Optimal Parameter Selection Technique for a Neural Network Based Local Thresholding Method

Article · July 2010

DOI: 10.13176/11.146

Mohammed J Islam
Shahjalal University of Science and Technology

Majid Ahmadi
University of Windsor

Yasser Alginahi

Journal of Pattern Recognition Research 1 (2010) 69-94

Received March 23, 2009. Revised February 10, 2010. Accepted July 23, 2010.

Optimal Parameter Selection Technique for a Neural Network Based Local Thresholding Method

Mohammed J. Islam islam1l@uwindsor.ca
Majid Ahmadi ahmadi@uwindsor.ca
Maher A. Sid-Ahmed ahmed@uwindsor.ca
Electrical and Computer Engineering, University of Windsor,
401 Sunset Avenue, Windsor, ON N9B 3P4, Canada

Yasser M. Alginahi yginahi@taibahu.edu.sa
College of Computer Science and Engineering, Taibah University,
P.O. Box 30002, Madinah Munawarah, The Kingdom of Saudi Arabia

Abstract

Thresholding of a given image into binary image is a necessary step for most image analysis and recognition techniques. In document recognition application, success of OCR mostly depends on the quality of the thresholded image. Non-uniform illumination, low contrast and complex background make it challenging in this application. In this paper, selection of optimal parameters for Neural Network (NN) based local thresholding approach for grey scale composite document image with non-uniform background is proposed. NN-based local image thresholding technique uses 8 statistical and textural image features to obtain a feature vector for each pixel from a window of size $(2n + 1)$x$(2n + 1)$, where $n \geq 1$. An exhaustive search was conducted on these features and found pixel value, mean and entropy are the optimal features at window size 3x3. To validate these 3 features some non-uniform watermarked document images with known binary document images called base documents are used. Characters were extracted from these watermarked documents using the proposed 3 features. The difference between the thresholded document and base document is the noise. A quantitative measure Peak-Signal-to-Noise ratio (PSNR) is used to measure the noise. In case of unknown base document characters were extracted through the proposed 3 features and used in a commercial OCR to obtain the character recognition rate. The average recognition rate 99.25% and PSNR shows that the proposed 3 features are the optimal compare to the NN-based thresholding technique with different parameters presented in the literature.

Keywords: Image Features, Image Segmentation, Neural Network, OCR

1. Introduction

Giant steps have been taken in the last two decades, both in terms of technological supports and in software products to provide computerized Document Analysis System (DAS). Optical character recognition (OCR) contributes to this progress by providing techniques to convert large volumes of text to readable file automatically. There are so many papers and patents claiming recognition rates as high as 99.99%, this gives the impression that OCR problems seem to have been solved. However, the failure of some real applications shows that performance problems subsist on composite and degraded paper documents with non-uniform background. Non-uniform background is caused by watermarks and complex patterns used in printing security documents. Transforming composite documents with non-uniform background into electronic format in a form suitable for efficient storage, retrieval and interpretation continues to be a challenging problem [5, 11, 12].

Thresholding is a popular tool for image segmentation that tries to identify and extract the object from its background on which it is superimposed. In general, performance of the image thresholding technique depends on the type of document, image illumination, contrast and the complexity of the image background. There are many thresholding schemes published in the literature and selecting an appropriate one can be a difficult task. NN-based local thresholding method is considered to be one of the best solutions in document recognition application.

Trier and Jain [1] present a comparison of 4 global and 11 local thresholding techniques. In Sahoo et. al. [3] 20 global thresholding techniques were compared. Sezgin and Sankur [10] conducted an exhaustive survey on 40 bi-level thresholding techniques. From the techniques used in these above mentioned literatures, Niblack [2] local adaptive method and Otsu [4] method were found to be the best candidate in thresholding scheme.

Very few researchers have investigated the use of NN in thresholding of grey level images. This is because of computational cost of the NN-based thresholding which makes it unsuitable for online applications. Koker and Sari [8] used NN to automatically select a global threshold value for an industrial vision system based on the histogram of the image. In this method the histogram of the supervised training data (256 levels) is used as input and global threshold value that is determined by visually inspecting the histogram of the training image is used as target value. Using this procedure the training data is prepared for the application in different illumination environment. In training phase, the produced output is compared to the target output, calculates the errors, adjust the synaptic weights of the NN until an acceptable weight is achieved to minimize the error between produced and target outputs. Adnan Khashman [13] used a supervised neural network for document image binarization where local threshold values are used to train the network in order to obtain the optimum global threshold values. 86% recgnition rate is obtained based on visual inspection for degraded non-uniform historical document images.

Alginahi et. al. [5, 12] developed NN-based technique for thresholding composite digitized documents with complex background. This method uses 8 statistical and textural features of an image for each pixel at a window size 5x5. A modified NN-based thresholding technique was developed by Alginahi [11] very recently where the author claimed 5 parameters at window size 3x3 provides the best results. But the author did not present any justification in support of the claim. The main aim of this paper is to conduct an exhaustive search on these 8 features and find the optimal parameters at optimal window size.

A handful of binarization performance criteria [14] have been proposed, based on OCR performance criteria which is one of the most acceptable method. Therefore, the optimality is measured quantitatively by PSNR for known base documents and recognition rate for unknown base documents. To test the performance of the optimal parameters in NN-based thresholding techniques a comparative study is carried out. In our experimentation, Otsu's global thresholding [4], Niblack [2], HMM [15] and Kittler[16] were utilized as benchmark for this comparison. To validate the optimal parameters, thresholded image using the proposed technique will be compared to the Niblack [2] and Otsu [4] techniques in terms of yielding the character recognition rate.

The remainder of this paper is organized as follows: In Section 2, statistical texture features are defined. NN-based thresholding technique and it's observation, criticisms and objectives are discussed in Sections 3 and 4 respectively. Finally, experimental results are discussed in Section 5 and Section 6 concludes the paper.

2. Statistical Texture Features

Statistical features are extracted for each centered pixel in a window of size $(2n+1)$x$(2n+1)$, where $n \geq 1$. This window can be randomly placed at different locations by the user in the area of the image in which features need to be extracted . The extracted features are then used in neural classifier to train it for the recognition of a particular area in images of similar nature. The NN-based thresholding technique takes advantage of the image textural characteristics by considering the statistical descriptors in a neighborhood of pixels. The statistical and textural features are useful in characterizing the set of neighborhood values of pixels, which are related to its moments. These features are **pixel value**, **mean**, **standard deviation**, **smoothness**, **entropy**, **skewness**, **kurtosis** and **uniformity**. These are adapted from [5, 6, 11, 12].

1. **Pixel Value**
   The center pixel, $p(i,j)$ in a window of size $(2n+1)$x$(2n+1)$, where $n \geq 1$, was taken as the first feature.

2. **Mean**
   The mean, μ_{ij} of the pixel values in the defined window, estimates the value around the pixel in which central clustering occurs.

$$\mu_{ij} = \frac{1}{(2n+1)^2} \sum_{x=i-n}^{i+n} \sum_{y=j-n}^{j+n} p(x,y) \tag{1}$$

3. **Standard Deviation**
   The standard deviation, σ_{ij} is the estimate of the mean square deviation of grey pixel value $p(x,y)$ from its mean value μ_{ij}. Standard deviation describes the dispersion within a local region. The standard deviation is defined as:

$$\sigma_{ij} = \frac{1}{(2n+1)} \sqrt{\sum_{x=i-n}^{i+n} \sum_{y=j-n}^{j+n} (p(x,y) - \mu_{ij})^2} \tag{2}$$

4. **Smoothness**
   Relative smoothness, R_{ij} is a measure of grey level contrast that can be used to establish descriptors of relative smoothness.

$$R_{ij} = 1 - \frac{1}{1 + \sigma_{ij}^2} \tag{3}$$

5. **Entropy**
   Entropy, h_{ij} can also be used to describe the distribution variation in a region. Overall entropy of neighborhood pixels in the window can be calculated as:

$$h_{ij} = -\sum_{k=0}^{L-1} Pr_k (log_2 Pr_k) \tag{4}$$

   Where Pr_k is the probability of the k-th grey level in the range [0 1], which can be calculated as $Z_k/(2n+1)^2$, Z_k is the total number of pixels with the k-th grey level and L is the total number of grey levels in the window.

6. **Skewness**
Skewness, S_{ij} characterizes the degree of asymmetry of a pixel distribution in the specified window around its mean. Skewness is a pure number that characterizes only the shape of the distribution.

$$S_{ij} = \frac{1}{(2n+1)^2} \sum_{x=i-n}^{i+n} \sum_{y=j-n}^{j+n} \left[\frac{p(x,y) - \mu_{ij}}{\sigma_{ij}} \right]^3 \tag{5}$$

7. **Kurtosis**
Kurtosis, K_{ij} measures the peakness or flatness of a distribution relative to a normal distribution. The conventional definition of kurtosis is:

$$K_{ij} = \left\{ \frac{1}{(2n+1)^2} \sum_{x=i-n}^{i+n} \sum_{y=j-n}^{j+n} \left[\frac{p(x,y) - \mu_{ij}}{\sigma_{ij}} \right]^4 \right\} - 3 \tag{6}$$

The last term (-3) makes the value zero for a normal distribution.

8. **Uniformity**
Uniformity, U_{ij} is a texture measure based on histogram and is defined as:

$$U_{ij} = \sum_{k=0}^{L-1} Pr_k^2 \tag{7}$$

U is maximum for which all grey levels are equal. Before computing any of the texture features described above, the pixel values of the image were normalized by dividing each pixel by 255 in order to achieve computational consistency.

3. NN-based Thresholding

The statistical textural features are useful in characterizing the set of neighborhood values of pixels. Alginahi [5, 11, 12] utilized these features value in preparing data for training a Multi-Layer Perceptron (MLP) NN which are obtained from window of size $(2n+1)$x$(2n+1)$, taken from various parts of one image and repeated over many images. Figure 1 shows a sample screen capture of the training data preparation. The first step to prepare data for training is to load an image then a point or pixel in the image is selected. It is supervised training data preparation method. Therefore, the user knows about the selected pixel whether it is background or foreground. Accordingly, the user clicks on the object or background buttons in order to calculate the feature vector of that point. The process is repeated for different random points in the same image and different images with complex backgrounds to get a wide variety of features. The 8 features and their corresponding target value (for background=1 and foreground=0) are stored in a file and then used as inputs to the NN to train the network to produce the weights needed for testing the classifier.

4. Observations, Criticism and Objectives

The NN-based image thresholding algorithm proposed by Alginahi uses 8 features in the paper [5, 12] at window size 5x5 and 5 features in the paper [11] at window size 3x3 as an input to the NN. But increase in features results in more computational cost. Furthermore, larger window size also increases the computational cost hence slower the feature extraction

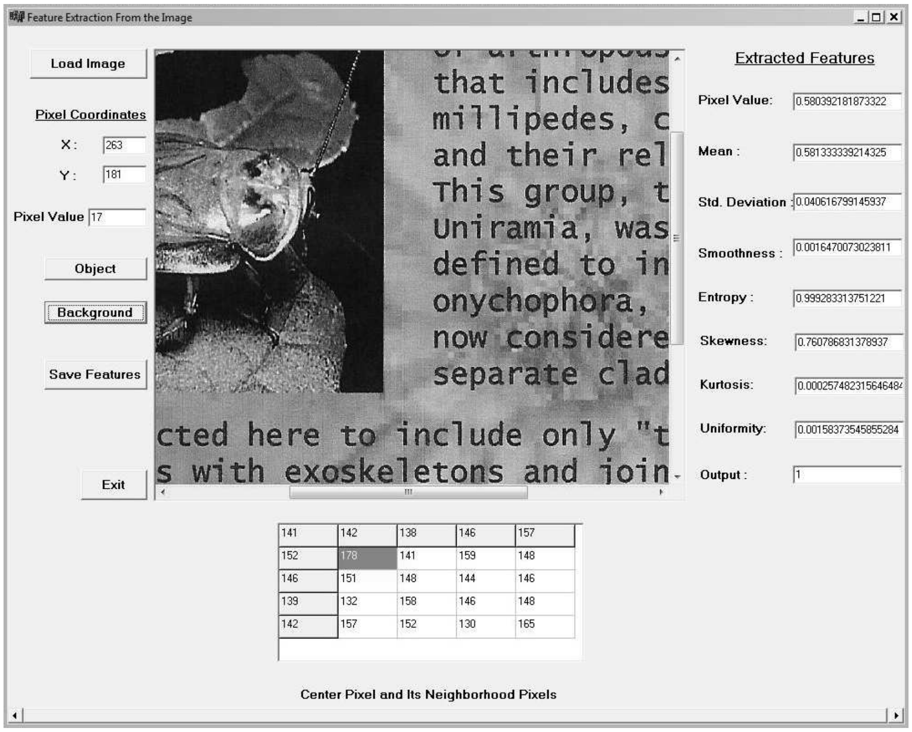


Fig. 1: Sample Training Data Preparation

process. In this paper, we are questioning the validity of the number of features and window size used in the method [12, 11], whether these 8 or 5 features with the window size of 5x5 or 3x3 are optimal in terms of yielding the best recognition rate and providing the best PSNR. Therefore, the main objective of this research is to conduct an exhaustive search to find and validate the optimal window size and the optimal feature combination among 8 features that provide the best recognition rate and PSNR when it is compared to other similar techniques.

4.1 Testing Setup

The main objective of this paper is to find the optimal window size and the optimal combination of features in terms of yielding the best recognition rate and best PSNR. To fulfill these objectives, rigorous methodological steps are followed. First of all, 8 features are labeled as 1, 2, 3, 4, 5, 6, 7 and 8 (see Section 2). Total 255 subsets are possible without repetition from this set of 8 features that are determined using the well known mathematical expression shown in equation 8.

$$^{n}C_{r} = \frac{n!}{r!(n-r)!} \tag{8}$$

For example, if we consider 7 features out of 8 features, 8 subsets are possible.

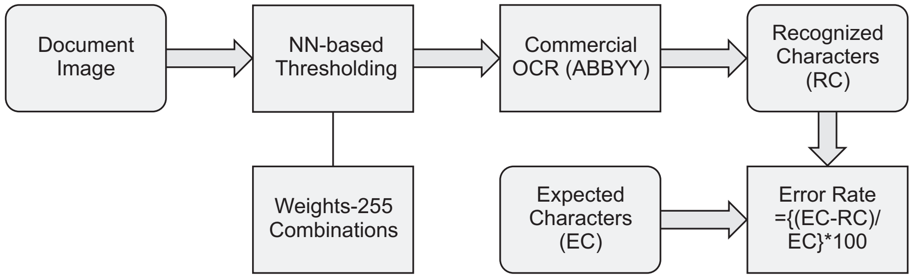
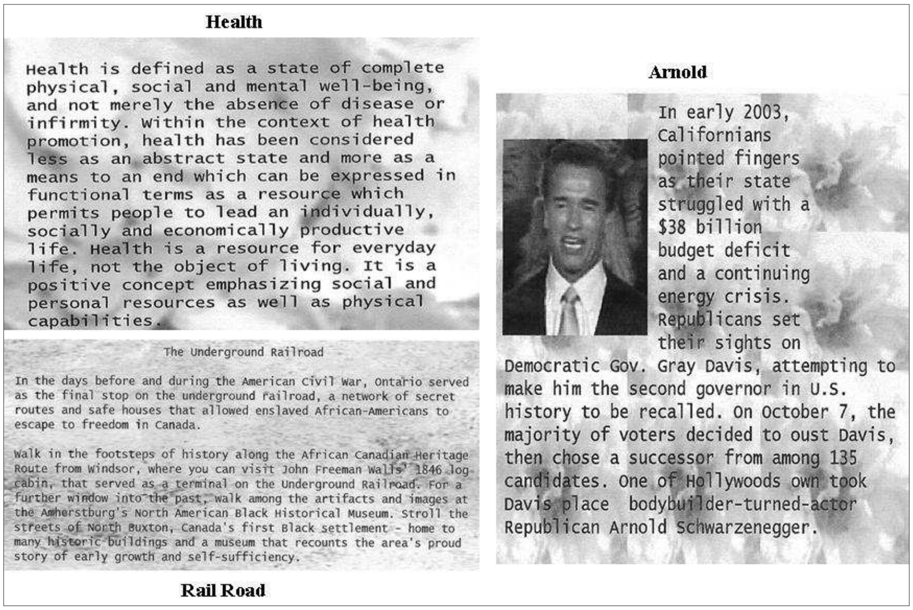


Fig. 2: Test Setup

For each combination, corresponding features are extracted from the main training dataset for a specific window size. The feature vector for this combination is used to train the NN and the corresponding weights are calculated. The weight vector is used to test the NN for thresholding the document image. The binary image is then passed to the commercial OCR for example ABBYY (Version 7.0) to obtain the recognized characters. The recognized characters are then compared to the expected characters. The expected characters database can be generated by directly feeding the document to the commercial OCR and correcting any misclassified characters from the original image just for testing purpose. The error rate is calculated from the expected characters and recognized characters for this combination. For each one of the 255 combinations, by following the same procedure, we can have 255 binary images as well as error rates when they are passed through ABBYY for a specific window size. The recognition rate for each subset is obtained with respect to different window size for example 3x3, 5x5, 7x7 and 9x9 and tabulated them in the grid shown in appendix A. The highest recognition rate and its corresponding feature combination and window size is determined. The same process is followed for more testing images and obtain the best feature combination for each image and then select the best among best combinations. The flow chart of the test setup is shown in Figure 2.

5. Experimental Results

The NN-based local thresholding technique is applied to several images for preparing the training dataset as well as for testing. Total 120 synthetically produced images were used in that experimentation. About 25% images were used for generating training dataset using all 8 features and rest of them were used for testing purpose. Since 255 different combinations are available with 8 features, therefore for each testing image total 255*4=1020 different binary images are available. These images are passed through the commercial OCR and corresponding character recognition rates are obtained. The highest recognition rate represents the best feature combination. The whole image set is divided into three groups; simple, moderate and complex and testing was done in all groups in terms of the complexity of the background and resolution. Sample document image with complex background and its corresponding segmented images are shown in Figure 3 and 4. For each document image we have 1020 binary images as well as 1020 recognition rates. In this experiment initially we have chosen one simple image of 654 characters shown in Figure 3 (Rail Road). The recognition rate corresponding to the subsets for each window size is tabulated in the grid shown in appendix A. Three subsets are found best in terms of their recognition rate and they are very close to each other (1 or 2 characters difference) in different window sizes.


Health

Health is defined as a state of complete physical, social and mental well-being, and not merely the absence of disease or infirmity. Within the context of health promotion, health has been considered less as an abstract state and more as a means to an end which can be expressed in functional terms as a resource which permits people to lead an individually, socially and economically productive life. Health is a resource for everyday life, not the object of living. It is a positive concept emphasizing social and personal resources as well as physical capabilities.

Arnold

In early 2003, Californians pointed fingers as their state struggled with a $38 billion budget deficit and a continuing energy crisis. Republicans set their sights on Democratic Gov. Gray Davis, attempting to make him the second governor in U.S. history to be recalled. On October 7, the majority of voters decided to oust Davis, then chose a successor from among 135 candidates. One of Hollywoods own took Davis place bodybuilder-turned-actor Republican Arnold Schwarzenegger.

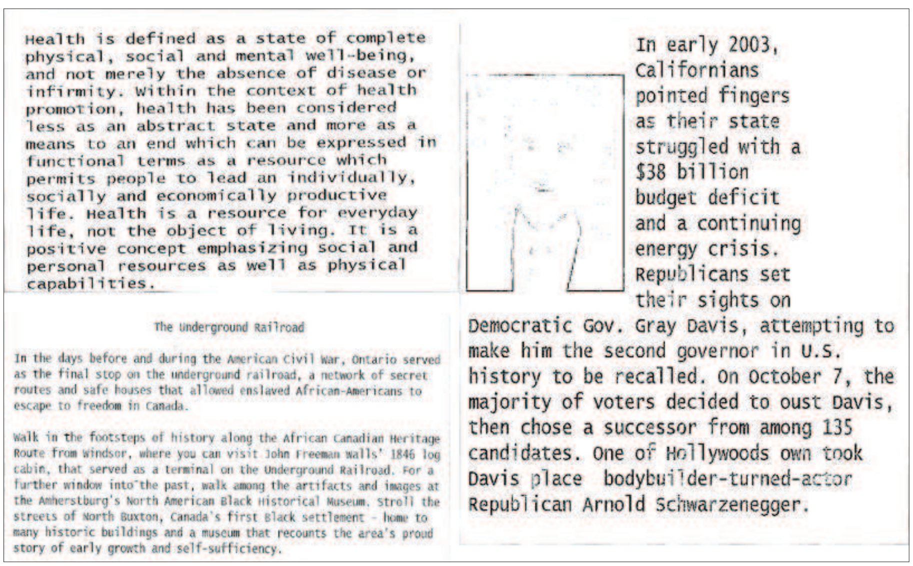
The Underground Railroad

In the days before and during the American Civil War, Ontario served as the final stop on the underground railroad, a network of secret routes and safe houses that allowed enslaved African-Americans to escape to freedom in Canada.

Walk in the footsteps of history along the African Canadian Heritage Route from Windsor, where you can visit John Freeman Walls' 1846 log cabin, that served as a terminal on the Underground Railroad. For a further window into the past, walk among the artifacts and images at the Amherstburg's North American Black Historical Museum. Stroll the streets of North Buxton, Canada's first Black settlement - home to many historic buildings and a museum that recounts the area's proud story of early growth and self-sufficiency.

Rail Road

Fig. 3: Sample Document Image


Health is defined as a state of complete physical, social and mental well-being, and not merely the absence of disease or infirmity. Within the context of health promotion, health has been considered less as an abstract state and more as a means to an end which can be expressed in functional terms as a resource which permits people to lead an individually, socially and economically productive life. Health is a resource for everyday life, not the object of living. It is a positive concept emphasizing social and personal resources as well as physical capabilities.

In early 2003, Californians pointed fingers as their state struggled with a $38 billion budget deficit and a continuing energy crisis. Republicans set their sights on Democratic Gov. Gray Davis, attempting to make him the second governor in U.S. history to be recalled. On October 7, the majority of voters decided to oust Davis, then chose a successor from among 135 candidates. One of Hollywoods own took Davis place bodybuilder-turned-actor Republican Arnold Schwarzenegger.

The Underground Railroad

In the days before and during the American Civil War, Ontario served as the final stop on the underground railroad, a network of secret routes and safe houses that allowed enslaved African-Americans to escape to freedom in Canada.

Walk in the footsteps of history along the African Canadian Heritage Route from Windsor, where you can visit John Freeman Walls' 1846 log cabin, that served as a terminal on the Underground Railroad. For a further window into the past, walk among the artifacts and images at the Amherstburg's North American Black Historical Museum. Stroll the streets of North Buxton, Canada's first Black settlement - home to many historic buildings and a museum that recounts the area's proud story of early growth and self-sufficiency.

Fig. 4: Sample Thresholded Image

The original image is then passed through the commercial OCR ABBYY and the recognition rate is compared with the results obtained from the binary images using best 3 subsets. The same process is repeated for 2 more images and results are tabulated in Table 1. Although we have obtained the recognition rate for window sizes 3x3, 5x5, 7x7 and 9x9, for comparison purpose we have shown the sample output in Table 1 only for window size 5x5, because our work is based on the paper [12] that describes the NN-based thresholding techniques and found 98% correct recognition rate at window size 5x5. Therefore, our intention is to show how our method performs for the same window size.

Table 1: Comparative Statement of Recognition Rates (%) - Using Binary and Original Image

Image	Binary Image - Features Subset			Original Image
	{1, 2, 5}	{1, 2, 6}	{1, 5, 6}	ABBYY (7.0)
Health	99.79	99.79	99.58	99.37
Rail	99.85	99.54	99.54	99.24
Arnold	99.51	99.01	96.30	96.79
Average	**99.72**	**99.45**	**98.07**	**98.47**

It is observed from Table 1 that feature subsets {1, 2, 5}, {1, 2, 6} and {1, 5, 6} provides the best results and these recognition rates are comparable to different window sizes. Then from these three combinations, two combinations {1, 2, 5} and {1, 2, 6} are picked based on their average recognition rate and we follow the same procedure for more images. The results are shown in Table 2 for window size 5x5.

Table 2: Comparative Statement of Recognition Rates (%) - Features Combination

Image	Features Subset	
	{1, 2, 5}	{1, 2, 6}
Niagara	100	99.41
Chretien	99.77	99.53
George	99.52	99.28
Volcano	100	99.71
Cats	99.66	99.66
Average	**99.79**	**99.52**

Considering all types of images and their corresponding recognition rate the combination of three selected features "**pixel value (1), mean (2) and entropy (5)**" was found to be the best combination with recognition rate of 99.79%. The same process is applied to more testing images for different window sizes using the best feature subset. For each window size the binary image is generated and is saved in a bit-map (BMP) file. The thresholding time is also recorded and is shown in Table 3 for sample images.

The computational time for processing one pixel centered on the different window size is shown in Table 5.

The processing time includes feature extraction, NN testing and finally classification of the pixel into object or background. The processing time for each pixel and character recognition rate with respect to different window size is shown in Figure 5 and 6 respectively.

From Tables 4, 5 and Figures 5 and 6, it is observed that window size does not play a significant role especially in character recognition rate. However, increase in window size results in higher computational costs hence slower the segmentation process. Therefore, comparing the computation costs and recognition rate, window size 3x3 is found to be the optimal in this application using NN-based local thresholding approach. This research indicated that 3 features used (Pixel value, Mean and Entropy) outperforms the 8 features utilized by Alginahi [5, 12] and 5 features utilized by the same author [11]. The processing time for 3, 5 and 8 features is tabulated in Table 6 for an image of size 689x723 at window size 3x3.

Table 3: Thresholding Time (Sec)

Image	**Image Size**	**Window Size** 3x3	5x5	7x7	9x9
Arnold	689x723	6.66	7.66	9.11	11.15
Health	543x759	5.60	6.57	7.77	9.44
Rail	529x1191	8.38	9.91	11.62	14.10
Cat	463x877	5.58	6.35	7.64	9.14
George	693x712	7.01	8.00	9.47	11.37
Volcano	492x886	5.96	6.85	8.25	9.84
Niagara	723x845	7.32	9.92	11.45	15.13
Chretien	499x726	5.59	6.32	7.46	9.36
Health	479x660	5.13	5.51	6.29	7.52
Mercury	538x799	6.43	7.94	8.19	10.33
Urban	680x966	10.11	11.08	12.42	15.73
Insect	869x1084	14.40	16.02	18.21	22.43

Table 4: Window Size Vs. Correct Recognition Rate (%)

Image	**Total Chars**	**Window Size - Misclassification** 3x3	5x5	7x7	9x9
Arnold	405	0	3	1	3
Health	476	2	2	2	0
Rail	654	0	1	1	2
Cat	585	0	1	1	1
George	414	0	1	1	1
Volcano	384	0	1	0	0
Niagara	339	1	0	0	0
Chretien	426	3	0	1	0
Health	383	0	0	0	0
Mercury	301	0	0	0	0
Urban	692	0	0	0	0
Insect	487	4	20	6	8
Total	5546	10	29	13	15
Correct (%)	100	**99.82**	**99.48**	**99.77**	**99.73**

Table 5: Window Size Vs. Process Time for each pixel

Window Size	**Process Time (μs)**
3x3	13.5
5x5	15.7
7x7	18.5
9x9	22.6

Table 6: Number of Features Vs. Process Time (Sec)

No. of Features	**Process Time (Sec)**
3 (Proposed)	6.60
5 (Alginahi [11])	8.47
8 (Alginahi [5, 12])	9.03

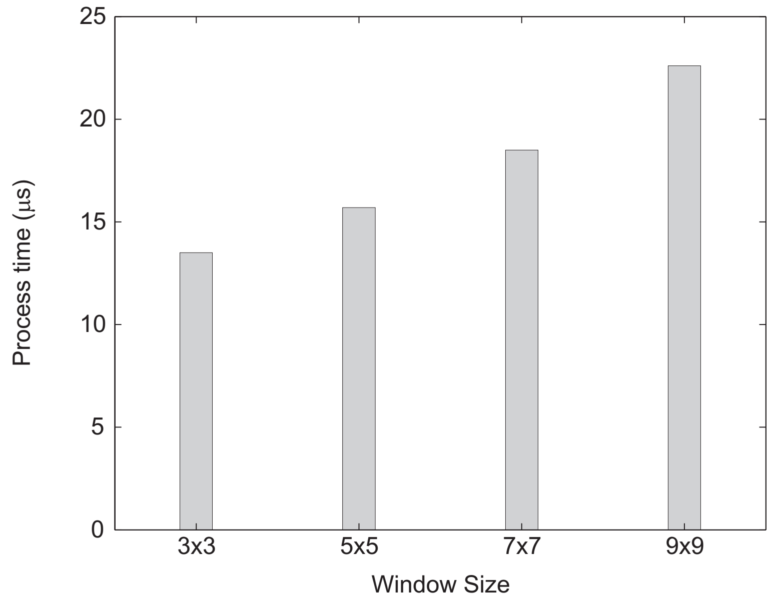


Fig. 5: Window Size Vs. Process Time (μs) for each pixel

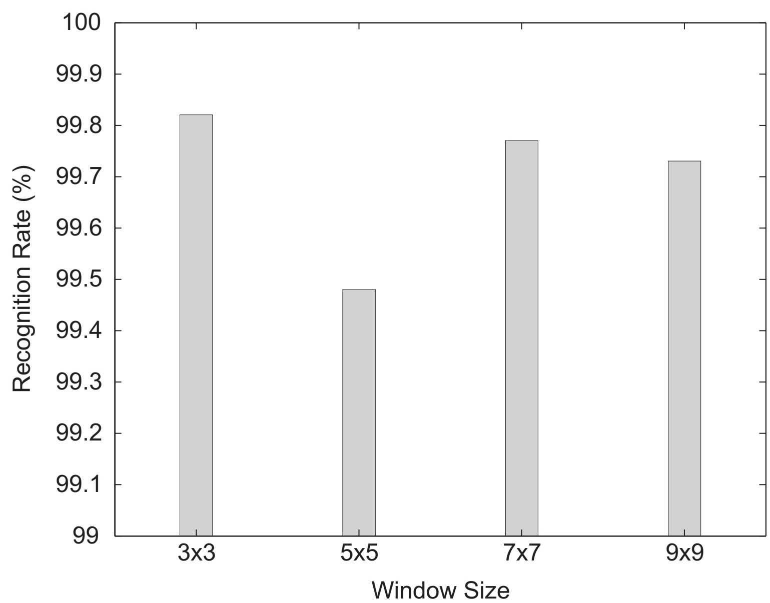


Fig. 6: Window Size Vs. Character recognition rate (%)

5.1 Optimal Features Validation

To validate the optimal features and the window size a known binary document image (black text in white background) called base document is used. That document is added to the simple watermark image to the complex watermark image. In each case the watermarked image is thresholded using 3, 5 and 8 features. The difference between the thresholded image and base document is the noise. PSNR is used to measure the noise in each case and it is tabulated in Table 7 for 4 sample images.

Sample base document, watermark image, watermarked document image and thresholded image using the proposed optimal features are shown in Figures 7-20.

Table 7: PSNR Measurement for Different Parameters

Image	Image Size	No. of Features	No. of Error Pixels	PSNR-dB
		3 (Proposed)	4	49.5575
Fig. 17	426x848	5 [11]	48088	8.7577
		8 [5, 12]	22077	12.1387
		3 (Proposed)	0	∞
Fig. 18	426x848	5 [11]	37187	9.8741
		8 [5, 12]	18230	12.9702
		3 (Proposed)	3992	19.5661
Fig. 19	426x848	5 [11]	22079	12.1383
		8 [5, 12]	18649	12.8715
		3 (Proposed)	1573	24.2698
Fig. 20	659x638	5 [11]	34527	10.8555
		8 [5, 12]	29443	11.5472

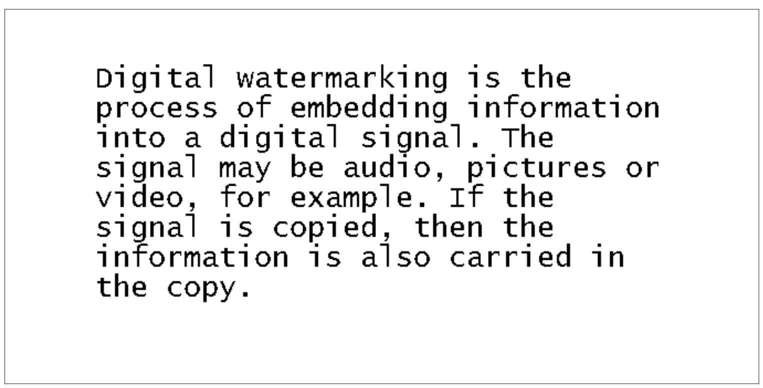

Digital watermarking is the process of embedding information into a digital signal. The signal may be audio, pictures or video, for example. If the signal is copied, then the information is also carried in the copy.

Fig. 7: Sample Base Document 1

Digital watermarking is the process
of embedding information into a
digital signal. The signal may be
audio, pictures or video, for
example. If the signal is copied,
then the information is also carried
in the copy.

In visible watermarking, the
information is visible in the picture
or video. Typically, the information
is text or a logo which identifies
the owner of the media. The image on
the right has a visible watermark.
When a television broadcaster adds
its logo to the corner of transmitted
video, this is also a visible
watermark.

Fig. 8: Sample Base Document 2

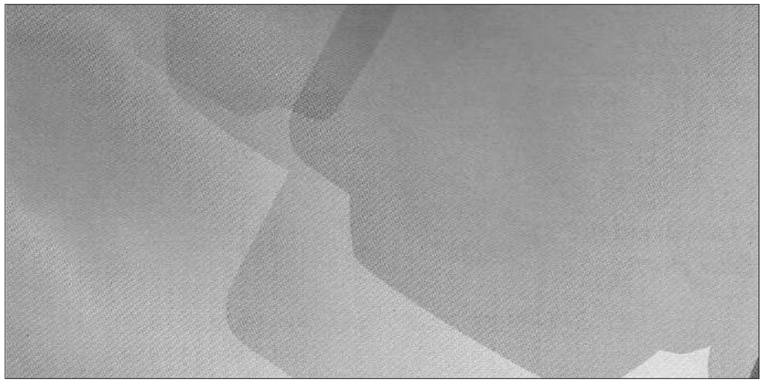


Fig. 9: Sample Watermark 1

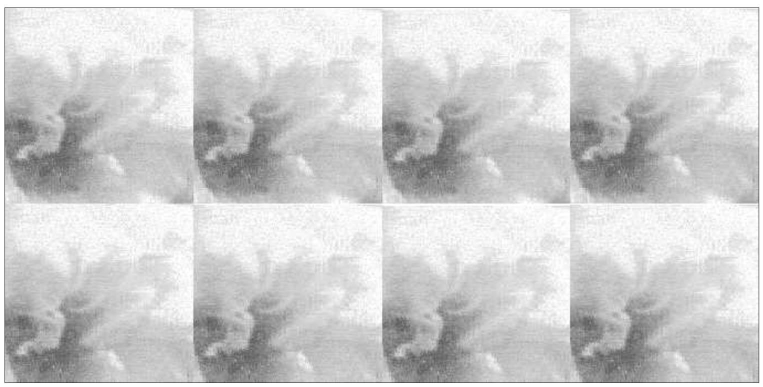


Fig. 10: Sample Watermark 2

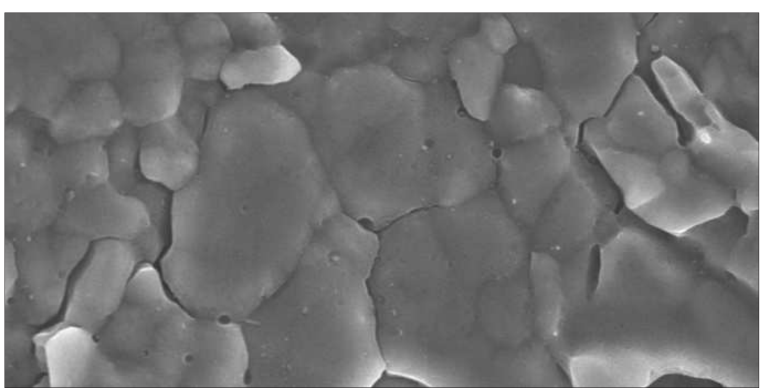


Fig. 11: Sample Watermark 3

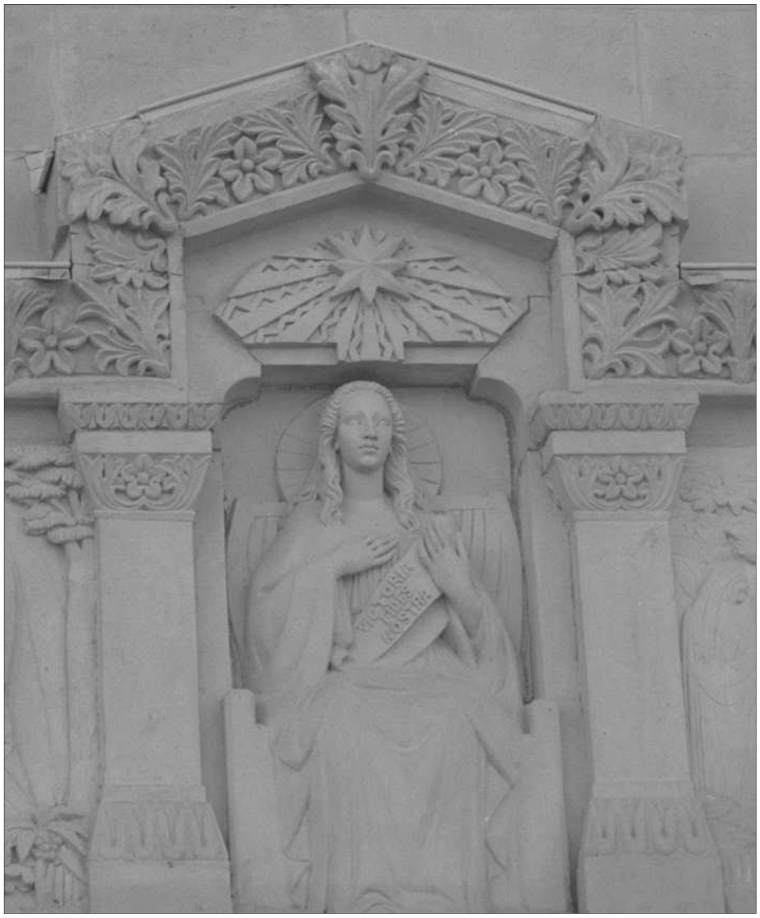


Fig. 12: Sample Watermark 4

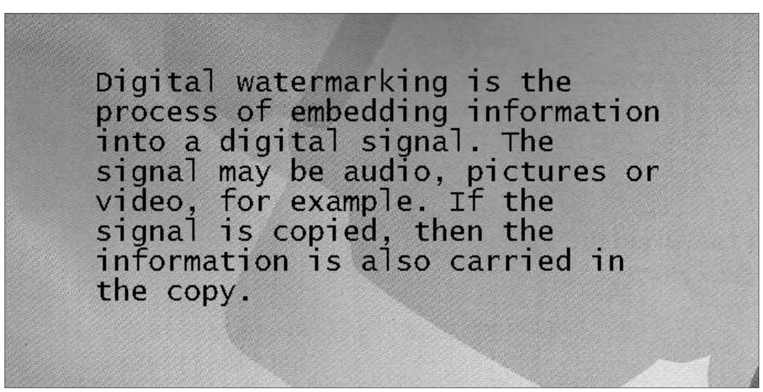


Fig. 13: Sample Document Watermarked Image 1

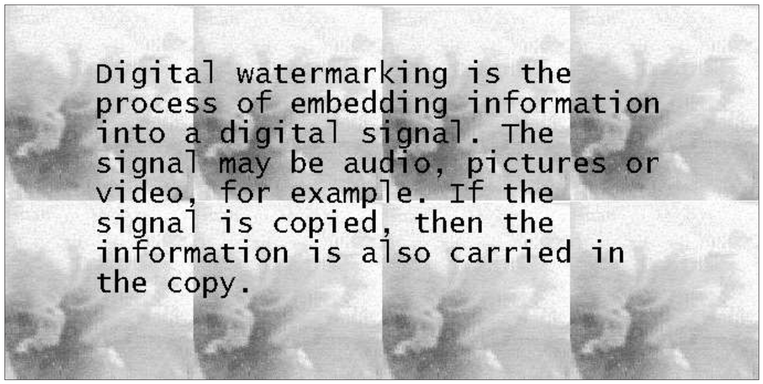


Fig. 14: Sample Document Watermarked Image 2

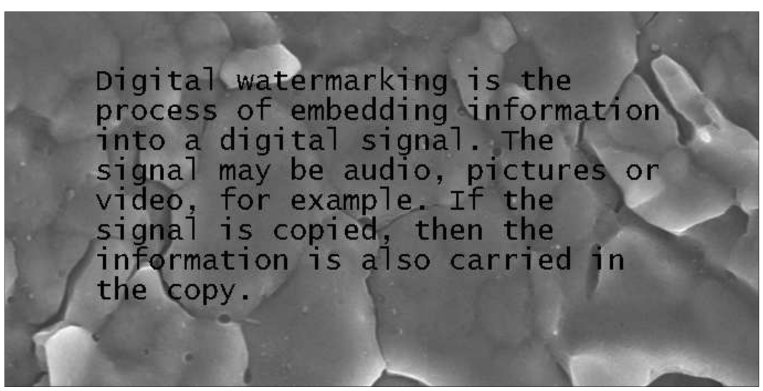


Fig. 15: Sample Document Watermarked Image 3

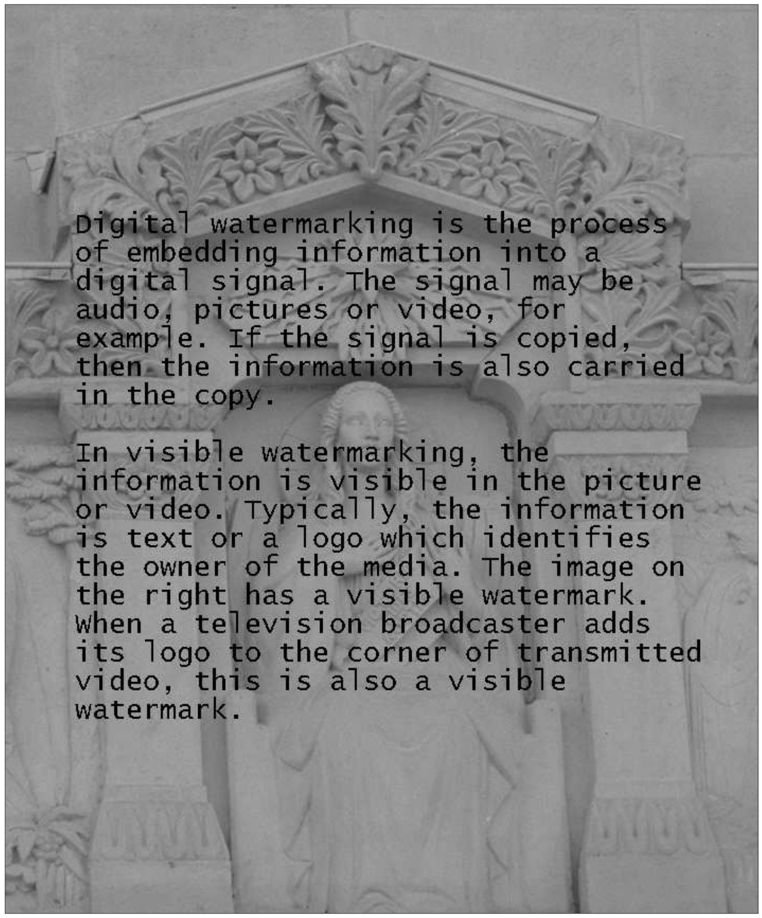


Fig. 16: Sample Document Watermarked Image 4

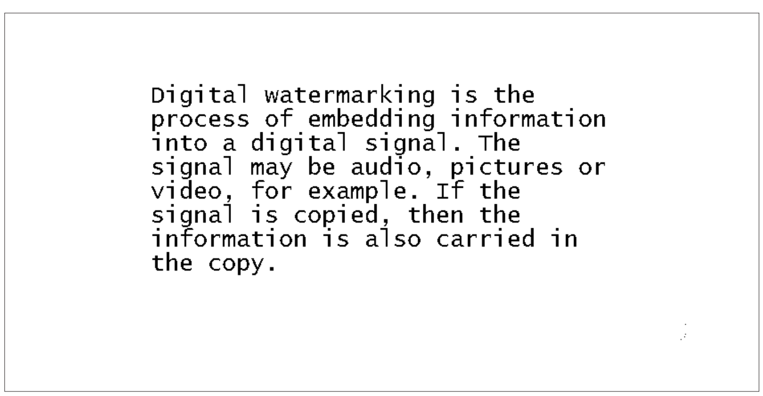


Fig. 17: Thresholded Image 1 Using Proposed 3 Features

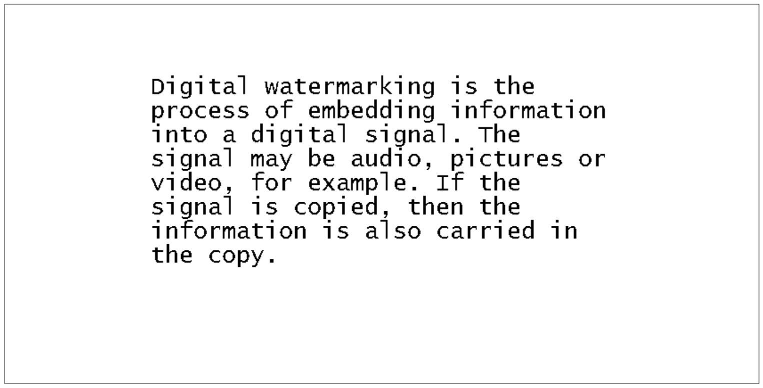


Fig. 18: Thresholded Image 2 Using Proposed 3 Features

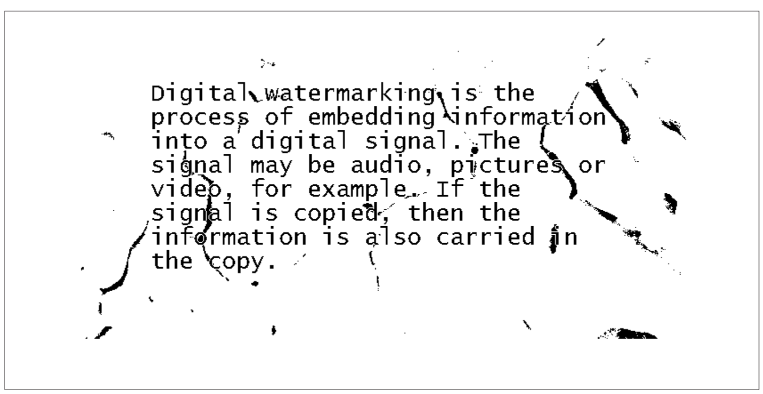
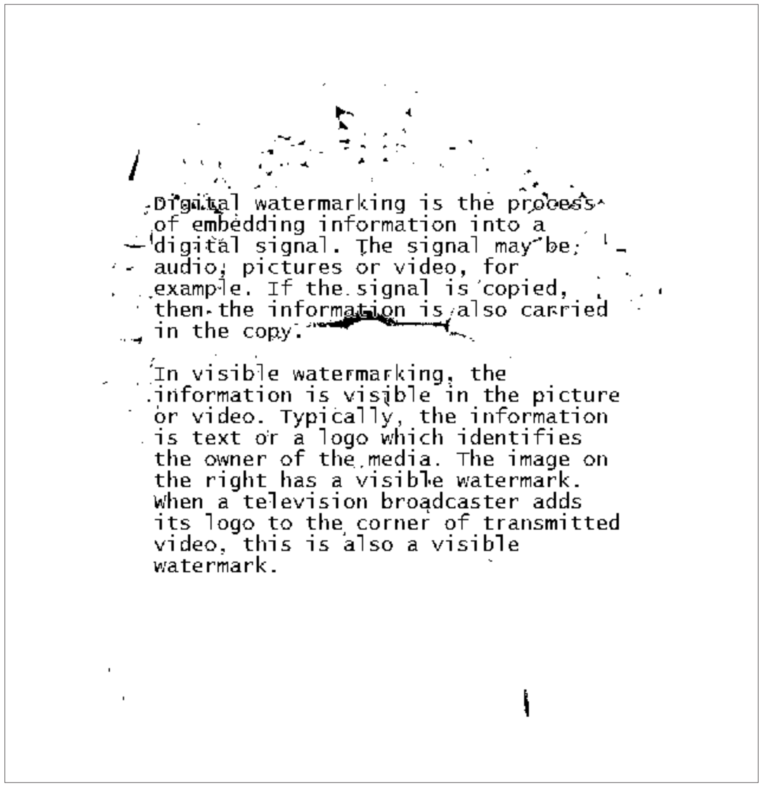


Fig. 19: Thresholded Image 3 Using Proposed 3 Features


Digital watermarking is the process
of embedding information into a
digital signal. The signal may be
audio, pictures or video, for
example. If the signal is copied,
then the information is also carried
in the copy.

In visible watermarking, the
information is visible in the picture
or video. Typically, the information
is text or a logo which identifies
the owner of the media. The image on
the right has a visible watermark.
When a television broadcaster adds
its logo to the corner of transmitted
video, this is also a visible
watermark.

Fig. 20: Thresholded Image 4 Using Proposed 3 Features


Recognized	Expected
• in early 2003, Californians pointed fingers as their state struggled with a $38 billion budget deficit and a continuing energy crisis. Republicans set their sights on Democratic Gov. Gray Davis, attempting to make him the second governor in U.S. history to be recalled. On October 7, the majority of voters decided to oust Davis, then chose a successor from among 135 candidates. One of Hollywoods own took Davis place bodybuilder-turned-actor Republican Arnold Schwarzenegger.	• In early 2003, Californians pointed fingers as their state struggled with a $38 billion budget deficit and a continuing energy crisis. Republicans set their sights on Democratic Gov. Gray Davis, attempting to make him the second governor in U.S. history to be recalled. On October 7, the majority of voters decided to oust Davis, then chose a successor from among 135 candidates. One of Hollywoods own took Davis place bodybuilder-turned-actor Republican Arnold Schwarzenegger

Fig. 21: OCR output: From Binary Image (Proposed) vs. Expected

Using the window size 3x3 and the optimal feature combination the binary image is passed through ABBYY and sample OCR output and the expected output is shown in Figure 21.

To validate the performance of the proposed combination the output from the original image is compared to the expected output that is shown in Figure 22. The output of the proposed NN-based thresholding technique using 3 best features at window size 3x3 is compared to Otsu [4] and Niblack [2] method and their outputs are shown in Figures 23, 24 and Table 8 for a document image of 405 characters (Fig. 3 - Arnold).

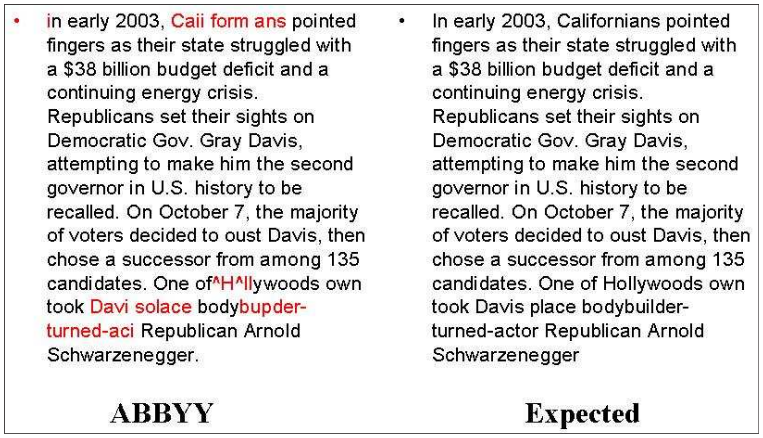


Fig. 22: OCR output: From Original Image vs. Expected

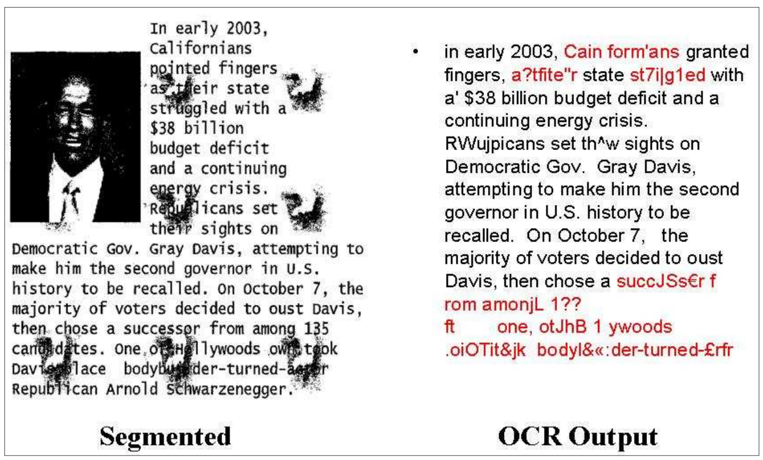


Fig. 23: OCR output: From Binary Image (Using Otsu [4])

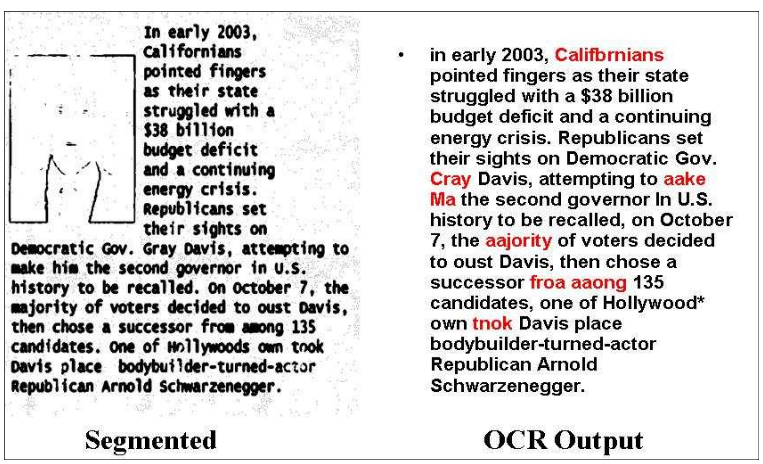


Fig. 24: OCR output: From Binary Image (Using Niblack [2])

Table 8: Recognition Rate Comparison- Proposed, Niblack [2] and Otsu [4]- (Fig. 3- Arnold)

Techniques	Misclassification (chars)	Correct Recognition Rate (%)
Proposed	0	100
Niblack	11	97
Otsu	59	85

Using these three features (Pixel value, Mean and Entropy) and window size 3x3, segmentation is performed for simple, moderate and complex images. Another sample segmented image using proposed features is compared with hidden Markov model (HMM) [15]-based binarization, Otsu [4] and Kittler [16] method that is shown in Figure 25.

Fig. 25: Segmentation Performance Comparison: Proposed, HMM[15], Kittler [16] and Otsu [4]

From this figure it is obvious that Kittler [16] and HMM [15] based segmentation has almost identical performance. But the proposed one is better than that.

Then the segmented image is passed through the ABBYY commercial OCR to obtain the overall simulation results for around 15000 characters. The simulation result is shown in Table 9.

Table 9: Performance of the Proposed Features (Pixel, Mean, Entropy)

Technique	Tested Chars	Recognition Rate (%)
Proposed, 3 features	**15000**	**99.25**
ABBYY [5] - Original Image	14600	96
Alginahi [5], 8 features	14600	98
Alginahi [11], 5 features	Not Mentioned	98.3

6. Conclusions

An exhaustive search was conducted to develop an efficient thresholding technique based on the work reported in [5, 11, 12]. In this paper, a method is devised to find the optimal parameters and optimal window size in NN-based thresholding technique to segment the non-uniform and composite document images. Three features **pixel value**, **mean** and **entropy** and window size 3x3 are found to be optimal in this technique in terms of PSNR, computational costs and correct recognition rate. To measure the performance of the proposed method, the segmented image obtained by the optimal features was compared with HMM [15], Otsu[4], Niblack[2] and Kittler[16]- based segmented image. For known base documents and watermark images, the noise is measured in terms of PSNR from the segmented image obtained using proposed 3 optimal features, 5 features [11] and 8 features [12] and is compared. The result shows the proposed method outperforms the reference methods. Then the optimal features were used to segment the same set of document image files used in [5] and applied to a commercial OCR ABBYY and 99.25% recognition rate is achieved for around 15000 characters. Our quantitative study shows that recognition rate obtained using the proposed method is much higher than the methods found in the literature.

Acknowledgments

This research has been financially supported by Ontario Student Assistance Program through Ontario Graduate Scholarship (OGS) and Research Center for Integrated Microsystems (RCIM) of University of Windsor. The financial support of these organizations is greatly acknowledged with appreciation.

References

[1] O.D Trier and A.K. Jain, *Goal-directed evaluation of binarization methods.* IEEE Trans. on Pattern Recognition And Machine Intelligence, Vol. 17, No. 12, pp. 1191-1201, 1995.

[2] W. Niblack, *An introduction to Digital Image Processing.* Prentice Hall,Eaglewood Cliffs, NJ, 1986.

[3] P.K. Sahoo, S. Soltani, A.K.C. Wong, *A Survey of thresholding techniques.* Computer vision, Graphics and image Processing, Vol. 41, pp. 233-260, 1988.

[4] N. Otsu, *A Threshold Selection Method From Gray Level Histograms.* IEEE Trans. On Systems, Man and Cybernetics, Vol. 9, pp. 62-66, 1979.

[5] Y. Alginahi, *Computer Analysis of Composite Documents with Non-uniform Background.* PhD Thesis, Electrical and Computer Engineering, University of Windsor,Windsor, ON, 2004.

[6] M.A. Sid-Ahmed, *Image Processing Theory, Algorithms and Architectures.* McGraw-Hill, New York, 1995.

[7] R.C. Gonzalez, and R.E. Woods, *Digital Image Processing.* Prentice-Hall, New Jersy, 2002.

[8] R. Koker and Y. Sari, *Neural Network Based Automatic Threshold Selection for an Industrial Vision System.* Proc. Int. Conf. on Signal Processing, pp. 523-525, 2003.

[9] N. Papamarkos, *A Technique for Fuzzy Document Binarization.* Proc. Of the ACM Symposium on Document Engineering, pp. 152-156, 2001.

[10] M. Sezgin and B. Sankur, *Survey over image thresholding techniques and quantitative performance evaluation.* Journal of Electronic Imaging, Vol. 13, No. 1, 2004.

[11] Y. Alginahi, *Thresholding and character recognition in security documents with watermarked background.* Proc. Int. Conf. on Digital Image Computing: Techniques and Applications (DICTA'08), pp. 220-225, 2008.

[12] Y. Alginahi, M.A. Sid-Ahmed and M. Ahmadi, *Local thresholding of composite documents using Multi-layer Perceptron Neural Network.* The 47-th IEEE International Midwest Symposium on Circuits and Systems, pp. I209- I212, 2004.

[13] A. Khashman, B. Sekeroglu, *Document image binarization using a supervised neural network.* International Journal of Neural Systems, Vol 18, No 5, pp. 405-418, 2008.

[14] Y.J. Zhang, *A survey on evaluation methods for image segmentation.* Pattern Recognition Vol 29, pp. 13351346, 1996.

[15] S. Huang, M. Ahmadi, M.A. Sid-Ahmed, *A hidden Markov model-based character extraction method.* Pattern Recognition, Vol 41, pp. 2890-2900, 2008.

[16] J. Kittler, J. Illingworth, *Minimum error thresholding.* Pattern Recognition, Vol 19, pp. 41-47, 1986.

Appendix A. Recognition Rate- Grid-based Method

This table represents the recognition rate of a sample document image (figure 3- Rail Road) using a commercial OCR ABBYY 7.0. For 8 features 255 different subsets are available. Each subset is used to obtain the binary image at window size 3x3, 5x5, 7x7 and 9x9. The binary image is then passed through the commercial OCR ABBYY 7.0 and the obtained recognition rate is tabulated in the following Tables 10, 11 and 12. Therefore, total 255*4=1020 binary images as well as 1020 recognition rates are available for one document image.

Table 10: Recognition rate in various window size (%)

Features Subset	Recognition Rate 3x3	5x5	7x7	9x9	Features Subset	Recognition Rate 3x3	5x5	7x7	9x9
{1}	0	0	0	0	{1,4,5}	98.32	99.24	98.93	98.32
{2}	0	0	0	0	{1,4,6}	74.01	77.06	78.59	77.83
{3}	0	0	0	0	{1,4,7}	76.03	75.54	74.01	74.47
{4}	0	0	0	0	{1,4,8}	74.47	74.01	73.24	72.48
{5}	0	0	0	0	**{1,5,6}**	**99.24**	**99.54**	**99.08**	**98.93**
{6}	0	0	0	0	{1,5,7}	98.32	98.47	98.01	96.79
{7}	0	0	0	0	{1,5,8}	94.95	95.41	94.65	94.8
{8}	0	0	0	0	{1,6,7}	82.42	84.71	83.94	81.65
{1,2}	98.93	99.24	99.39	98.47	{1,6,8}	83.18	81.65	79.36	78.59
{1,3}	95.72	95.11	95.87	95.26	{1,7,8}	84.71	83.94	82.42	81.65
{1,4}	95.26	94.65	94.95	94.5	{2,3,4}	93.88	89.3	87.77	87
{1,5}	98.78	99.08	99.24	98.93	{2,3,5}	75.54	77.06	76.3	75.54
{1,6}	98.93	99.08	98.62	98.47	{2,3,6}	89.3	90.83	92.35	91.59
{1,7}	94.8	94.04	95.11	95.26	{2,3,7}	83.18	84.71	82.42	81.65
{1,8}	92.35	93.88	94.5	94.04	{2,3,8}	87.77	89.3	90.06	86.24
{2,3}	95.11	95.87	94.8	94.65	{2,4,5}	85.47	87.77	83.94	81.65
{2,4}	94.5	95.41	95.26	95.11	{2,4,6}	86.24	84.74	87.47	85.62
{2,5}	0	0	0	0	{2,4,7}	83.18	81.65	80.12	82.42
{2,6}	0	0	0	0	{2,4,8}	77.83	77.06	75.54	74.01
{2,7}	91.59	92.35	94.04	93.88	{2,5,6}	0	0	0	0
{2,8}	80.12	81.65	83.94	82.42	{2,5,7}	91.59	92.35	93.88	90.06
{3,4}	90.06	89.3	84.71	88.53	{2,5,8}	90.83	89.3	90.06	88.53
{3,5}	85.47	84.71	86.24	83.18	{2,6,7}	94.04	93.88	95.11	91.59
{3,6}	84.71	86.24	83.94	85.47	{2,6,8}	91.59	90.83	91.13	92.05
{3,7}	86.24	87	87.77	85.47	{2,7,8}	77.83	77.06	76.30	77.22
{3,8}	82.42	87.77	83.94	85.47	{3,4,5}	75.54	74.01	72.48	73.24
{4,5}	61.77	69.42	64.07	66.36	{3,4,6}	77.06	75.54	76.30	76.76
{4,6}	62.54	69.36	69.42	67.13	{3,4,7}	89.60	89.3	89.14	88.99
{4,7}	65.6	67.89	68.65	67.13	{3,4,8}	83.18	81.65	83.94	83.18
{4,8}	70.18	68.65	66.36	69.42	{3,5,6}	64.83	66.36	65.60	63.30
{5,6}	0	0	0	0	{3,5,7}	69.42	69.89	70.18	70.03
{5,7}	83.18	84.71	88.53	85.47	{3,5,8}	77.22	77.06	76.30	75.23
{5,8}	83.94	81.65	82.42	83.18	{3,6,7}	74.77	74.01	74.31	75.23
{6,7}	88.53	86.24	85.47	84.71	{3,6,8}	71.71	72.48	74.01	73.70
{6,8}	71.71	77.06	75.54	74.77	{3,7,8}	72.48	70.95	72.17	71.25
{7,8}	77.83	75.54	77.06	76.3	{4,5,6}	70.95	69.42	68.81	69.11
{1,2,3}	96.79	95.41	96.64	95.87	{4,5,7}	85.02	84.71	83.94	85.47
{1,2,4}	96.18	94.95	95.87	95.41	{4,5,8}	78.59	77.06	77.83	77.06
{1,2,5}	**100**	**99.85**	**99.85**	**99.69**	{4,6,7}	72.48	70.95	71.71	72.17
{1,2,6}	**99.24**	**99.54**	**99.08**	**98.01**	{4,6,8}	74.16	72.48	72.17	72.78
{1,2,7}	96.18	95.11	95.72	93.88	{4,7,8}	69.11	69.42	68.65	70.18
{1,2,8}	95.41	95.26	96.18	95.72	{5,6,7}	85.32	84.71	85.47	85.02
{1,3,4}	96.02	94.95	96.18	95.87	{5,6,8}	76.30	77.06	78.29	75.54
{1,3,5}	95.57	96.18	96.02	96.18	{5,7,8}	72.17	72.48	73.24	70.95
{1,3,6}	94.34	93.88	92.35	90.83	{6,7,8}	69.72	69.42	71.71	68.65
{1,3,7}	94.19	92.35	91.59	90.83	{1,2,3,4}	94.50	93.88	94.19	93.12
{1,3,8}	80.12	84.71	83.94	82.42	{1,2,3,5}	93.12	92.35	92.66	91.59

Table 11: Recognition rate in various window size (%)

Features Subset	Recognition Rate 3x3	5x5	7x7	9x9	Features Subset	Recognition Rate 3x3	5x5	7x7	9x9
{1,2,3,6}	93.58	94.04	93.12	93.58	{2,4,6,8}	71.71	67.89	69.42	70.95
{1,2,3,7}	95.41	94.5	93.88	93.58	{2,4,7,8}	61.77	65.60	63.30	61.01
{1,2,3,8}	95.11	94.65	94.19	95.41	{2,5,6,7}	64.83	67.43	69.42	68.65
{1,2,4,5}	95.11	95.72	95.41	94.95	{2,5,6,8}	66.36	67.13	67.89	70.18
{1,2,4,6}	91.59	92.35	92.05	91.13	{2,5,7,8}	67.13	64.83	68.65	63.30
{1,2,4,7}	92.35	91.59	92.51	91.13	{2,6,7,8}	66.36	65.29	64.83	64.07
{1,2,4,8}	98.01	98.32	98.47	97.71	{3,4,5,6}	69.42	66.36	67.89	64.83
{1,2,5,6}	99.24	99.39	99.54	98.78	{3,4,5,7}	73.24	72.48	70.95	70.18
{1,2,5,7}	98.17	99.08	98.47	97.71	{3,4,5,8}	70.95	69.42	69.11	67.89
{1,2,5,8}	97.71	97.55	97.25	98.01	{3,4,6,7}	64.83	66.36	67.89	68.04
{1,2,6,7}	97.71	98.47	98.01	98.32	{3,4,6,8}	69.42	67.13	64.83	63.30
{1,2,6,8}	98.78	98.17	97.71	97.40	{3,4,7,8}	70.18	70.95	69.42	67.89
{1,2,7,8}	99.24	98.78	98.47	97.55	{3,5,6,7}	67.89	68.65	66.36	67.13
{1,3,4,5}	96.94	96.18	95.41	94.65	{3,5,6,8}	74.01	72.48	69.42	73.24
{1,3,4,6}	96.18	95.72	96.64	95.87	{3,5,7,8}	69.42	74.01	66.36	64.83
{1,3,4,7}	96.18	95.41	95.72	94.65	{3,6,7,8}	72.48	69.42	70.95	67.89
{1,3,4,8}	96.64	96.02	95.41	94.19	{4,5,6,7}	68.20	66.36	67.89	65.60
{1,3,5,6}	92.35	93.88	93.12	90.83	{4,5,6,8}	70.18	67.89	69.42	66.36
{1,3,5,7}	94.19	94.65	95.11	94.50	{4,5,7,8}	68.65	69.11	67.89	68.20
{1,3,5,8}	92.35	92.66	93.12	92.51	{4,6,7,8}	66.36	64.83	61.77	63.30
{1,3,6,7}	93.88	93.12	92.66	92.35	{5,6,7,8}	70.95	69.42	67.89	67.13
{1,3,6,8}	91.59	92.35	90.06	92.05	{1,2,3,4,5}	91.59	90.83	90.06	88.99
{1,3,7,8}	92.35	91.59	91.13	90.83	{1,2,3,4,6}	90.06	89.30	91.59	89.76
{1,4,5,6}	95.11	94.65	95.41	94.19	{1,2,3,4,7}	89.45	90.06	89.30	90.67
{1,4,5,7}	91.59	92.35	92.05	93.12	{1,2,3,4,8}	93.43	91.59	92.05	90.98
{1,4,5,8}	93.58	93.12	92.66	93.43	{1,2,3,5,6}	90.83	87.77	89.30	88.53
{1,4,6,7}	91.59	90.83	91.13	90.06	{1,2,3,5,7}	89.30	87.00	88.53	89.45
{1,4,6,8}	91.90	92.66	90.83	91.28	{1,2,3,5,8}	90.83	91.59	93.12	92.51
{1,4,7,8}	92.05	90.83	90.06	89.76	{1,2,3,6,7}	88.53	87.92	89.30	89.76
{1,5,6,7}	93.12	92.35	92.05	91.28	{1,2,3,6,8}	87.77	86.24	87.00	87.31
{1,5,6,8}	90.83	91.28	91.59	90.67	{1,2,3,7,8}	89.76	88.84	90.98	90.06
{1,5,7,8}	88.53	89.30	88.99	88.07	{1,2,4,5,6}	94.19	93.12	92.51	92.05
{1,6,7,8}	86.54	86.24	85.93	85.47	{1,2,4,5,7}	92.05	91.59	90.98	92.51
{2,3,4,5}	68.65	69.42	67.89	66.36	{1,2,4,5,8}	88.07	87.77	88.99	88.53
{2,3,4,6}	74.01	72.48	70.95	69.42	{1,2,4,6,7}	87.77	86.24	85.78	86.70
{2,3,4,7}	77.83	77.06	76.30	74.77	{1,2,4,6,8}	88.07	87.16	87.77	86.39
{2,3,4,8}	74.77	74.01	72.48	73.39	{1,2,4,7,8}	87.00	86.39	88.38	87.92
{2,3,5,6}	76.76	75.54	74.31	73.24	{1,2,5,6,7}	91.74	91.59	90.83	91.13
{2,3,5,7}	64.83	68.65	63.30	61.77	{1,2,5,6,8}	94.19	93.88	93.12	93.58
{2,3,5,8}	63.30	68.20	68.81	64.83	{1,2,5,7,8}	90.83	90.06	88.38	91.28
{2,3,6,7}	69.42	72.94	72.48	68.81	{1,2,6,7,8}	86.24	84.71	86.70	85.63
{2,3,6,8}	74.01	72.17	71.10	70.64	{1,3,4,5,6}	90.52	90.06	88.53	88.99
{2,3,7,8}	72.48	71.71	70.95	69.42	{1,3,4,5,7}	92.97	91.44	89.76	90.52
{2,4,5,6}	69.42	71.10	72.48	71.71	{1,3,4,5,8}	88.53	86.70	87.77	87.31
{2,4,5,7}	70.95	69.27	71.71	73.24	{1,3,4,6,7}	90.83	89.76	89.30	88.53
{2,4,5,8}	67.89	66.06	68.65	69.42	{1,3,4,6,8}	87.77	86.24	87.00	87.16
{2,4,6,7}	70.95	68.65	67.89	67.13	{1,3,4,7,8}	87.00	85.93	85.47	85.17

Table 12: Recognition rate in various window size (%)

Features Subset	**Recognition Rate** 3x3	5x5	7x7	9x9	**Features Subset**	**Recognition Rate** 3x3	5x5	7x7	9x9
{1,3,5,6,8}	88.53	87.46	88.38	87.46	{1,2,3,4,6,8}	73.24	71.41	72.94	72.17
{1,3,5,7,8}	90.06	88.84	87.92	87.46	{1,2,3,4,7,8}	76.15	75.54	74.46	73.24
{1,3,6,7,8}	92.51	91.13	90.06	90.52	{1,2,3,5,6,7}	80.12	77.06	75.54	73.70
{1,4,5,6,7}	86.39	84.71	85.78	85.47	{1,2,3,5,6,8}	74.77	74.01	73.39	72.32
{1,4,5,6,8}	89.76	88.23	88.53	88.99	{1,2,3,5,7,8}	75.54	76.15	74.01	74.77
{1,4,5,7,8}	87.00	86.24	84.71	85.47	{1,2,3,6,7,8}	81.65	78.29	80.12	79.20
{1,4,6,7,8}	90.98	89.60	90.52	90.06	{1,2,4,5,6,7}	72.78	71.41	70.18	73.09
{1,5,6,7,8}	85.78	85.47	86.24	84.86	{1,2,4,5,6,8}	68.65	70.95	70.03	71.41
{2,3,4,5,6}	74.01	77.06	72.48	74.01	{1,2,4,5,7,8}	74.77	73.55	74.16	75.69
{2,3,4,5,7}	75.23	74.01	73.24	72.48	{1,2,4,6,7,8}	76.15	74.92	74.01	73.85
{2,3,4,5,8}	77.83	77.06	75.23	73.70	{1,2,5,6,7,8}	80.28	78.75	78.29	79.36
{2,3,4,6,7}	74.01	75.54	72.48	76.15	{1,3,4,5,6,7}	77.83	77.06	76.30	76.61
{2,3,4,6,8}	72.48	70.95	71.41	70.18	{1,3,4,5,6,8}	69.42	70.95	68.65	67.89
{2,3,4,7,8}	67.89	69.42	70.18	70.80	{1,3,4,5,7,8}	68.35	69.42	68.04	67.13
{2,3,5,6,7}	74.01	71.71	69.42	68.81	{1,3,4,6,7,8}	66.36	68.96	67.89	68.65
{2,3,5,6,8}	71.41	72.17	72.48	69.72	{1,3,5,6,7,8}	67.89	70.03	70.34	71.41
{2,3,5,7,8}	75.54	73.24	71.56	72.94	{1,4,5,6,7,8}	72.78	72.02	74.46	70.64
{2,3,6,7,8}	66.36	69.27	68.04	70.34	{2,3,4,5,6,7}	64.83	66.36	67.89	68.65
{2,4,5,6,7}	68.65	66.36	70.95	70.64	{2,3,4,5,6,8}	70.03	67.89	68.65	67.13
{2,4,5,6,8}	64.83	67.89	64.07	67.13	{2,3,4,5,7,8}	77.83	76.15	74.77	77.22
{2,4,5,7,8}	72.48	70.18	71.25	69.57	{2,3,4,6,7,8}	73.55	74.46	73.09	72.17
{2,4,6,7,8}	70.64	69.42	67.89	66.06	{2,3,5,6,7,8}	72.78	72.02	71.10	70.64
{2,5,6,7,8}	74.77	71.71	71.41	70.18	{2,4,5,6,7,8}	67.89	69.42	66.06	67.58
{3,4,5,6,7}	70.95	69.42	71.41	70.18	{3,4,5,6,7,8}	71.71	74.16	73.39	71.41
{3,4,5,6,8}	64.83	67.89	66.36	68.65	{1.2.3.4.5.6.7}	90.83	90.06	90.52	88.84
{3,4,5,7,8}	77.06	74.16	75.69	72.63	{1,2,3,4,5,6,8}	87.46	86.24	86.70	85.93
{3,4,6,7,8}	74.77	72.94	74.01	72.78	{1,2.3.4.5.7.8}	93.12	92.35	91.44	90.98
{3,5,6,7,8}	71.41	69.72	67.89	66.36	{1,2,3,4,6,7,8}	92.66	91.59	90.67	90.52
{4,5,6,7,8}	64.83	67.89	66.36	68.35	{1,2,3,5,6,7,8}	87.46	86.70	85.78	85.47
{1,2,3,4,5,6}	80.12	77.06	75.54	74.77	{1,2,4,5,6,7,8}	88.99	88.38	87.92	87.00
{1,2,3,4,5,7}	74.77	74.01	72.48	70.95	{1,3,4,5,6,7,8}	90.21	88.99	87.46	86.09
{1,2,3,4,5,8}	76.30	74.77	73.09	72.48	{2,3,4,5,6,7,8}	69.42	70.95	66.36	69.88
					{1,2,3,4,5,6,7,8}	90.06	88.53	86.70	89.45