Who are you? – A Wearable Face Recognition System to Support Human Memory

Yuzuko Utsumi, Yuya Kato, Kai Kunze, Masakazu Iwamura, Koichi Kise
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Naka, Sakai Osaka 599-8531, Japan
{yuzuko,masa,kise}@cs.osakafu-u.ac.jp , {katou,kai}@m.cs.osakafu-u.ac.jp

ABSTRACT

Have you ever experienced that you cannot remember the name of a person you meet again? To circumvent such an awkward situation, it would be great if you had had a system that tells you the name of the person in secret. In this paper, we propose a wearable system of real-time face recognition to support human memory. The contributions of our work are summarized as follows: (1) We discuss the design and implementation details of a wearable system capable of augmenting human memory by vision-based realtime face recognition. (2) We propose a 2 step recognition approach from coarse-to-fine grain to boost the execution time towards the social acceptable limit of 900 [ms]. (3) In experiments, we evaluate the computational time and recognition rate. As results, the proposed system could recognize a face in 238 ms with the the cumulative recognition rate at the 10th rank was 93.3 %. Computational time with the coarse-to-fine search was 668 ms less than that without coarse-to-fine search and the results showed that the proposed system has enough ability to recognize faces in real time.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

Keywords

Face recognition, coarse-to-fine recognition, wearable system

1. INTRODUCTION

Have you ever experienced that you cannot remember the name of a person you meet again? In such a situation, didn't you think how good it would have been if you had had a system that tells you the name of the person in secret? We propose a wearable system of real-time face recognition, that can help you in such situations and can keep a log about people you interacted with during the day to augment your memory and help you remember better. Multiple of unobtrusive systems augmenting human memory have been proposed [2], yet to our knowledge nobody implemented and experimented with a system doing real time face recognition on a portable display. What would be the requirements of the system? To circumvent awkward situations, first, the system should recognize faces unnoticed by bystanders or the person you are meeting. Second, we need high recognition accuracy, as using the wrong name might be fatal. The third one is computational time. According to Thrope et. al., it takes an average of 445 [ms] for a person to recognize and react to a complex visual scene [8]. This is an important limit. To enable seamless social interaction our system should perform the face recognition task at least in 2 x 445 [ms] (the minimum time it could take before a person recognizes that his/her name was not remembered due to visual clues). Additionally, the system should work for all contacts the user might encounter at least during one day. According to Cross et. al., we interact with around 10-12 people daily (numbers based on people working and living in the United States [3]). Unfortunately, we could not find any information on how much these numbers vary and how difficult it is to predict which people an average user might encounter during the day.

The first requirement could be satisfied if you always wear such a device and the recognition result can be reported only to you. Our solution is a head mounted display, which can let the user see the candidates of the person (face recognition results). The second and third requirements are closely related to each other and there are two strategies to satisfy them. One is to employ a fast classifier. An advantage of this strategy is the ease of implementation since it is already fast enough. An example is a wearable face recognition system to assist a visually impaired person employing the Eigenface and the LDA on a low-spec hardware[5]. The main drawback of the strategy is that fast classifiers are generally not accurate enough, especially with a large database containing many face images. Thus satisfying the second requirement with a large database is not easy for a fast classifier. The other is to employ an accurate classifier. Although its discriminant ability is satisfactory, this kind of classifiers are generally slow and their computational time is often proportional to the number of candidates. Therefore, in order to use such a classifier for a real-time system, a reduction in computational time is unavoidable. In the current paper, we take the latter strategy. We employ Affine hull based face recognition method[1] shown to be highly accurate in the literature. The secret to reduce the computational time

AH'13, March 07 - 08 2013, Stuttgart, Germany

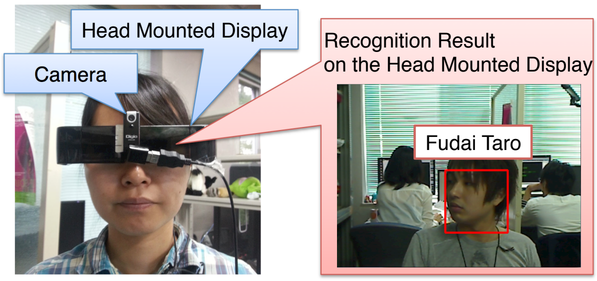

Figure 1: The proposed wearable face recognition system

is the coarse-to-fine approach. We propose a coarse-to-fine method suitable for the classifier, which preserves the accuracy with less computational time.

The contributions of our work are summarized as follows:

1. We discuss the design and implementation details of a wearable system capable of augmenting human memory by vision-based realtime face recognition.
2. We propose a 2 step recognition approach from coarse-to-fine grain to boost the execution time towards the social acceptable limit of 900 [ms].
3. In experiments, we evaluate the computational time and recognition rate. As results, the proposed system could recognize a face in 238 ms with the the cumulative recognition rate at the 10th rank was 93.3 %. Computational time with the coarse-to-fine search was 668 ms less than that without coarse-to-fine search and the results showed that the proposed system has enough ability to recognize faces in real time.

2. SYSTEM OVERVIEW

A proposed wearable face recognition system consists of a camera, a head mounted display (HMD) and a computer, shown in Fig. 1. The camera is LOAS MCM-14 with a resolution of 1280 × 1024 pixels and rate of 30 frames per second(fps). The HMD is EPSON BT-100 and the resolution of the display is 960 × 540. We mount the camera on the HMD so it captures a scene close to the user's field of view. The captured video is displayed on the HMD and sent to the computer. Face detection, tracking and recognition are performed on the video and recognition results are superimposed on the video displayed on the HMD.

The procedure of the proposed face recognition system is shown in Fig. 2. First, the system detects faces from the captured video. Subsequently, the system tracks the detected face, as our face recognition method requires multiple images for recognition. Thus, when the system collects a large enough number of face images, face recognition is applied to the collected face images. A recognition result is sent to the HMD and shown to the user. In the current implementation, face detection is executed when the face tracking fails and the face recognition is executed when a certain number of face images are collected. Details follow.

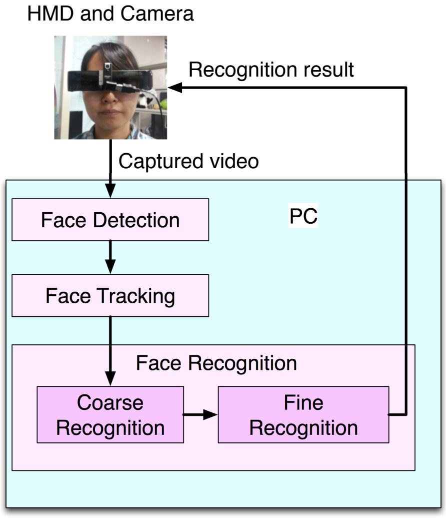

Figure 2: Procedure of the proposed face recognition system

In order to execute face recognition in real time, we need fast face detection. In the proposed system, we employ the cascade face detector using the Adaboost algorithm by Viola et al. [9]. We use the fast face tracking method[4] to trail the detected faces, which can track a face with up to 100 fps. Tracking results contain not only frontal faces but also slightly turned faces. After tracking faces, the face regions are cropped from raw images. The system recognizes a face with the cropped images. A feature vector is created directly from pixel values of a cropped face image. For example, when the resolution of the cropped face image is 20x20, the dimensionality of the feature vector is 400.

Recognition is divided into two parts; coarse recognition and fine recognition. In the coarse recognition, the system uses the average of feature vectors. The system calculates the average of feature vectors for each person in the database. As soon as there are enough images recorded for recognition, the system calculates the average over them. The system matches the database and the collected images by comparing the two averages of feature vectors. Then the top N of matching results in the coarse recognition are used for fine face recognition. After the coarse recognition, the Affine hull based face recognition method is applied to the top N candidates selected in the coarse recognition process in the fine recognition [1]. Finally the top M recognition results in the fine recognition are displayed on the HMD to let the user see the candidates of the person.

3. SYSTEM EVALUATION

3.1 Evaluation environments

In this section, we show the face recognition performance of the proposed system. We made a face database for evaluating the proposed system. The face database consists of the VidTIMIT Audio-Video Dataset [7], the Honda/UCSD Video Database [6] and a database from people working at our department. We merged those databases in order to increase the number of faces for testing. We used 42 people from the VidTIMIT database consisting of color videos of 512 × 384 pixel, and 20 from the Honda/UCSD Video Database consisting of gray videos of 640 × 480. Our database consists of color videos from 14 people taken by the same camera as we later use for recognition. The total number of the people in the merged dataset is 76.

We applied face tracking to these videos and cropping face regions from videos. The size of face images were normalized to 20 × 20 [pixel] and we converted these images into 8 bits gray scale images. Some cropped exampled images from our dataset are shown in Fig. 3. 8 bits gray scale pixel values are used as feature vectors.

Figure 3: Examples of database images

We stored 200 images for each person. A query sequence consists of 16 sequential images obtained by tracking. 16 feature vectors are calculated for the 16 sequential images. Then, they are used for recognition as a query sequence. The number of people for queries is also 76. The PC used for evaluation had core i7 2600K 3.40GHz CPU and 8GB RAM. The OS of the PC was Windows 7 Enterprise 64bits.

3.2 Coarse-to-fine face recognition

In order to evaluate the efficiency of face recognition, we execute recognition experiments. First of all, we evaluate the coarse recognition. We just executed the coarse recognition using the average feature vectors and calculate the top N cumulative recognition rates. The recognition rates are shown in Fig. 4. The horizontal axis shows the rank and the vertical axis shows the cumulative recognition rates. The graph shows that the cumulative face recognition rate at at the 40th rank was 100%. Next, we evaluated the recognition results and the computational time of the proposed system. We changed the number of people that is chosen in coarse recognition, and calculated the recognition rates and computational time. Recognition rates and computational time are shown in Figs. 5 and 6. In both graphs, horizontal axes show the number of people chosen in coarse recognition and vertical axes show the rank-1 recognition rates and computational time for recognition. >From Fig. 5, recognition rates went up when the the number of people chosen in coarse recognition was between 10 and 20. When the number of people was more than 20, recognition rates were uniform. This showed that the system can recognize faces in the same accuracy as the recognition without the coarse-to-fine recognition when the number of people chosen in the coarse recognition is 20. Figure 6 shows that computational time went up when the number of people chosen in the coarse recog-

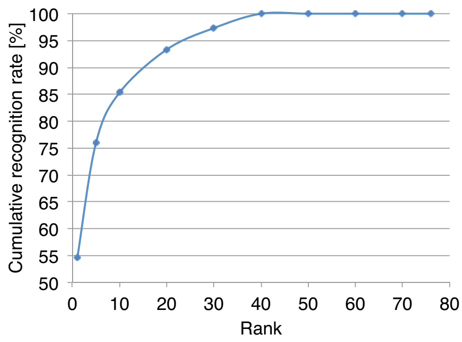

Figure 4: Cumulative recognition rates for the coarse face recognition

nition. The computational time was linear to the number of people chosen in coarse recognition. The computational time when the number of people chosen in coarse recognition was 20 was 238 ms. The computational time without coarse-to-fine recognition was 906 ms. This showed that the proposed method can recognize faces in 668 ms less than the method without coarse-to-fine recognition with keeping the recognition rate. Figure 7 shows the cumulative recognition rates of the system when the top 20 candidates were selected in the coarse recognition, which shows that the proposed system can show 10 candidates of the person with 93.3% of accuracy.

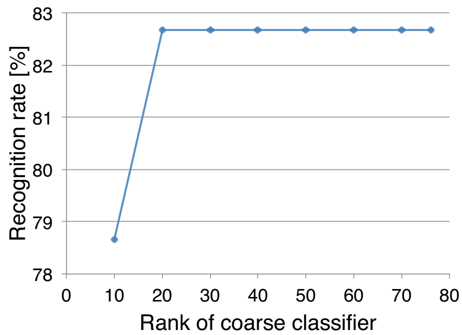

Figure 5: Recognition rates of the coarse-to-fine recognition method

3.3 Overall computational time of the proposed system

We also evaluate computational time for detection and tracking and confirm that the proposed method can recognize a face in real time. We used 43 videos from the VidTIMIT Audio-Video Dataset and 20 videos from the Honda/UCSD Video Database. The detection and tracking were executed with these videos, and calculated the average time for detection and tracking. The average time for detecting a face was 119 ms and the average time for tracking a face on a image was 8 ms. Capturing a image took 33 ms (30 fps). Thus it took 33 × 16 = 528 ms to collect 16 face images and recognize faces 238 ms at fastest with high recognition accuracy. If the system execute face recognition, it took

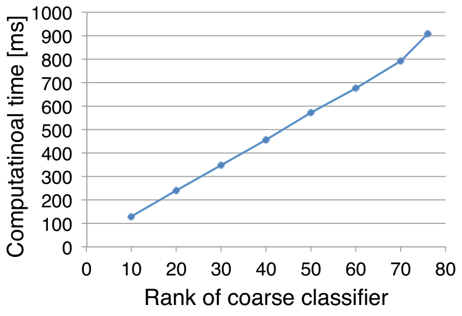

Figure 6: Total computational time of the coarse and fine recognition processes

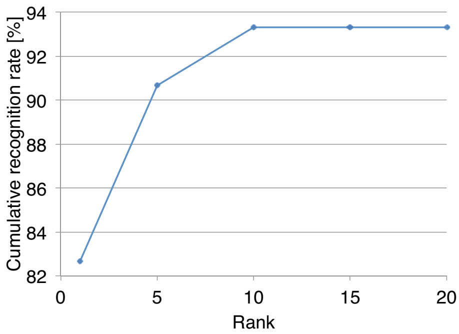

Figure 7: Cumulative recognition rates for the coarse-to-fine face recognition

$119 + 528 + 238 = 885$ ms to recognize a face. This show that the proposed system is enough to achieve real-time face recognition.

4. CONCLUSION

This paper presented a fast wearable face recognition system by using coarse-to-fine recognition. In experimental results, we showed the system could perform well regarding the processing speed given by the human visual perception [8] with a speed of 668 ms and the accuracy was 93.3% for a dataset with 10 people. Thus, the proposed system is effective.

For future work we want to make our system more mobile. For the tests we used a desktop PC, yet we are planning to implement it on a mobile phone. This gives us additional performance problems, that could be partially solved by a client-server architecture (introducing some additional delay). How feasible this is has to be tested.

Additionally, in current implementation, we just detect and track only one face. It is not useful when the user meet more than two people. The system should detect faces while the system tracks a face. The last one is the accuracy of face recognition. In this paper, the the cumulative recognition rate at the 10th rank was 93.3 %. We need to improve the recognition rate to put the proposed system to practical use. A strategy is using additional information. For example, we can use social network information for recognition. When you meet an acquaintance in working time, the probability that you meet the person's colleagues would increase, and when you meet a friend at a university, the probability that you meet university friends and professors would increase. If the system gets social network information from social networking sight like Facebook and applies the information to the face recognition, the recognition rate will increase.

5. REFERENCES

[1] Hakan Cevikalp and Bill Triggs. Face recognition based on image sets. In *Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2567–2573, 2010.

[2] Vicka R Corey. The memory glasses: subliminal vs. overt memory support with imperfect information. In *Proceedings of the Seventh IEEE International Symposium on Wearable Computers (ISWC 03)*, volume 1530, pages 17–00.

[3] Robert L Cross and Andrew Parker. *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business Press, 2004.

[4] Joao F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *Proceedings of European Conference on Computer Vision*, pages 702–715, 2012.

[5] Sreekar Krishna, Greg Little, John Black, and Sethuraman Panchanathan. A wearable face recognition system for individuals with visual impairments. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pages 106–113, 2005.

[6] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang, and David Kriegman. Visual tracking and recognition using probabilistic appearance manifolds. In *Computer Vision and Image Understanding*, volume 99, pages 303—331, 2005.

[7] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. In *Lecture Notes in Computer Science (LNCS)*, volume 5558, pages 199–208, 2009.

[8] Simon Thorpe, Denis Fize, Catherine Marlot, et al. Speed of processing in the human visual system. *nature*, 381(6582):520–522, 1996.

[9] Paul Viola and Michel J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, May 2004.