A New Face Recognition Algorithm using Bijective Mappings

Christine Podilchuk, Ankur Patel, Ashwath Harthattu, Saket Anand, Richard Mammone
WISE Lab, Center for Advanced Information Processing, Rutgers University
96 Frelinghuysen Road. Piscataway, NJ 08855-1390
{chrisp1, ankurrp, aithal, anands, mammone}@caip.rutgers.edu

Abstract

A new face recognition algorithm is proposed which is robust to variations in pose, expression and illumination. The framework is similar to the ubiquitous block matching algorithm used for motion estimation in video compression but has been adapted to compensate for illumination differences. One of the key differentiators of this approach is that unlike traditional face recognition algorithms, the image data representing the face or features extracted from the facial data is not used for classification. Instead, the mapping between the probe and gallery images given by the block matching algorithm is used to classify the faces for recognition. Once the mappings are found for each gallery image, the degree of bijectivity that each mapping produces is used to derive the similarity scores for recognition.

1. Introduction

Face recognition generally falls into two categories [1] [2] [3], appearance-based techniques and feature-based techniques. Feature-based methods rely on the detection and identification of certain key points on the face such as the eyes, the nose, the mouth, etc. Once the features are detected, these techniques are invariant to differences in pose, illumination, background, and other conditions. Recent techniques based on constructing 3D models of the human face [5] [6] are related to feature-based techniques in that they rely on extraction of key features in order to register the captured image data to the generated 3D model. This explicit modeling of the face allows for handling illumination, pose and expression problems in practice; however the key problems for the above techniques are accurate and automatic detection of feature points and the ease of construction of a satisfactory model.

Appearance-based approaches to face recognition rely on the entire image directly without the use of detecting and identifying fiduciary features in the face, or any intermediate models. A popular and successful class of appearance-based approaches considers an image as a high-dimension vector, i.e., a point in high-dimensional vector space, and transforms the image into a lower dimensional representation for classification. Different approaches based on this idea are PCA (Principle Component Analysis) [7], Fisher's Discriminant Analysis [8], ICA (Independent Component Analysis) [9] and SVM (Support Vector Machines) [10] among others. One weakness of such approaches is the need for a large set of training data. Such methods can also be very sensitive to variations in pose, lighting and other differences between the training and test data.

We propose a new technique for face recognition that incorporates the advantages found in both feature-based and appearance-based approaches. The method is appearance-based in that it uses all of the facial data captured and does not depend on accurate detection of key features while discarding other facial information. Like feature-based approaches, this technique is robust to lighting- and pose-variations through the use of a block-based transformation that corrects for lighting and pose differences. The new method does not rely on training data either to build a reliable model or to train the classifier. Finally, the system is fully automatic and does not require any manual feature detection either for training or recognition.

2. Proposed Algorithm

The block matching algorithm used for motion estimation in current video coding standards such as MPEG [11] is the basic framework for the registration algorithm in order to compensate for variations in pose and expression between the captured probe images and the stored gallery images in a face recognition system. The mapping between probe and gallery images can be applied in a "forward" direction where the mapping is found which converts the probe image into the gallery image as well as in a "backward" direction where the mapping is found that converts the gallery image into the probe image. The forward and backward mappings are not simply the inverse of one another. The block-matching algorithm was first introduced to perform motion estimation and compensation for video compression in order to take advantage of temporal correlations between video frames by estimating the current frame from the previous frame. In a hybrid video coder with traditional motion compensation, motion estimation is performed by matching blocks in the current frame to blocks in the previously reconstructed frame [11]. An estimate of the current block is obtained by

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

searching blocks in the previously encoded (or original image) frame in a predetermined search area. The block-matching algorithm is used for motion estimation between two video frames for compression and in our case, the block matching algorithm is used for disparity estimation between the probe image and each gallery image. Strictly speaking, there is no concept of time between the gallery and probe image as in the video compression problem nor is there a concept of disparity between two camera views as found in the stereo correspondence problem. However, the basic idea is the same in that we are trying to estimate a disparity map or correspondence between the probe image and test image. In addressing face recognition, the key differentiator for our approach is that we expect the disparity map between the correctly matched faces to have significantly different properties than the disparity maps found for mismatched faces. We use the properties of the disparity fields found in mapping the probe image to the gallery images for classification instead of the original facial images or features extracted from the faces. The block-matching algorithm assumes simple translational motion or disparity within the image plane that is constant over a small block size. A straightforward variation of the BMA is the full search algorithm (FS) or exhaustive search algorithm that finds the best match by exhaustively searching every pixel or sub-pixel location within a predetermined search range. The image to be represented is partitioned into distinct blocks, and a match is found for each block within a specified search area in the search image.

The most commonly used cost functions for the block-matching algorithm are the mean square error (MSE) or L2 norm and the mean absolute difference (MAD) or L1 norm [11]. We generally prefer the L1 norm because it is more robust to outliers and computationally less expensive. We found in practice, that the MAD also provides better recognition performance than the MSE. In order to make the BMA robust to changes in illumination and pose we modify the MAD cost function to include a multiplicative term *a(i,j)* for illumination variations, so that the modified block matching criterion becomes:

$$E(d_i,d_j)=\sum_{\substack{search\\region}}\left|Y(i,j)-a(i+d_i,j+d_j)X(i+d_i,j+d_j)\right| \qquad (1)$$

where Y is the image to be represented, X is the search image to be mapped into the image Y, and a is assumed to be a constant over a small region *(MxN)* and is solved by least squares i.e.

$$Y(i,j)=a(i+d_i,j+d_j)X(i+d_i,j+d_j) \qquad (2)$$
$$0\le i\le M-1$$
$$0\le j\le N-1$$

Note that the block size for illumination and disparity do not have to be identical. For the images that we have processed with a resolution of 256x256, we have found that an illumination block size of 8x8 and a disparity block size of 16x16 yields good results.

Let $Y'(k)$ represent a vector of length L $(M\times N)$ of the concatenated block $Y(i,j)$ and $X'(k+d_k)$ represent a vector of length L of the concatenated block $X(i+d_i,j+d_j)$. The least squares solution for a can be expressed as

$$a(i+d_i,j+d_j)=\frac{\left\langle Y'(k),X'(k+d_k)\right\rangle}{\left\langle X'(k+d_k),X'(k+d_k)\right\rangle} \qquad (3)$$

In order to compensate for illumination variations between the two images, we modify the traditional similarity metric to incorporate a multiplicative term that is constant over a small area, solved using least squares and represents illumination differences between the images.

3. Recognition Principle

A bijective function (one-to-one correspondence or bijection) is a function that is both injective ("one-to-one") and surjective ("onto"). More formally, a function $f: X\rightarrow Y$ is bijective if for every y in the codomain Y there is exactly one x in the domain X with $f(x) = y$.

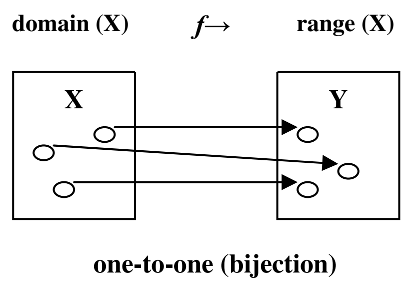

Figure 1: Bijective Function

This concept of a bijective function can be extended to the mapping (vector field) that is obtained using the block-matching algorithm to find correspondence between the probe and gallery images.

Let us consider the case in which the probe image is the searched image and the gallery image is the image to be represented. A function 'f' is found which maps blocks in the probe image to blocks in the gallery image,

$$Y=f(X) \qquad (4)$$

where X is the probe image, Y is the gallery image and f represents the mapping between probe and gallery image.

For each distinct block Y_i $(1 \leq i \leq p)$, we generate a vector representing the location of the best match found in X in a predetermined search region resulting in a disparity field of vectors for every block location. There are three scenarios for the vector field representations. A block in X uniquely represents only one block in Y that we refer to as one-to-one correspondence. A block in X matches multiple blocks in Y that we refer to as one-to-many correspondence and a block in X does not represent any block in Y that we refer to as one-to-none correspondence. The main idea behind our approach is that we expect the mapping between two images containing a matched face to have a higher percentage of one-to-one correspondence than the mapping between mismatched faces. Because the block-matching algorithm is based on non-overlapping blocks only in the image to be represented *(Y)*, we measure one-to-one correspondence on a pixel level or the percentage of the entire image X (probe image) that has been matched to Y (gallery image) as a one-to-one correspondence. This is illustrated in figure 2 where the shaded region represents the one-to-one correspondence and the dark region represents the one-to-many correspondence.

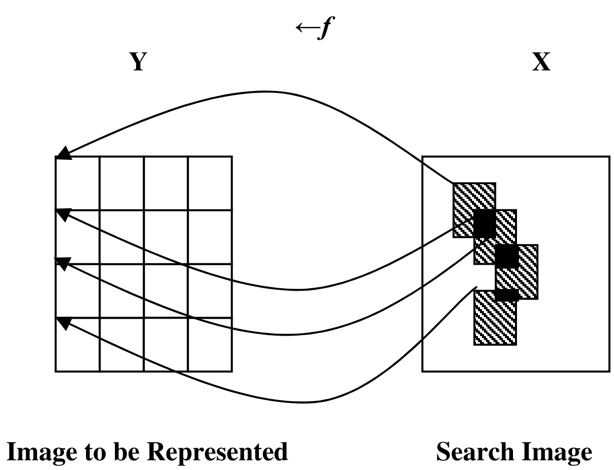

Figure 2: Mapping between the search image and the image to be represented

The measure of bijectivity is given by

$$M_f = \frac{\eta[\{f(X_1) \cup ... \cup f(X_p)\} - \{f(X_1) \cap ... \cap f(X_p)\}]}{\eta[X]} \qquad (5)$$

where p is the total number of distinct blocks in Y, the numerator in equation (5) is the total number of pixels having one-to-one correspondence in the search image and the denominator in equation (5) is the total number of pixels in the search image. Figure 3 illustrates the difference in the mappings between the gallery and probe of the same subject in Figures 3a and 3b, and the gallery and probe of the different subjects as illustrated in Figures 3a and 3c. The vector fields mapping the probe to the gallery are shown in Figures 3d and 3e. A gray-level representation of the vector field characteristics of Figures 3d and 3e are illustrated in Figure 3f and 3g where the white areas correspond to one-to-one mappings, the black areas correspond to one-to-none mappings and the gray areas correspond to one-to many mappings. The degree of bijectivity for the probe and gallery of the same subject is 52.91% as illustrated in Figure 3f and that for the probe and gallery of different subject is 28.05% as illustrated in Figure 3g. Note that the disparity vector field is uniform and smooth for the matched faces while it is random for the mismatched faces. The percentage of bijective mappings is much greater for the gallery and probe pair of the same subject then for the mismatched subjects. Likewise, the reconstructed face using the corresponding vector fields results in a better reconstruction for the matched faces.

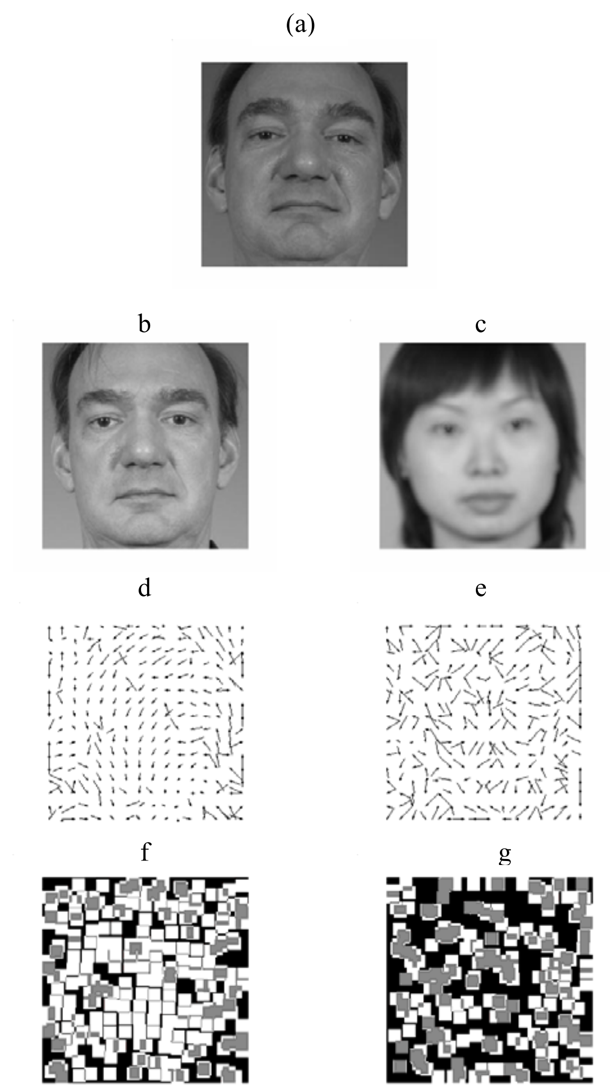

Figure 3: Bijective mapping between correctly matched and incorrectly matched pair of gallery and probe images

4. Similarity Score

The measure of bijectivity is performed in a forward and reverse direction where the mapping in the forward direction maps the probe image into the gallery image and the mapping in the reverse direction maps the gallery image into the probe image. These mappings are not inverses of each other and produce two measures of bijectivity that when combined, produce better similarity scores and improved recognition performance.

The similarity score is given by a linear combination of the forward and backward measure of bijectivity.

$$S_{BMA} = \lambda_1 M_f + \lambda_2 M_b \qquad (6)$$

where M_f is the forward bijectivity (mapping probe image to gallery image), M_b is the backward bijectivity (mapping gallery image to probe image) and λ_1, λ_2 are scalar weights between 0 and 1. We found the optimum weights to be approximately 0.5 indicating that the errors between the forward and backward mappings are statistically independent.

It is observed that the scores obtained by the BEE baseline algorithm [12] and the proposed algorithm, are uncorrelated. In order to take the advantage of both algorithms, the scores were normalized (7) using the min-max normalization and linearly fused (8).

$$s_k^{'} = \frac{s_k - \min}{\max - \min} \qquad (7)$$

$$S_{final} = \beta_1 S_{BMA} + \beta_2 S_{PCA} \qquad (8)$$

where S_{BMA} is the score obtained by the proposed algorithm, S_{PCA} is the BEE baseline score [12] and β_1, β_2 are scalar weights between 0 and 1. We found the optimum weights of β_1 and β_2 to be approximately 0.85 and 0.15 respectively.

5. Results

Initial results for the algorithm have been obtained using the Face Recognition Grand Challenge (FRGC) database collected by the University of Notre Dame [12].

The FRGC comprises six experiments to measure the performance of 2D and 3D facial recognition algorithms. We show our algorithm's performance on the FRGC (version 1 and 2) Experiment 1. Experiment 1 consists of indoor controlled single still images versus indoor controlled single still images. Although training data was also available, our algorithm does not require a training stage and does not use this additional data. Also, our algorithm is fully automatic and does not require the eye coordinates that is also available with the data.

The data set for FRGC version 1 contains 366 training images, 152 gallery and 608 probe images. The probe set consists of image captured from each person up to 8 different time-lapsed sessions. The data set for FRGC version 2 contains 16028 controlled query and target images and 12776 training images. For FRGC version 2 Experiment 1, the target and the query sets are identical. It should be noted that the training set provided by FRGC (version 1 and 2) is never used in the experiments for the proposed algorithm.

Figure 4 shows the ROC performance on FRGC version 1 Experiment 1. The images from the FRGC database were cropped to 256x256 to extract the facial region. The experiments were carried out using a block size of 16x16 pixels and a search region of 48x48 pixels, while the a's were calculated using 8x8 pixel blocks. The algorithm took approximately 0.8 sec on a 3GHz P4 machine with 1GB RAM to obtain the bijectivity score between a single probe and a single gallery image. The proposed method achieves a False Rejection Rate of 6.85% at a False Acceptance Rate (FAR) of 0.1%, which is a significant improvement over the BEE baseline algorithm. The BEE baseline is a PCA algorithm and the distance measure used in the nearest neighbor classifier is Whitened Cosine [12].

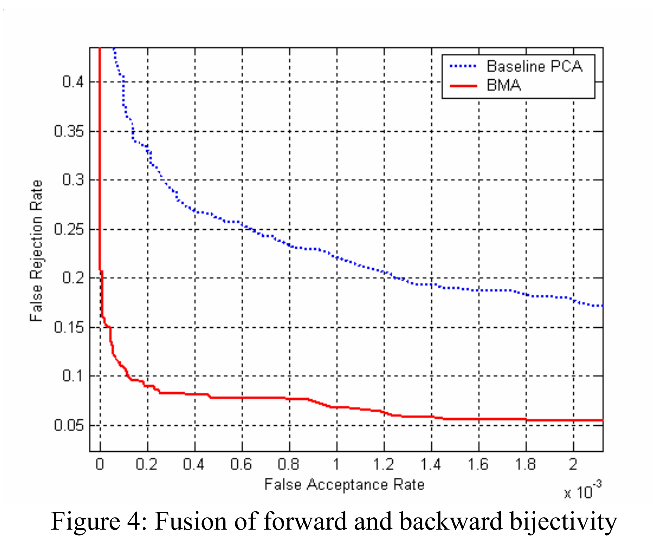

Figure 4: Fusion of forward and backward bijectivity estimation

The ROC curve in Figure 5 shows that the proposed algorithm results in a lower false rejection rate at a false acceptance rate of 0.1% for higher resolution images

(256x256) compared to lower resolution images (128x128) at additional computational cost due to an increase of four times the data.

Figure 6 ROC curve shows the improvement achieved by the fusion of BMA on the higher resolution (256x256) image and PCA scores on experiments performed on the FRGC version 1 data set.

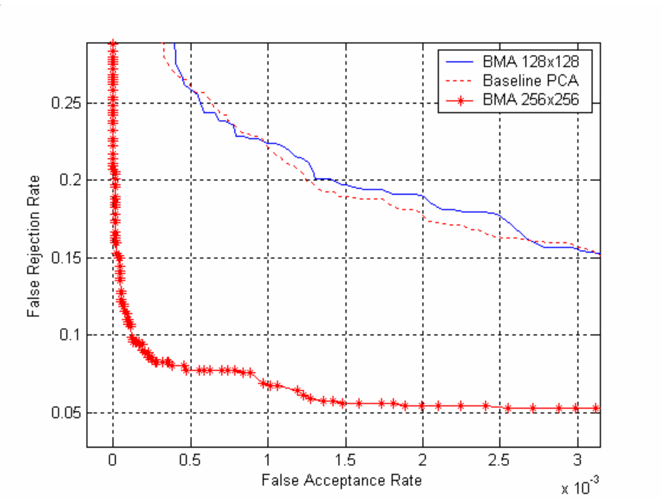

Figure 5: Effect of Image resolution on the performance of BMA

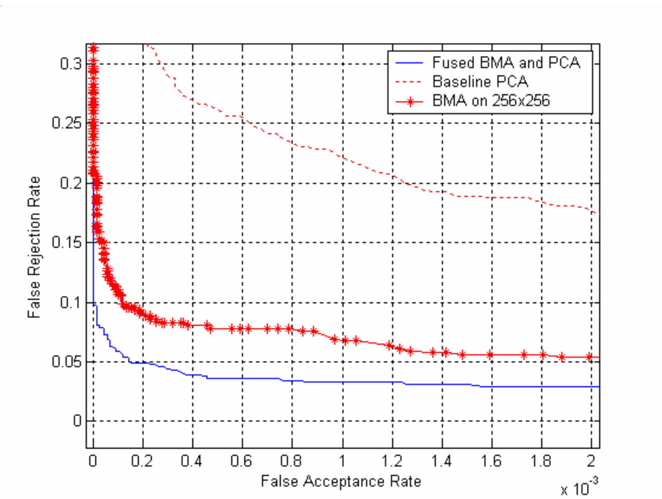

Figure 6: Fusion of BMA and PCA for FRGC version 1

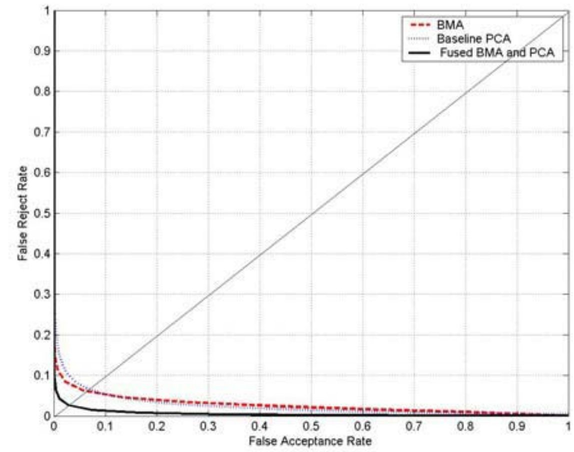

Figure 7: ROC curves for BMA and PCA for FRGC version 2 database

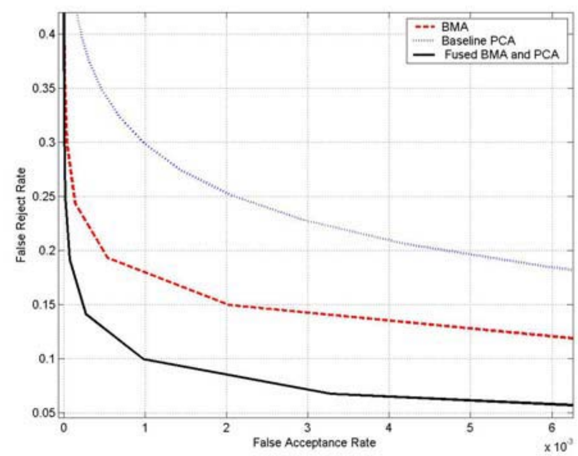

Figure 8: Fusion of BMA and PCA for FRGC version 2

Figure 7 shows the ROC performance on FRGC version 2 Experiment 1. The image size, block size and search regions are the same as in version 1, experiment 1. The BEE baseline algorithm is PCA with whitened cosine.

Figure 8 shows the ROC performance where at 0.1% FAR, baseline PCA yields approximately 30% false reject rate, the new BMA yields approximately 18% false reject rate and a fused PCA-BMA yields 10% false reject rate.

6. Conclusion and Future Work

Initial results show that the face recognition algorithm using bijective mappings works well under varying pose, expression and illumination changes. We have found that performance degrades for drastic illumination conditions (outdoor lighting, deep shadows) and drastic pose variations (greater than 60 degrees). Future work includes improving the disparity estimation for more difficult illumination and pose conditions including a more accurate model for illumination than the current piecewise constant multiplicative model.

Significant improvement was achieved by using higher resolution images at the expense of increase in computational complexity. Another direction for future research includes reducing the complexity of the disparity estimation algorithm. Directions aimed at reducing the complexity include reducing the number of candidates searched for each block, reducing the number of computations for the block matching score, disparity estimation to reduce the search window size and tree-search strategies for large databases.

Acknowledgments

The authors would like to thank David Herrington and Jim Wayman for their support and technical guidance. This research was sponsored by Technical Support Working Group (TSWG) under the contract No. DAAD05-03-C-0043.

References

[1] R. Chellapa, C.L. Wilson and S.S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proceedings of the IEEE*, Vol. 83. No. 5, May. 1995, pp. 705-741.

[2] W. Y. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: A literature survey", *UMD CfAR Technical Report CAR-TR-948,* 2000.

[3] R. Brunelli and T. Poggio, "Face recognition: features versus templates", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, Oct. 1993, pp. 1042-1052.

[4] T. Kanade, "Computer Recognition of Human Faces", Birkhäuser Verlag, Stuttgart Germany, 1977.

[5] J. Xiao, S. Baker, I. Matthews and T. Kanade, "Real-time combined 2D+3D active appearance models", *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern recognition*, Vol. 2, 27 June – 2 July 2004, pp. 535-542.

[6] V. Blanz and T.Vetter, "Face recognition based on fitting a 3D morphable model", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, Sept. 2003, pp. 1063-1074.

[7] M. Turk and A. Pentland, "Eigenfaces for recognition", *J. Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.

[8] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July. 1997, pp. 711-720.

[9] M.S. Bartlett, J.R. Movellan and T.J Sejnowski, "Face recognition by independent component analysis", *IEEE Trans. On Neural Networks*, Vol. 13, No. 6, Nov. 2002, pp. 1450-1464.

[10] K. Jonsson, J. Kittler, Y.P. Li and J. Matas, "Support vector machines for face authentication", *In T. Pridmore and D. Elliman, editors, BMVC'99,* 1999, pp. 534-553.

[11] T. Sikora, "MPEG digital video-coding standards", *IEEE Signal Processing Magazine*, Vol. 14, No. 5, Sept. 1997, pp. 82-100

[12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005*