IEEE International Symposium
on
Signal Processing and Information Technology

December 12-15, 2012 - Ho Chi Minh City- Vietnam

ISSPIT 2012

The 11th IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2012, is a premiere technical forum for researchers in the fields of signal processing and information technology. ISSPIT 2012 will include state-of-the-art oral, poster sessions, and tutorials related to the key areas outlined below. Accepted papers will be published in the Proceedings of IEEE ISSPIT 2012. A contest for the Best Paper Award will be held and an award will be given.

The 12th IEEE International Symposium on
Signal Processing and Information Technology (ISSPIT 2012)
December 12-15, 2012 - Ho Chi Minh City- Vietnam

Program:

Wednesday, December 12
12:00pm – 6:00pm Registration (The Entrance Hall of A Building)
6:00pm – 9:00pm Reception **Room: 4th Floor, C Building**

Saturday, December 15

10:00am – 10:30am (Room: 1st corridor connecting A and B Building) Coffee Break	
10:30am – 12:15pm	
IT6: Object detection **Room: Meeting Room B (B Building)** **Session Chair: Đinh Dũng**	**SP5: Information processing** **Room: A104 (1st Floor, A Building)** **Session Chair: Trong-Tu Bui**
Automatic Extraction of Common Action Characteristics Tran Thang Thanh, Fan Chen, Bac Le, and Kazunori Kotani ***Real-time Video-based Fall Detection using Motion Gradients and Shape Features*** Huu-Hung Huynh, Trong-Nguyen Nguyen, and Jean Meunier ***Face recognition from video under uncontrolled lighting condition*** Thi-Thanh-Hai Tran ***Illumination Compensation for Face Tracking on Smart Phones*** Hoang-Nam Bui, Soo-Hyung Kim, and In-Seop Na	***Ordered-Subsets Acceleration of Radio Interferometric Calibration: OS-SAGE calibration algorithm*** Sanaz Kazemi, Sarod Yatawatta, and Saleem Zaroubi ***Non-negative Factorization of Low-Rank Matrices: Application to Spectral Time Series*** John Gruninger, Hoang Dothe, and Xuemin Jin ***A Case Study of Connect6 Game FPGA-based Implementation Using the Multi-turn Prediction Algorithm*** Trong-Thuc Hoang, Vi-Thuy Tran, Thanh Le, Minh-Triet Luu, Cao-Quyen Tran, Xuan-Thuan Nguyen, and Trong-Tu Bui ***3D Reconstruction using Kinect Sensor and Parallel Processing on GPU*** Thuong Le-Tien, Thai Phu Ho, and Viet Dai Tran
12:15pm – 1:45pm (Room: The Entrance Hall of A Building) Luncheon	
1:45pm – 3:30pm	
SP6: Communication Networks **Room: Meeting Room B (B Building)** **Session Chair: Thanh Huu Nguyen**	**SP7: Image processing** **Room: A104 (1st Floor, A Building)** **Session Chair: Thuong Tien Le**
Real-Time Communications for Distributed Control System Nguyen Trong Cac and Nguyen Van Khang ***Communications delay model for Real-Time Distributed Control System*** Nguyen Trong Cac and Nguyen Van Khang ***An Improved Image Retrieval Algorithm For JPEG2000 Compressed Images*** *Amani Chaker*, Mounir Kaaniche, and Amel Benazza-Benyahia	***On the Complexity Reduction of Airborne Ultrasonic Image Generation*** Dorel Aiordachioaie ***Fragile Fingerprinting using Permutation Code for Digital Image*** Ta Minh Thanh, Munetoshi Iwakiri, and Nguyen Trung Thanh ***Efficient Differential Prediction Error based Intra Mode Decision Algorithm for Video Coding*** Bharanitharan Karunanithi and Chen Bo Wei

Face Recognition from Video under Uncontrolled Lighting Condition

Thi-Thanh-Hai Tran
International Research Institute MICA
HUST – CNRS / UMI – 2954 – INP Grenoble
1 Dai Co Viet Hanoi Vietnam
thanh-hai.tran@mica.edu.vn

***Abstract*—This paper concerns the problem of face recognition from video under uncontrolled lighting condition. To face with illumination change, we propose to inspire the idea that pre-processes input images in order to represent them robust to illumination change [1]. We then use embedded Hidden Markov Model (EHMM), a famous model for face recognition to identify faces [2]. The main reason that we would like to study and experiment this model is that it allows us to represent structure of face images that makes more explicit face representation than numeric face descriptors. The traditional EHMM was applied to face recognition from still images. In our paper, we deal with face recognition from video. Therefore, we propose to combine the recognition result obtained from several frames to make our decision more confident. This improves significantly the recognition rate. We have trained our model and tested with two face databases, the one is the Yale-B database and the other one is created by MICA Institute.**

Keywords—illumination; EHMM; face recognition

I. Introduction

Face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. Actually, this topic remains still very active because of its usefulness in both military and civil applications (e.g. information security, access control, law enforcement, surveillance, image understanding) [3-5].

A general statement of the face recognition problem in computer vision can be formulated as follows: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces.

Although many successful results have been achieved, there are still a lot of open problems to be resolved, for example recognition from outdoor facial images, recognition from non-frontal facial images, recognition at low false accept/alarm rates, understanding why males are easier to recognize than females, integration of morphable models into face recognition performance.

The work presented in this paper deals with the problem of face recognition from video under difficult lighting condition. The main contributions of this paper are:

- We detect faces automatically from images using Haarlike features and Cascaded Adaboost classifier [6]. The face recognizer will be then applied only on face candidates that allows reducing the computational time.
- We propose to use the EHMM for face recognition. However, to deal with illumination changes, we normalize input images using a technique proposed recently in [1]. By this way, the face recognizer is more robust to lighting condition than the traditional method.
- As human often moves his/her head during image acquisition, we propose to make decision of recognition not only from a unique image but from several consecutive ones. This improves significantly the recognition rate. We call this technique majority vote.

II. Realated Works

Face recognition works have been surveyed in some papers [3-5, 7-9]. In general, a configuration of generic face recognition consists of 3 main phases: face detection, feature extraction and face recognition. Each phase is an issue that opens a lot of applications. In this paper, we are not ambitious to review all works in the domain. However, we will analyze two works that our method is based on [2], [1].

Hidden Markov Model (HMM) has been successfully used for speech and action recognition. In [2], the authors proposed an Embedded Hidden Markov Model (EHMM), which is a 2D HMM, to model face structure. This model has been shown to be more efficient than the traditional HMM or 2D pseudo HMM in face recognition. The recognition rate of the method reaches to 98% when training and testing with the Olivetti Research Ltd. Database (ORL). However, the EHMM works only under some following assumptions: i) faces need to be frontal; ii) lighting does not change during training and testing phase; iii) face regions have to be detected from image.

One of the most difficulties when doing face recognition is illumination change. In [1], Tan et Triggs have proposed a solution to solve this by pre-processing input images which consists of different steps: gamma correction, Difference of Gaussian (DoG) filtering and contrast equalization. This pre-processing is applied on the input image in order to render it insensitive to illumination. With this pre-processing phase, recognition rate has been improved significantly even under strong uncontrolled lighting. After pre-processing, the authors introduced Local Ternary Patterns (LTP) and Gabor filter for face description and a modified method based on Kernel Linear Discriminant Analysis (KLDA) for face recognition.

Although [1] has improved the method so that it is more robust to illumination change, the proposed method works only with still images. Tested images are images containing only face regions cropped from general images. In addition, this approach used implicit features (textured features) to represent faces. Face is a structured object (composed of eyes, mouth, and nose), so texture, color and shape cannot represent all information of the face. The approach based on EHMM allows representing object in a structured manner so it is more explicit and can deal with partial occluded faces. With these reasons, in this paper, we propose to deal with illumination change using the method [1] and apply EHMM for face recognition [2].

III. PROPOSED APPROACH

A. General framework for face recognition

We propose a framework for face recognition that composes of 2 main phases: (i) learning and (ii) recognition.

- **Learning:** This phase will take face samples, pre-process them and put into EHMM process to train an EHMM for each person to be recognized.
- **Recognition:** It consists of detecting face window candidate from a given image, pre-process it and find the EHMM that best matches. The recognition decision will be taken based on majority voting.

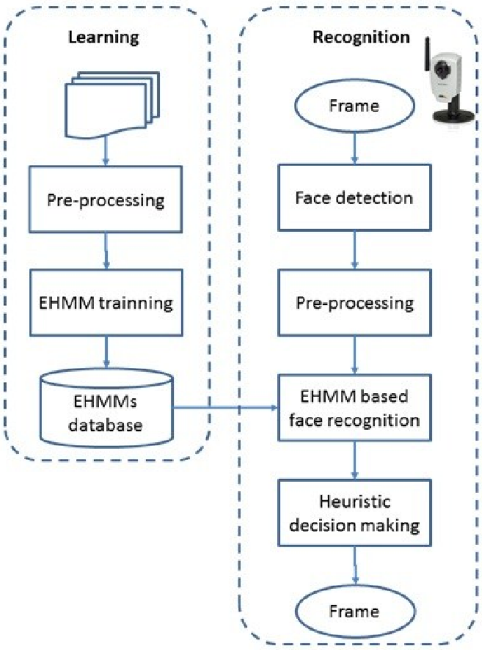

Figure 1. Proposed framework for face recognition from video.

As we can see in the Fig. 1, the pre-processing step is common in both training and recognition phases. In the following, we will describe in more detail each step in the framework of face recognition.

B. Pre-processing image - Illumination normalization

As we would like to work under uncontrolled lighting conditions, illumination normalization is necessary to make the face recognition robust to illumination change. In [1] the authors have proposed a method to handle it by applying a pre-processing chain before face recognition (Fig. 2).

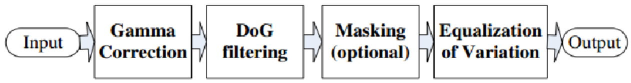

Figure 2. The stages of image pre-processing pipeline.

It incorporates a series of stages designed to counter the effects of illumination variations, local shadowing and highlights while preserving the essential elements of visual appearance. In the following, we briefly describe this pre-processing chain.

Gamma detection: The gamma correction allows enhancing the local dynamic range of the image in the dark or shadowed regions while compressing it in bright regions and at highlights.

Difference of Gaussian Filtering: As mentioned in [1], gamma correction cannot remove the influence of overall intensity gradients such as shading effects. The authors indicated that the use of DoG (Difference of Gaussian) can solve this. Therefore, a DoG filter will be applied to the image corrected by gamma correction step.

C. Face detection

Detecting positions at which faces appear in an image before recognizing them is a step that should be applied in order to reduce the recognition time. Face detection is a popular problem that has been solved successfully in the literature using Haar-like features for face representation and Cascaded Adaboost classifier [6].

In this paper, we use this method to detect face regions from images. The face detection consists of 2 phases: (i) learning and (ii) detecting. The learning phase takes face and non-face images as positive and negative examples to train the Cascaded Adaboost classifier. Given an input image, a popular sliding technique is used and all sliding windows will be classified into two classes: face and non-face using the trained Cascaded Adaboost classifier.

D. Face recognition based on Embedded Hidden Markov Model

For face recognition, there exist a lot of methods including geometric feature-based methods, template-based methods, and model-based methods [7-9]. Face is a particular object which has significant facial features like hair, forehead, eyes, nose and mouth. Furthermore, these features appear in a natural order. Therefore, a face can be modeled by one-dimensional HMM.

In [2], the authors have shown that EHMM does better than traditional HMM. Therefore, in this work, we propose to use EHMM for face recognition.

In the approach for face recognition using EHMM, each input image will be converted into gray scale image which is input of the EHMM based recognizer. In this paper, the input image is not the entire gray scale image, but only the face candidate detected from section III. C, which is pre-processed in section III. A.

Structure of an EHMM: The structure of an EHMM is explained in Fig. 3. The elements of an EHMM are:

- The number of super states N_0 and the set of super states $S_0 = \{S_{0,i}\}$, $1 \le i \le N_0$
- The initial super state distribution, $\mathbf{\Pi_0} = \{\pi_{0,i}\}$ where $\pi_{0,i}$ are probabilities of being in state *i* at time zero
- The super state transition probability matrix $\mathbf{A_0} = \{a_{0,ij}\}$ where $a_{0,ij}$ is the probability of transitioning from super state *i* to super state *j*

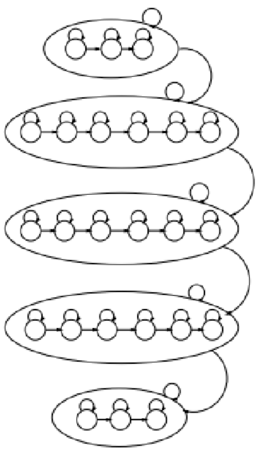

Figure 3. EHMM for face recognition.

- The parameters of the embedded HMMs include:

— The number of embedded states in the k^{th} super stage, $N_1^{(k)}$, and the set of embedded states, $S_1^{(k)} = \{S_{1,j}^{(k)}\}$

— The initial state distribution $\mathbf{\Pi_1^{(k)}} = \{\pi_{1,i}^{(k)}\}$ where $\pi_{1,i}^{(k)}$ are the probabilities of being in state i of super state k at time zero.

— The state transition probability matrix $\mathbf{A_1^{(k)}} = \{a_{1,jk}^{(k)}\}$ that species the probability of transitioning from state *k* to state *j*

— The state probability matrix $B^{(k)} = \{b_i^{(k)}(O_{t0,t1})\}$ for the set of observations where $O_{t0,t1}$ is the observation vector at row t_0 and column t_1. b_i has the following form:

$$b_i^{(k)}(\mathbf{O}_{t_0,t_1}) = \sum_{m=1}^{M} c_{im}^{(k)} N(\mathbf{O}_{t_0,t_1}, \mu_{im}^{(k)}, \mathbf{U}_{im}^{(k)})$$

where $1 \le i \le N_1^{(k)}$, $c_{im}^{(k)}$ is the mixture coefficient for the m^{th} mixture in the state *i* of the super state *k*. N is a Gaussian pdf.

So $\Lambda^{(k)} = \{\Pi_1^{(k)}, A_1^{(k)}, B^{(k)}\}$ is the set of parameters that define the super state k^{th}. An EHMM is defined as the triplet: $\lambda = (\Pi 0, A0, \Lambda)$.

Observation vectors: As proposed in [2], the observation vector is computed from each face window as follows. We slide a PxL window from left to right and top to bottom. The overlap between adjacent windows is M lines in the vertical direction and Q columns in the horizontal direction. For each PxL sub-window, we compute the 2D-DCT. The six coefficients of 2D-DCT will be used to compose the observation vector.

Training face model: To recognize someone, we need to learn the EHMM for this person. A set of images representing different instances of the same face will be used to train the EHMM. The input of this algorithm is a sequence of observation vectors. This sequence will be next segmented in N0 super states, then data corresponding to each super states are uniformly segmented from left to right into N1(k) states. The training processes uses Viterbi and K-Means to learn the EHMM.

Face recognition: Given a face candidate, we compute the sequence of observation vectors and this sequence will be input into a doubly embedded Viterbi recognizer. The model with the highest likelihood is selected and this model reveals the identity of the known face.

Heuristic decision making using majority vote: The face recognizer presented in the section 3.4.4 works with inputs as still images. In our context, we can capture a sequence of face images. Therefore, we propose a strategy of majority vote for making recognition decision as follows. We apply the recognizer on each frame in the sequence and count the number of the recognition results. The face having the biggest number of times recognized by the system will be revealed. This strategy allows improving significantly the recognition rate of the system than using only one image.

IV. Experimental Results

The experiments will be carried out to answer to the two following questions:

- Is the pre-processing pipeline proposed in [1] still good for EHMM based face recognizer?
- Is the proposed method good for online face recognition in real condition?

A. Dataset and evaluation measurement

We prepare two databases to evaluate our proposed method:

1) **Extended Yale-B database:** The Yale-B database is a de facto standard database for studies of recognition under variable lighting over the past decade. It contains 38 subjects at 64 illumination conditions. Half of the images were used in training, the other half were used for testing. Although there are only 38 people with a little of variability of expression, ageing, and this database is a challenging database because of extreme lighting condition. The **Fig. 4** illustrates some images in this database.

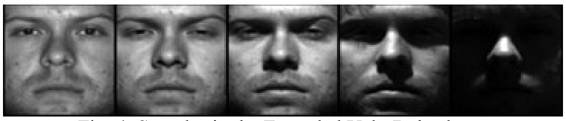

Fig. 4. Samples in the Extended Yale-B database.

2) **MICA database:** As the extended Yale-B database contains only face images that does not allow us to test our framework for face recognition from real condition (sequence of images taken under different lighting conditions), to evaluate the performance of the proposed method in case of video, we build a database of video as follows

- The number of people to be trained and recognized: 15 people aging from 20 to 30 years old. The people can bring or not glasses. The rotation of each person is about 40 degree.
- Condition of image acquisition: indoor under neon lighting, outdoor at different time of summer days (morning, mid-day, afternoon)
- The complex background (indoor office, outdoor)

The Fig. 5 illustrates some examples of the MICA database.

Figure 5. Illustration of the MICA database.

3) **Performance measurement:** The recognition rate will be computed to evaluate the performance of the proposed method. We will analyze this measure to show the robustness of the method with illumination change and the improvement of recognition rate from video w.r.t from still images.

B. Experimental results

To simplify the training and testing, we have developed a graphical interface with multiple options for face recognition: training from a database, recognition from still image, recognition from a video, recognition from camera. As presented above, we will answer to the two following questions:

Is the pre-processing pipeline proposed in [1] still good for EHMM based face recognizer?

As the method based on EHMM presented in [2] has been only evaluated with still images, we test then our framework also with still images taken from the Extended Yale-B database. In addition, as working with images, we do not apply the majority vote technique. The main difference between our method with [1] and [2] is: [2] uses EHMM for face recognition, it does not apply any illumination normalization on the input images; [1] does not use EHMM model but apply illumination normalization on the input images; our method uses EHMM in [2] while applying the illumination normalization on the input images proposed.

We found that with illumination normalization we obtain good results under uncontrolled lighting condition. We test and obtain 98.6% recognition rate for Yale-B database. EHMM has been tested with the ORL database and obtained 98% of recognition rate. But without normalization phase, it obtains only 56.7% in term of recognition rate with the Yale-B database. We can see a significant improvement of recognition rate when illumination normalization has been applied (Tab. 1). This performance is similar to the method [1].

TABLE I. PERFORMANCE COMPARISON OF TRADITIONAL EHMM AND EHMM WITH PRE-PROCESSING

Database	**Method in [2]** (EHMM without illumination normalization)	**Method in [1]** (Local Ternary Patterns (LTP) and Gabor filter for face description and a modified method based on Kernel Linear Discriminant Analysis (KLDA)	**Our method** (EHMM with illumination normalization)
Extended Yale - B	56.7%	98.6%	99%

The Fig. 6 shows an example in which without illumination normalization step, the method EHMM presented in [2] cannot recognize face in a hard lighting condition.

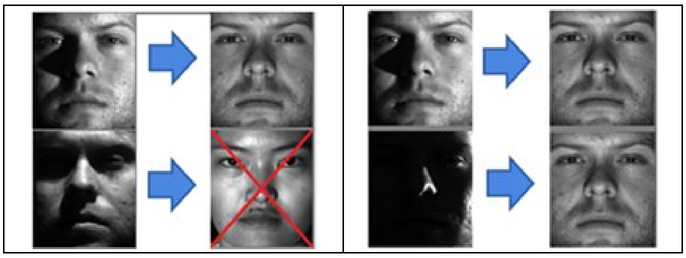

Figure 6. At the left, the method EHMM presented in [2] cannot recognize face in a hard lighting condition (in the second row). At the right, our method using EHMM with an illumination normalization step can recognize faces, even under very hard lighting condition (in the second row).

Is the proposed method good for online face recognition in real condition?

With the conclusion that our proposed method could obtain the similar performance w.r.t the method based on [1], we would like to verify if with majority voting technique, we can deal with face recognition from a sequence of images under uncontrolled lighting condition.

With MICA database, we have obtained 91% recognition rate when testing frame by frame (see some example in the Fig. 7). We have also recognized the same people online from webcam with answer after a certain time with these 15 people and obtain 99 % recognition rate (Table II).

TABLE II. PERFORMANCE COMPARISON OF PROPOSED METHOD WITH STILL IMAGES AND FROM VIDEO

Database	From still images	From Video (using majority voting)
MICA	91%	99%

V. CONCLUSIONS

In this paper, we proposed a method for face recognition which is robust to illumination change thank to a pre-processing phase. We showed that the pre-processing pipeline is still efficient for other model of face recognition as EHMM. In addition, to deal with moving face, we proposed a heuristic decision making by combining results after a certain time. This is reasonable in real-life when we want to use the system for access checking, as we cannot determine the best shot of the person for recognition, the use of several shots is suitable. In the future, we will validate the method with more databases and deploy it for automatic building access checking.

ACKNOWLEDGEMENTS

The research leading to this paper was supported by the National Project B2011-01-052 "*Study and develop an object detection and recognition system in smart and perceptive environments*" and the Nafosted project 102.01-2011.17 "*Some Advanced Statistical Learning Techniques for Computer Vision*". We would like to thank the projects and people involved in these projects.

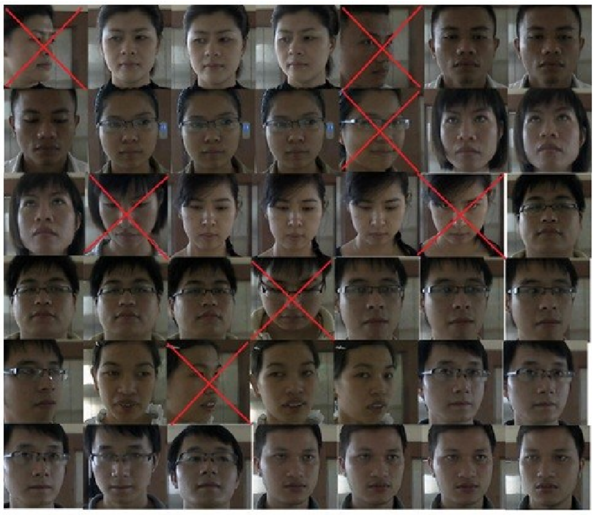

Figure 7. Frame by frame recognition results in outdoor environment of MICA database. As we trained the EHMMs only with frontal faces, the algorithm cannot recognize neither inclined nor declined faces.

REFERENCES

[1] X.Tan and B.Triggs, "Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions", IEEE Transactions on Image Processing, 2010. 19(6): p. 1635-1650.

[2] Ara V. Nefian and Monson H. Hayes III, "Face recognition using an embedded HMM", in IEEE Conference on Audio and Visual-based Person Authentication. 1999.

[3] A. F. Abate, et al., "2D and 3D face recognition: A survey", Pattern Recognition Letters, 2007. 28, : p. 1885–1906.

[4] W. Zhao, et al., "Face recognition: A literature survey", Journal ACM Computing Surveys 2003. 35(4): p. 399 - 458.

[5] X. Tan, et al., "Face recognition from a single image per person:A survey", Pattern Recognition, 2006. 39: p. 1725–1745

[6] P. Viola and M. J. Jones, "Robust Real-time Object Detection", International Journal of Computer Vision, 2001.

[7] J. Wright, et al., "Robust Face Recognition via Sparse Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence 2009. 31(2): p. 210- 227

[8] P.J. Phillips, et al., "Overview of the face recognition grand challenge", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR. 2005. p. 947- 954.

[9] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006. 28(12): p. 2037 - 2041.