MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Robust Facial Feature Extraction Using Embedded Hidden Markov Model for Face Recognition under Large Pose Variation

Ping-Han Lee[1], Yun-Wen Wang[1], Jison Hsu[2], Ming-Hsuan Yang[3] and Yi-Ping Hung[1]
[1]Dept. of Computer Science and Information Engineering, Nation Taiwan University
[2]PENPOWER Technology Ltd., Taiwan
[3]Honda Research Institute
contact email: hung@csie.ntu.edu.tw

Abstract

We propose an algorithm for extracting facial features robustly from images for face recognition under large pose variation. Rectangular facial features are retrieved via the by-products of an embedded Hidden Markov Model (HMM) which decodes an observed face image into a state sequence. While an HMM is able to segment images into features at a fixed pose, multiple HMMs are trained for each individual to robustly extract features under large pose variation. Using the extracted features of each individual, appearance models based on subspaces are constructed for face identification and verification. The effectiveness of the proposed approach is validated through empirical studies against numerous methods using the CMU PIE database. Our experiments demonstrate that the proposed approach is able to extract facial features robustly, thereby rendering superior results in identification and superior performance in verification under large pose variation.

1. Introduction

Face recognition is one of the most active research areas in computer vision with numerous applications including identification which matches an image against a set of registered images in a database for unique label, and verification which confirms whether the claimed identity of an given image is true or not. Numerous face recognition algorithms have been proposed and a thorough review on this subject can be found in [10]. Generally speaking, these algorithms can be categorized into holistic and feature-based approaches. While the holistic methods, e.g., the Eigenface [9] and Fisherface [1] methods, have demonstrated their potentials in face recognition, they do not work well when the pose of a test image varies significantly from the ones registered in the database.

Diametric to the holistic approaches, algorithms based on Hidden Markov Model (HMM) [6] [5] [7] [2] [4] exploit the fact that facial features can be observed in a sequential order even under appearance variation caused by in-plane/out-of-plane rotation or alignment error. In [6] a face pattern is divided into several overlapping, horizontal strips and these sub-images are modelled by a one-dimensional HMM model. Following [6], better identification results were achieved by using the coefficients of the two-dimensional discrete cosine transform (2D-DCT) as feature vectors [5] instead of the pixel intensity values. The one-dimensional HMM was extended to the pseudo two-dimensional HMM [7] that, as a compromise of computational complexity as well as model capability, consists of a vertical top-to-bottom one-dimensional Markov chain with super states and each super state in turn contains a horizontal left-to-right one-dimensional Markov chain. Further improvements on top of [7] were obtained by using the 2D-DCT representation [2], and a variation of [7] with lower computational complexity, the embedded HMM, was reported in [4].

While recent findings have demonstrated the success of these HMM-based methods in face identification [2] at fixed pose, these algorithms are not effective for face verification. Given an image, a trained HMM always finds a state sequence that best accounts for the observations with no regard to whether the person of that image is an legitimate user or an imposter. The reason being that face images consist of similar features and a trained HMM simply strives its best to account for the given observations. Despite its weakness in verification, a learned HMM can robustly decode an observed face image into states corresponding to certain facial regions. As will be clear later in this paper, some of these regions (states) correspond to visually salient facial features (e.g., eyes and nose) while the others match other components of a human face (e.g., foreheads and cheeks). Since localization of features plays a crucial role in any feature-based method for face recognition, a robust facial feature extraction is of great value for this as well as other applications. Motivated by this, we develop an HMM-based method for extracting facial features in spite of pose variation. We demonstrate the merits of our algorithm by empirical studies against numerous methods using with the CMU PIE database.

2. Embedded Hidden Markov Model

An embedded HMM consists of a Markov chain with super states, and each super state in turn is modelled by another Markov chain with a set of embedded states. While the super states are used to model two-dimensional data along one

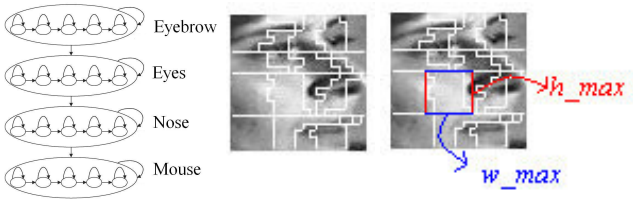
direction (e.g., top to bottom for image analysis), the embedded states are used to model along the other direction (e.g., left to right for image analysis). Figure 1(a) shows the structure of an HMM with 4 super states and 5 embedded states. An embedded HMM is defined as the triplet $\lambda = (\Pi_0, A_0, \Lambda)$, with a set of super states S_0, where Π_0 is the initial state distribution, A_0 is the super state transition probability matrix, and $\Lambda = \{\Lambda^{(1)}, \Lambda^{(2)}, ..., \Lambda^{(N_0)}\}$ contains a set of parameters of embedded HMMs, with each $\Lambda^{(k)} = (\Pi_1^{(k)}, A_1^{(k)}, B^{(k)})$. The training algorithm is based on the classic Viterbi algorithm and more details can be found in [7] [3] [4].

3. Feature Extraction under Pose Variation

Face recognition methods such as [7] [4] have applied the embedded HMMs to account for an observed image O with a probabilistic measure, $P(O|\lambda)$. In this work we utilize the by-product of the embedded HMM, which segments an observed image into regions based on state transitions via the Viterbi algorithm [3] [4], to extract facial features. We also use 2D-DCT coefficients to represent observation vectors with sampling windows overlapped in both directions to better model their neighborhood relations.

Although an embedded HMM is able to handle certain amount of variation caused by pose variation, it is not able to accurately account for observations undergoing large pose change. To deal with such situations, we apply the K-means algorithm to group face images of each person into K clusters based on their poses. Within each cluster, an HMM is trained to accurately account for observed images as their pose variation is limited. That is, for every person in the database we train K HMMs where each HMM is responsible for decoding face images within limited pose variation. Meanwhile, each trained HMM is able to segment each training image into facial regions based on the decoded state sequence. Figure 1(b) illustrates the decomposition of an observed face image, into 4 super states and the corresponding embedded states therein. Note that some of decoded states match visually salient features such as eyes and noses, while others represent facial components (e.g., cheeks). Pixels in the same region are estimated with the same embedded state as well as their super state, and each of these regions is considered as a facial feature. In our experiments, we have 20 features for each person (as a result of using 20 states in the embedded HMM).

To extract sub-images for each feature of every person, we first compute the center point, maximum width (w_max) and height (h_max) of the corresponding facial regions decoded by the trained HMMs. One example is shown in Figure 1(c). Sub-images of each facial feature are extracted based on the average of the width (w), height (h) and center points. For each feature of each person, we perform principle component analysis (PCA) on the extracted sub-images to obtain the approximated PCA subspace L. A feature can be modelled by $\{w, h, H, L\}$, where $H = \{h_1, h_2, ..., h_K\}$ is a set of HMMs that are responsible for segmenting images at different pose.



(a) (b) (c)

Figure 1. (left)Structure of an embedded HMM. (middle)Decoded state of an image. (right)Their maximal width and height of a sub-image for one feature.

4. Feature-Based Face Recognition

In our algorithm, the distance between an unknown person with a given image I and person p is the average distance between the features of I and the corresponding features of person p. For each feature f, we obtain K sub-images I_{fk} according to center points of facial regions segmented by K different HMMs, as well as width w and height h of feature f obtained in the training phase. We then compute the L2 distance, $d(I_{fk}, L_{fp})$, from I_{fk} to its projection on the subspace L_{fp} of feature f belonging to person p. Since in this approach a PCA subspace is trained for each feature of every person, we dub this approach as the **individual PCA** in the rest of this paper.

As our training set encompasses face images taken from different pose and each HMM is responsible for certain pose with limited variation, for each feature f the distance $d(I_{fk}, L_{fp})$ will be particularly small for one HMM. Hence we segment an observed image with the state sequence of the HMM with minimum average distance of all features. The distance between a given face I with the person p can be computed according to (1), where N is the number of features for person p.

$$D(I, p) = \min_k \frac{1}{N} \sum_f d(I_{fk}, L_{fp}). \quad (1)$$

Given this distance metric $D(I, p)$, the task of face recognition can be implemented straightforwardly. For face identification, the true identity is the person p with minimum distance $D(I, p)$. For face verification, a threshold is set based on the distance $D(I, p)$ with minimum recognition error.

5. Experiments and Results

We validate the proposed method with experiments using the CMU PIE [8] data set which consists of images of 68 people acquired at 13 different poses in the yaw (left to right) direction. These images are manually cropped and normalized to 64×64 pixels as some shown in Figure 2. For each person, 5 images (at pose {c, d, e, h, m}) are used for tests and the remaining 8 samples are used for training. There are 4 super states and 5 embedded states in each embedded HMM. Each observation consists of 8×8 image blocks and is represented by the first 10 2D-DCT coefficients. The observation windows overlap 5 pixels with each other in both horizontal and vertical directions.

To handle large pose variation, we train 3 embedded HMMs for each person. The training images are clustered automatically using the K-means algorithm. The method that trains multiple embedded HMMs for each person for recognition is referred as **E-HMM-based Extraction I**. As the K-means algorithm does not usually cluster images according to their pose perfectly, we also manually group training images to see whether better results can be achieved with ground truth clustering results. This method is referred as **E-HMM-based Extraction II**. In our experiments, the training samples of each person are clustered into the sets according to their pose such as {a, b, i}, {j, k} and {f, g, l} in Figure 2.

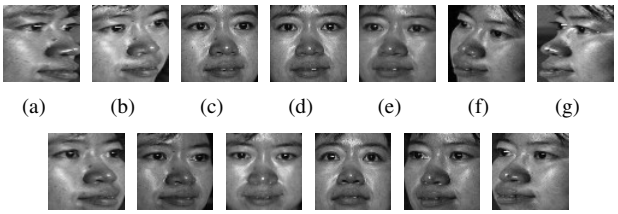
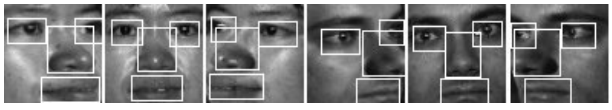


Figure 2. Cropped and normalized faces of one person in the CMU PIE database.

The baseline algorithm for our experimental comparisons is the holistic **Eigenface** method [9]. Instead of constructing one single PCA using images of all persons, we also experiment with the idea of individual PCA subspace for each person (referred as the **Individual Eigenface** method)). For comparisons, we also evaluate the **Embedded HMM** method [4] with the CMU PIE data set. For feature-based face recognition, 20 facial features (resulting from 4 super states and 5 embedded states for each super states) are extracted in our **E-HMM-based Extraction I** and **E-HMM-based Extraction II** methods for experiments. In the **Uniform Extraction** approach, 20 facial features are uniformly extracted from images as presented in Figure 3(a). In the **Manual Extraction** method, we meticulously manually crop 4 facial features (two eyes, nose and mouse) from each face image (so that the localization error is negligible) as shown in Figure 3(b). For all feature-based approaches, we use the distance metric (1) for experiments. The image size for the Eigenface methods is 64×64 pixels, and each sub-images of facial feature are normalized to 12×12 pixels for experiments.

(a) 20 uniformly extracted facial features.



(b) 4 manually extracted facial features.

Figure 3. Extracted facial features by two methods.

We carry out two face recognition tasks: identification and verification. The experiment results of face identification are shown in the second column of Table 1. It is evident that the accuracy is considerably improved by the use of multiple subspaces in the **Individual Eigenface** and **E-HMM-based I**, and **E-HMM-based II** methods with the corresponding distance metrics. Note that the proposed algorithm (where the features are automatically extracted) has almost the same identification rate with the one in which features are manually extracted without localization error.

Table 1. Experimental results in identification and verification.

	Identification	Verification	
	Accuracy(%)	Hit (%)	ERR(%)
Eigenface	80.59	58.18	11.28
Individual Eigenface	96.18	77.09	5.73
Embedded HMM	99.71	4.71	28.25
Uniform Extraction	92.94	76.40	6.74
Manual Extraction	100	84.83	5.02
E-HMM-based Extraction I	**98.53**	**86.47**	**4.35**
E-HMM-based Extraction II	**99.71**	**94.17**	**2.54**

The results with verification experiments are shown in the last two columns of Table 1. The third column is the hit rate when false alarm rate (FAR) equals 1%, and the fourth column is the equal error rate when false reject rate (FRR) equals FAR. It is clear that our algorithms outperform the other methods by large margins. The Receiver Operating Characteristic (ROC) curves of the above-mentioned methods are shown in Figure 4.

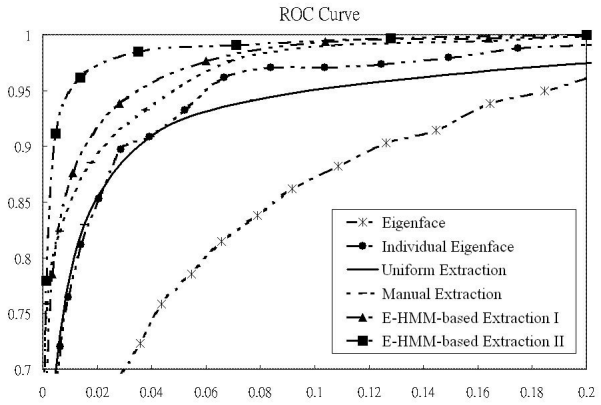


Figure 4. ROC curves of the evaluated algorithms in verification tests.

In both identification and verification experiments, the method using manually cropped features (**Manual Extraction**) yields better results than holistic **Eigenface** and **Individual Eigenface** methods, as well as the feature-based approach with features uniformly extracted (**Uniform Extraction**). Note that the identification is performed on a close set and with no imposter rejection mechanism. The identification task in this case is simpler and thus these methods have smaller error rates in identification than verification tasks. This is evident in the **Embedded HMM** method which has

good performance in identification tests but very poor results in verification experiments. Surprisingly the proposed algorithm (**E-HMM-based Extraction I**) whose facial features are extracted automatically, outperforms the method (**Manual Extraction**) with manually cropped facial features in verification tests. The main reason is that our method exploits the merits of the individual embedded HMMs. In our method test images of legitimate users are segmented consistently with the training face images via the trained HMMs, and thus the localization errors of extracted features are relatively small (Figure 5). On the other hand for any imposter, the localization errors of extracted features in a test images will be larger since these HMMs were trained specifically for each legitimate person in the database (Figure 5). The effect of large localization error is subsequently amplified when computing the distance of a test image to to a person p in (1), while the localization errors for both legitimate users and impostors are the same in the method with manually cropped features. In other words, each HMM in our method is tuned for each person and penalizes any impostors, thereby rendering better results in verification tasks. When the images are manually clustered according to their pose without error, the proposed approach (**E-HMM-based Extraction II**) achieves the best results. This suggests that more elaborated clustering algorithms other than K-means may be employed to further improve our **E-HMM-based Extraction I** method.

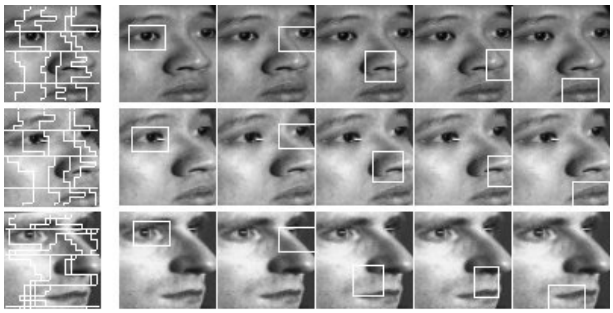


Figure 5. Some results of extracted facial features using our method. The first column shows the decoded states by our embedded HMMs. The extracted features of an image in training phase are presented in the first row. The extracted features obtained in test phase for a legitimate user and an imposter are shown on the second and third rows, respectively. Note that we show only 5 among 20 facial features.

6. Concluding Remarks and Future Work

We have proposed a novel algorithm for extracting facial features robustly with applications in identification and verification. The proposed method utilizes the decoded states from an embedded HMM to extract rectangular features under large pose variation. Individual subspace is constructed for each feature to account for appearance variation, and the associated distance metric helps in finding the best matched feature in the training set. We have demonstrated the merits of our algorithm in extracting facial features for identification and verification experiments, with comparisons to numerous methods. Owning to our person-specific feature extraction method which results in small localization errors for legitimate users and large ones for impostors, our algorithm outperforms other methods even when features are manually cropped with no localization errors.

The proposed method extracts rectangular features via the embedded HMMs without utilizing shape information. Our future work will incorporate the shape information of for feature extraction. We will also extend the person-specific feature extraction algorithm so that it can robustly extract facial features of any face image.

7. Acknowledgments

This work is supported in part under grants NSC-94-2213-E-002-027 and 95-EC-17-A-02-S1-032.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] S. Eickeler, S. Muller, and G. Rigoll. High performance face recognition using pseudo 2-d hidden markov models. *European Control Conference*, 1999.

[3] S. Kuo and O. Agazzi. Keyword spotting in poorly printed documents using pseudo 2-d hidden Markovmodels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):842–848, 1994.

[4] A. V. Nefian and M. Hayes. An embedded HMM-based approach for face detection and recognition. volume 6, pages 3553–3556, 1999.

[5] A. V. Nefian and M. H. Hayes. Hidden markov models for face recognition. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 2721–2724, 1998.

[6] F. Samaria. Face segmentation for identification using hidden Markov models. In *British Machine Vision Conference*, pages 399–408, 1993.

[7] F. Samaria. *Face recognition using hidden markov model*. PhD thesis, Engineering Department,University of Cambridge, October 1994.

[8] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[9] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[10] W. Y. Zhao, R. Chellappa, J. P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. Technical Report CAR-TR-948, Center for Automation Research, University of Maryland, 2000.