FACE IDENTIFICATION FROM ONE SINGLE SAMPLE FACE IMAGE

Hung-Son Le, Haibo Li

Digital Media Lab, Dept. of Applied Physics and Electronics,
Umeå University, SE-901 87 Umeå, Sweden

ABSTRACT

This paper is addressing a challenging face recognition problem: Face Identification From One Single Face Image. We present a novel approach to face identification, which is capable to identify a person from face images that are significantly different from the sample image in terms of illumination, camera view angles and expressions. The approach is based on a new measurement of dissimilarity between the two face images. A person is identified based on the smallest dissimilarity, which is the summation of the dissimilarities of all pairs of observations extracted from the face image on both vertical and horizontal directions. Our experiment results tested on both the AR face database and the CMU PIE face database show that the proposed method outperforms the PCA, LDA, LFA based approaches.

1. INTRODUCTION

The aim of a computer based face identification system is to identify a person from his/her digitized facial image. Such an automatic system has a lot of practical applications, for instance, personal record retrieval, access control, security monitoring and surveillance. Within the past decade, many face recognition systems have emerged that are capable of achieving high recognition rate of over 95% accuracy under controlled condition. However, there are limitations in the use of face identification at critical places such as border control station, e.g. the airport, mainly due to the availability of true facial images of the wanted people and the requirement of real-time processing. In our study we make a minimal assumption that *only one sample face image of the wanted person is available.* This can be formed as a technical problem: *Face Identification From One Single Face Image.* Even if a single face image is available it was mostly taken in an uncontrolled environment. It is very challenging to most face identification techniques. This is due to the environmental condition, and the used camera as well as its set-up when the sample image was taken differs from those during identification. To deal with the facial appearance changes between the sample face image and the captured one in the field, alignment techniques have to be used [1, 2]. A typical way is to synthesize a new view from the sample face image by means of 3D morphable model [3]. The problem is that the quality of synthesized view is affected by the accuracy of the employed 3D model. Furthermore most of current face recognition algorithms require some manual initialization and the complexity due to image synthesis is too high to be used in real-time applications as in the airports.

In this paper we propose a novel and fast method to compute the dissimilarity between two face images. It is an extension of our previous work [4], which is based on Hidden Markov Model. Nevertheless the current approach is based on a new set of simple rules to measure the distance between two observation blocks in the sequence extracted from face images on both vertical and horizontal directions. The proposed method was tested with the images taken at near frontal faces under varying expression, illumination, partially occlusion from the AR face database [5] and the images taken at different poses from the CMU PIE face database [6]. The results were compared to PCA, LDA and LFA based approaches.

This work was supported by the Tactile Video project financed by grants from EU Objective One Northern Sweden.

2. ALGORITHM

The dissimilarities between images are computed from all pairs of observations extracted on both vertical and horizontal directions. We consider the scheme over the vertical direction first. The procedure on the horizontal direction is straightforward (the same procedures are applied, but on the transposed image). Assume we have a set of R people to be identified. In order to do face recognition, we perform the following steps:

2.1. Observation vector generation.

The vector sequence generation from image I of width W and height H, is carried out as following:

- First, overlapping vertical strips are extracted from I by horizontally scanning from left to right. Each strip has the width w_s and the height of I's height H.

0-7803-8554-3/04/$20.00

The amount of vertical overlap dw between consecutive strips is permitted up to one pixel less than strip width. Number of strips per image is

$$N_S = 1 + \lfloor (W - w_s)/(w_s - dw) \rfloor \qquad (1)$$

- Each strip is divided into overlapping blocks of height h_b. The amount of overlap dh between consecutive blocks is permitted up to one pixel less than the block height. Then new feature blocks are computed based on the difference of two consecutive blocks. Number of feature blocks per strip is

$$N_b = \lfloor (H - h_b)/(h_b - dh) \rfloor \qquad (2)$$

- The extracted feature blocks are then normalized

$$b_i = \frac{B_i - \mu}{\sigma^2} \qquad (3)$$

  where μ and σ^2 are the mean value and the variance of the extracted feature block B_i respectively, whereas $\{i : 1 \leq i \leq N_b\}$.

- The normalized feature blocks are arranged columnwise to form the observation vectors.

We apply the Haar wavelet decomposition [7] to the image as in [4] to reduce the dimension of observation vector.

2.2. Recognition

First of all observation vectors are generated from all the sample images to constitute a "codebook". Each vector in the codebook is indexed by a triple $\{i, j, r\}$ to address that the vector was extracted from the i^{th} block of j^{th} strip of the sample image of r^{th} person.

Each unknown face image, which is going to be identified, is compared against reference images of all subjects in the system. The dissimilarity between the test image and the reference images of all subjects must be computed. The selection of the identified person is based on the smallest dissimilarity. We will focus to the computation of dissimilarity between the unknown image and the reference image of r^{th} person.

The input test image is transformed by Haar wavelet and then divided into overlapping vertical strips. These vertical strips are considered sub-images and these sub-images can be treated independently. In other words the recognition of one face image is based on the recognition of its constituent sub-images. The following part will give more details on the processing and computation for each strip (or sub-image). The observation vectors from the strip are generated as described in 2.1. Let the position of the unknown observation vector in the strip be i_u and let the ordering number of the strip in which the vector belongs to, be j_u. The unknown observation vector is searched in the codebook against those vectors that have indices satisfied:

$$\begin{cases} i \in \{i_u, i_u \pm 1, i_u \pm 2, \ldots, i_u \pm n_v\} \\ j \in \{j_u, j_u \pm 1, j_u \pm 2, \ldots, j_u \pm n_h\} \\ r \in \{1, 2, \ldots, R\} \end{cases} \qquad (4)$$

Parameters n_v and n_h are used to limit the search range in the codebook. We could extend the searching region to deal with the expression variant or big rotation or translation of the face in the image, but it would introduce higher computational cost. Euclidean distances between the vectors are computed and the one with minimum value is assumed the best match. Let the indices of the best match vector found from codebook be $\{i_f, j_f, r_f\}$.

2.2.1. *Forming observation sequences according to subject*

The observation vectors extracted from a strip will form observation sequence(s). The similarity between the unknown subject to each subject $\{r : 1 \leq r \leq R\}$ in the system is computed based on the extracted observations. The way of labelling an observation vector is context dependant of both the indices of the best match vector $\{i_f, j_f, r_f\}$ and the index r of the subject in considering:

- The observation vector is given a label of observation o_{i_f} if $r_f = r$, else o_0 otherwise.
- Those observation vectors got the same j_f index form observation sequence O^{j_f}.
- Those O^{j_f} with length $T > 2$ are broken down to pairs of observations, $\{O_1^{j_f}, O_2^{j_f}, \ldots, O_{T-1}^{j_f}\}$, where $O_t^{j_f} = \{o_t, o_{t+1}\}$ and $1 \leq t \leq T - 1$

2.2.2. *Measurement of Dissimilarity*

The dissimilarity, denoted as Ds is computed from a pair of observations as following:

- If one of the observations is o_0: $Ds = N_O$
- If both two observations are o_0: $Ds = 2 * N_O$
- For a pair $[o_i, o_j]$, where $i, j > 0, |i - j| = 1, Ds = 0$
- For a pair $[o_i, o_j]$, where $i, j > 0, |i - j| > 1$
  Let $M = max(i, j); m = min(i, j)$
  If $i < j then : Ds = (M - m - 0.5)/(R * N_O)$
  Else $Ds = (M - m)/(R * N_O)$
  where as N_O is the number of labels, $N_O = N_b + 1$

The smaller the dissimilarity computed from a pair $O_t^{j_f}$, the more likely the subsequence $O_t^{j_f}$ belongs to the j_f^{th} sub-image of the r^{th} person. The sum of dissimilarity over all pairs and all strips

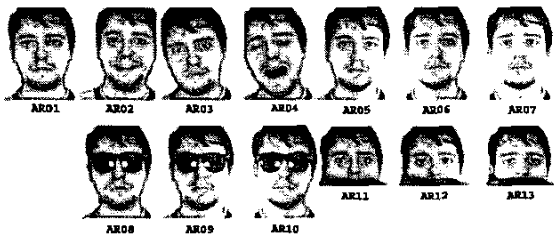

Fig. 1. Examples of images from the AR database.

$$\sum_{i=1}^{N_S}\sum_{j_f,t} Ds(O_t^{j_f}) \tag{5}$$

gives dissimilarity between the testing image and the reference image of the r^{th} person.

The above procedure can be considered as computing dissimilarity over the vertical direction. We do horizontal computation using the same steps but on the transposed images. The sum of Ds over all pairs of observations on both two directions in considering the r^{th} person's context gives the final dissimilar score for that person. The dissimilarity of the unknown face image to the images of all subjects must be computed. For computation against each r^{th} subject, the observation vectors are re-labelled as in 2.2.1 and the sum of dissimilarity over all observation subsequences and all strips on both two directions are computed as in 2.2.2. The selection of the identified person is based on the smallest dissimilarity.

3. EXPERIMENTAL RESULTS

We have tested our system on both the AR and the CMU PIE face databases.

The publicly available AR database [5] contains images of 126 individuals. The subjects were recorded twice at a 2-week interval. During each session 13 conditions with varying facial expressions, illumination and occlusion were captured. Those images of those people who have pictures taken in both 2 sessions were selected for the test described in this paper. That is a total of 120 people (65 males and 55 females), 26 images per each person, giving a total of 3120 images. Before using the database for the recognition experiment, the unnecessary background area in the original images should be removed. We process all neutral images $AR01$ first. All that images were normalized (using bilinear interpolation) to the same size and the same *distance between two eyes* roughly and then cropped to include the face area mainly. Since there are no big differences in face position, pose and scale in those images of the same person taken in the same session, the other images of types $AR02$, ..., $AR13$ can be pre-processed in the same manner applied to its corresponding neutral image $AR01$. The images were converted to gray and processed with histogram equalization. Most of the scarf region in the images of types $AR11$, $AR12$, $AR13$ were removed. Fig. 1 shows examples for different conditions.

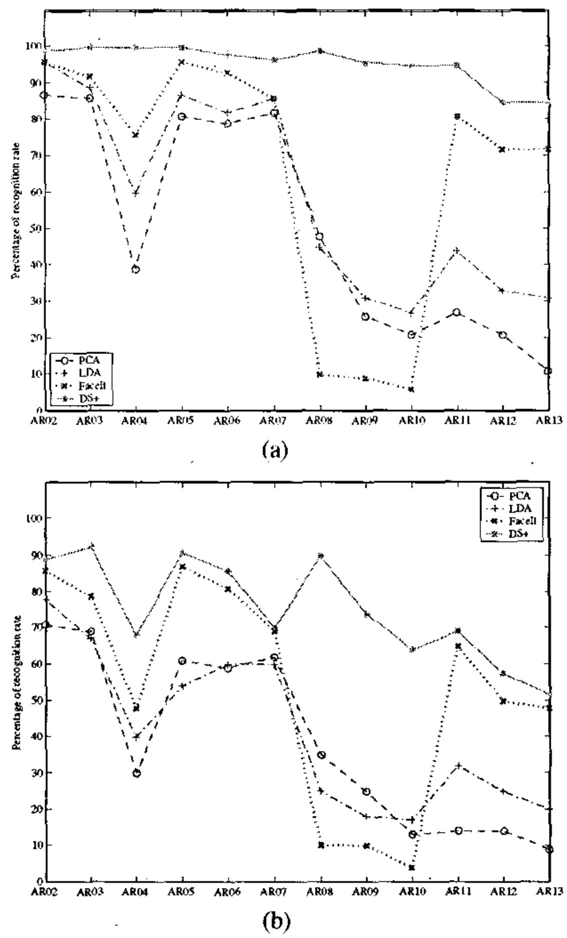

Fig. 2. Recognition rate (%) for expression, illumination, occlusion coupled with illumination variations of a) AR images recorded in the 1st session b) "duplicate" images recorded in the 2nd session.

Neutral images of type $AR01$ from the first recording session were used as sample images. All the images of other types $AR02 \ldots AR13$ from both 2 sessions were used for testing. We used the same parameter setting as in [4]: (1) Observation vectors are generated from transformed image with block size of 5x4. (2) The step for vertical and horizontal scanning is set at 1 pixel each. (3) Parameters n_v and n_h decided the search range in the codebook are set to 3.

Our system works in the Matlab environment running on PIII 1Ghz PC/Win2000. The time to train, in reality

Table 1. Recognition rate (%) for PIE images across pose variations

β	-32	0	0	16	31	44	44
α	2	2	1.9	2	2	2	13
pose	c11	c27	c07	c05	c37	c25	c02
FaceIt	75	sample	93	93	62	6	3
DS+	97	sample	98	100	82	61	69

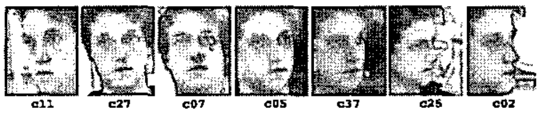

Fig. 3. Examples of images from the CMU PIE database.

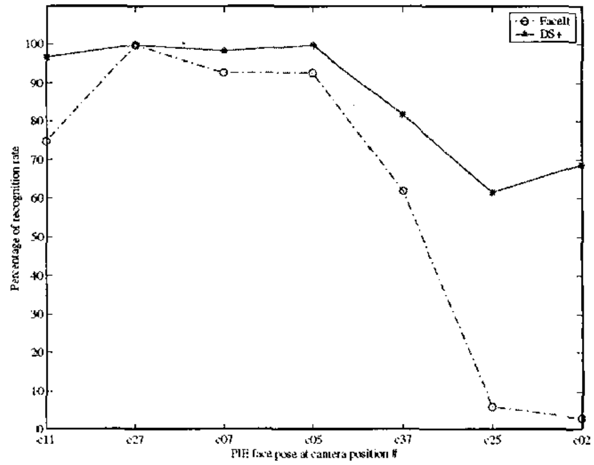

Fig. 4. Recognition rate (%) for PIE images across pose variations.

to form the codebook, for a person and the time to recognize an image are 0.25 second and 0.85 second respectively. These time results already included the time of 0.14 second for Haar wavelet transforms to get the coefficients matrix at resolution of 28x23 from the pre-processed image of size 224x184. Fig. 2 shows the recognition rate results of the proposed method, denoted as DS+ and comparative results of other methods reported in [2], which are based on the PCA, LDA, LFA approaches.

The proposed scheme was also tested with images taken from different poses. The data set for our testing is CMU PIE database that contains a total of 41,368 images taken from 68 individuals. The subjects were imaged in the CMU 3D Room using a set of 13 synchronized high-quality color cameras and 21 flashes. The resulting images are 640x480 in size, with 24-bit color resolution. The images of all subjects across several poses were selected for the experiment. PIE Feature Point Pose Labels were used for extraction, alignment and normalization of face region. Fig. 3 shows examples for different conditions. The same parameters setting as in previous experiment with AR database was used. Frontal images of the pose $c27$ were used for training. The images of other poses $c02$, $c05$, $c07$, $c11$, $c25$, $c37$ were used for testing. Tab. 1 and Fig. 4 show the recognition rate results of the proposed method, denoted as DS+ and comparative results of other method reported in [2], which are based on LFA approach.

4. CONCLUSION

A low complexity yet efficient face recognition system has been introduced. The present work exploited new way of extracting observations sequences and applying some simple but efficient rules to compute the dissimilarity between two images. The proposed method works efficiently with a single example image per person and even without precise facial localization and registration. For the considered facial databases, the results of the proposed scheme show substantial improvement over the results obtained from some other methods. The proposed algorithm is robust to single variation of facial expression, illumination, partially occlusion; and the system can recognize duplicate image and small pose variant image tolerably.

5. REFERENCES

[1] A.M. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[2] Ralph Gross, Jianbo Shi, and Jeffrey Cohn, "Quo vadis face recognition?," Tech. Rep. CMU-RI-TR-01-17, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, June 2001.

[3] Volker Blanz and Thomas Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, 2003.

[4] Hung-Son Le and Haibo Li, "Simple 1D Discrete Hidden Markov Models for Face Recognition," in *Proc. of Int. workshop, VLBV 2003 (visual content processing and representation)*. Sep 2003, LNCS 2849, Springer-Verlag Heidelberg.

[5] A.M. Martinez and R. Benavente, "The AR face database," Tech. Rep. 24, CVC Tech, 1998.

[6] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. of the 5th Int. Conf. on Automatic Face and Gesture Recognition*, 2002.

[7] Stephane G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1989, vol. 2.