Video Based Face Recognition Using Graph Matching

Gayathri Mahalingam and Chandra Kambhamettu

Video/Image Modeling and Synthesis (VIMS) Laboratory,
Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA

Abstract. In this paper, we propose a novel graph based approach for still-to-video based face recognition, in which the temporal and spatial information of the face from each frame of the video is utilized. The spatial information is incorporated using a graph based face representation. The graphs contain information on the appearance and geometry of facial feature points and are labeled using the feature descriptors of the feature points. The temporal information is captured using an adaptive probabilistic appearance model. The recognition is performed in two stages where in the first stage a Maximum a Posteriori solution based on PCA is computed to prune the search space and select fewer candidates. A simple deterministic algorithm which exploits the topology of the graph is used for matching in the second stage. The experimental results on the UTD database and our dataset show that the adaptive matching and the graph based representation provides robust performance in recognition.

1 Introduction

Face recognition has long been an active area of research, and numerous algorithms have been proposed over the years. For more than a decade, active research work has been done on face recognition from still images or from videos of a scene [1]. A detailed survey of existing algorithms on video-based face recognition can be found in [2] and [3]. The face recognition algorithms developed during the past decades can be classified into two categories: holistic approaches and local feature based approaches. The major holistic approaches that were developed are Principal Component Analysis (PCA) [4], combined Principal Component Analysis and Linear Discriminant Analysis (PCA+LDA) [5], and Bayesian Intra-personal/Extra-personal Classifier (BIC) [6].

Chellappa *et al.* [7] proposed an approach in which a Bayesian classifier is used for capturing the temporal information from a video sequence and the posterior distribution is computed using sequential importance sampling. As for the local feature based approaches, Manjunath and Chellappa [8] proposed a feature based approach in which features are derived from the intensity data without assuming any knowledge of the face structure. Topological graphs are used to represent relations between features, and the faces are recognized by matching the graphs. Ersi and Zelek [9] proposed a feature based approach where in a statistical Local

Feature Analysis (LFA) method is used to extract the feature points from a face image. Gabor histograms are generated using the feature points and are used to identify the face images by comparing the Gabor histograms using a similarity metric. Wiskott *et al.* [10] proposed a feature based graph representation of the face images for face recognition in still images. The face is represented as a graph with the features as the nodes and each feature described using a Gabor jet. The recognition is performed by matching graphs and finding the most similar ones. A similar framework was proposed by Ersi *et al.* [11] in which the graphs were generated by triangulating the feature points.

Most of these approaches focused on image-based face recognition applications. Various approaches to video-based face recognition have been studied in the past, in which both the training and test set are video sequences. Video-based face recognition has the advantage of using the temporal information from each frame of the video sequence. Zhou *et al.* [12] proposed a probabilistic approach in which the face motion is modeled as a joint distribution, whose marginal distribution is estimated and used for recognition. Li [13] used the temporal information to model the face from the video sequence as a surface in a subspace and performed recognition by matching the surfaces. Kim *et al.* [14] recognized faces from video sequences by fusing pose-discriminant and person-discriminant features by modeling a Hidden Markov Model (HMM) over the duration of a video sequence. Stallkamp *et al.* [15] proposed a classification sub-system of a real-time video-based face identification system. The system uses K-nearest neighbor model and Gaussian mixture model (GMM) for classification purposes and uses distance-to-model, and distance-to-second-closest metrics to weight the contribution of each individual frame to the overall classification decision.

Liu and Chen [16] proposed an adaptive HMM to model the face images in which the HMM is updated with the result of identification from the previous frame. Lee *et al.* [17] represented each individual by a low-dimensional appearance manifold in the ambient image space. The model is trained from a set of video sequences to extract a transition probability between various poses and across partial occlusions. Park and Jain [18] proposed a 3D model based approach in which a 3D model of the face is used to estimate the pose of the face in each frame and then matching is performed by extracting the frontal pose from the 3D model. Xu *et al.* [19] proposed a video based face recognition system in which they integrate the effects of pose and structure of the face and the illumination conditions for each frame in a video sequence in the presence of multiple point and extended light sources. The pose and illumination estimates in the probe and gallery sequences are then compared for recognition applications.

In this paper, we propose a novel graph based approach for image-to-video based face recognition which utilizes the spatial and temporal characteristics of the face from the videos. The face is spatially represented by constructing a graph using the facial feature points as vertices and labeling them with their feature descriptors. A probabilistic mixture model is constructed for each subject which captures the temporal information. The recognition is performed in two stages where in the first stage the probabilistic mixture model is used to prune the search space using a MAP rule. A simple deterministic algorithm that uses cosine similarity measure is used to compare the graphs in the second stage. The probabilistic models are updated with the results of recognition from each frame of the video sequence, thus making them adaptive. Section 2 explains our procedure in constructing the graphs and the adaptive probabilistic mixture models for each subject. The two stage recognition is explained in section 5.

2 Face Image Representation

In this section, we describe our approach in extracting the facial feature points and their descriptors which are used in the spatial representation of the face images. Every face is distinguished not by the properties of individual features, but by the contextual relative location and comparative appearance of these features. Hence it is important to identify those features that are conceptually common in every face such as eye corners, nose, mouth, etc. In our approach, the facial feature points are extracted using a modified Local Feature Analysis (LFA) technique, and extracted feature points are described using Local Binary Pattern (LBP) [20], [21] feature descriptors.

2.1 Feature Point Extraction

The Local Feature Analysis (LFA) proposed by Penev and Atick [22] constructs kernels, which are basis vectors for feature extraction. The kernels are constructed using the eigenvectors of the covariance matrix of the vectorized face images. LFA is referred to as a local method since it constructs a set of kernels that detects local structure; e.g., nose, eye, jaw-line, and cheekbone, etc. The local kernels are optimally matched to the second-order statistics of the input ensemble [22]. Given a set of n d-dimensional images $x_1, \ldots, x_n$, Penev and Atick [22] compute the covariance matrix C, from the zero-mean matrix X of the n vectorized images as follows:

$$C = XX^T. \tag{1}$$

The eigenvalues of the covariance matrix C are computed and the first k largest eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_k$, and their associated eigenvectors $\psi_1, \ldots, \psi_k$ to define the kernel K,

$$K = \Psi \Lambda \Psi^T \tag{2}$$

where $\Psi = [\psi_1 \ldots \psi_k]$, $\Lambda = diag(\frac{1}{\sqrt{\lambda_r}})$.

The rows of K contain the kernels. These kernels have spatially local properties and are ”topographic” in the sense that the kernels are indexed by spatial location of the pixels in the image, *i.e.*, each pixel in the image is represented by a kernel from K. Figure 1(a) shows the kernels corresponding to the nose, eye, mouth and cheek positions. The kernel matrix K transforms the input image matrix X to the LFA output $O = K^T X$ which inherits the same topography as the input space.

Hence, the dimension of the output is reduced by choosing a subset of kernels, M, where M is a subset of indices of elements of K. These subsets of kernels are considered to be at those spatial locations which are the feature points of the face image. Penev and Atick [22] proposed an iterative algorithm that uses the mean reconstruction error to construct M by adding a kernel at each step whose output produces the maximum reconstruction error,

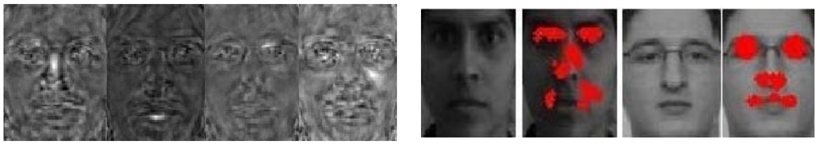
$$\arg\max_{x} \langle ||O(x) - O^{rec}(x)||^2 \rangle \tag{3}$$

where $O^{rec}(x)$ is the reconstruction of the output $O(x)$.

Although mean reconstruction error is a useful criterion for representing data, it does not guarantee an effective discrimination between data from different classes as the kernels selection process aims at reducing the reconstruction error for the entire image and not the face region. Hence, we propose to use the Fisher's linear discriminant method [23] to select the kernels that characterize the most discriminant and descriptive feature points of different classes. We compute the Fisher scores using the LFA output O. Fisher score is a measure of discriminant power which estimates how well different classes of data are separated from each other, and is measured as the ratio of variance between the classes to the variance within the classes. Given the LFA output $O = [o_1 \dots o_n]$ for c classes, with each class having n_i samples in the subset χ_i, the Fisher score of the x^{th} kernel, $J(x)$ is given by

$$J(x) = \frac{\sum_{i=1}^{c} n_i (m_i(x) - m(x))^2}{\sum_{i=1}^{c} \sum_{o \in \chi_i} (o(x) - m_i(x))^2} \tag{4}$$

where $m(x) = \frac{1}{n} \sum_{i=1}^{c} n_i m_i(x)$ and $m_i(x) = \frac{1}{n_i} \sum_{o \in \chi_i} o(x)$. The kernels that correspond to high Fisher scores are chosen to represent the most discriminative feature points of the image. Figure 1(b) shows the set of feature points extracted using the Fisher scores.



(a) $K(x, y)$ derived from a set of 315 images

(b) The first 100 feature points extracted from the training images

Fig. 1. 1(a) shows $K(x, y)$ at the nose, mouth, eye, and cheeks and 1(b) shows the feature points extracted (best viewed in color).

2.2 Feature Description with Local Binary Pattern

A feature descriptor is constructed for each feature point extracted from an image using Local Binary Pattern (LBP).

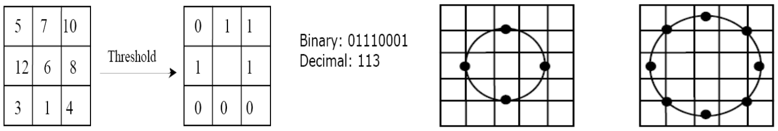
The original Local Binary Pattern (LBP) operator proposed by Ojala *et al.* [20] is a simple but very efficient and powerful operator for texture description. The operator labels the pixels of an image by thresholding the $n \times n$ neighborhood of each pixel with the value of the center pixel, and considering the result value as a binary number. Figure 2(a) shows an example of the basic LBP operator and figure 2(b) shows a $(4, 1)$ and $(8, 2)$ circular LBP operator. The histogram of the labels of the pixels of the image can be used as a texture descriptor. The grey-scale invariance is achieved by considering a local neighborhood for each pixel, and invariance with respect to scaling of the grey scale is achieved by considering just the signs of the differences in the pixel values instead of their exact values. The LBP operator with P sampling points on a circular neighborhood of radius R is given by,

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{5}$$

where

$$s(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases} \tag{6}$$

Ojala *et al.* [21] also introduced another extension to the original operator which uses the property called *uniform patterns* according to which a LBP is called uniform if there exist at most two bitwise transitions from 0 to 1 or vice versa. Uniform patterns can reduce the dimension of the LBP significantly which is advantageous for face recognition.



(a) The Basic LBP operator

(b) $(4, 1)$ and $(8, 2)$ circular LBP operator

Fig. 2. The basic LBP operator and the circular LBP operator

In our experiments, we use $LBP^{u2}_{8,2}$ operator which denotes a uniform LBP operator with 8 sampling pixels in a local neighborhood region of radius 2. A 5×5 window around the pixel is chosen as the neighborhood region and a feature vector of length 59 is obtained.

3 Image Graph Construction

The most distinctive property of a graph is its geometry, which is determined by the way the vertices of the graph are arranged spatially. Graph geometry plays an important role in discriminating the graphs of different face images. In our approach, the graph geometry is defined by constructing a graph with constraints imposed on the length of the edges between a vertex and its neighbors.

Considering that we extract around n feature points from each face image, at least $n!$ graphs can be generated for each image. Evaluating this number of graphs for each probe image would be very computationally expensive. Hence, a graph generating procedure that generates a unique graph with the given set of vertices is proposed. At each iteration, vertices and edges are added to the graph in a Breadth-first search manner and considering a spatial neighborhood distance for each vertex. This generates a unique graph for a set of feature points. The procedure to generate a graph given a set of vertices is given as follows;

```
1 Pick a random vertex $v$ from the list of vertices of the graph.
2 Add $v$ to the end of the queue $q$.
3 While $NOT$ all the vertices have been visited
      Pick a vertex $u$ from the front of the queue $q$.
      If $u$ is not visited
          Find the Neighbors $N$ of $u$ who are within a Euclidean distance.
          Add $N$ to the queue $q$.
          Mark $u$ as visited
      endif
  endwhile
```

The idea behind representing face images using graphs is mainly due to the spatial properties of the graph, as a graph can represent the inherent shape changes of a face and also provide a simple, but powerful matching technique to compare graphs.

4 Probabilistic Graph Appearance Model

The appearance of a graph is another important distinctive property and is described using the feature descriptors of the vertices of the graph. An efficient as well as effective description of the appearance of the vertices of the graphs is required in order construct a graph appearance model that elevates the distinctive properties of the face of an individual. Modeling the joint probability distribution of the appearance of the vertices of the graphs of an individual produces an effective representation of the appearance model through a probabilistic framework. Since the model is constructed using the feature descriptors, it is easy to adapt the model to the changes in the size of the training data for the individual. Given N individuals and M training face images, the algorithm to learn the model is described as follows:

1. Initialize N model sets.
2. For each training image I_c^j, (j^{th} image of the c^{th} individual)
   a. Extract the feature points (as described in Subsection 2.1).

   b. Compute feature descriptors for each feature point (as described in Subsection 2.2).
   c. Construct Image graphs (as described in Subsection 3).
   d. Include the graph in the model of the c^{th} individual.
3. Construct the appearance model for each individual using their model sets.

In our approach, a probabilistic graph appearance model is generated for each subject and is used for training purposes. Given a graph $G(V, E)$, where V is the list of vertices in the graph, and E the set of edges in the graph, the probability of G belonging to a model set (subject) k is given by,

$$R_k = \max_n P(G|\Phi_n) \tag{7}$$

where $P(G|\Phi_n)$ is the posterior probability, and Φ_n is the appearance model for the n^{th} subject constructed using the set of feature descriptors F of the set of vertices of all the graphs of the subject. The appearance model Φ_n is constructed by estimating the joint probability distribution of the appearance of the graphs for each subject. R_k is called the Maximum a Posteriori (MAP) solution. In our approach, we estimate the joint probability distribution of the graph appearance model for each subject using the Gaussian Mixture Model (GMM) [24] which can efficiently represent heterogeneous data, the dominant patterns which are captured by the Gaussian component distributions.

Given a training face database containing images of L subjects and each subject having at least one image in the training database, the set of feature descriptors X for each subject to be used to model the joint likelihood of the subject will be a $(m \times f) \times t$ distribution, where m is the number of images for each subject, f is the number of feature points extracted for each image and t the dimension of the feature vector (in our case, it is 59 and is reduced to 20). To make the appearance model estimation more accurate and tractable, we use the Principal Component Analysis (PCA) to reduce the dimensionality of the feature vectors.

Each subject in the database is modeled as a GMM with K Gaussian components. The set of feature descriptors X of each subject is used to model the GMM of that individual. Mathematically, a GMM is defined as:

$$P(X|\theta) = \sum_{i=1}^{K} w_i.N(X|\mu_i, \sigma_i). \tag{8}$$

where

$$N(X|\mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}}.e^{-\frac{(X-\mu_i)^2}{2\sigma^2}}$$

are the components of the mixture, $\theta = \{w_i, \mu_i, \sigma_i^2\}_{i=1}^{K}$ includes the parameters of the model, which includes the weights w_i, the means μ_i, and the variances σ_i^2 of the K Gaussian components.

In order to maximize the likelihood function $P(X|\theta)$, the model parameters are re-estimated using the Expectation-Maximization (EM) technique [25]. The

EM algorithm is an iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. In ML estimation, we wish to estimate the model parameters for which the observed data are the most likely:

$$\theta^c = arg \max_{\theta} P(X|\theta). \tag{9}$$

At each iteration of the EM algorithm the missing data are estimated with the current estimate of the model parameters, and the likelihood function is maximized with assumption that the missing data are known. For more details about the EM algorithm see [25].

5 Adaptive Matching and Recognition

In this section, we describe our two stage matching procedure to adaptively match every frame of the video sequence and the trained appearance models and the graphs. In the first stage of the matching process, a MAP solution is computed for the test graph using the trained appearance models. The MAP solution is used to prune the search space for the second stage of matching. A subset of individuals' appearance model and their trained graphs are selected based on the MAP solution. This subset of appearance models are used in the second stage of matching process. In the second stage, a simple deterministic algorithm that uses the cosine similarity measure and the nearest neighborhood classifier to find the geometrical similarity of the graphs is proposed. The GMM is adapted with the result of recognition from each frame of the test video sequence. We use the likelihood score and the graph similarity score to decide on the correctness of the recognition and update the appropriate GMM. The recognition result of a frame is considered correct if the difference between the highest likelihood score and the second highest likelihood score is greater than a threshold. A similar difference in graph similarity scores is also computed to support the decision. This measure of correctness is based on the same idea as Lowe [26], that reliable matching requires the best match to be significantly better than the second-best match. For a given test sequence, the difference in the likelihood scores and the difference in the similarity scores are computed and the GMM is updated if these values are greater than a threshold. Given an existing GMM Θ_{old} and observation vectors O from the test sequence, the new GMM is estimated using the EM algorithm with Θ_{old} as the initial values. The entire matching procedure is given as follows;

1. For *each frame f in the video sequence*
   a. Extract the facial feature points and their descriptors from *f*.
   b. Reduce the dimension of feature descriptors using the projection matrix from training stage.
   c. Construct the image graph G.
   d. Obtain the probability of G belonging to each appearance model, and select the k model sets with highest probability. k is 10% in our experiments.

e. Obtain the similarity scores between G and the graphs of k individuals.

f. Update the appropriate appearance model based on the likelihood score and similarity score.

2. Select the individual with the maximum number of votes from all the frames.

The algorithm to find the spatial similarity between two graphs is given as follows;

1. For each vertex v in the test graph with a spatial neighborhood W, a search is conducted over W (in the trained graph) and the best matching feature vertex u is selected, such that

$$S_{vu} = \frac{f_v \cdot f_u}{|f_v||f_u|} \tag{10}$$

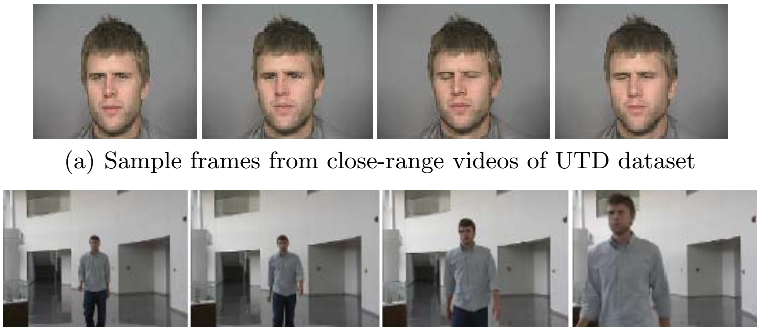
   where f_v and f_u are the feature vectors of v and u respectively, and S_{vu} is the similarity score between v and u.
2. Repeat step 1 with neighbors of v and so on until all the vertices have been matched. The sum of the similarity scores of all the vertices gives the measure of similarity between the two graphs.

6 Experiments

In order to validate the robustness of the proposed technique, we used a set of close range and moderate range videos from the UTD database [27]. The database included 315 subjects with high resolution images in various poses. The videos included subjects with neutral expression and also walking towards the camera from a distance. We also generated a set of moderate range videos (both indoor and outdoor) with 6 subjects. Figure 3 shows sample video frames from the UTD dataset and figure 4 shows sample video frames from our dataset.

In the preprocessing step, the face region is extracted from the image, normalized using histogram equalization technique and are resized to 72×60 pixels. 150 features were extracted and a LBP is computed for each feature point. PCA is performed on the feature vectors to reduce the dimension from 59 to 20 (with nearly 80% of the non-zero eigenvalues retained). A graph is generated for each face image with a maximum spatial neighborhood distance of 30 pixels. A graph space model is constructed for each subject using GMM with 10 Gaussian components.

During the testing stage, in order to mimic the practical situation, we consider a subset of frames in which an individual appear in the video and use it for testing purposes. We randomly select an individual and a set of frames that include the individual. The preprocessing and the graph generation procedure similar to those performed in the training stage are applied to each frame of the video sequence. The likelihood scores are computed for the test graph and the GMMs and the training graphs are matched with the test graph to produce similarity scores, and the appropriate GMM is updated using the similarity and



(a) Sample frames from close-range videos of UTD dataset

(b) Sample frames from moderate-range videos of UTD dataset

Fig. 3. Sample video frames from the UTD video dataset

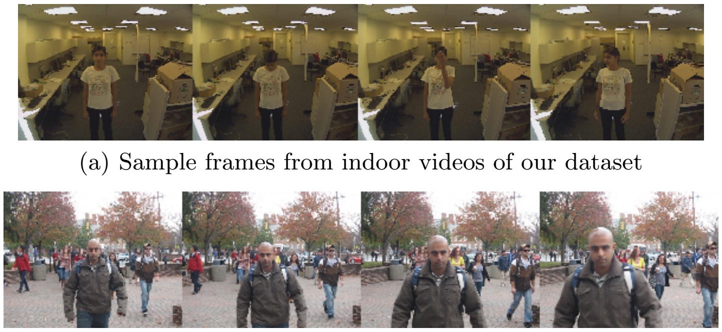
likelihood scores. The threshold is determined by the average of the difference in likelihood scores and similarity scores between each class of data. Though the threshold value is data dependent, the average proves to be an optimum value.

Table 1. Comparison of the error rates with different algorithms

	HMM	AGMM	Graphs	AGMM+Graphs
UTD Database (close-range)	24.3%	24.1%	23.2%	20.1%
UTD Database (moderate-range)	31.2%	31.2%	29.8%	25.4%
Our Dataset	8.2%	3.4%	2.1%	1.1%

The performance of the algorithm is compared with video-based recognition algorithm in [16] which handles video-to-video based recognition. The algorithm in [16] performs eigen analysis on the face images and uses an adaptive Hidden Markov Model (HMM) for recognition. We also test the performance of the system with only the adaptive graph appearance model (AGMM) and the appearance model with the graph model sets (AGMM+Graphs). The results are tabulated in the Table 1. Figure 5 shows the Cumulative Match Characteristic curve obtained for various algorithms (HMM, AGMM and AGMM+Graphs) on the UTD dataset.

From the error rates we can see that the performance of our approach is definitely promising when compared with the other approaches. The account of spatial and temporal information together improves the performance of the recognition process. The number of images in the training dataset played an important role in the performance, as it is evident from the error rates. The close-range videos of the UTD database has lower error rates than the moderate-range videos. This is due to the reason that the frame of the video sequence mostly contains the face region thus gathering more details of the facial features than

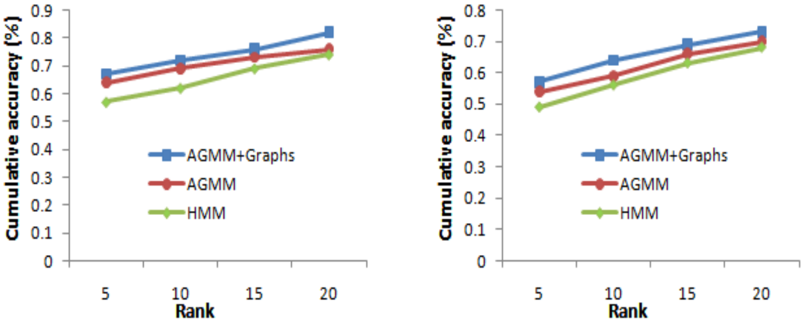


(a) Sample frames from indoor videos of our dataset

(b) Sample frames from outdoor videos of our dataset

Fig. 4. Sample video frames from our video dataset

the moderate-range videos. The number of training set images for each subject played a role in the performance. The UTD dataset included 3 training images for each subject whereas our dataset included at least 5 training images. The algorithm shows a high recognition rate when experimented on our dataset as it can be seen for the error rates. Though there were limited number of subjects in the dataset, the videos in the dataset included both indoor and outdoor videos taken using a PTZ camera which is mainly used for surveillance. The system provided a better performance with both indoor and outdoor videos which has different illumination, pose changes and in moderate range.



(a) CMC curve for close-range videos of UTD database

(b) CMC curve for moderate-range videos of UTD database

Fig. 5. Cumulative Match Characteristic curves for close-range and moderate-range videos

The system performs better as a video based face recognition system than a still image based face recognition system, due to the wealth of temporal information available from the video sequence and the effective use of it by the proposed adaptive probabilistic model. As a still image based face recognition, an image with a frontal pose of the face yields better performance than non frontal pose image. Thus, pose of the face image plays a role in the recognition. Also, the system's performance is affected by the comparison of a single high resolution image with a low resolution frame in a still image based face recognition system. Thus the adaptive matching technique combined with the graph based representation is significantly an advantage in matching images with videos.

From our experiments, we found that changing the value of the parameters did not significantly change the performance of the system and the values that we used tend to be the optimum. For example, increasing the maximum Euclidean distance between two vertices of a graph to a value greater than the width or length of the image will have no effect as the graph will always be connected as the distance between two vertices will never be greater than these values.

7 Conclusion

In this paper, we proposed a novel technique for face recognition from videos. The proposed technique utilizes both the temporal and spatial characteristics of a face image from the video sequence. The temporal characteristics are captured by constructing a probabilistic appearance model and a graph is constructed for each face image using the set of feature points as vertices of the graph and labeling it with the feature descriptors. A modified LFA and LBP were used to extract the feature points and feature descriptors respectively. The appearance model is built using GMM for each individual in the training stage and is adapted with the recognition results of each frame in the testing stage. A two stage matching procedure that exploits the spatial and temporal characteristics of the face image sequence is proposed for efficient matching. A simple deterministic algorithm to find similarity between the graphs is also proposed. Our future work will handle video sequences involving various pose of the faces, different resolutions, and video-to-video based recognition.

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces: a survey. Proceedings of the IEEE **83** (1995) 705–741
2. Wang, H., Wang, Y., Cao, Y.: Video-based face recognition: A survey. World Academy of Science, Engineering and Technology **60** (2009)
3. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey (2000)
4. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86
5. Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. Journal of the Optical Society of America **14** (1997) 1724–1733
6. Moghaddam, B., Nastar, C., Pentland, A.: Bayesian face recognition using deformable intensity surfaces. In Proceedings of Computer Vision and Pattern Recognition (1996) 638–645

7. Zhou, S., Krueger, V., Chellappa, R.: Probabilistic recognition of human faces from video. Computer Vision and Image Understanding **91** (2003) 214–245
8. Manjunath, B.S., Chellappa, R., Malsburg, C.: A feature based approach to face recognition (1992)
9. Ersi, E.F., Zelek, J.S.: Local feature matching for face recognition. In Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (2006)
10. Wiskott, L., Fellous, J.M., N.Kruger, Malsburg, C.V.D.: Face recognition by elastic bunch graph matching. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997) 775–779
11. Ersi, E.F., Zelek, J.S., Tsotsos, J.K.: Robust face recognition through local graph matching. Journal of Multimedia (2007) 31–37
12. Zhou, S., Krueger, V., Chellappa, R.: Face recognition from video: A condensation approach. In Proc. of fifth IEEE Internation Conference on Automatic Face and Gesture Recognition (2002) 221–228
13. Li, Y.: Dynamic face models: construction and applications. Ph.D. Thesis, University of London (2001)
14. Kim, M., Kumar, S., Pavlovic, V., Rowley, H.A.: Face tracking and recognition with visual constraints in real-world videos. CVPR (2008)
15. Stallkamp, J., Ekenel, H.K.: Video-based face recognition on real-world data (2007)
16. Liu, X., Chen, T.: Video-based face recognition using adaptive hidden markov models. CVPR (2003)
17. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Visual tracking and recognition using probabilistic appearance manifolds. Computer Vision and Image Understanding **99** (2005) 303–331
18. Park, U., Jain, A.K.: (The 2 nd international conference on biometrics, seoul, korea, 2007 3d model-based face recognition in video)
19. Wu, Y., Roy-Chowdhury, A., Patel, K.: Integrating illumination, motion and shape models for robust face recognition in video. EURASIP Journal of Advances in Signal Processing: Advanced Signal Processing and Pattern Recognition Methods for Biometrics (2008)
20. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition (1996) 51–59
21. Ojala, T., Pietikainen, M., Maenpaa, T.: A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. Second International Conference on advances in Pattern Recognition, Rio de Janeiro, Brazil (2001) 397–406
22. Penev, P., Atick, J.: Local feature analysis: A general statistical theory for object representation. Network: Computation in Neural Systems **7** (1996) 477–500
23. Mika, S., Rtsch, G., Weston, J., Schlkopf, B., Mller, K.R.: Fisher discriminant analysis with kernels (1999)
24. McLachlam, J., Peel, D.: Finite mixture models (2000)
25. Redner, R.A., Walker, H.F.: Mixture densities, maximum likelihood and the em algorithm. SIAM Review **26** (1984) 195–239
26. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
27. O'Toole, A., Harms, J., Hurst, S.L., Pappas, S.R., Abdi, H.: A video database of moving faces and people. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 812–816