Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Real-Time Surveillance System by Use of the Face Understanding Technologies

Yi-Tsung Chien, Yea-Shuan Huag, Sheng-Wen Jeng, Yao-Hong Tasi, and Hung-Xin Zhao

Advanced Technology Center, Computer & Communications Research Laboratories,
Industrial Technology Research Institute, Hsinchu, Taiwan 310, R.O.C.
TonyChien@itri.org.tw

Abstract. This paper introduced a human surveillance system which integrated the face understanding technologies to recognize personal identities in real time. We proposed a coarse-to-fine strategy to find out faces quickly and accurately. The global features of faces are used to reduce the search areas while the local ones are utilized to refine the faces' positions. Tracking and maintaining these faces all the time also help the detection on next frame. Finally, facial features are estimated and the results of before frames are combined to make the system more stable and robust to unexpected human activities. More than 60 persons of an on-line experiment were trained in the natural environment. Each of them was effectively recognized in terms of the system correctness and performance.

1 Introduction

Among many personal features, face is the most acceptable and intuitive one for human perception. Recently, algorithms proposed to search, tracking, and recognize faces are more and more robust and effective. This makes the face recognition systems trustworthy for high security demand applications, such as entrance access, video surveillance, etc. Besides, because digital cameras are gradually popular, identity authentication is applied to electronic devices to provide personalized control and entertainment, such as ATM, PDA, cell phone, etc. But integrating these face understanding technologies to reach a real-time and robust system is a challenge.

McKenna and Gong [1] described work aimed at performing face tracking, detection, and recognition in more unconstrained environments. Steffens et al. [2] proposed a system called PersonSpotter to perform face tracking, finding, and recognition in real-time. Clippingdale and Ito [3] also presented a prototype system which recognizes faces in video by estimating sizes, angles, and feature points. The system is computationally expensive and hard to approach a real-time performance. Cruz-Llanas et al. [4] aimed to analyze the performance of two different state-of-art automatic face recognition systems. Choi et al. [5] proposed a novel template matching technique using multiple mean faces to achieve robust face detection and recognition. Zhang et al. [6] presented a novel system which contains two modules: (1)

eye detection based upon a hybrid neural method and (2) face recognition using the dual eigenspaces method.

We have successfully integrated technologies of face tracking, face detection, and face recognition to authenticate personal identity in real time. This paper consists of five sections. The kernel technologies are described in Section 2. System integration and some experimental results are presented in Section 3 and 4 respectively. Finally, Section 5 gives our conclusions.

2 Kernel Technologies

2.1. Face Tracking

The face tracking technology [7] finds out Regions of Interest (ROIs), i.e. probable positions of humans and tracks them all the time. We use the Uniform Color System (UCS) to segment skin-color areas, and confirm them with motion information. After this process, static objects in the background with skin-like color are eliminated. Sometimes the rest areas are piecewise and involve non-face places of skin color, e.g. hands or clothes. We combined some of the areas to be the face candidates and verified them by some heuristic rules, such as shape, contour, and color distribution. Then the ROIs are compared with positions of former faces reserved in the face history. The overlay areas and intensities are correlated to understand that they are the new user or the known users. A result is shown in the Fig. 1.

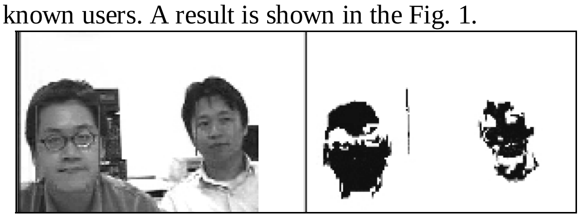


Fig. 1. A result of the face tracking algorithm (Black blocks are skin-color regions and red rectangles are combined ROIs.)

2.2. Eyes Detection

The process of extracting eye-analogues is composed of a sequence of morphological operations [8, 9]. But some of them are eyes-like patterns in image. Therefore, some symmetric properties, e.g. shape and orientation are applied to check the coupled eye-analogue. Finally, all features of the faces are measured and verified by a two classes neural network which is modeled by a sufficient amount of faces and non-faces images. As showed in the Fig. 2, the algorithm can handle face rotation and scale of size but may make false detection on eyebrows.

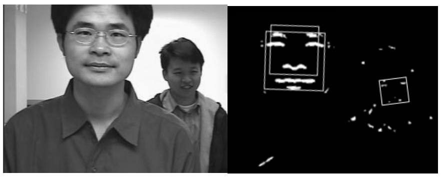


Fig. 2. A result of the eyes detection algorithm (The white rectangle are detected faces.)

2.3. Iris Detection

We utilized the deformable template matching (DTM) [10] to search the precise positions of the iris. The iris template is defined as the combination of one circle which represents the iris with three degree of freedoms (DOFs) and two parabolas which represent the eyelids with six DOFs respectively (see Fig. 3). We proposed some effective heuristic rules to reduce the deformable ranges and maximized two energy functions to locate the eyes' positions correctly. The first energy function summarizes the differences between inner and outer gray levels of the iris template. The second energy function accumulates values on the eyelids' line template.

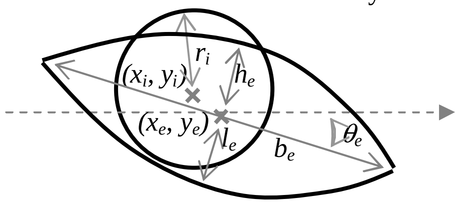


Fig. 3. Deformable template of the iris detection algorithm

2.4. Face Recognition

Our face recognition algorithm uses a RBF-based neural network [11]. Fig. 4 illustrates the structure of the network. Each enroller maintains his own subnet and can be recognized to certain one depending on likelihood responses of the network. We proposed an optimized training approach by reducing a classification-oriented error function [12]. Let x be the training data, and the error function E is given by:

$$E(x) = 1 - e^{(O_T - O_M)/A(n)} \tag{1}$$

$$O_M = \max_{1 \le k \le K} O_k$$

$$O_k = \sum_{j=1}^{k_n} \lambda_{kj} O_{kj}$$

$$O_{kj} = \exp(-\sum_{i=1}^{N} w_{kji}(x_i - C_{kji})^2)$$

where T denotes the genuine class of x, M denotes the output class having the maximum likelihood to x. $A(n)$ is positive and monotonically-decreasing value. According to the generalized delta rule, C_{kji} (kernel center), λ_{kj} (output weight), and w_{kji} (feature weight) can be updated as follows:

$$\nabla C_{kji} = -\alpha(n)\frac{\partial E}{\partial C_{kji}} \quad (2)$$
$$\nabla \lambda_{kj} = -\beta(n)\frac{\partial E}{\partial \lambda_{kj}}$$
$$\nabla w_{kji} = -\gamma(n)\frac{\partial E}{\partial w_{kji}}$$

where $\alpha(n)$, $\beta(n)$, and $\gamma(n)$ are three positive and monotonically-decreasing learning rates. These updating rules are further modified by allowing feature updating not only when an input sample x is classified incorrectly but also when x is classified correctly with low confidence.

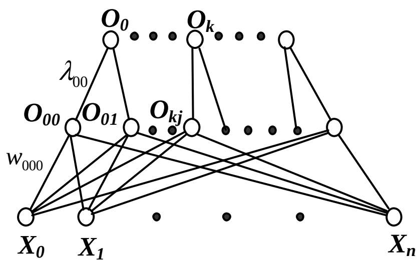


Fig. 4. RBF-based neural network

3 System Integration

We have developed a system with two phases: the authentication phase and the enrollment phase. They are illustrated in Fig. 5.

The authentication phase surveys the whole view and recognizes people identities. The skin color model locates the face candidates and the tracker maintains their movements continuously. This method reduces the error rate and speed-ups the execution of the face detection which depends on the sizes of the ROIs. The face detection step uses two strategies to enhance the abilities of search range and accuracy. The first strategy uses the eyes detection algorithm which watches profiles of eyes in the ROIs to locate approximate positions. The second strategy utilizes the iris detection algorithm which uses detailed features and refines the positions to obtain more precise ones. Giving the nearly eyes' locations from the strategy one also greatly reduces the computing time of the strategy two. To reach real-time calculation, the face locker picks up one person who stands closest to the camera at the initialization. Our system would only recognize this one's identity in the sequential frames unless he or she leaves away and then an initialization happens. Before doing the recognition, face normalization and color compensation [13] are utilized. We also design a robust method (see following section) to combine recognition results of before decisions to make sure of the personal identity.

The face enrollment phase collects facial features and trains the recognition model. Some criteria are proposed to estimate qualities of the faces and leave out the bad ones. These criteria include the face size, position, rotation angle, lighting condition, camera focus, and appearance variance.

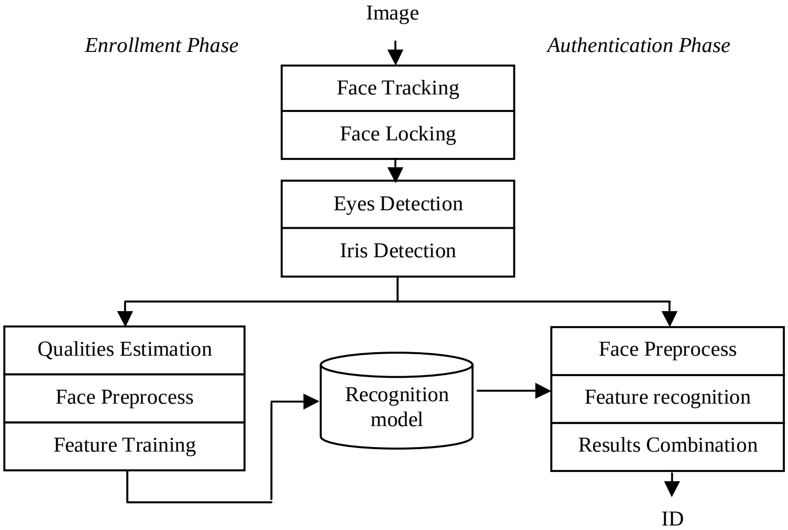


Fig. 5. Processes of the integrated system which includes the authentication phase and the enrollment phase (right flow and left flow respectively)

4. Experiments

4.1. Experiment I: Performance of Space Domain Integration

We combine some of the face understanding technologies as mentioned above to see the integration robustness and process time. Fig. 6 has an experimental result.

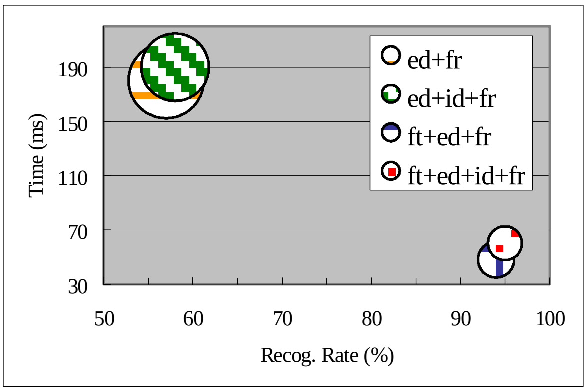


Fig. 6. The five technologies including face tracking (ft), eyes detection (ed), iris detection (id), and face recognition (fr) are combined in four ways to compare their performance. The experiment runs in a PC of 2.4 GHz CPU and 512MB RAM. Each of the 50 users sat naturally in the indoor environment and was taken 30-60 images of 320x240 pixels for training and 70-100 images for testing. (Centers of the circles are the results and radiuses mean time variations of the integrated methods.)

4.2. Experiment II: Improvement of Time Domain Integration

We compare four methods of results combination:

1 Max-Score method recognizes the person who has the maximum of recognition scores over an interval of time.

$$\underset{j}{Arg}\ \underset{i,j}{Max}\ p_j(x_i) \tag{3}$$

where $p_j(x_i)$ is the recognition score of an enrollee *j* at frame x_i.

2 Sum-Score method recognizes the enrollee who has the maximal summation of recognition scores over a number of frames.

$$\underset{j}{Arg}\ \underset{j}{Max}(\sum_{i} p_j(x_i)) \tag{4}$$

3 Wei-Score method recognizes the user who has the maximal weighed summation of recognition scores over an interval of time.

$$\underset{j}{Arg}\ \underset{j}{Max}(\sum_{i} (w_j(x_i)p_j(x_i))) \tag{5}$$
$$w_j(x_i) = p_{j^*}(x_i) - \underset{j \neq j^*}{Max}\ p_j(x_i)$$
$$p_{j^*}(x_i) = \underset{j}{Max}\ p_j(x_i)$$

4 Sum-Vote method recognizes the person who has most votes over a number of frames.

$$\underset{j}{Arg}\,\underset{j}{Max}\sum_i v_j(x_i) \qquad (6)$$

$$v_j(x_i) = \begin{cases} 0 & if\ p_j(x_i) > threshold \\ 1 & otherwise \end{cases}$$

where $v_j(x_i)$ is the vote of an enrollee *j* at frames x_i.

An experimental result is shows in Fig. 7. It shows that for any frame length, Wei-Score method has the best recognition rate than other methods. That's because the result of our recognition model is more correct if the difference between the maximal score and the next ones is larger. Fig. 7 also shows that accumulating results of more frames can always increase the recognition rate. But it makes system paying more memories, calculations, and waiting time when system initializes.

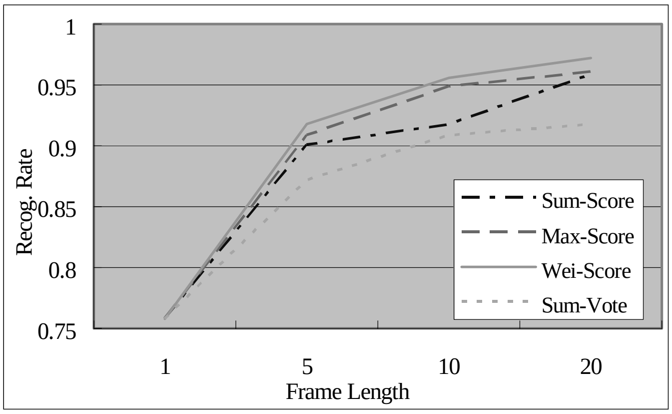
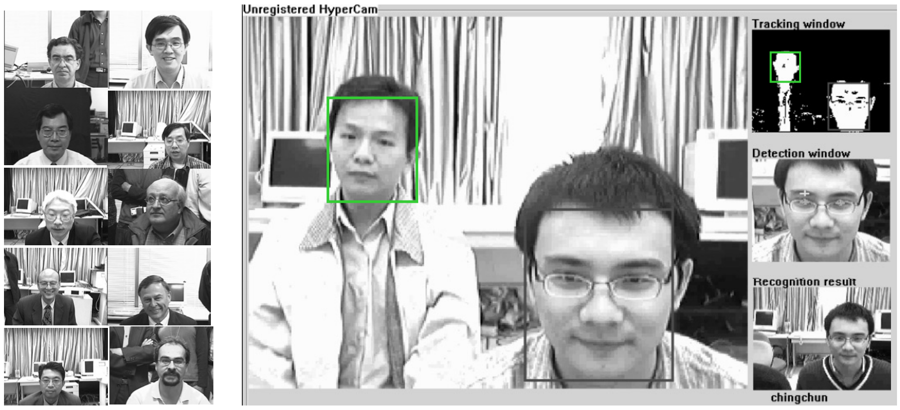


Fig. 7. An experiment compares four methods of results combination. There are 50 enrollees who have natural expressions and poses in the laboratory. We fist take 32 images for training and then another 100 images for testing. Note that increasing frame length would decrease decision times.

4.3. Experiment III: Real Case

An on-line experiment runs in a PC with 866MHz CPU and 256MB RAM. Each of 60 persons enrolls 7-12 faces for training and is recognized directly. The system can process about 8-12 frames per second in image size of 320x240 pixels. People are recognized correctly under uniform variation of lighting condition, facial expression, and rotation of head on image plane. Fig. 8 (a) presents parts of the enrolled images and Fig. 8 (b) shows an example of the experimental results.



(a) (b)

Fig. 8. (a) shows some enrolled images of real-case experiment. (b) demonstrates a example of the integrated system. The left image presents the surveyed person by a red rectangle (the front one) and the others by green rectangles. The right top image depicts the ROIs as white regions. On the right middle image, rough position of the eyes is detected as yellow crosses and is refined to the position of cyan circles. The right bottom shows the enrolled image and name of the person.

5. Conclusions

We have proposed a real-time surveillance system which tracks, detects, and recognizes human faces in natural indoor environment. The system uses a coarse-to-fine strategy to locate faces' positions quickly and precisely. The color segmentation, eyes detection, and iris detection are utilized sequentially. Each of them works at the nearby regions which are obtained from forward process. These faces are also tracked continuously to reduce the searching regions on next frame. Finally, facial features are extracted by the relation to the eyes' positions and the identities of the persons are recognized by referencing before results. Our experiments show that the system works quite stable and robust to tolerate variations of lighting, head size, facial expression, and rotation on image plane.

Acknowledgements

This work is a partial result of the project number A311XS1A60 which conducted by ITRI under sponsorship of the Ministry of Economic Affairs, R.O.C.

References

1. McKenna, S. J., Gong, S.: Non-intrusive Person Authentication for Access Control by Visual Tracking and Face Recognition. Int. Conf. on Audio-and Video-based Biometric Person Authentication (1997) 177-184
2. Steffens, J., Elagin, E., Neven, H.: PersonSpotter-fast and Robust System for Human Detection, Tracking and Recognition. Int. Conf. of IEEE on Automatic Face and Gesture Recognition (1998) 516-521
3. Clippingdale, S., Ito, T.: A Unified Approach to Video Face Detection, Tracking and Recognition. Int. Conf. on Image Processing (1999) 662-666
4. Cruz-Llanas, S., Ortega-Garcia, J., Martinez-Torrico, E., Gonzalez-Rodriguez, J.: Comparison of Feature Extraction Techniques in Automatic Face Recognition Systems for Security Applications. IEEE Int. Carnahan Conf. on Security Technology (2000) 40-46
5. Choi, J., Lee, S., Lee, C., Yi, J.: A Real-time Face Recognition System Using Multiple Mean Faces and Dual Mode Fiserfaces. IEEE Int. Symposium on Industrial Electronics (2001) 1686-1689
6. Zhang, D., Hui, P., Jie, Z., Pal, S. K.: A Novel Face Recognition System Using Hybrid Neural and Dual Eigenspaces Methods. IEEE Transactions on Systems, Man and Cybernetics (2002) 787-793
7. Zhao, H. X., Huang, Y. S.: Real-Time Multiple-Person Tracking System. Int. Conf. on Pattern Recognition (2002) 897-900
8. Han, C. C., Liao, H. Y., Yu, G. J., Chen, L.H.: Fast Face Detection via Morphology-based Pre-processing. Int. Conf. on Image Analysis and Processing (1997)
9. Chow G., Li, X.: Towards a System for Automatic Facial Feature. Pattern Recognition (1993) 1739-1755
10. Jeng, S. W.: Speed up Shape Extraction of Eyes and Mouth in a Face by Pre-locating the Facial Feature Points. Conf. on Computer Vision Graphics and Image Processing, Taiwan R.O.C. (2002) 74-80
11. Lampariello, F., Sciandrone, M.: Efficient Training of RBF Neural Networks for Pattern Recognition. IEEE Trans. On Neural Networks (2001) 1235-1242
12. Huang, Y. S., Tsai, Y. H.: An RBF-Based Pattern Recognition Method by Competitively Reducing Classification-Oriented Error. Int. Conf. on Pattern Recognition (2002) 180-183
13. Huang, Y. S., Tsai, Y. H.: Robust Face Recognition with Light Compensation. IEEE Pacific-Rim Conf. on Multimedia (2001) 237-244