Integration of face and hand gesture recognition

Yung-Wei Kao[1], Hui-Zhen Gu[1], and Shyan-Ming Yuan[1,2]
[1] *Department of Computer Science and Engineering*
National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan
[2] *Department of Computer Science*
Asia University, Lioufeng Rd., Wufeng , Taichung County, Taiwan
{ywkao, hcku}@cs.nctu.edu.tw, smyuan@cis.nctu.edu.tw

Abstract

Face recognition and hand gesture recognition technologies have been developed separately for many years. Usually they are treated as independent systems. In this paper, we integrate the face and hand gesture recognition. We claim that the face recognition rate can be improved by hand gesture recognition. Also, we propose a security elevator scenario. Finally, we simulate this security elevator scenario by PCA method, based on the ORL database, and show that the face recognition rate and overall accuracy is improved after integration. We believe that this is a general method to integrate two recognition engines, not only for face and hand gesture recognition.

1. Introduction

Face recognition and hand gesture recognition technologies have been developed a lot separately for many years [1][2][3]. Also, the face recognition has been used to be the authentication mechanism for security surveillance system [4]. Although there are many researches adopted these two recognition systems into some applications, such as robot application [5], in these researches, the face and hand gesture recognition system are regarded as independent functions. In this paper, we discuss the benefit to integrate two recognition systems.

For face or hand gesture recognition, there are a lot of technologies can be adopted as the recognition engine, such as PCA (Principle Component Analysis) [6], HMM (Hidden Markov Model) [7], AdaBoost [8], ANN (Artificial Neural Network) [9], etc. Although the recognition technologies are different, most of them follow the basic concept of pattern recognition, that is, to find several linear/nonlinear lines to distinguish the testing/training data into several clusters. The number of clusters depends on the number of patterns defined in the application. Usually, the less clusters cause higher recognition rate. In this paper, we claim that if the number of clusters can be dynamically reduced, the overall recognition rate can be enhanced. We also implement a simulation to show that the result of hand gesture recognition is able to eliminate the number of face cluster, and enhance the recognition rate of face recognition.

This paper is organized as follows: chapter 2 describes the system overview of the security elevator scenario. Chapter 3 reviews the PCA method. Chapter 4 shows the system model. Chapter 5 illustrates the evaluation of our simulation program. Finally, the conclusion is drawn.

2. System Overview

The basic pattern recognition technique is to derive several liner/nonlinear lines to separate the feature space into multiple clusters. For example, Figure 1 has nine clusters, from C1 to C9. Assume that these nine clusters belong to different classes.

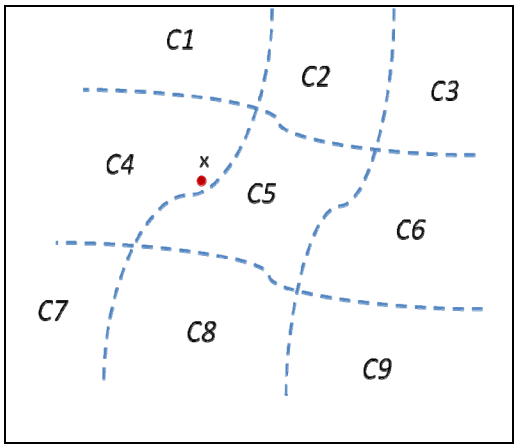

Figure 1. Nine clusters in feature space

If point x belongs to C5 in Figure 1 actually, it will be recognized to be in C4 incorrectly. However, if we

978-0-7695-3407-7/08 $25.00 © 2008 IEEE
DOI 10.1109/ICCIT.2008.74

dynamically reduce some impossible clusters by

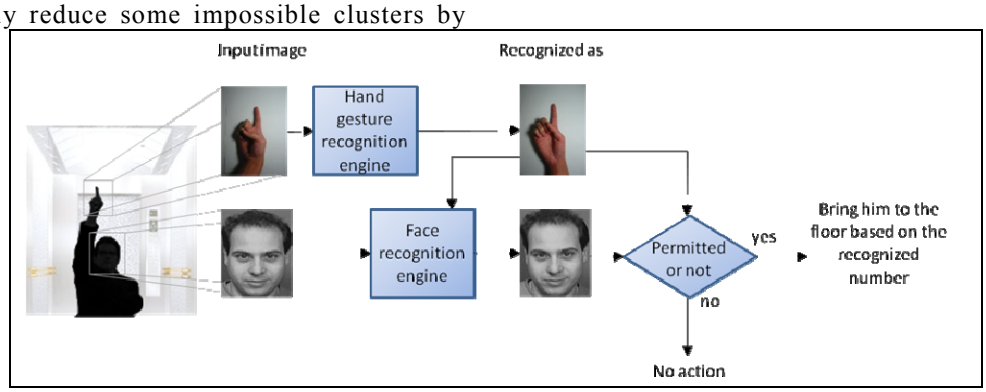

Figure 3. System overview

additional information, the result may be different. For example, in Figure 2, there are five clusters remaining.

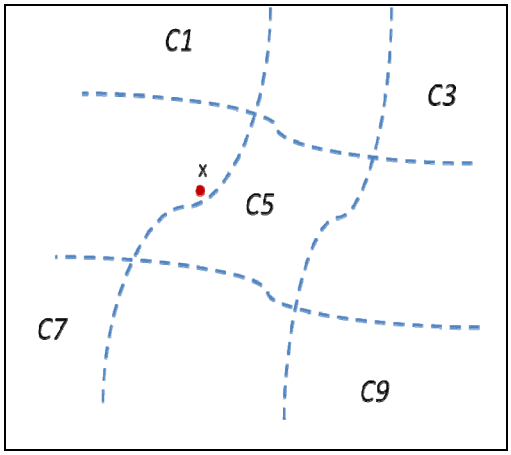

Figure 2. Five clusters in feature space

In this case, although x belongs to no cluster, the distance between x and C5 is the smallest, so it may be correctly clustered into C5. This example provides the intuition of how the recognition rate can be improved.

Let's imagine a simple scenario that there is a security elevator. In this elevator, there is no floor button for pressing. The decision for bringing someone to some floor is taken depends on his face and his hand gesture. The hand gesture indicates the floor he want to reach, and his face is used to decide that is this person permitted to reach the floor indicated or not based on his hand gesture. For such a security system, the recognition rate, especially for face recognition, is very important to perform the security control.

Now, we focus on this security elevator scenario mentioned, and develop a simulation system to evaluate the performance of face and hand gesture recognition integration. The system overview is shown in Figure 3.

In this system, the input hand gesture and face images are extracted first. After that, these two images are processed by the hand gesture and face recognition engine partially simultaneously. After the hand gesture recognition result is produced, the face recognition engine eliminates the impossible candidates based on the recognized hand gesture dynamically, and figure out which is the recognized person. Finally, we check that is this person permitted to reach the floor indicated or not based on his hand gesture. If he is permitted, then the elevator will bring him to this floor, otherwise, the elevator takes no action.

We assume that the security elevator can reach from the 1st floor to the 9th floor. Hence, we defined nine hand gesture symbols, from one to nine, based on the American Sign Language [10]. The hand gesture symbols are listed in table 1. On the other hand, the permitted floor number for each person is based on the face and hand gesture mapping. For example, table 2 is the mapping we generate randomly. Here, we assume that the 1st floor is available for everyone.

Table 1. Nine hand gesture patterns

1	2	3	4	5
				
6	7	8	9	
				

Table2. Face and hand gestures mapping example

Face	Available Hand Gestures

The integration point takes place at the final stage of face recognition engine. The integration process is shown in figure 4. First of all, we take the recognized hand gesture as input, and check the available hand gesture mapping. If this hand gesture is not available for any candidate, then this candidate will be eliminated before the recognition. After that, the remaining processes of traditional face recognition are conducted.

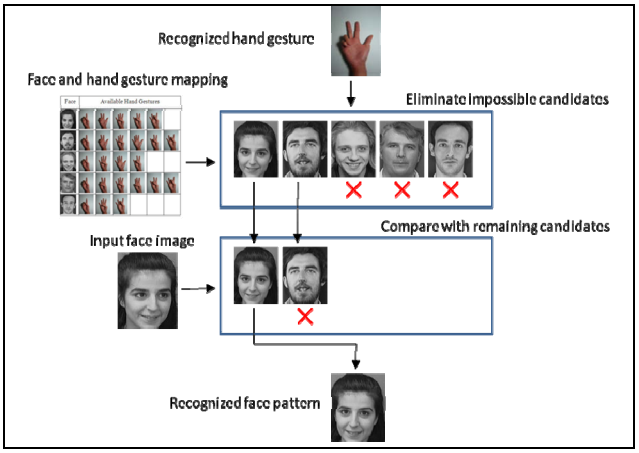

Figure 4. Integration process

Take Figure 4 for example, without the elimination, the average correctly guess rate is 1/5. On the other hand, with the elimination method, the average correctly guess rate is 1/2, which is 2.5 times higher than the previous one. This is the main idea that how the face recognition rate can be improved.

However, there are still risks of this elimination method. For example, in Figure 5, if we incorrectly eliminate the expected face candidate, the result of the face recognition must be wrong. This kind of error can be resulted from the wrong recognition result of hand gesture recognition engine, or the hand gesture is in fact unavailable for this person. Hence, the first recognition method chosen to conduct the elimination is very important. We choose hand gesture recognition to be the first recognition method, because it is much easier to be a high recognition rate method.

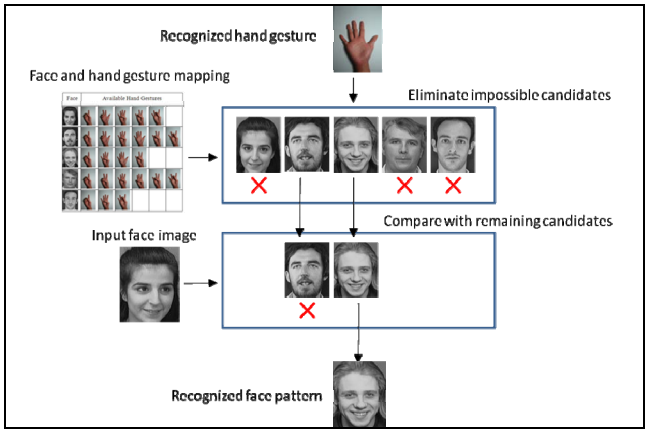

Figure 5. Integration process with error

The face recognition accuracy is analyzed in table 4. We discuss the accuracy of face recognition after integration if the expected face candidate is eliminated or not, and the result of original face recognition is correct / incorrect. We can see that, when the candidates are correctly eliminated, and the original face recognition is correct, the recognition result with elimination method can be correct. Also, if the expected face pattern is not eliminated, although the original face recognition result is incorrect, the result of face recognition after integration still has the chance to be correct.

Table 3. Face recognition accuracy analysis

Correct / Incorrect type	Accuracy of face recognition after integration
1. EC + FC	Correct
2. EC + FI	Unknown
3. EI + FC	Incorrect
4. EI + FI	Incorrect

EC/ EI: Eliminated candidates don't / do include expected pattern.
FC / FI: The result of original face recognition is correct / incorrect

3. PCA (Principal Component Analysis)

This section reviews the PCA method [6], which has been widely used in applications such as face recognition and image compression. PCA is a common technique for finding patterns in data, and expressing

the data as eigenvector to highlight the similarities and differences between different data. The following steps summarize the PCA process.

1. Let $\{D_1, D_2, \ldots D_M\}$ be the training data set. The average Avg is defined by:

$$Avg = \frac{1}{M}\sum_{i=1}^{M} Di \qquad (1)$$

2. Each element in the training data set differs from Avg by the vector $Y_i = D_i - Avg$. The covariance matrix Cov is obtained as:

$$Cov = \frac{1}{M}\sum_{i=1}^{M} Yi \cdot Yi^T \qquad (2)$$

   Since the covariance matrix Cov is square, we can calculate the eigenvectors and eigenvalues for this matrix.

3. Choose M' significant eigenvectors of Cov as E_K's, and compute the weight vectors W_{ik} for each element in the training data set, where k varies from 1 to M'.

$$W_{ik} = E_k^T \bullet (D_i - Avg), \ \forall\, i, k \qquad (3)$$

Based on PCA, many face recognition techniques have been developed, such as *eigenfaces* [1]. The following steps summarize the *eigenface* recognition process:

1. Initialization: Acquire the training set of face images $I_1, I_2, \ldots I_M$. Calculate each face difference vector from the average face Avg by (1), and the covariance matrix Cov is obtained by (2). Then compute the eigenvectors E_k of Cov, which define the face space. Finally, compute the weights W_{ik} by (3) for each image in the training set.
2. Input querying: When a new testing face image is encountered, calculate a set of weights W_{testK} depending on the same steps mentioned above. The weights W_{testk} forming a vector $T_p = [w_1, w_2, \ldots, w_{M'}]^T$ describes the contribution of each *eigenface* in representing the input face image
3. Recognition: A simplest technique to classify the weight pattern is to compute the minimum distance of $W_{testK.}$ from $T_{P.}$ It means that the test image can be classified to be in class p when $min(Dp) < \Theta i$, where $Dp = \| W_{testK} - T_P \|$ and Θi is the threshold.

Figure. 6 shows a simplified version of face space to illustrate the projecting results of three training face images W_1, W_2, W_3 and a testing image W_{testk}. We can recognize $W_{testk.}$ as one of the three known individuals W_1, W_2 and W_3 by the projecting distance between W_{testk} with each training images. In this case, there are two *eigenfaces* e_1, e_2 to construct the face space. The distance between W_{testk} and W_2 is larger than the threshold Θ_i, they are not considered to be the same person consequently. Furthermore, the projecting location of W_{testk} in the face space is more close to the projecting location of W_1 than W_2. Therefore, we believe that $W_{testk.}$ and W_1 are the same person.

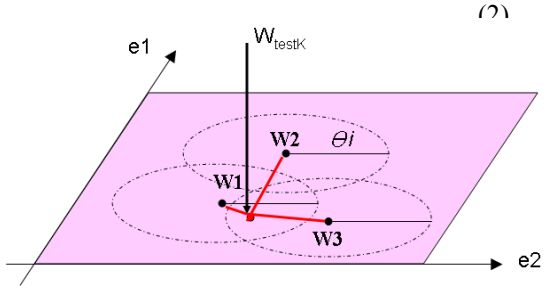

Figure 6. A simplified version of face space

4. System Model

In this section, we analyze our system from the probability point of view. Because the hand gesture recognition rate will not be influenced after integration, we only focus on the face recognition rate improvement and the overall accuracy improvement.

Assume that the number of face patterns is Nf, the average correctly guess rate is $(Nf)^{-1}$. However, if the expected value of the number to eliminate face candidates by recognized hand gesture is E_δ, then the correctly guess rate can be improved to $(Nf - E_\delta)^{-1}$. This provides the intuition of the concept that the face recognition rate can be improved by eliminating the face candidates.

Let the recognition rate of original face recognition engine be Pf, and the face recognition rate after integration be Pf'. Moreover, let the recognition rate of hand gesture recognition engine be Ph.

In this paper, the overall accuracy is calculated only when both the face and hand gesture recognition results are correct. That is:

$$Po = Pf \times Ph \ ; \ Po' = Pf' \times Ph$$

So that, if Pf' is enhanced, the Po' can also be improved. Also, because $Pf' \le 1$, so that the overall recognition rate Po' is impossible to exceed the hand

gesture recognition rate, $Po' \leq Ph$. Moreover, if we want $Po' > Pf$, then:

$$Pf' \times Ph > Pf, \text{ so } Ph > Pf / Pf'$$

Because $Pf' \leq 1$, so that $Ph > Pf$. We know that $Pf' \geq Po'$, so if we want $Pf' > Pf$, the hand gesture recognition rate must be higher than the original face recognition rate, which makes sense because face patterns are usually more complicated than hand gesture patterns.

If this integration concept is extended into other applications, it should be noticed that the recognition rate of the first recognition must be higher than the recognition rate of the second recognition method. Otherwise, after integration, the overall accuracy or the recognition rate of the second recognition method will not gain any improvement.

5. Evaluation of simulation

We simulate the secure elevator scenario with C++ program. Both the face and hand gesture recognition are performed by the basic PCA method. The face or hand detection is not the topic we focus on, so we use different database for face and hand gesture. In the experiment, the face database we used is ORL database, [11] which contains forty face patterns and ten images per pattern.

Table 4 and figure 7 illustrate the recognition rate of original face recognition, integrated face recognition, and hand gesture recognition. We analyze the recognition rate if there are one to five training images of these three recognition method. In general, *Ph* is always higher than *Pf* and *Pf'*, and *Pf'* almost outperforms than *Pf*. However, we may notice that if *Pf* is high enough, the *Pf'* is even a little bit lower than *Pf*.

Table 4. Recognition rate of *Pf*, *Pf'*, and *Ph*

	# of training data				
	1	2	3	4	5
Pf	0.703	0.82	0.873	0.905	0.963
Pf '	0.757	0.83	0.885	0.915	0.96
Ph	0.907	0.933	0.958	0.953	0.983

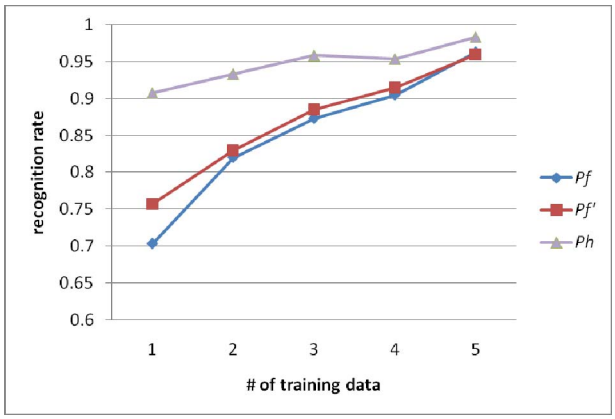

Figure 7. Recognition rate of *Pf*, *Pf'*, and *Ph*

Table 5 and figure 8 illustrate the recognition rate of the original and integrated overall accuracy. We can see that, the same, *Po* is almost higher than *Po'*. Also, we may notice that *Po'* is able to be higher than 94% under five training data.

Table 5. Recognition rate of *Po*, and *Po'*

	# of training data				
	1	2	3	4	5
Po	0.638	0.765	0.836	0.862	0.947
Po'	0.687	0.774	0.848	0.872	0.944

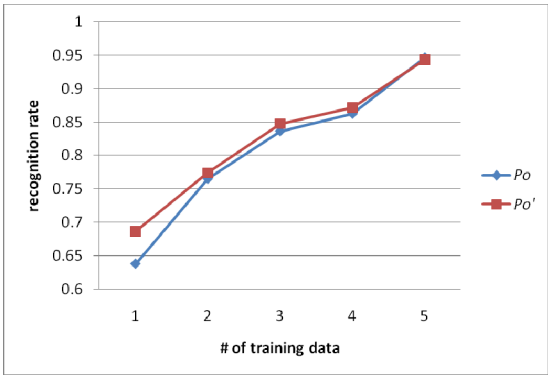

Figure 8. Recognition rate of *Po*, and *Po*

6. Conclusion

In this paper, we integrate two recognition systems: face and hand gesture recognition. We claim that the face recognition rate can be improved after the integration. During the integration, the result of hand gesture recognition and the available hand gesture mapping is used to eliminate face candidates dynamically. Also, we introduce a security elevator scenario, and simulate this scenario by PCA method.

The result of simulation shows that both face recognition rate and overall accuracy can be improved.

Although the simple PCA method is used in the simulation, we believe that other recognition engines, which are based on the similar candidate concept, are also able to be benefited by our integration method.

Acknowledgments

This work was mainly supported by National Science Council grant NSC96-2221-E-009-142: pub/sub-based P2P platform and its applications, and partially supported under grant NSC96-2520-S-009-007-MY3.

7. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books.

[1] Turk, M. A. and A. P. Pentland, "Face recognition using eigenfaces", in Proc. IEEE Conf. CVPR, June 3-6, 1991, pp.586-591.

[2] Brunelli, R. and T. Poggio, "Face recognition: features versus templates," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.15, No.10, pp.1042-1052, 1993.

[3] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition", Tech. Rep. TR-94-03a, Mitsubishi Electric Research Laboratories, Cambridge Research centre, Cambridge, Mass, USA, 1995.

[4] A.M. Burton, S. Wilson, M. Cowan, V. Bruce, "Face recognition in poor-quality video: Evidence From Security Surveillance", Psychological Science, Vol. 10, No. 3, May 1999, pp. 243-248.

[5] Triesch J. and Malsburg C. V. D, "A Gesture Interface for Human-Robot Interaction", Proc. of 3th IEEE Intl. Conf. on Automatic Face and Gesture Recognition, pp. 546-551. 1998.

[6] Fukunaga, Keinosuke, "Introduction to Statistical Pattern Recognition," Elsevier,1990.

[7] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models", IEEE ASSP Mag., pp 4--16, Jun. 1986.

[8] Yoav Freund and Robert E. Schapire, "A short introduction to boosting", Journal of Japanese Society for Artificial Intelligence, 14(5):771--780, September, 1999.

[9] Gurney K., *An Introduction to Neural Networks*, UCL Press, 1997, ISBN 1 85728 503 4.

[10] American Sign Language, http://www.lifeprint.com/

[11] ORL database, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html