

end
4. Repeat steps 2-3 until $p = N1$.

5. Select next text region $r_i = [r_{min}, r_{max}, c_{min}, c_{max}]$ from R .
6. Repeat steps 2-5 until all text regions are refined.

End

3.5.2 Combining unprocessed regions

The proposed method works by dividing an image of size 240x320 into 50x50 sized blocks till the completion of phase 3.3. Hence the remaining 40 rows and 20 columns will be left unprocessed after phase 3.3. As the unprocessed rows and columns may also contain text information they need to be processed for further refinement of detected text regions. In this step, the detected text regions are further refined by processing and adding adjacent unprocessed area to cover missed portion containing small text. During processing only *average contrast* feature value $(f_2+f_4+f_6+f_8) / 4$ is computed from unprocessed area, and if the *average contrast* value is greater than or equal to the threshold 50, then the entire size of the unprocessed region is combining with adjacent detected text regions for refinement. The heuristic value 50 is chosen empirically based on experiments and has produced good results.

The proposed methodology is robust and performs well for different sizes of font and image resolution. The block size is an important design parameter whose dimension must be properly chosen to make the method more robust and insensitive to variation in size, font and its alignment. The approach also detects nonlinear text regions and results are presented in the next section. However, the method detects larger text regions than the actual size when the image background is more complex containing trees, vehicles, and other details from sources of outdoor scenes. The method takes about 6 to 10 seconds of processing time based on the complexity of background contained in the image.

4. RESULTS AND ANALYSIS

The proposed methodology for text region detection and extraction has been evaluated for 100 indoor and outdoor low resolution natural scene display board images (having 2400, 50x50 size blocks) with complex backgrounds. The experimental tests were conducted for most of the images containing Kannada text and few containing English text and results were highly encouraging. The experimental results of processing a typical display board image containing text with varying background is described in section 4.1. And the results of processing several other display board images dealing with various issues and the overall performance of the system are reported in section 4.2.

4.1 Text Region Extraction of a typical display board image

A display board image of size 240x320 given in figure 3a, containing text information having smaller and bigger Kannada characters and complex backgrounds such as building walls, doors, and uneven lighting conditions is initially preprocessed to suppress the background using a high pass filter in the DCT domain. The experimental values of applying a high pass filter for the first 8x8 DCT block of image in figure 3a are given in Table 1.



FIGURE 3: a) Original image b) Background Suppressed image c) Detected text region before post processing d) Refined text region after post processing

TABLE1: Preprocessing of first 8x8 block of an image in figure 3a

Step1: 8x8 first image block	Step2: 8x8 first DCT block
159 159 159 159 159 159 159 159	1.2821 -0.0019 0.0007 -0.0006 0.0004 -0.0002 0.0001 0.0002
159 159 159 159 159 159 159 159	-0.0062 0.0009 -0.0001 0.0001 0.0001 -0.0003 0.0002 -0.0002
160 160 160 160 160 160 160 161	-0.0015 0.0013 -0.0004 0.0003 -0.0004 0.0004 -0.0003 -0.0000
160 160 160 160 160 160 160 161	-0.0001 -0.0011 -0.0002 -0.0000 -0.0002 0.0003 -0.0001 0.0003
160 160 160 160 161 161 161 162	-0.0006 0.0003 -0.0001 0.0003 0.0001 -0.0004 0.0002 -0.0001
160 160 161 161 161 161 162 162	0.0007 0.0002 0.0004 -0.0003 0.0000 0.0002 -0.0001 -0.0001
161 161 161 161 161 161 162 162	0.0007 -0.0002 -0.0003 -0.0003 0.0000 0.0001 0.0001 0.0000
161 161 161 161 161 161 161 161	0.0009 -0.0003 0.0004 0.0002 0.0001 -0.0004 0.0000 -0.0000
Step 4: 8x8 first preprocessed image block after applying inverse DCT	Step3: 8x8 first DCT block after applying high pass filter
0 0 0 0 0 0 0 0	0 -0.0019 0.0007 -0.0006 0.0004 -0.0002 0.0001 0.0002
0 0 0 0 0 0 0 0	-0.0062 0.0009 -0.0001 0.0001 0.0001 -0.0003 0.0002 -0.0002
0 0 0 0 0 0 0 1	-0.0015 0.0013 -0.0004 0.0003 -0.0004 0.0004 -0.0003 -0.0000
0 0 0 0 0 0 0 1	-0.0001 -0.0011 -0.0002 -0.0000 -0.0002 0.0003 -0.0001 0.0003
0 0 0 0 1 1 1 2	-0.0006 0.0003 -0.0001 0.0003 0.0001 -0.0004 0.0002 -0.0001
0 0 1 1 1 1 2 2	0.0007 0.0002 0.0004 -0.0003 0.0000 0.0002 -0.0001 -0.0001
1 1 1 1 1 1 2 2	0.0007 -0.0002 -0.0003 -0.0003 0.0000 0.0001 0.0001 0.0000
1 1 1 1 1 1 1 1	0.0009 -0.0003 0.0004 0.0002 0.0001 -0.0004 0.0000 -0.0000

The experimental values in Table 1 demonstrate that the constant gray level values in the range 159-162 in the first 8x8 block of the image in figure 3a have been compressed to a narrow range 0-2 after preprocessing. Hence, the transform function given in equation 1 that attenuates DC component as given in step3 has performed well in suppressing most of the constant background. The processed image after applying background suppression for all blocks is shown in figure 3b, where most of the unwanted details are removed. And only gray level discontinuities belonging to text and edges remain for further image analysis. Hence, the extracted texture features such as *homogeneity* and *contrast* from such preprocessed images aid classification decisions and help in increasing detection rate. All 24 50x50 image blocks of preprocessed image in figure 3b from which texture features are extracted are shown in figure 3e. Table 2 summarizes the extracted features.

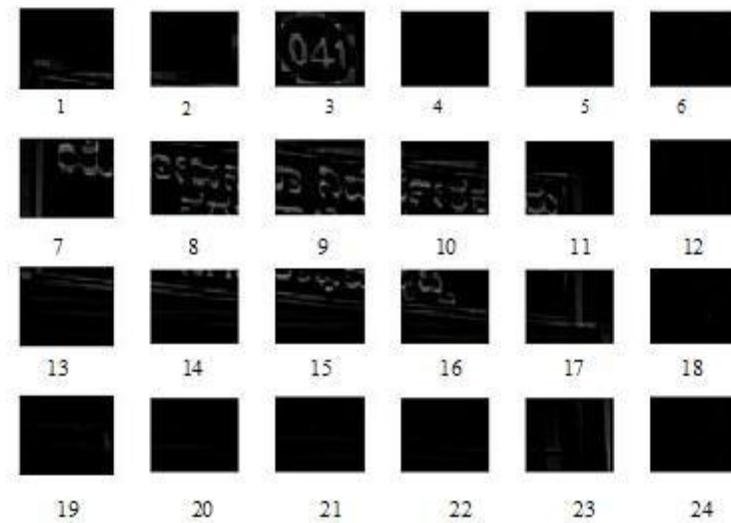


FIGURE 3e: 24 50x50 blocks of preprocessed image in figure 3b.

TABLE2: Feature Matrix D containing extracted features of image of size 240x320 given in figure 3b.

block	r_{min}	r_{max}	c_{min}	c_{max}	f1	f2	f3	f4	f5	f6	f7	f8
1	1	50	1	50	0.35563	5.7518	0.27407	40.044	0.29799	33.25	0.27741	33.556
2	1	50	51	100	0.37939	3.9122	0.30362	25.075	0.32683	21.255	0.31086	22.454
3	1	50	101	150	0.23571	145.05	0.21203	256.51	0.26159	115.69	0.21793	224.92
4	1	50	151	200	0.5486	0.060408	0.4624	0.14452	0.49457	0.11714	0.46736	0.13869
5	1	50	201	250	0.55128	0.056735	0.47864	0.11308	0.51388	0.081837	0.48201	0.11058
6	1	50	251	300	0.41841	0.21367	0.35536	0.33944	0.40281	0.19714	0.35614	0.33882
7	51	100	1	50	0.23662	80.981	0.20995	134.79	0.2649	59.38	0.21274	121.14
8	51	100	51	100	0.29437	108.49	0.23146	266.22	0.27296	167.09	0.24404	213.76
9	51	100	101	150	0.27832	94.329	0.21101	271.21	0.24637	182.34	0.23078	220.13
10	51	100	151	200	0.26405	70.58	0.1898	185.71	0.22773	130.81	0.20226	169.87
11	51	100	201	250	0.35865	31.801	0.29939	78.564	0.35312	47.617	0.31177	60.969
12	51	100	251	300	0.45093	0.17143	0.38476	0.26801	0.4281	0.1451	0.37781	0.27239
13	101	150	1	50	0.31607	5.8076	0.21355	52.554	0.2322	44.147	0.22656	42.147
14	101	150	51	100	0.29954	31.962	0.19948	90.116	0.21792	57.519	0.2132	71.612
15	101	150	101	150	0.29219	48.188	0.19208	171.58	0.2114	135.96	0.20367	154
16	101	150	151	200	0.2646	50.764	0.16705	134.08	0.19386	90.476	0.18203	118.49
17	101	150	201	250	0.30584	18.546	0.24364	49.318	0.30325	29.852	0.25069	43.421
18	101	150	251	300	0.47275	0.26082	0.42262	0.45981	0.4802	0.2702	0.43067	0.41733
19	151	200	1	50	0.38193	2.4163	0.32077	3.747	0.35694	1.4843	0.32784	3.3561
20	151	200	51	100	0.4088	0.095102	0.32114	0.87568	0.33529	0.80796	0.32327	0.8284
21	151	200	101	150	0.37253	0.22653	0.2801	1.3925	0.2939	1.2329	0.28104	1.3711
22	151	200	151	200	0.42539	0.073673	0.31561	1.3961	0.32448	1.3386	0.31156	1.3282
23	151	200	201	250	0.22828	48.116	0.19393	53.1	0.26798	3.7524	0.19778	49.633
24	151	200	251	300	0.52383	0.15612	0.48039	0.2045	0.5277	0.083878	0.48225	0.19992

The experimental values in Table 2 demonstrate the values of coordinates and texture features extracted at 0, 45, 90 and 130 degree orientations of all 24 blocks shown in figure 3e. As per figure 3e only 9 blocks numbered 3, 7-11, and 14-16 are text blocks and remaining blocks are non text blocks. The classification phase recognizes 6 blocks numbered 3, 7-10 and 16 as text blocks as their *homogeneity* and *contrast* values satisfy discriminant functions given in equations 10 and 11. And their coordinate values are recorded into feature vector *B* as shown in Table 3. The blocks 11, 14, and 15 which contain small text does not satisfy discriminant functions and are not recognized as text blocks. But they are handled properly during merging and post processing phases of the proposed method to improve the system performance.

TABLE 3: Vector *B* showing coordinate values of identified text blocks

r_{min}	r_{max}	c_{min}	c_{max}
1	50	101	150
51	100	1	50
51	100	51	100
51	100	101	150
51	100	151	200
101	150	151	200

The methodology merges identified text blocks and detects a single text region whose coordinates are stored into vector *R* as shown below. The detected text region also covers blocks 14 and 15 during merging which were undetected after classification phase thus improving detection accuracy. And the corresponding extracted text region of the image is shown in figure 3c.

$$R = [1 \ 150 \ 1 \ 200]$$

Where $r_{min} = 1, \ r_{max} = 150$
 $c_{min} = 1, \ c_{max} = 200$

The extracted text region now undergoes post processing phase, where during step1, the adjacent blocks 11 and 17 in right column are combined with the detected text region as their average *contrast* values satisfy the first condition in *step 3.2 of algorithm2*. It is also noted that the non text block 17 is falsly accepted. And during step2, the adjacent unprocessed areas are not

combined as they do not contain text information. Hence, the final refined text region coordinates are now shown below;

$$R = [1 \ 150 \ 1 \ 250]$$

Where $r_{min} = 1, \ r_{max} = 150$
 $c_{min} = 1, \ c_{max} = 250$

The performance of the system indicating detection rate before and after post processing is described in Table 4.

TABLE 4: System Performance showing detection rate before and after post processing image in figure 3a

Total no of blocks tested(# of text blocks)	Correctly detected text blocks before/ after post processing	Falsy detected text blocks before / after post processing	Missed text blocks before / after post processing	Detection rate in % before/ after post processing
24 (9)	6 / 9	0 / 01	03 / 00	66.6 / 100

The output of the system described in Table4 brings out the following details after post processing;

- Detection Rate** = (Number of text blocks correctly detected/ Number of text blocks tested) * 100
 = (9/9) * 100 = 100%
- False Acceptance Rate (FAR)** = (Number of text blocks falsy detected/ Number of text blocks tested) * 100
 = (01/09)*100 = 11%
- False Rejection Rate (FRR)** = (Number of missed text blocks / Number of text blocks tested) * 100
 = (00/09)*100 = 0%

4.2 Text Region Extraction: An experimental analysis dealing with various issues

The text detection and extraction results of testing several different low resolution natural scene display board images dealing with various issues are shown in figures 4-5. And the corresponding detailed analysis is presented in Table5.

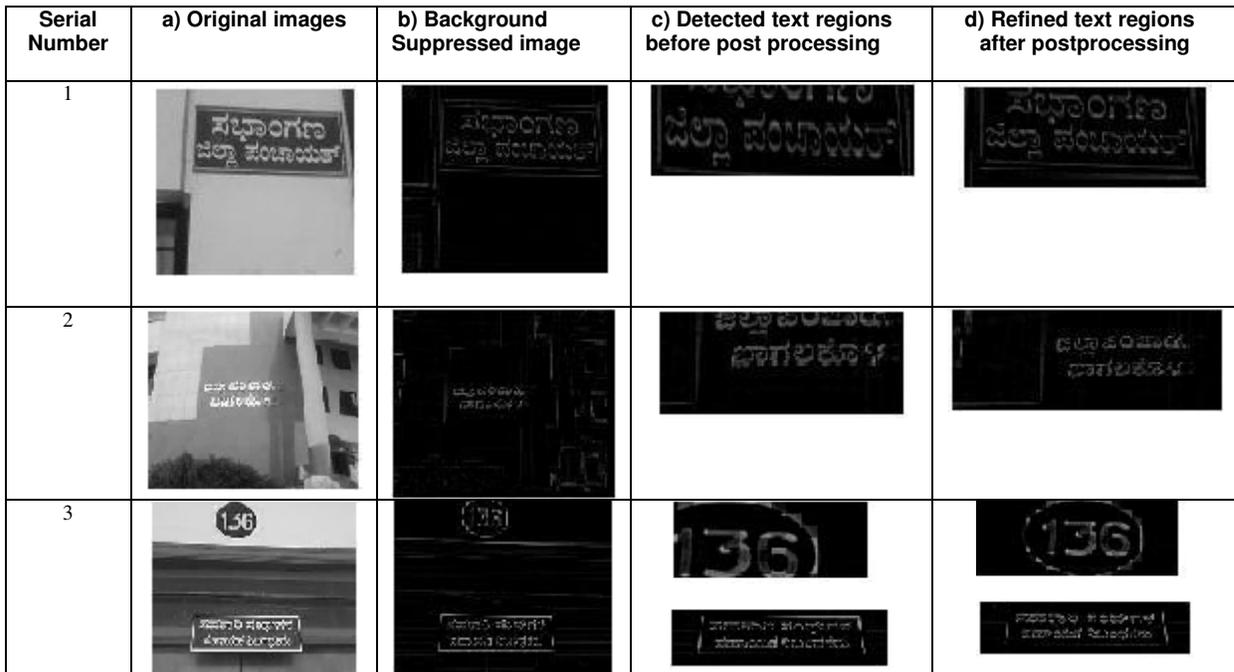


FIGURE 4: Text Extraction results of processing 3 low resolution natural scene images dealing with complex backgrounds and multiple text regions

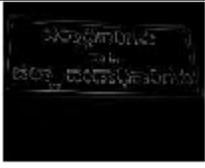
Serial Number	a) Original images	b) Background Suppressed image	c) Detected text regions before post processing	d) Refined text regions after postprocessing
1				
2				
3				
4				
5				
6				

FIGURE 5: Text Extraction results of processing 6 low resolution natural scene images dealing with various issues

TABLE 5: The performance of the system of processing different images given in figures 4-5 dealing with various issues

Input Image	Total no of blocks tested(# of text blocks)	Correctly detected text blocks	Falsly detected text blocks	Missed text blocks	Detection rate (After post processing) %	Description
Figure 4-1a	24(10)	10	00	00	100	Extraction of smaller and bigger font text and elimination of unwanted background.
Figure 4-2a	24(7)	07	03	00	100	Robustness of the method in detecting correct text region by processing an image, containing text and background information, such as building walls, windows, trees and pillars etc., and uneven lighting conditions
Figure 4-3a	24(6)	06	00	00	100	Detecting multiple text regions by processing

						an image, containing complex background and uneven lighting conditions. The area containing numerals is also detected as text region.
Figure 5-1a	24(15)	15	00	00	100	Text extraction of processing natural scene image containing text of varying size, font, and alignment with varying background.
Figure 5-2a	24(12)	12	00	00	100	
Figure 5-3a	24(9)	9	00	00	100	
Figure 5-4a	24(07)	04	04	03	57.14	Text extraction result of processing an image containing non linear text and complex background
Figure 5-5a	24(08)	05	07	03	62.50	Detection of non linear text containing complex background.
Figure 5-6a	24(13)	13	00	00	100	Extraction of English text from an image with complex background

The proposed methodology has produced good results for natural scene images containing text of different size, font, and alignment with varying background. The approach also detects nonlinear text regions. Hence, the proposed method is robust and achieves an overall detection rate of 96.6%, and a false reject rate of 3.4% is obtained for 100 low resolution display board natural scene images. The method is advantageous as it uses only two texture features for text extraction. The advantage lies in less computation involved in feature extraction and classification phases of the method. The reason for false reject rate is the low contrast energy of blocks containing minute part of the text, which is too weak for acceptance to classify blocks as text blocks. However, the method has detected larger region than the actual size of the text region, when display board images with more complex backgrounds containing trees, buildings and vehicles are tested. One such result of an example is shown in 3rd row of figure 5. The system is developed in MATLAB and evaluated for 100 low resolution natural scene images on Intel Celeron (1.4GHz) computer. And it was observed that the processing time lies in the range of 6 to 10 seconds due to varying background.

As the texture features such as *homogeneity* and *contrast* used in the method does not capture language dependent information, the method can be extended for text localization from the images of other languages with little modifications. To explore such possibilities the performance of the method has been tested for localizing English text without any modifications as illustrated in 6th row of figure 5. But, the thorough experimentation is not carried out for various images containing English and other language text. The use of different block sizes can also be experimented to improve the detection accuracy and reduce false acceptance rate. The overall performance of the system of testing 100 low resolution natural scene display board images dealing with various issues is given in Table 6. The system performance is also pictorially depicted in figure 6.

TABLE 6: Overall System Performance

Total no of blocks tested(# of text blocks)	Correctly detected text blocks	Falsy detected text blocks (FAR)	Missed text blocks (FRR)	Detection rate %
2400 (1211)	1169	19 (1.5%)	42 (3.4%)	96.6 %

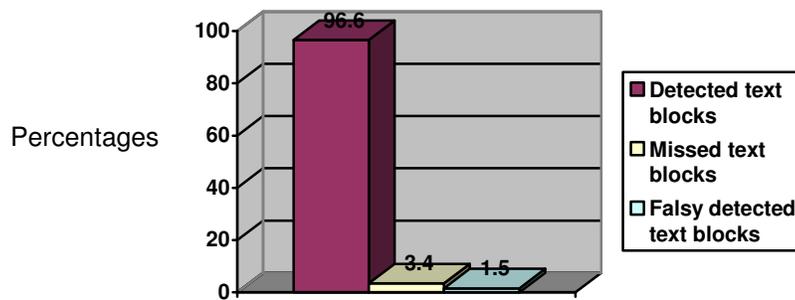


FIGURE 6: Overall results of proposed model for text localization

5. CONCLUSIONS AND FUTURE WORK

The effectiveness of the method that uses texture analysis for text localization from low resolution natural scene display board images is presented. The texture features *homogeneity* and *contrast* have performed well in detection and segmentation of text region and are the ideal choice for degraded noisy natural scene images, where the connected component analysis techniques are found to be inefficient. The intelligent post processing method has improved detection accuracy to a greater extent. The proposed method is robust and has achieved a detection rate of 96.6% on a variety of 100 low resolution natural scene images each of size 240x320.

The proposed methodology has produced good results for natural scene images containing text of different size, font, and alignment with varying background. The approach also detects nonlinear text regions. However, it detects larger text regions than the actual size when the background in the image is more complex containing trees, vehicles, and other details from sources of outdoor scenes for some images. The system is developed in MATLAB and evaluated for 100 low resolution natural scene images on Intel Celeron (1.4GHz) computer. And it was observed that the processing time lies in the range of 6 to 10 seconds due to varying background.

As the texture features such as *homogeneity* and *contrast* used in the method does not capture language dependent information, the method can be extended for text localization from the images of other languages with little modifications. The performance of the method has been tested for localizing English text, but needs further exploration.

6. REFERENCES

1. Tollmar K. Yeh T. and Darrell T. "IDEixis - Image-Based Deixis for Finding Location-Based Information", In Proceedings of Conference on Human Factors in Computing Systems (CHI'04), pp.781-782, 2004.
2. Natalia Marmasse and Chris Schamandt. "Location aware information delivery with *comMotion*", In Proceedings of Conference on Human Factors in Computing Systems, pp.157-171, 2000.
3. Eve Bertucci, Maurizio Pilu and Majid Mirmehdi. "Text Selection by Structured Light Marking for Hand-held Cameras" Seventh International Conference on Document Analysis and Recognition (ICDAR'03), pp.555-559, August 2003.
4. Tom yeh, Kristen Grauman, and K. Tollmar. "A picture is worth a thousand keywords: image-based object search on a mobile platform", In Proceedings of Conference on Human Factors in Computing Systems, pp.2025-2028, 2005.

5. Abowd Gregory D. Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton, "CyberGuide: Amobile context-aware tour guide", *Wireless Networks*, 3(5):421-433, 1997.
6. Fan X. Xie X. Li Z. Li M. and Ma. "Photo-to-search: using multimodal queries to search web from mobile phones", In proceedings of 7th ACM SIGMM international workshop on multimedia information retrieval, 2005.
7. Lim Joo Hwee, Jean Pierre Chevallet and Sihem Nouarah Merah, "SnapToTell: Ubiquitous information access from camera", *Mobile human computer interaction with mobile devices and services*, Glasgow, Scotland, 2005.
8. Yu Zhong, Hongjiang Zhang, and Anil. K. Jain. "Automatic caption localization in compressed video", *IEEE transactions on Pattern Analysis and Machine Intelligence*, 22(4):385-389, April 2000.
9. Yu Zhong, Kalle Karu, and Anil. K. Jain. "Locating Text in Complex Color Images", *Pattern Recognition*, 28(10):1523-1535, 1995.
10. S. H. Park, K. I. Kim, K. Jung, and H. J. Kim. "Locating Car License Plates using Neural Networks", *IEE Electronics Letters*, 35(17):1475-1477, 1999.
11. V. Wu, R. Manmatha, and E. M. Riseman. "TextFinder: An Automatic System to Detect and Recognize Text in Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1224-1229, 1999.
12. V. Wu, R. Manmatha, and E. R. Riseman. "Finding Text in Images", In *Proceedings of ACM International Conference on Digital Libraries*, pp. 1-10, 1997.
13. B. Sin, S. Kim, and B. Cho. "Locating Characters in Scene Images using Frequency Features", In proceedings of International Conference on Pattern Recognition, Vol.3:489-492, 2002.
14. W. Mao, F. Chung, K. Lanm, and W. Siu. "Hybrid Chinese/English Text Detection in Images and Video Frames", In *Proceedings of International Conference on Pattern Recognition*, Vol.3:1015-1018, 2002.
15. A. K. Jain, and K. Karu. "Learning Texture Discrimination Masks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):195-205, 1996.
16. K. Jung. "Neural network-based Text Location in Color Images", *Pattern Recognition Letters*, 22(14):1503-1515, December 2001.
17. K. Y. Jeong, K. Jung, E. Y. Kim and H. J. Kim. "Neural Network-based Text Location for News Video Indexing", In *Proceedings of IEEE International Conference on Image Processing*, Vol.3:319-323, January 2000.
18. H. Li. D. Doerman and O. Kia. "Automatic Text Detection and Tracking in Digital Video", *IEEE Transactions on Image Processing*, 9(1):147-156, January 2000.
19. H. Li and D. Doermann. "A Video Text Detection System based on Automated Training", In *Proceedings of IEEE International Conference on Pattern Recognition*, pp.223-226, 2000.
20. W. Y. Liu, and D. Dori. "A Proposed Scheme for Performance Evaluation of Graphics/Text Separation Algorithm", *Graphics Recognition – Algorithms and Systems*, K. Tombre and A. Chhabra (eds.), *Lecture Notes in Computer Science*, Vol.1389:359-371, 1998.
21. Y. Watanabe, Y. Okada, Y. B. Kim and T. Takeda. "Translation Camera", In *Proceedings of International Conference on Pattern Recognition*, Vol.1: 613-617, 1998.
22. I. Haritaoglu. "Scene Text Extraction and Translation for Handheld Devices", In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2:408-413, 2001.
23. K. Jung, K. Kim, T. Kurata, M. Kourogi, and J. Han. "Text Scanner with Text Detection Technology on Image Sequence", In *Proceedings of International Conference on Pattern Recognition*, Vol.3:473-476, 2002.
24. H. Hase, T. Shinokawa, M. Yoneda, and C. Y. Suen. "Character String Extraction from Color Documents", *Pattern Recognition*, 34(7):1349-1365, 2001.
25. Ceiline Thillou, Silvio Frreira and Bernard Gosselin. "An embedded application for degraded text recognition", *EURASIP Journal on applied signal processing*, 1(1):2127-2135, 2005.

26. Nobuo Ezaki, Marius Bulacu and Lambert Schomaker. "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", In Proceedings of 17th International Conference on Pattern Recognition (ICPR'04), IEEE Computer Society vol.2:683-686, 2004.
27. Chitrakala Gopalan and Manjula. "Text Region Segmentation from Heterogeneous Images", International Journal of Computer Science and Network Security, 8(10):108-113, October 2008.
28. Rui Wu, Jianhua Huang, Xianglong Tang and Jiafeng Liu. "A Text Image Segmentation Method Based on Spectral Clustering", IEEE Sixth Indian Conference on Computer Vision, Graphics & Image Processing, 1(4): 9-15, 2008.
29. Mohammad Osiur Rahman, Fouzia Asharf Mousumi, Edgar Scavino, Aini Hussain, and Hassan Basri. "Real Time Road Sign Recognition System Using Artificial Neural Networks for Bengali Textual Information Box", European Journal of Scientific Research, 25(3):478-487, 2009.
30. Rajiv K. Sharma and Amardeep Singh. "Segmentation of Handwritten Text in Gurmukhi Script", International Journal of Image Processing, 2(3):13-17, May/June 2008.
31. Amjad Rehman Khan and Zulkifli Mohammad. "A Simple Segmentation Approach for Unconstrained Cursive Handwritten Words in Conjunction with the Neural Network", International Journal of Image Processing, 2(3):29-35, May/June 2008.
32. Amjad Rehman Khan and Zulkifli Mohammad. "Text Localization in Low Resolution Camera", International Journal of Image Processing, 1(2):78-86, 2007.
33. R.M. Haralick, K. Shanmugam and I. Dinstein. "Textural Features for Image Classification", IEEE Transactions Systems, Man, and Cyber-netics, 3(6):610-621, 1973.

Lossless Grey-scale Image Compression using Source Symbols Reduction and Huffman Coding

C. SARAVANAN

cs@cc.nitdgp.ac.in

Assistant Professor, Computer Centre, National Institute of Technology,
Durgapur, West Bengal, India, Pin – 713209.

R. PONALAGUSAMY

rpalagu@nitt.edu

Professor, Department of Mathematics, National Institute of Technology,
Tiruchirappalli, Tamilnadu, India, Pin – 620015.

Abstract

Usage of Image has been increasing and used in many applications. Image compression plays vital role in saving storage space and saving time while sending images over network. A new compression technique proposed to achieve more compression ratio by reducing number of source symbols. The source symbols are reduced by applying source symbols reduction and further the Huffman coding is applied to achieve compression. The source symbols reduction technique reduces the number of source symbols by combining together to form a new symbol. Therefore, the number of Huffman code to be generated also reduced. The Huffman code symbols reduction achieves better compression ratio. The experiment has been conducted using the proposed technique and the Huffman coding on standard images. The experiment result has analyzed and the result shows that the newly proposed compression technique achieves 10% more compression ratio than the regular Huffman coding.

Keywords: Lossless Image Compression, Source Symbols Reduction, Huffman Coding.

1. INTRODUCTION

The image compression highly used in all applications like medical imaging, satellite imaging, etc. The image compression helps to reduce the size of the image, so that the compressed image could be sent over the computer network from one place to another in short amount of time. Also, the compressed image helps to store more number of images on the storage device [1-4].

It's well known that the Huffman's algorithm is generating minimum redundancy codes compared to other algorithms [6-11]. The Huffman coding has effectively used in text, image, video compression, and conferencing system such as, JPEG, MPEG-2, MPEG-4, and H.263 etc. [12]. The Huffman coding technique collects unique symbols from the source image and calculates its probability value for each symbol and sorts the symbols based on its probability value. Further, from the lowest probability value symbol to the highest probability value symbol, two symbols combined at a time to form a binary tree. Moreover, allocates zero to the left node and one to the right node starting from the root of the tree. To obtain Huffman code for a particular symbol, all zero and one collected from the root to that particular node in the same order [13 and 14].

2. PROPOSED COMPRESSION TECHNIQUE

The number of source symbols is a key factor in achieving compression ratio. A new compression technique proposed to reduce the number of source symbols. The source symbols combined together in the same order from left to right to form a less number of new source symbols. The source symbols reduction explained with an example as shown below.

The following eight symbols are assumed as part of an image, 1, 2, 3, 4, 5, 6, 7, 8. By applying source symbols reduction from left to right in the same sequence, four symbols are combined together to form a new element, thus two symbols 1234 and 5678 are obtained. This technique helps to reduce 8 numbers of source symbols to 2 numbers i.e. 2^n symbols are reduced to $2^{(n-2)}$ symbols. For the first case, there are eight symbols and the respective Symbols and Huffman Codes are 1-0, 2-10, 3-110, 4-1110, 5-11110, 6-111110, 7-1111110, 8-1111111. The proposed technique reduced the eight symbols to two and the reduced Symbols and Huffman codes are 1234-0, 5678-1.

The minimum number of bits and maximum number of bits required to represent the new symbols for an eight bit grayscale image calculated. The following possible combinations worked out and handled perfectly to ensure the lossless compression. The following are few different possible situations to be handled by source symbols reduction.

If all symbols in the four consecutive symbols are 0, i.e. 0 0 0 0, then the resulting new symbol will be 0.

If the four consecutive symbols are 0 0 0 1, then the resulting new symbol will be 1.

If the four consecutive symbols are 0 0 1 0, then the resulting new symbol will be 1000.

If the four symbols are 0 1 0 0, then the resulting new symbol will be 1000000.

If the four symbols are 1 0 0 0, then the resulting new symbol will be 1000000000.

If the four symbols are 255 255 255 255, then the resulting new symbol will be 255255255255.

The average number L_{avg} of bits required to represent a symbol is defined as,

$$L_{avg} = \sum_{k=1}^L l(r_k) p_r(r_k) \quad (1)$$

where, r_k is the discrete random variable for $k=1,2,\dots,L$ with associated probabilities $p_r(r_k)$. The number of bits used to represent each value of r_k is $l(r_k)$. The number of bits required to represent an image is calculated by number of symbols multiplied by L_{avg} [5].

In the Huffman coding, probability of each symbols is 0.125 and $L_{avg} = 4.175$.

In the proposed technique, probability of each symbol is 0.5 and $L_{avg}=1.0$.

The L_{avg} confirms that the proposed technique achieves better compression than the Huffman Coding.

From the above different possible set of data, the following maximum and minimum number of digits of a new symbol formed by source symbols reduction calculated for an eight bits grey-scale image. The eight bits grey-scale image symbols have values ranging from 0 to 255. The minimum number of digits required to represent the new symbol could be 1 digit and the maximum number of digits required to represent the new symbols could be 12 digits. Therefore, if the number of columns of the image is multiples of four, then this technique could be applied as it is. Otherwise, the respective remaining columns (1 or 2 or 3 columns) will be kept as it is during the source symbols reduction and expansion.

Four rows and four columns of eight bits grey-scale image having sixteen symbols considered to calculate required storage size. To represent these 16 symbols requires $16 \times 1 \text{ byte} = 16$ bytes storage space. The proposed source symbol reduction technique reduces the 16 symbols into 4 symbols. The four symbols require $4 \times 4 \text{ bytes} = 16$ bytes. Therefore, the source symbols data and the symbols obtained by the source symbols reduction requires equal amount of storage space. However, in the coding stage these two techniques make difference. In the first case, sixteen symbols generate sixteen Huffman codes, whereas the

proposed technique generates four Huffman codes and reduces L_{avg} . Therefore, the experiment confirms that the source symbols reduction technique helps to achieve more compression.

The different stages of newly proposed compression technique are shown in figure 1. The source image applied by source symbols reduction technique then the output undergoes the Huffman encoding which generates compressed image. In order to get the original image, the Huffman decoding applied and an expansion of source symbols takes place to reproduce the image.

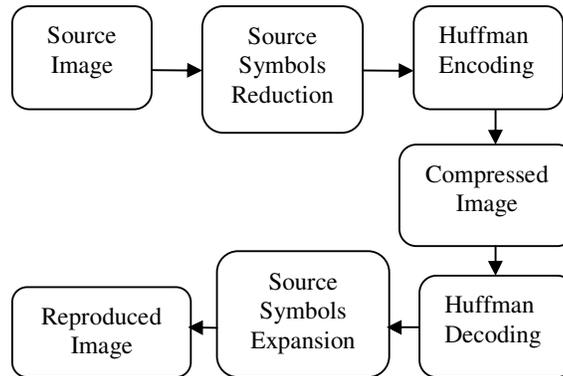


FIGURE 1: Proposed Compression Technique

Five different test images with different redundancy developed for experiment from 0% to 80% in step size of 20% i.e 0%, 20%, 40%, 60%, and 80% redundancy. The Huffman coding could not be applied on data with 100% redundancy or single source symbol, as a result 100% redundancy is not considered for the experiment. The test images with 16 rows and 16 columns will have totally 256 symbols. The images are 8 bit grey-scale and the symbol values range from 0 to 255. To represent each symbol eight bit is required. Therefore, size of an image becomes $256 \times 8 = 2048$ bit. The five different level redundancy images are applied the Huffman coding and the proposed technique. The compressed size and time required to compress and decompress (C&D) are noted.

3. EXPERIMENT RESULTS

Following table 1 shows the different images developed for the experiment and corresponding compression results using the regular Huffman Coding and the proposed technique. The images are increasing in redundancy 0% to 80% from top to bottom in the table.

IMAGE	Huffman Coding	SSR+HC Technique
	Compressed size (bits)	Compressed size (bits)
	2048	384
	1760	344
	1377	273
	944	188
	549	118

TABLE 1: Huffman Coding Compression Result

The experiment shows that the higher data redundancy helps to achieve more compression. The experiment shows that the proposed compression technique achieves more compression than the Huffman Coding. The first image has 0% redundancy and its compressed image size is 2048 bit using the Huffman coding whereas the proposed compression technique has resulted compressed image of size 384 bit. No compression takes place for the first image using Huffman coding, where as the proposed technique achieved about 81% compression.

For all images the compressed size obtained from the proposed technique better than the Huffman coding. The proposed compression technique achieves better compression. The results obtained from the present analysis are shown in figure 2.

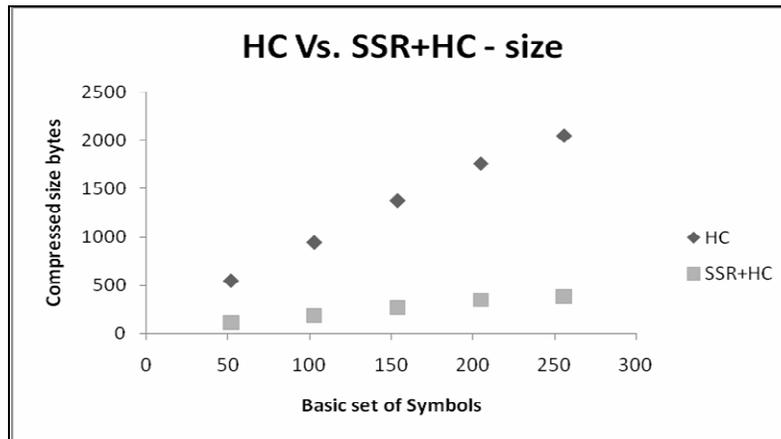


FIGURE 2: Compressed Size comparisons

Table 2 shows the comparison between these two techniques. Compression Ratio (CR) is defined as

$$CR = \frac{\text{Original size}}{\text{Compressed size}} \quad (2)$$

Redundancy	Huffman Coding	SSR+HC Technique
	Compression Ratio	Compression Ratio
0%	1.0000	5.3333
20%	1.1636	5.9535
40%	1.4873	7.5018
60%	2.1695	10.8936
80%	3.7304	17.3559

TABLE 2: Compression Ratio versus Time

From the result of the experiment it is found that the two compression techniques are lossless compression technique, therefore the compression error not considered. The following figure 3 compares the compression ratio of the experiment. From the figure it is observed that the proposed technique has performed better than the Huffman Coding. The proposed technique shows better compression ratio for the images having higher redundancy when compared with the images of lower redundancy.

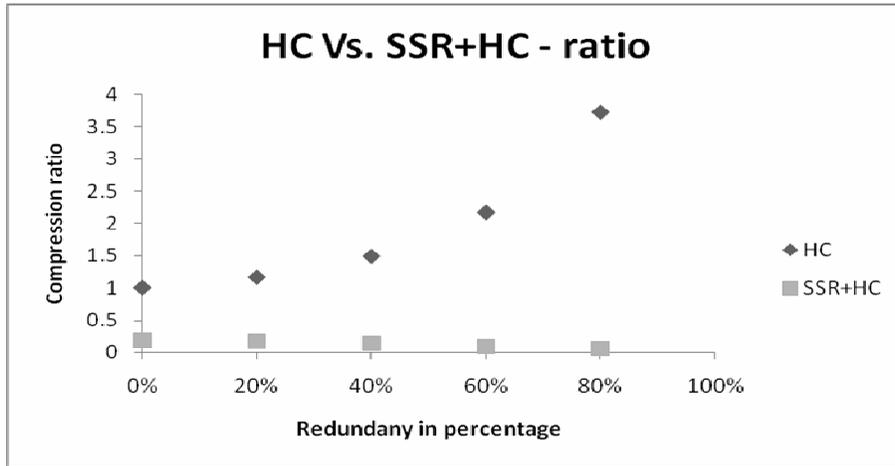


FIGURE 3: Compression ratio comparisons

In the real time, images are usually having higher data redundancy. Hence, the proposed technique will be suitable for the user who desires higher compression. Moreover, standard gray scale images considered for testing. The standard images require 65,536 bytes storage space of 256 rows and 256 columns. The image is eight bit gray scale image.

The standard images applied using the two the compression techniques and standard JPEG compression technique. The compression size of the experiment is noted. The following figure 4 is one of the source image used for the experiment and figure 5 is the reproduced image using the proposed technique.

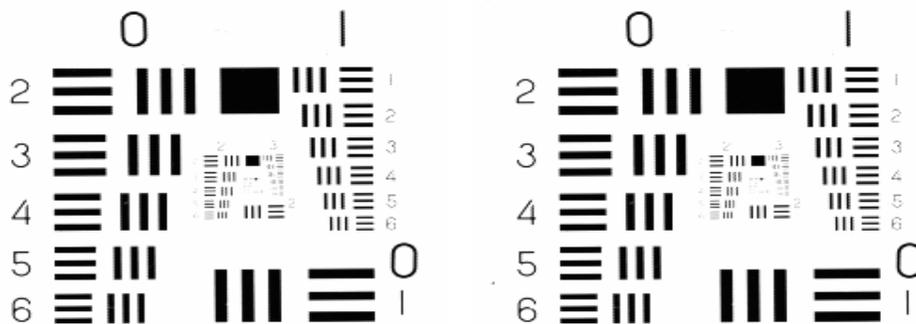


FIGURE 4: Source image chart.tif

FIGURE 5: Reproduced image chart.tif

Table 3 shows the compression result using Huffman coding, and the proposed technique for one of the standard image *chart.tif*. The proposed technique has achieved better compressed size than the Huffman coding. The source symbols reduction and expansion takes more time if the number of symbols are higher. Hence, the newly proposed technique is suitable to achieve more compression.

Source Image	Huffman Coding	SSR+HC Technique
Size (bits)	Compressed size (bits)	Compressed size (bits)
5,128,000	1,015,104	54,207

TABLE 3: Compression test results for chart.tif

4. CONCLUSIONS

The present experiment reveals that the proposed technique achieves better compression ratio than the Huffman Coding. The experiment also reveals that the compression ratio in Huffman Coding is almost close with the experimental images. Whereas, the proposed compression technique Source Symbols Reduction and Huffman Coding enhance the performance of the Huffman Coding. This enables us to achieve better compression ratio compared to the Huffman coding. Further, the source symbols reduction could be applied on any source data which uses Huffman coding to achieve better compression ratio. Therefore, the experiment confirms that the proposed technique produces higher lossless compression than the Huffman Coding. Thus, the proposed technique will be suitable for compression of text, image, and video files.

5. REFERENCES

1. Gonzalez, R.C. and Woods, R.E., Digital Image Processing 2nd ed., Pearson Education, India, 2005.
2. Salomon, Data Compression, 2nd Edition. Springer, 2001.
3. Othman O. Khalifa, Sering Habib Harding and Aisha-Hassan A. Hashim, Compression using Wavelet Transform, Signal Processing: An International Journal, Volume (2), Issue (5), 2008, pp. 17-26.
4. Singara Singh , R. K. Sharma, M.K. Sharma, Use of Wavelet Transform Extension for Graphics Image Compression using JPEG2000 Framework, International Journal of Image Processing, Volume 3, Issue 1, Pages 55-60, 2009.
5. Abramson, N., Information Theory and Coding, McGraw-Hill, New York, 1963.
6. Huffman, D.A., A method for the construction of minimum-redundancy codes. Proc. Inst. Radio Eng. 40(9), pp.1098-1101, 1952.
7. Steven Pigeon, Yoshua Bengio — A Memory-Efficient Huffman Adaptive Coding Algorithm for Very Large Sets of Symbols — Université de Montréal, Rapport technique #1081.
8. Steven Pigeon, Yoshua Bengio — A Memory-Efficient Huffman Adaptive Coding Algorithm for Very Large Sets of Symbols Revisited — Université de Montréal, Rapport technique #1095.
9. R.G. Gallager — Variation on a theme by Huffman — IEEE. Trans. on Information Theory, IT-24(6), 1978, pp. 668-674.
10. D.E. Knuth — Dynamic Huffman Coding — Journal of Algorithms, 6, 1983 pp. 163-180.
11. J.S. Vitter — Design and analysis of Dynamic Huffman Codes — Journal of the ACM, 34#4, 1987, pp. 823-843.
12. Chiu-Yi Chen; Yu-Ting Pai; Shanq-Jang Ruan, Low Power Huffman Coding for High Performance Data Transmission, International Conference on Hybrid Information Technology, 2006, 1(9-11), 2006 pp.71 – 77.
13. Lakhani, G, Modified JPEG Huffman coding, IEEE Transactions Image Processing, 12(2), 2003 pp. 159 – 169.
14. R. Ponalagusamy and C. Saravanan, Analysis of Medical Image Compression using Statistical Coding Methods, Advances in Computer Science and Engineering: Reports and Monographs, Imperial College Press, UK, Vol.2., pp 372-376, 2007.

An intelligent control system using an efficient License Plate Location and Recognition Approach

Saeed Rastegar

s.rastegar@stu.nit.ac.ir

*Faculty of Electrical and Computer Engineering,
Signal Processing Laboratory Babol Noshirvani
University of Technology
Babol, P.O. Box 47135-484, IRAN*

Reza Ghaderi

r_ghaderi@nit.ac.ir

*Faculty of Electrical and Computer Engineering,
Babol Noshirvani University of Technology
Babol, P.O. Box 47135-484, IRAN*

Gholamreza Ardeshir

g.ardeshir@nit.ac.ir

*Faculty of Electrical and Computer Engineering,
Babol Noshirvani University of Technology
Babol, P.O. Box 47135-484, IRAN*

Nima Asadi

nima228@gmail.com

*Faculty of Electrical and Computer Engineering,
Babol Noshirvani University of Technology*

Abstract

This paper presents a real-time and robust method for license plate location and recognition. After adjusting the image intensity values, an optimal adaptive threshold is found to detect car edges and then the algorithm uses morphological operators to make candidate regions. Features of each region are to be extracted in order to correctly differentiate the license plate regions from other candidates. It was done by analysis of percentage of Rectangularity of plate. Using color filter makes the algorithm more robust on license plate localization (LPL). The algorithm can efficiently determine and adjust the plate rotation in skewed images. The Binary unit uses Otsu method to find the optimal adaptive threshold corresponding to the intensity of image. To segment the characters of the license plate, a segmentation algorithm based on the profile is proposed. In the following, an optical character recognition (OCR) engine has then been proposed. The OCR engine includes characters dilation, resizing input vector of Artificial Neural Network (ANN). To recognize the characters on the plates, Multi layer Perceptron (MLP) has been used and compared with Hopfield, Linear Vector Quantization (LVQ) and Radial Basis Function (RBF). The results show that MLP outperforms. According to the results, the performance of the proposed system is better even in case of low-quality images or images with illumination effects and noise.

Keywords: license plate recognition (LPR), Rectangularity percentage, optical character recognition (OCR), neural network.

1. INTRODUCTION

Automatic license plate recognition (LPR) has been a practical technique in the past decades [1-6]. One type of intelligent transportation system (ITS) technology is the automatic license plate recognition (ALPR) which can distinguish each car as unique by recognizing the characters of the license plates. In ALPR, a camera captures the vehicle images and a computer processes them and recognizes the information on the license plate by applying various image processing and optical pattern recognition techniques. Prior to the character recognition, the license plates must be separated from the background vehicle images. This task is considered as the most crucial step in the ALPR system, which influences the overall accuracy and processing speed of the whole system significantly. Since there are problems such as poor image quality, image perspective distortion, other disturbance characters or reflection on vehicle surface, and the color similarity between the license plate and the background vehicle body, the license plate is often difficult to be located accurately and efficiently. Researchers have found many diverse methods of license plate localization [7-12]. Edge density and background color have been used to detect a number of plate locations, in (Sherin et al, 2008), according to the characteristics of the number plate. In (Wenjing Jia, 2007), mean shift algorithm and color filter were used. (Rodolfo and Stefano, 2000) devised a method based on vector quantization (VQ). VQ image representation is a quadtree representation using a specific coding mechanism and it gives a system some hints about the contents of the image regions that boost location performance. In (Wang and Lee, 2003), (Bai et al, 2003), (Bai and Liu, 2004) the presence of abundant edges, especially vertical edges, in license plate regions is due to the presence of characters used to generate the candidate regions for classification. Combined with geometrical properties of license plates, good performance is obtained by the algorithms even dealing with some deficient license plates (Bai and Liu, 2004). Other popular methods focus on detecting the frames, e.g. the Hough Transform which is used widely in (Yanamura et al, 2003). In addition to the algorithms based on gray-level image processing, color information of license plates also plays an important role in license plates localization, where the unique color or color combination between the license plates and vehicle bodies are considered as the key feature to locate the license plates. In order to determine the exact color at a certain pixel, neural network classifier (Wei et al, 2001; Kim et al, 2002) and genetic algorithm (Kim et al, 1996) are widely used. License plates (LP) are identified as image areas with high intensity rather than dark lines or curves. Therefore, the location is handled by looking for rectangular regions in the image containing maxima of response to these line filters, which is computed by a cumulative function (Luis, Jose, Enrique, Narucuso, & Narucuso, 1999). Plate characters can be directly identified by scanning through the input image and looking for portions of the image that were not linked to other parts of it. If a number of characters are found to be in a straight line, they may make up a license plate (Lim, Yeo, Tan, & Teo, 1998). Fuzzy logic has been applied to solve the problem of locating license plates by (Zimic, Ficzk, Mraz, and Virant, 1997). The authors made some intuitive rules to describe the license plate, and gave some membership functions for the fuzzy sets "bright" and "dark", and "bright and dark sequence" to get the horizontal and vertical plate positions. However, this method is sensitive to license plate color and brightness and needs much processing time. Using color features to locate license plate has been studied by (Zhu, Hou, and Jia 2002) & (Wei, Wang, and Huang, 2001), but these methods are not robust enough to the different environments. We propose a new approach using an automated license plate location and recognition that overcomes most of the problems with previous approaches. The mechanism is able to deal with difficulties raised from illumination variance, noise distortion, and complex and dirty backgrounds. Numerous captured images including various types of vehicles with different lighting and noise effects have been handled. The remainder of the paper is organized as follows. Section 2 presents the proposed model and describes the procedures of the system. Section 3 describes the learning and recognition processes. Section 4 explains the experimental settings, and the experimental results on different testing sets will be reported in this section. In section 4, also the architecture parameters will be described. Finally, a conclusion and discussion are presented in Section 5.

2. THE PROPOSED ALGORITHM

According to the feature characteristics of car license plates, we considered that the license plates are rectangular with blue regions in left containing eight alphanumeric formats. Fig. 1 shows an example of different located license plates. Our intelligent access control system based on license plate recognition can be broken

down into the following block diagram, as illustrated in Fig. 2. Alternatively, this progression could be viewed as the reduction or suppression of unwanted information from the information carrying signal. Here is a video sequence containing vast amounts of irrelevant information, to abstract symbols in the form of the characters of a license plate.



FIGURE 1: Examples of different located license plates.

2.1. Captured RGB image

Experiments have been carried out using a camera to capture the approaching vehicle image. Various vehicle images with different camera positions and different orientations have been processed. Current Iranian cars have license plates with a white background and a blue region at the left. Each car is specified by eight characters, seven digits and one Persian alphabet. Images have been taken by "canon powershot A520".(640 * 460 by pixels) .

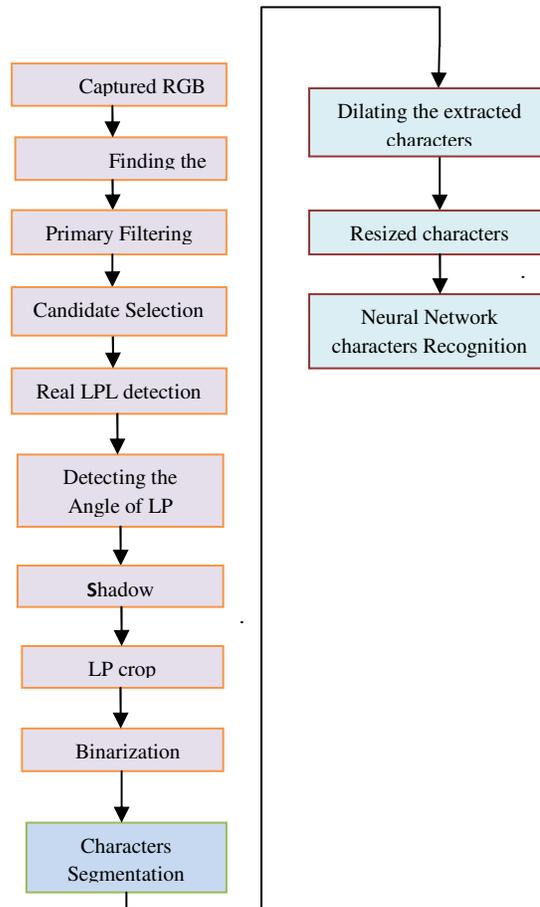


FIGURE 2: Block diagram of the proposed framework.

2.2. Finding the contours

In the algorithm, gradient operators [13-14] have been used to detect edges of grey level images which form primary regions. Then a morphological operator is used to dilate adjacent regions. Fig.3 shows the output of this step of the algorithm. Images of cars were taken from different distances. The distance is the important parameter in the proposed algorithm. In some instances, due to adhesion of the plate to the body of the car, Sobel operators cannot detect the plate region. The problem of adhesion can be removed according to different distances of car from the camera; a suitable edge detector is used. This unit uses sobel operator has been used for distances less than 3.5 meters, and For distances longer than 3.5 meters a structure element $[-1 \ 0 \ 1]$ has been applied to the image. Fig.4. describes the explained problem and how the proposed method solves it.

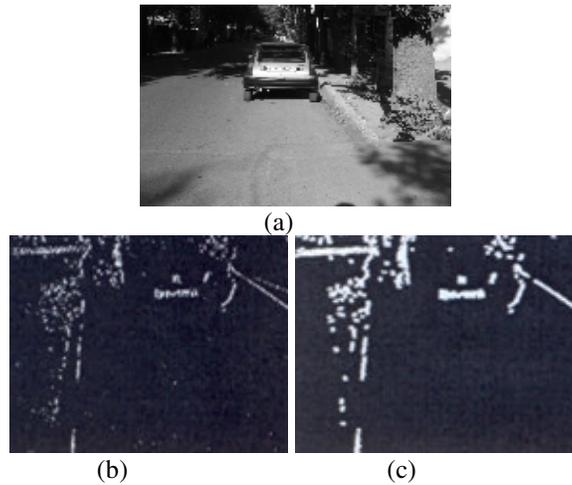


FIGURE 3: (a) grey image, (b) binary image after edge detections and (c) dilation using morphological operator.

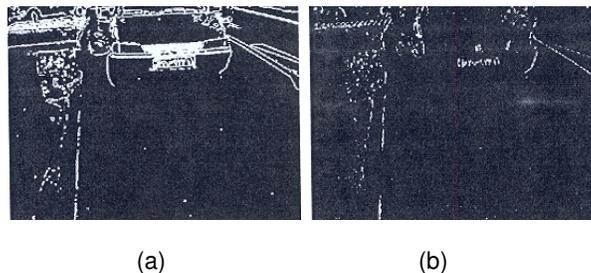


FIGURE 4: (a) adhesion of car plate to other excessive parts of the car and (b) applying proposed method to solve the problem.

2.3. Primary filtering

Selection of candidate regions has been done based on two criteria. This is the reason why the speed of computation increases the speed of computation. The regions with sizes smaller or bigger than (the) threshold and region in which the aspect ratio is smaller than 5, are removed at this step. Note that the efficient cluster-Otsu method is used to make binary images. Fig.5. illustrates the operation of finding candidate regions on the image which resulted in previous step.

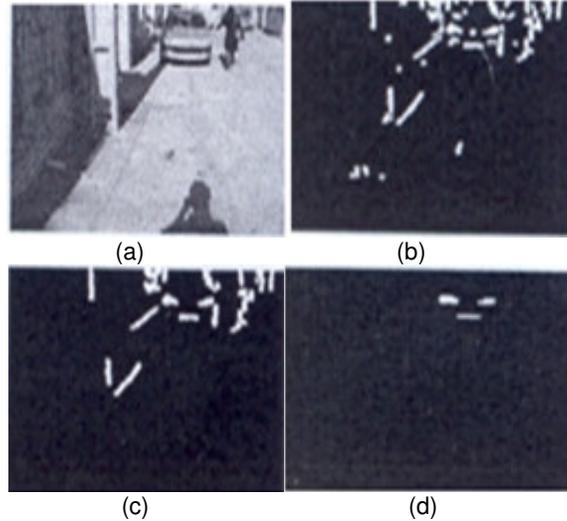


FIGURE 5: (a) primary image, (b) regions with vertical edges, (c) removed regions, less than the threshold and (d) removed additional regions, by applying two terms.

2.4. Candidate regions

Extracted regions after primary filtering are considered as candidate regions.

2.5. Real LPL detection

In this unit real license plate localization is determined among candidate regions. It includes two stages. In the first stage rectangularity percentage of the candidate regions is attained and the color filter is used to the rest of the candidate regions as a second stage.

2.5.1 Rectangularity percentage

In this section four criteria have been used to determine the rectangularity percentage. The first criterion is the 'aspect ratio' of width to height: for a real license plate, this Value is five. However, when we have a pan or tilt condition for the plate this value is less than five. The 'Perimeter ratio' is used as the second criterion. For a rectangle, the parameter ratio is one. It is a good criterion to estimate the percentage of rectangularity on candidate regions. The third criterion is 'Extend'. Each shape with a completely filled box, can be used as a suitable feature for calculating the percentage of rectangularity. Finally, the fourth criterion is Solidity. Rectangular is the smallest convex polygon that can surround objects. Oghlidian distances from ideal values of these criteria are calculated at each step and then normalized by Gaussian filter. Fig.6. describes the explained procedure. The result of this stage is shown in Fig.7.

2.5.2. Color filter

Color is another feature which is used in the LP system. There are many color models which can be used in LPL unit. It's clear that, color model type varies according to the characteristics of car LP in each country. For instance, yellow background is used in Italian or Egyptian LP. Current Iranian car plates have a blue region in the left. So, finding a color space which is sensitive to blue color is important. After many tests on different LP conditions are done YCbCr model is applied in the second stages of the LPL unit.

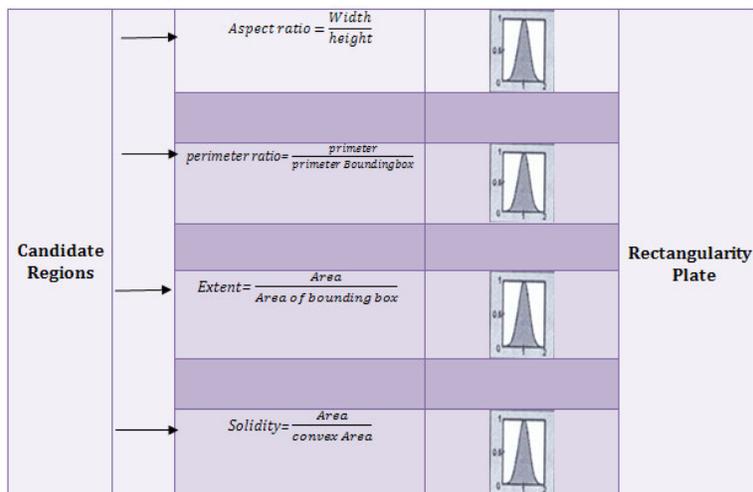


FIGURE 6: First stage of the license plate localization (LPL) unit.

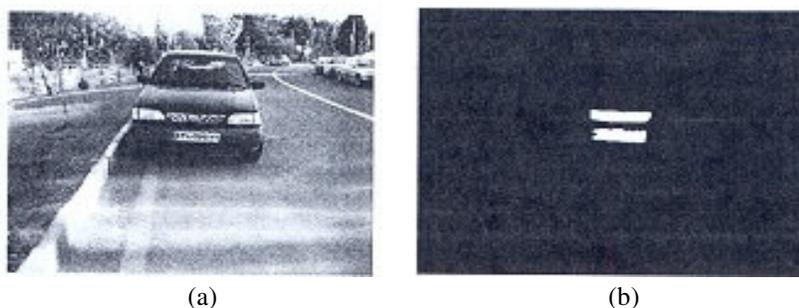


FIGURE 7: (a) input image and (b) output of the first stage of LPL unit.

2.5.2.1 YCbCr space

The CIEYCbCr is a scaled and an offset version of the YUV color space. The YUV color space is the basic color space used by the composite color video standards. The YUV signal separates the intensity Y from the color components UV that correspond to the hue and saturation aspects of the model. Therefore the UV components, also called chrominance, are usually subsampled in image applications. The YUV parameters in terms of RGB are:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.299 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

the YCbCr results from YUV as following

$$Y = 0.299R + 0.587G + 0.114B \quad (2)$$

$$Cb = B - Y \tag{3}$$

$$Cr = R - Y$$

2.5.2.2 Blue color filter

In this work, Gaussian filter has been used to detect blue region in the LP. Before applying Gaussian filter to Cr and Cb components of the image, YCbCr model has been applied. Statistical methods are used to determine the variance and average parameters needed in Gaussian filter. Also, color components are obtained by testing thirty car images in different conditions. For Gaussian filter we have:

$$f(x, y) = \exp\left[\frac{-1}{2} (X - m) C_x^{-1} (X - m)^T \right] \tag{4}$$

$$m = [\overline{C_b} \quad \overline{C_r}]$$

Where X is the color components vector in Cartesian that coordinates with the input image. μ , is the average vector and C_x , is covariance matrix. Fig.8. shows the results of applying the color filter to the image.

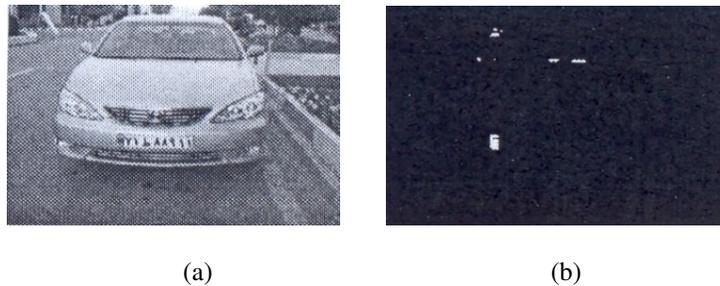


FIGURE 8: (a) input grey image and (b) extraction of the blue region from the RGB image by applying the color filter.

Using the blue color filter on two detected regions on the plate region in figure 8(b), the real plate is detected.

2.6. Detecting the angle of LP

One of the problems confronted in LP images is rotation of images which makes recognition of characters on the plate difficult. To estimate the rotation angle, moment of region scan be used. Degree of rotation is considered respected to camera diaphragm as:

$$\alpha = \frac{1}{2} \cdot \tan^{-1} \left[\frac{2m_{1,1}}{m_{2,0} - m_{0,2}} \right] \tag{5}$$

$$\mu_{i,j} = \sum_{x=1}^H \sum_{y=1}^W (x - \bar{x})^i (y - \bar{y})^j \tag{6}$$

Where $\mu_{i,j}$, is (i, j)th moment in x, y axes, α is the rotation value respected to x axis, H is the height of plate and W is the width of LP(it).After finding the angle, rotation adjustment will be done. New position assigned to each pixel is:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \quad (7)$$

Fig.9. shows the image frame before and after the rotation adjustment.



FIGURE 9: (a) the image before rotation and (b) rotated image after rotation adjustment.

2.7. Shadow removal

The Shadow in images can be divided into regular and irregular. In the irregular shadow, LP is exposed to different illustrations. Frequency components are the main keys for the detection of the shadow. Frequency components of characters are much more than shadow components. Bothat morphology transform has been used to smooth the image. Its Structure element (SE) is a square element with length of $K=0.25H$ in which H is the height of LP.

2.8. License plate crop

The proposed approach is based on the vertical gradient operation. By applying $[-1 \ 1 \ 0]$ operator, gradient amplitude is obtained for each grey pixel. Profile of the image, $f(x, y)$, is:

$$projection(x) = E_R = \sum_{y=1}^W \left| \frac{\delta f(x, y)}{\delta x} \right| \quad (8)$$

Where y shows the column of grey image and W is the width of LP. For an image with $W \times H$ size, $H \times 1$ is assigned to E_R .

Excess part of the LP is removed by applying a threshold value about maximum value of E_R .an efficient crop is done by using the following equation:

$$\begin{aligned} R_{\max} &= \arg \text{Max} \{ E_R \} \\ \Rightarrow \{ R_{\text{high}} &= \{ R \mid E_R = 0.2E_{\max}, R = 1 \rightarrow R_{\max} \\ \Rightarrow \{ R_{\text{bottom}} &= \{ R \mid E_R = 0.2E_{\max}, R = R_{\max} \rightarrow H \end{aligned} \quad (9)$$

Where E_{\max} and R_{\max} are the maximum energy and related line respectively. R_{high} is the highest line and R_{bottom} is the lowest line to be cropped.

2.9. Binary unit

Before character segmentation, plate characters are converted to binary code. To convert the image, a threshold value is determined by using efficient cluster-Otsu method [20].

2. 10. Characters Segmentation

To segment characters, vertical projection is obtained by summation of columnar pixels which is defined by:

$$P_y(y) = \sum_x Image(x, y) \tag{10}$$

The pseudocode shown in Fig. 10 explains the method used to segment characters of the license plates [2]. Fig. 11 shows the results of the character segmentation.

```

For every column sum the pixels {
If (sum = 0) consider that a separating line;
}
Get the area between the separating lines;
Take the largest 8 areas to be the segments for the 8 characters;
    
```

FIGURE 10: Pseudocode for character segmentation

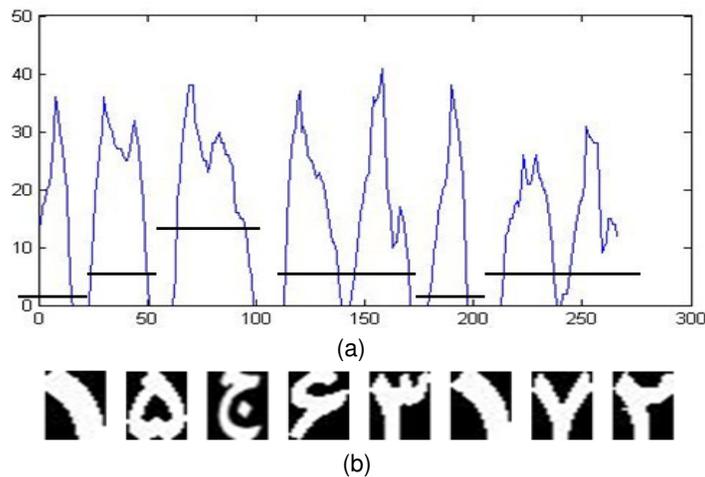


FIGURE 11: (a) vertical profile and (b) characters segmentation.

2.11. Dilating the character image

In this step of algorithm, a dilation operation is applied to the resultant segmented characters to enhance the image.

2.12. Resized characters

Before applying the neural networks, the characters are re-sized to ' 20 * 30 '.

3. NEURAL NETWORK AND LEARNING DATASET

Twenty different images for each character were taken and used for constructing a learning dataset, which all had a standard size. Aspect Ratio of zero pixels to whole pixels, Top-down projection and/or left right projection could have been used to extract the feature vector for each character and so. However, for the sake of simplicity, in this work the feature vector for each character was obtained by summing the Pixels in each row and column. The ANN unit in this work uses MLP network after testing different networks (Table1).

4. EXPERIMENTAL RESULTS

To show the efficiency of the proposed system, experiments have been done on numerous captured images including various types of vehicles with different lighting and noise effects. Also, experiments have been carried out using different camera positions and different orientations. To validate the results, each experiment is repeated 5 times and the result is reported as an average over these 5 repetitions. Moreover, in this section we compare the performance of the proposed method with some other methods: line sensitive filters (Luis et al., 1999), row-wise and column-wise DFTs (Parii, Claudio, Lucarelli, & Orlandi, 1998), and edge image improvement (Ming et al., 1996 & Zheng et al., Zheng, Zhao, & Wang, 2005), and fuzzy logic approach (Zimic et al., 1997) Works well under the assumptions that the Majority of plates are white and black characters, while most of the Italian license plates are yellow with black characters. Therefore, these three methods are not employed in our comparative experiments. The method proposed in Zheng et al. (2005) has some drawbacks. It has much computational time; some characters in this method are relative to the estimated size of the license plate. So, the method works well in case of that the license plates in images have the same size. Besides, no handling of rotational adjustment is considered. The "line sensitive filters" method consists of three steps" subsampling image, applying line sensitive filters, and looking for rectangular regions containing maxima of response. The "row-wise and column-wise DFTs" method involves four steps: decomposing expected harmonics by using horizontal DFT on the image, averaging the harmonics in the spatial frequency domain, finding the horizontal strip of the image containing the plate by maximizing the energy, and finding the Vertical position of the plate in the same way by using vertical DFT on the candidate strips. "The edge image improvement" method contains five steps: extracting the edge image using Sobel operators, computing the horizontal projections of the edge image, calculating the medium range of the edge density level, eliminating the highest and lowest portion of the horizontal projections to simplify the whole image, and finding the candidates of the license plates.

Two sets of vehicle images are used in our experiments. The first set has 170 images and they are captured on gates at different camera positions and orientations. The second set has 150 images, which are captured in the shadow of strong sunlight near a road, lighting effects, plate damage, dirties and complex backgrounds. The dataset used in the experiments consists of 340 characters. 205 characters were used for training and 135 characters for test. Table 1 shows the results of four neural networks on this dataset. According to these results, MLP network has been chosen due to its better performance. Precision Correct in table 1 is obtained by using the following equation:

$$Percent\ Correct = [test\ sample - \left(abs \left(\frac{A - LL}{2} \right) \right)] * 100 \quad (11)$$

Where A is the output of the network resulted from simulation and LL is the real value of output. All the mentioned methods have been applied to the two sets and the results have been shown in Table 2. From this table, it can be seen that the proposed method outperforms the other four methods. The high location rates on the three sets reveals robustness and efficiency of the proposed method. For the other methods,

most misallocated plates happened with images containing noise effects, rotations, and low-resolutions, and the images captured in string sunlight or under the gloomy light condition.

Number	Networks	Percent Correct (%)	Training Times
1	LVQ	76	24.4
2	Hopfield	83	30.4
3	RBF	97.6	27.4
4	MLP	99.2	28.18

Table 1: a Comparison on the efficiency of different networks in LP recognition:

The average processing time for the 5 stages of the license plate location in the proposed method has been shown in Table 3. Most of the elapsed time used is for the license plate extraction and rotation adjustment. The computational time needed by the five methods has been shown in Table 4, when they run on Pentium-four 2.4 GHz, 512 MB RAM PC.

Number	Methods	Datasets	Plates not Detected	Location rates (%)
1	Line sensitive filters , Luis et al. (1999)	Dataset1	10	94.2
		Dataset2	14	91.1
2	Row-wise and column -Wise DFTs, Parii et al. (1998)	Dataset1	8	5.7
		Dataset2	11	92.6
3	Edge image improvement Ming et al. (1996)	Dataset1	7	96.2
		Dataset	8	94.6
4	Zheng et al. method, Zheng et al. (2005)	Dataset1	5	97.1
		Dataset	8	95.2
5	The proposed method	Dataset	0	100
		Dataset	1	99.4

Table 2: Comparison of location rates of the five methods on two sets:

License plate Extraction	skew angle and rotation adjustment	Crop LP	shadow attenuation	LP-character segmentation	OCR	total time
0.93	1.13	0.12	0.1	0.28	0.14	2.7

Table 3: The processing time for each stage of the proposed method (s)

number	Method	Time(s)
1	Line sensitive Filters DFTs	2.1
2	Row-wise and column-wise	2.44
3	Edge image Improvement	1.97
4	Zheng et al. method	5.04
5	The Proposed method	2.3

Table 4: Comparison of LP location computational time of the five methods(s)

5. CONCLUSION

In this paper, we have proposed an intelligent vehicle-access control system based on an efficient license plate location and recognition approach. The multilayer perceptron neural network (MLPNN) is selected as a powerful tool to perform the recognition process. The proposed system is successfully implemented in experimental settings that are representative for of many practical simulations. The experimental results demonstrate that the proposed technique is capable of locating a CAR LICENSE PLATE to infer the position of a part of interest effectively. It has been shown that the system is able to handle images at low-resolutions and images containing noise. Furthermore, experimental results have proved that the proposed system is able to locate and recognize LABELS even in damaged patterns. A study of the different parameters of the training and recognition phases showed that the proposed system reaches promising results in most cases and can achieve high success rates. A great effect of our method in license plate location has been confirmed by the experiments.

References

- [1] Jianbin Jiao, Qixiang Ye & Qingming Huang(2009). A configurable method for multi-style license plate recognition. Pattern Recognition Letter.
- [2] Sherin M. Youssef, Shaza B & AbdelRahman(2008). A smart access control using an efficient license plate location and recognition approach. In proceeding of Expert Systems with Applications, 34, 256–265.
- [3] Wenjing Jia_, Huaifeng Zhang & Xiangjian He(2007). Region-based license plate detection. In proceeding of network and computer application,30,1324-1333.
- [4] C.N.E. Anagnostopoulos, I.E. Anagnostopoulos, V. Loumos, E. Kayafas, A license plate-recognition algorithm for intelligent transportation system applications, IEEE Trans. Intell. Transp. Syst. 7 (2006) 377–391.
- [5] H. Zhang, W. Jia, X. He, Q. Wu, Learning-based license plate detection using global and local features, in: International Conference on Pattern Recognition,2006, pp. 1102–1105.
- [6] F. Martn, M. Garca, L. Alba, New methods for automatic reading of VLP's (Vehicle License Plates), in: Proceedings of IASTED International Conference on SPPRA, 2002, Available from: <http://www.gpi.tsc.uvigo.es/pub/papers/sppra02.pdf>.
- [7] Lim, B.L., Yeo, W.Z., Tan, K.Y., Teo, C.Y., 1998. A novel DSP based real-time character classification and recognition algorithm for car plate detection and recognition. In: Fourth Internat. Conf. on Signal Process., vol. 2, pp.1269–1272.

- [8] Luis, S., Jose, M., Enrique, R., Narucuso, G., 1999. Automatic car plate detection and recognition through intelligent vision engineering. In: Proc. IEEE 33rd Internat. Carnahan Conf. on Security Tech., pp. 71–76.
- [9] Ming, G.H., Harvey, A.L., Danelutti, P.1996. Car number plate detection with edge image improvement, In: Fourth Internat. Symp. On Signal Processing and its Applications, vol.2, pp. 597–600.
- [10] Parisi, R., Claudio, E.D., Lucarelli, G., Orlandi, G., 1998. Car plate recognition by neural networks and image processing.In: Proc. IEEE Internat. Symp. on Circuits and Systems,vol. 3, pp. 195–198.
- [11] Kamat, Varsha, & Ganesan (1995). An efficient implementation of hough transform for detecting vehicle license plates using DSP's. In Proceedings of real-time technology and applications (pp. 58–59).
- [12] Lim, B. L., Yeo, W.Z., Tan, K. Y., & Teo, C. Y. (1998). A novel D.S.P based real-time character classification and recognition algorithm for car plate detection and recognition. In Proceedings of the fourth international conference on signal process, Vol. 2, (pp. 1269–1272).
- [13] Chandra Sekhar Panda, Srikanta Patnaik(2009).Filtering Corrupted Image and Edge Detection in Restored Grayscale Image Using Derivative Filters.International Journal of Image Processing (IJIP). Volume 3, Issue 3-(pp: 105-119).
- [14] Raman Maini, Himanshu Aggarwal.Raman Maini, Himanshu Aggarwal,(2009).Study and Comparison of Various Image Edge Detection Techniques International Journal of Image Processing (IJIP). Volume 3, Issue 1(pp: 1-11).
- [15] Luis, S., Jose, M., Enrique, R., Narucuso, & Narucuso, G. (1999). Automatic car plate detection and recognition through intelligent vision engineering. In Proceedings of IEEE 33rd international carnahan conference on security technology. (pp. 71–76).
- [16] Ming, G. H., Harvey, A. L., & Danelutti, P. (1996). Car number plate detection with edge image improvement. In Proceedings of the 4th international symposium on signal process and its applications. Vol. 2,(pp. 597–600).
- [17] Parii, R., Claudio, E. D., Lucarelli, & G., Orlandi, G. (1998). Car plate recognition by neural networks and image processing. In Proceedings of IEEE international symposium on circuits and systems. Vol. 3, (pp.195–198).
- [18] Park, S. H., Kim, K. I., Jung, K., & Kim, H. J. (1999). Locating car license plate using neural networks. Electronics Letters, 35(17), 1475–1477.
- [19] Rodolfo, Z., & Stefano, R. (2000). Vector quantization for license plate location and image coding. IEEE Transactions Industrial Electronics.47(1),159–167.
- [20] Wei, W., Wang, M. J., Huang, & Z. X. (2001). An automatic method of location for number plate using color features. In Proceedings of the international conference on image process. Vol. 1, (pp. 782–785).
- [21] Zheng, D., Zhao, Y., & Wang, J. (2005). An efficient method of license plate location. Pattern Recognition Letters, 26, 2431–2438.
- [22] Tabassam Nawaz, Syed Ammar Hassan Shah Naqvi, Habib ur Rehman , Anoshia Faiz(2009).Optical Character Recognition System for Urdu (Naskh Font)Using Pattern Matching Technique. International Journal of Image Processing (IJIP). Volume 3, Issue 3-(pp: 92-104).



COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA