Face Recognition Using the Moving Window Classifier

M. S. Hoque and M. C. Fairhurst
Electronic Engineering Laboratory,
University of Kent, Canterbury,
Kent CT2 7NT, United Kingdom.
{msh4,mcf}@ukc.ac.uk

Abstract

The Moving Window Classifier(MWC) has previously been proposed as an efficient scheme for text recognition applications. In this paper, the potential of the MWC algorithm in face recognition is investigated. To maintain the memory requirements of the classifier within acceptable practical limits, the concept of bit-plane encoding is utilized. The experimental results reported show very encouraging performance for both the schemes.

1   Introduction

Automatic recognition of facial images is known to be a challenging problem. Potential applications in this field include video surveillance, criminal identification, bank card and credit card user identification, and many others, and many algorithms have been proposed. For example, Kanade [5] presented a scheme using features based on ratios of geometric distances. Similar feature sets were also used by Brunelli *et. al.* [2]. Turk and Pentland [13] presented a scheme based on eigenfaces, while Samaria [11] presented an HMM based approach which was later extended using a pseudo-2D HMM. Lawrence *et. al.* [7] proposed a scheme combining Self-Organizing Maps with a convolutional neural network. Lin's [8] method uses a probabilistic decision based neural network, and Lucas [9] introduced a simple n-tuple based method designated the continuous n-tuple classifier. This last technique is similar in principle to the technique reported here, and shares the advantages of simplicity in concept and implementation.

The Moving Window Classifier (MWC) has been proposed [3] as an efficient scheme in text recognition. The algorithm is fundamentally based on the n-tuple scheme and works on the raw bitmap image. The n-tuple scheme has been successfully applied to many diverse pattern recognition applications [10], and here the MWC structure is investigated as a means for efficient recognition of simple face images. However, all n-tuple based systems are susceptible to huge memory space requirements, especially when processing gray scale patterns, and this problem is explicitly addressed here by invoking the principle of bit-plane decomposition.

A complete face recognition system generally involves image capture, localization and segmentation of the face from the image, detection of eye, nose, lips etc., feature extraction, face identification, post-identification processing, and so on. Although frontal face images are mainly considered, the face profile is also often used as complementary

information. Additional information such as gender, race, etc. can also be used effectively in the face recognition process. The scope of this paper is limited to identification using only frontal face images, in the absence of any additional information. The face images are pre-segmented and cropped with limited variations in tilt, rotation, and scale.

2 The Moving Window Classification Scheme

The Moving Window Classifier(MWC) is a modified version of the established n-tuple classifier. In an n-tuple classifier, the n-tuples are formed by selecting multiple sets of n distinct locations from a pattern space. Each n-tuple thus sees an n-bit feature derived from the pattern. For classification, a pattern is assigned to that class for which the number of matching features found in the training set is maximum. The training process, therefore, requires counting the number of times different features are seen by individual n-tuples. To ensure equal likelihood of all classes, the counts are normalized so that the maximum possible score remains the same for all classes.

In the MWC scheme, a window smaller than the image is defined, and only a portion of the image is visible through this window. The n-tuples are connected to this window and assign scores corresponding to the likelihood of the pattern viewed belonging to the individual classes. The window is then shifted and classification is carried out for the new part image visible through the window. Thus, the window is moved left to right and top to bottom in single pixel displacement steps until the entire image is covered and part classification is carried out for all different window positions. A decision fusion stage then combines these partial classification scores and, accordingly, assigns a class label to the test image as a whole. Further details can be found in [3].

3 Bit-plane Decomposition

The size of the memory space required by a typical n-tuple based scheme is σ^n units per tuple per class, where σ is the number of distinct gray levels and n is the size of the tuples. It can readily be seen that this can become excessively large even with a fairly small number of gray levels. There are several measures possible to moderate this space requirement. One is the use of a sparse array [4], but this decreases the processing speed significantly. An alternative is to reduce the number of gray levels. Although $\sigma = 256$ is very common, reducing σ by using a suitable thresholding algorithm may not degrade the image significantly and these reduced gray-scale images can still be used in recognition operations. However, too great a reduction in σ can cause significant loss of information and this often adversely affects the performance of the classifier.

Bit-plane decomposition, although introduced by Schwarz [12] as a means of data compression, can also be used to handle the memory space problem faced by the n-tuple based systems. The basic idea is to decompose an image into a collection of binary images ($\sigma = 2$) without losing any information contained in the original image.

For bit-plane decomposition, the gray levels of the gray-scaled image are represented in binary. Therefore, for 'σ' possible distinct gray levels, each pixel of the image is represented by a k $(= \lceil \log_2 \sigma \rceil)$ bit binary code. The image is decomposed into k layers where layer 'i' is composed of the ith bits of the gray level values. Thus, layer '0' is formed by collecting all the *least significant bits* (LSB) of the binary coded gray-scale image. Figure

Figure 1: A 256 level gray-scale image

layer 7 (MSB) layer 6 layer 5 layer 4
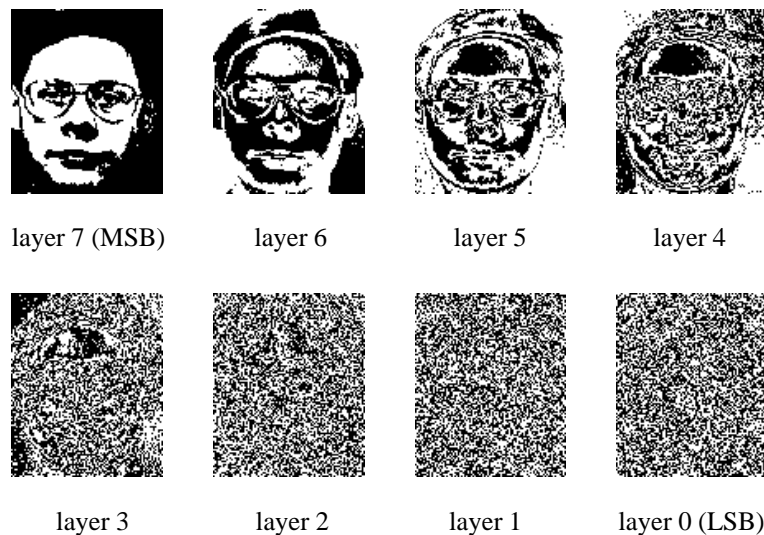
layer 3 layer 2 layer 1 layer 0 (LSB)

Figure 2: Decomposition of the image in Figure 1 (binary coding)

2 illustrates the 8 layers of the gray-scaled image shown in Figure 1, and the visually significant layers are readily apparent. The lower order layers only contribute to the fine detail of the original image and do not carry any significant information, especially when considered from the viewpoint of the recognition process.

One major limitation of binary coding of the gray levels is that a small change in the gray levels may affect the decomposed layers considerably. An alternative approach is to use Gray-coding to represent the gray-levels of the pixels instead of binary coding. Figure 3 illustrates the corresponding decomposition of the image in Figure 1 using the Gray-code.

4   Experimental Results

The 'ORL Database of Faces' [1] was used in all the experiments. The database contains ten different images of each of 40 distinct subjects (of both genders) taken over a period of 2 years. In all the images, the subjects are in an upright, frontal position with some variation in lighting, facial expression, facial details etc. The resolution of each image is 112×92 pixels, with 256 gray levels per pixel. The database is randomly divided into

| layer 7 (MSB) | layer 6 | layer 5 | layer 4 |

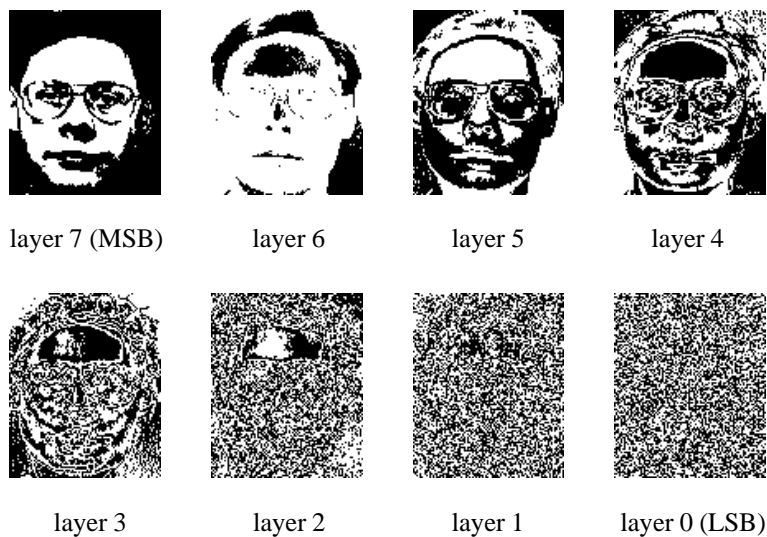| layer 3 | layer 2 | layer 1 | layer 0 (LSB) |

Figure 3: Decomposition of the image in Figure 1 (Gray coding)

two disjoint sets for each subject. The first part is used for training and the remainder for testing. This database was chosen because of its general accessibility and widely reported adoption,but it should be noted that no preprocessing operations were carried out on the images for the experiments reported here.

The two criteria used to evaluate the classifier performance here are the error rates and classification speed. The error rates are expressed as percentages and are obtained by dividing the number of misclassified images by the total number of test images. Speed measures are the average time needed for classification per image and are obtained by measuring the total classification time for all test images after excluding system overheads such as system initialization, file access, and so on. All the results reported here are the arithmetic mean of at least 20 test runs under different randomized parameters as appropriate (for example, different train/test divisions, different n-tuple mapping, etc.). The following sections describe the experiments carried out and the corresponding results obtained.

4.1   Recognition of Raw Gray-scaled Images

The memory space needed to implement an n-tuple based classifier is a function of the number of gray levels (see Section 4.5) and this imposes a practical limit on the maximum number of gray levels usable for a given n-tuple size. Table 1 shows recognition error rates achievable using 2, 4 and 8 gray levels per pixel. Both the n-tuple and the MWC classifier are configured with 500 3-tuples or 4-tuples. In the case of the MWC, the size of the moving window is 106×92 pixels. It can be readily seen that as the number of gray level increases, the accuracy of the classification system also improves. It is also observed that the MWC generates optimal performance with relatively smaller n-tuple sizes. This is very significant because n is a principal factor in determining the requirement for memory space. If only the optimal performances are considered, then for 8 gray levels per pixel,

Gray levels per pixel,	Error Rate (in %)			
	n-tuple		MWC	
σ	n=3	n=4	n=3	n=4
8	5.10	5.08	3.07	3.29
4	5.81	5.60	4.07	4.21
2	12.33	12.14	11.26	11.29

Table 1: Error rates in recognizing raw gray-scale images

Classification Scheme	Trainset (in samples per class)								
	1	2	3	4	5	6	7	8	9
MWC	26.4	16.0	8.5	4.9	3.1	3.1	1.9	1.1	0.4
Continuous n-tuple	26.5	16.9	8.2	4.8	3.4	3.2	2.4	1.8	0.6
n-tuple	27.1	17.6	10.1	6.8	5.1	4.2	3.5	3.0	2.0
Eigenface	38.6	20.9	18.2	15.4	10.5	n/a	n/a	n/a	n/a
SOM+CN	30.0	17.0	11.8	7.1	3.8	n/a	n/a	n/a	n/a

Table 2: Error rates for different trainset sizes

the error rate is 5.08% for the straight n-tuple scheme, whereas it falls to 3.07% when the MWC scheme is used. Experiments with more gray levels were not carried out because of the need for an impractically large memory space.

4.2   Effect of the Training Set Size on Performance

The number of images per class used in the training phase plays a significant role in the classifiers' performance. Too small a number leads to undertraining, making the classifier unable to generalize the diversity among patterns in a given class. On the other hand, too many training samples may lead to over-generalization and thus diminish the ability to distinguish between different classes. The ideal situation strikes a balance between supplying maximal information and avoiding the loss of the distinctive differences inherent among separate classes.

In applications such as face recognition, in common with many other similar problem domains where only relatively few samples may often be available for training purposes, this study is very important. In the ORL dataset, there are 10 images per class. Therefore, 1 to 9 samples were used in training and the remaining images were used for testing. Table 2 shows the reduction in the classification error rates as the size of the training set increases. For comparative purposes, similar measures from a range of other classification schemes are also included. The data for the Eigenface and the SOM+CN methods are taken from [7] and the rest are generated locally. For the n-tuple and the MWC schemes, 500 3-tuples were used with 8-level images. For the continuous n-tuple scheme [9], 500 3-tuples were used with 256-level images.

It can be readily seen that there is a potential danger of under-training, and for optimal training, 8 to 9 images per class or even more should be used. It is not possible here completely to optimize training because of the limited number of images per class available in the database. In all the experiments reported in this paper, 5 images per class were used

Layer	Error Rates (in %)			
	n-tuple		MWC	
	Binary Encoding	Gray coded	Binary Encoding	Gray coded
layer-7 (MSB)	12.3	12.3	11.2	11.2
layer-6	7.9	9.6	5.9	6.5
layer-5	15.7	8.0	14.1	6.8
layer-4	43.2	13.0	26.0	9.3
layer-3	84.3	40.9	57.8	25.8
layer-2	96.1	80.5	91.5	53.5
layer-1	97.2	94.8	97.5	90.5
layer-0 (LSB)	97.7	97.0	95.8	96.8

Table 3: Recognition performance with the individual decomposed layers

in the training because all the schemes presented here for comparison with the proposed method used this degree of training.

4.3 Performance with Individual Decomposed Layers

Different layers of the decomposed image can be used independently for pattern recognition. Table 3 shows the performance achievable with the n-tuple and the MWC classifier by using individual layers. Both the schemes used 500 3-tuples and the MWC used a $106{\times}92$ window. The two different encodings used show different behavioural patterns. With the n-tuple scheme and binary encoding, the top 4 layers performed well with layer 6 giving the optimum performance. For Gray encoding, the top 5 layers are useful and layer 5 offers the best performance. The MWC scheme showed similar behaviour except that best performance was obtained from layer 6 for both encoding schemes. Therefore, if a single binarized image is desirable, then layer 6 under the binary encoding gives the optimum performance. The least significant bit layers, as indicated in Section 3, are simply noise from the recognition point of view and can be discarded from the classification process.

It is possible to use different window sizes and n-tuple sizes in the MWC implementation. Using a $110{\times}84$ window instead of the $106{\times}92$ window, layer 7 gives an error rate of 9.1% with 7-tuples. Since the introduction of bit-plane decomposition drastically reduces the memory requirement, larger n-tuple sizes can be used without adversely affecting resource limitations. In a similar way, different operating parameters (e.g., different window size, different n, etc.) can be used for different layers to achieve the optimum performance. The lowest error rates with the top 4 layers achieved in this study with MWC (for $n = 7$) are 9.1% with layer 7 using $110{\times}84$ window, 4.1% with layer 6 using $108{\times}88$ window, 9.3% with layer 5 using $110{\times}84$ window, and 21.0% with layer 4 using $104{\times}90$ window when binary encoding were used. Similarly, in the case of Gray coding, these values are 9.1% (with $110{\times}84$ window), 5.3% (with $108{\times}84$ window), 6.1% (with $108{\times}84$ window) and 8.3% (with $108{\times}84$ window) respectively.

Layers Combined	Error Rate (in %)			
	n-tuple		MWC	
	Binary encoded	Gray coded	Binary encoded	Gray coded
layer-7, layer-6	6.6	5.0	6.2	5.2
layer-7, layer-5	6.9	7.9	7.2	6.7
layer-7, layer-4	7.9	6.9	7.8	6.6
layer-6, layer-5	5.7	6.7	3.8	3.7
layer-6, layer-4	6.4	7.5	**3.3**	3.9
layer-5, layer-4	14.2	7.0	8.1	4.5
layer-7, layer-6, layer-5	5.7	4.5	5.7	4.5
layer-7, layer-6, layer-4	5.8	**3.8**	5.5	4.3
layer-7, layer-5, layer-4	6.1	6.1	6.5	5.4
layer-6, layer-5, layer-4	**4.5**	5.1	3.7	**3.5**
layer-7, layer-6, layer-5, layer-4	4.9	3.9	4.8	4.3

Table 4: Error rates in recognition using multiple layers

4.4 Using Mutiple Layers

To improve the overall performance, it is possible to involve multiples of the decomposed layers in the classification process. Individual layers are classified separately and their scores combined to make the final classification decision. Many different decision fusion strategies (e.g., Bayesian combination, majority voting, weighted majority voting, etc.) can be incorporated. In this implementation, the 'sum-rule' [6] was used for the decision fusion because of its high performance and simplicity. The n-tuple and the MWC classification scheme assign to a test image scores corresponding to its likelihood of belonging to a particular class. When multiple layers are used for classification, these scores are normalized and added and the test image is assigned to the class generating the highest sum.

Up to four of the most significant bit layers were used. Table 4 shows the recognition errors for all possible combinations of the four layers. The most striking improvements were experienced with the n-tuple scheme. It is found that error rates of 4.5% and 3.8% respectively for binary and Gray coding can be achieved using three layers and standard n-tuple scheme. This is a significant improvement from error rates achievable using a single decomposed layer and even from that achievable using 8-level gray scale images directly. The 3.8% error rate is comparable to, or better than, that achieved by many complex and more sophisticated face recognition schemes. With MWC, the performance is somewhat degraded compared with what was achieved using 8-level images. However, the figures of 3.3% and 3.5% are superior to all other schemes when tested on the ORL database, while it is noted that the implementation requires less than 2% of the memory space required for direct processing of 8-level images. It possible to further fine tune the decision fusion process by using a weighted combination scheme assigning more weight to layers having better discrimination capacity.

Image Representation	Space requirement (in units per tuple per class)
256 gray level	$\approx 4.3 \times 10^9$
16 gray level	65536
8 gray level	4096
4 gray level	256
bit-plane decomposed	$k \times 16$

k=number of layers used in classification; $k \leq 8$ for 256 gray levels

Table 5: Memory space requirement ($n = 4$)

4.5 Memory Space Requirements

For n-tuple implementation, the memory space requirement is a controlling factor determining the type of images that can be practically handled. The memory space required is a function of σ^n, where σ is the number of gray levels per pixel and n is the tuple size. Table 5 shows the memory space requirements for images of different gray-scale resolution. It can be seen that it is impractical to use the 256 level images directly. Even a restriction to 16 gray levels would not be usable when the number of n-tuples and number of classes are high. (All experiments reported here were carried out with 40 classes and 500 n-tuples). 8-level images are a poor compromise compared to the original, although this still requires significantly large memory space. Against this, when bit-plane decomposition is introduced, the memory requirement is substantially reduced ($\leq 2^n \lceil \log_2 \sigma \rceil$ instead of σ^n), even when multiple layers are used in the classification process.

4.6 Study of Comparative Classification Performance

This section describes a study of the performances achieved by 8 different classification schemes as tested on the ORL face database and compares these with the schemes reported here. The probabilistic decision-based neural network(PDBNN) is reported by Lin *et. al.*[8]. The SOM+CN scheme combines local image sampling, a self-organizing map neural network, and a convolutional neural network and was proposed by Lawrence *et. al.*[7]. The HMM based schemes are reported by Samaria [11], while the eigenface algorithm is from Turk *et. al.*[13]. The continuous n-tuple is another variation of the n-tuple scheme and was proposed by Lucas [9]. The results with a nearest neighbour classifier are also from [9]. It is evident that bit-plane decomposition not only solves the memory resource constraint, but also enables the simple n-tuple scheme to outperform other more complex schemes. Although bit-plane decomposition did not improve the performance for the MWC scheme (in fact, it marginally degrades it) this nevertheless outperforms all other schemes in recognition accuracy. The achieved savings in memory also makes this a very attractive option among the algorithms compared.

5 Conclusion

The Moving Window Classifier is presented here as a tool for face recognition. This scheme offers all the simplicity of the established n-tuple scheme but at the same time

Classification Algorithm	Recognition Error rates (in %)
PDBNN	4.0
SOM+CN	3.8
Top-down HMM	13.0
Pseudo-2D HMM	5.0
Eigenface	10.0
n-tuple (binary image)	11.6
Continuous n-tuple	3.8
1-NN	4.1
n-tuple (8 level gray)	5.1
n-tuple (with bit-plain decomposition)	3.8
MWC (8 level gray)	3.1
MWC (with bit-plain decomposition)	3.3

Table 6: Error rates with different algorithms

generates very accurate classification decisions. The memory space problem usually associated with this approach to image classification is solved using the bit-plane decomposition method. This decomposition scheme proves so efficient that even the direct n-tuple scheme outperformed other classification methods. It should also be noted that no preprocessing was carried out on the ORL face images for the experiments in this paper. Preprocessing operations such as contrast stretching, size and position normalization, and other image enhancement techniques can reduce the intra-class variability and as such, improved recognition accuracy can be expected but at a slower speed.

MWC has also been effectively used in handwriting recognition and has demonstrated several positive aspects of using this scheme. Handwriting images are usually binary in nature and therefore bit-plane decomposition is inappropriate in this case. However, this decomposition technique is a generic process and can be applied to many other task domains.

References

[1] AT&T Laboratories, Cambridge University, UK. The database can be downloaded from `ftp://ftp.uk.research.att.com/pub/data/` via anonymous ftp.

[2] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15:1042–1052, 1993.

[3] M. C. Fairhurst and M. S. Hoque. Moving window classifier: approach to off-line image recognition. *Electronics Letters*, 36(7):628–630, 2000.

[4] T. M. Jorgensen. Classification of handwritten digits using a ram neural net architecture. *Int. Journal of Neural Systems*, 8(1):17–25, 1997.

[5] T. Kanade. *Picture Processing by computer and recognition of human faces*. PhD thesis, Kyoto University, Japan, 1973.

[6] J. Kittler and M. Hatel. Improving recognition rates by classifier combination. In *Proceedings of 5th Int. Workshop on Frontiers in Handwriting Recognition*, pages 81–102, University of Essex, Colchester, UK., 1996.

[7] S. Lawrence, C. Lee Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113, 1997.

[8] S. Lin, S. Kung, and L. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Transaction on Neural Networks*, 8(1):114–132, 1997.

[9] S. Lucas. Face recognition with the continuous n-tuple classifier. In *Proceedings of the British Machine Vision Conference*, 1997.

[10] M. Morciniec and R. Rohwer. The n-tuple classifier: Too good to ignore. Technical Report NCRG/95/013, Dept. of Computer Science, Aston University, Birmingham, UK, 1995.

[11] F. S. Samaria. *Face Recognition using Hidden Markov Model*. PhD thesis, Cambridge University, Cambridge, U.K., 1994.

[12] J. W. Schwarz and R. C. Barker. Bit-plane encoding: A technique for source encoding. *IEEE Transaction on Aerospace and Electronic Systems*, 2(4):385–392, 1966.

[13] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.