Selection of Distinguish Points for Class Distribution Preserving Transform for Biometric Template Protection

Yi C. Feng and Pong C. Yuen

Department of Computer Science
Hong Kong Baptist University
{ycfeng, pcyuen}@comp.hkbu.edu.hk

Abstract. This paper addresses the biometric template security issue. Follow out previous work on class distribution transform, the proposed scheme selects the distinguish points automatically. By considering the geometric relationship with the biometric templates, the proposed scheme transforms a real-value biometric template to a binary string such that the class distribution is preserved and proved mathematically. The binary string is then further encoded using BCH and hashing method to ensure that the template protecting algorithm is non-invertible. Two face databases, namely ORL and FERET, are selected for evaluation and LDA is used for creating the original template. Experimental results show that by integrating the proposed scheme into the LDA (original) algorithm, the system performance can be further improved by 1.1% and 4%, in terms of equal error rate, on ORL and FERET databases respectively. The results show that the proposed scheme not only can preserve the original template discriminant power, but also improve the performance if the original template is not fully optimized.

Keywords: Biometric template security, face recognition, one-way transform, class distribution preserving.

1 Introduction

Biometric recognition is a reliable, robust and convenient way for person authentication [1,2,4]. With the growing use of biometrics, there is a rising concern about the security and privacy of the biometric template itself. Recent studies have shown that "hill climbing attacks" [5] on a biometric system are able to recover the biometric templates. As a result, templates have to be stored in an encrypted form to protect them. The matching step needs to be performed in encrypted domain because the decrypted templates are insecure. Different biometric template protection algorithms have been proposed in the last few years and can be categorized into two approaches.

The first approach is the cancelable template proposed by Ratha et al. [2] in 2001. The basic idea is that the system only uses a cancelable template, which is transformed by the original template using a one-way function. Therefore,

if the cancelable template is lost or stolen, the system may reset and re-issue another cancelable template using different set of parameters of the one-way function. In their recent study on cancelable biometric for fingerprint [3], the use of distortion function caused the FAR to increase by about 5% for a given FRR. While the invertibility of their transform is low, the discriminability of their cancelable template is not known. No analysis of invertibility vs. discriminability was reported in [3].

The second approach combined cryptographic methods with the biometric system, known as biometric cryptosystem. This approach consists of two stages [10]. First, the biometric template is transformed into a string. The second step generates a cryptographic key from the string using cryptographic technique. Some methods such as fuzzy commitment scheme [7] and fuzzy vault scheme [8,9], have been proposed for the second step. However, direct applying error correcting coding can only handle relatively small intra-class variations. In order to handle a relative large variation such as face template, Ngo et al. [10] proposed a new algorithm for the first step. They employed random projection and other appearance-based dimension reduction methods (PCA, LDA) to further extract the features. After that, a bit string is obtained using a thresholding technique. However, it is not clear whether the algorithm preserves the original template discriminability. Along this line, Feng and Yuen [11] proposed a new conceptual model called Class Distribution Preserving (CDP) Transform. A new scheme was proposed using a set of "distinguish point" and distance measurement. The "distinguish point" set is randomly generated, which will increase the randomness of the representation. A tri-state representation is proposed to overcome the distortion in transform, but the tri-state bit is hard to estimate.

In this paper, based on the concept of CDP transform, a new scheme is proposed to transform a real value biometric template into a bit string. In the proposed scheme, the set of "distinguish point" will be selected and the estimation of the tri-state bit is not required. The rest of the paper is organized as follows. Section 2 will give a review on the class distribution preserving transform. Section 3 will present our new proposed scheme while the experimental results will be reported in Section 4. Finally, Section 5 will give the conclusion.

2 Class Distribution Preserving Transform

The basic idea of class distribution preserving (CDP) transform [11] can be illustrated using Fig. 1. Consider a two class problem with class Ω_1 (data represented by "cross") and Ω_2 (data point represented by "circle"). Given distinguish points say B_1 and B_2 , for any real value feature vector (biometric template) V in d -dimensional space, it can be transformed into a k -dimensional ($k = 2$ in this case) tri-state bit string $[b_1, b_2]$, where

$$b_i = \begin{cases} 0 & : d(B_i, V) < th_i - \frac{r}{2} \\ \phi & : th_i - \frac{r}{2} \leq d(B_i, V) \leq th_i + \frac{r}{2} \\ 1 & : d(B_i, V) > th_i + \frac{r}{2} \end{cases} \quad (1)$$

Fig. 1. Illustration for the Class Distribution Preserving Transform

Although the scheme in [11] works, there are rooms for improvement. First, all distinguish points are randomly generated and therefore, may not be able to generate a set of optimal bit string. Second, the number of distinguish points depends on the number of class which may introduce a complexity burden when the class number is large. Finally, it is hard to estimate the parameter "r" in calculating the tri-state ϕ .

3 Our Proposed Scheme

The block diagram of the proposed system is shown in Fig. 2. It consists of enrollment and authentication phases. The template protection algorithm consists of three blocks, namely class distribution preserving transform, error correction coding and hashing. The error correction coding and hashing steps are used to ensure that the protecting algorithm is non-invertible. This paper mainly focuses on the class distribution preserving transform.

The proposed scheme follows the concept of class distribution preserving (CDP) transform [11], but removes the limitations in our previous work [11]. This paper makes two major changes. First, instead of follow the scheme in [11] for biometric identification system (one to many), the new scheme is designed for authentication system (one to one). In practice, most of the biometric systems perform authentication, such as border control in Hong Kong and China. Second, the distinguish point set is determined so as to preserve the class distribution after transformation. Details of our scheme are discussed as follows.

3.1 CDP Transform for Authentication System

Assume there are c classes $C_T = \{\Omega_1, \Omega_2 \dots \Omega_c\}$ with the cluster centers $\{M_1, M_2 \dots M_c\}$ respectively. For each class Ω_o , let r_o be the largest distance any two

Fig. 2. Block diagram of the proposed system

data points in Ω_o , For authentication system, we only need to solve a two-class problem. Therefore, in determining the distinguish point set for class Ω_o , we only need to consider the class Ω_o , and $C_T - \Omega_o = \{x \in C_T | x \notin \Omega_o\}$. So the problem is defined as follows. Given a data point (biometric template) V , we would like to find a transformation T which has the following two CDP properties, where d_1 and d_2 are distance functions in input and transformed space respectively and t_o is the threshold in the transformed domain.

– Property One

If $d_1(V, M_o) < r_o$, then $d_2(T(V), T(M_o)) \leq t_o$

This property shows that if V belongs to class Ω_o , after transformation, V also belongs to the same class.

– Property Two

If $d_1(V, M_o) > r_o$, then $d_2(T(V), T(M_o)) > t_o$

This property shows that if V does not belong to class Ω_o , after transformation, V will also not belong to the class Ω_o .

3.2 Determining Distinguish Points and Thresholds

In order to develop the CDP transformation for authentication system as discussed in Section 3.1, we need to determine the set of k distinguish points $\{B_1, B_2 \dots B_k\}$ and the responding thresholds $\{th_1, th_2 \dots th_k\}$ for each class. For simplicity, we will represent them in pairs $\{B_1, th_1\}$.

Let the data points (biometric templates) in each class are represented as $CT - \Omega_o = \{P_1, P_2 \dots P_m\}$ and $\Omega_o = \{Q_1, Q_2 \dots Q_n\}$ with center M_o . Each template will be transformed into a k -dimensional binary string. The proposed method consists of two steps. The first step finds the k directions where the distinguish points to be located while the positions as well as the threshold values will be determined in the second step.

Step 1. Using M_o as an origin, each template in $CT - \Omega_o$ can be represented by a unit vector. This step determines the k representative directions from the m unit vectors. To do that, our strategy is to determine the k unit vectors with the largest separation (i.e. largest angles between two vectors). The problem can be solved using a k -mean algorithm: At the beginning we find the first one v_1 which has the largest angles between the other unit vectors. Then find the second one v_2 which has the largest angle to v_1 . Find v_3 which has largest angles to v_1 and v_2 , and so on until we find k unit vectors. With the k unit vectors, classify the unit vectors in $CT - \Omega_o$ to k clusters, the i -th cluster is composed of nearest neighbors of v_i . ($i = 1, 2 \dots k$). The output of this step will be k clusters $\{G_1, G_2, \dots G_k\}$ in which $\{v_1, v_2 \dots v_k\}$ are a set of k unit vectors (from M_o).

Step 2. This step determines the exact positions of distinguish points in each direction (unit vector). The position of the i -th distinguish point along the direction of v_i can be written as $B_i = M_o + a_i v_i$, where a_i is a real value (parameter) and controls the position of B_i along v_i . The threshold th_i for the point B_i is then equal to $th_i = |a_i - r_o|$. Therefore, what we need to do is to estimate a_i .

The value of a_i can be positive or negative. $a_i < 0$ means that $B_i M_o$ is with the same direction of v_i and $th_i = |B_i M_o| + r_o$. If $a_i > 0$, $B_i M_o$ is in the opposite direction of v_i and $th_i = |B_i M_o| - r_o$. Here we give a definition of a_i as $a_i = \pm(2 + e_i)d$, ($i = 1, 2 \dots k$).

The parameter d here is the largest value of $|P_j M_o|$, ($j = 1, 2 \dots m$) and e_i ($i = 1, 2 \dots k$) here is random generated value between range $[0, 2]$. The rationale in determining the a_i will be discussed in next section.

3.3 Proof of the CDP Properties

Before discussing the proof, let's consider two scenarios in which a distinguish point is located in two positions with respect to a data point outside the class Ω_o as shown in Fig. 2.

Scenario One. Considering a distinguish point $\{B_i, th_i\}$ where B_i and P_j are located in the opposite direction with respect to Ω_o with center M_o as Fig. 3. In this case, $th_i = |B_i M_o| + r_o$. For any data point P_j in group G_i (after clustering in step 1) and Q_s ($s = 1, 2 \dots n$) in Ω_o , if θ_j is the angle between direction $B_i M_o$ and $M_o P_j$, then $\cos \theta_j > r_o / |P_j M_o|$.

The distance between the distinguish point B_i and P_j , Q_s , M_o can be written as

$$\begin{aligned}
|B_i P_j| &\geq |B_i M_o| + |P_j M_o| \cos \theta_j > |B_i M_o| + r_o = th_i; \\
|B_i Q_s| &\leq |B_i M_o| + |Q_s M_o| \leq |B_i M_o| + r_o = th_i; \\
|B_i M_o| &< |B_i M_o| + r_o = th_i.
\end{aligned} \tag{2}$$

That means, $|B_i P_j| > th_i$, $|B_i M_o| \leq th_i$ and $|B_i Q_s| \leq th_i$. So, using B_i as a distinguish point, the point P_j , Q_s and M_o are then transformed into bit $b_i(P_j) = 1$, $b_i(M_o) = 0$, $b_i(Q_s) = 0$. That means, data points within the class Ω_i will be transformed into same bit value while data points outside the class Ω_o will be transformed into opposite value.

Fig. 3. Distinguish one point from M_o with pair B_i, th_i

Scenario Two. In this scenario, B_i' (with threshold th_i') and P_j are located in the same direction with respect to Ω_o with center M_o as Fig. 3. In this case, $th_i = |B_i M_o| - r_o$. Then the angle between direction $B_i' M_o$ and $M_o P_j$, θ_j , will satisfy

$$\cos \theta_j > \frac{|P_j M_o|}{2|B_i' M_o|} + \frac{r_o}{|P_j M_o|} - \frac{r_o^2}{2|B_i' M_o| \cdot |P_j M_o|} \tag{3}$$

and

$$|B_i' M_o| > 2|P_j M_o|. \tag{4}$$

The distance between the distinguish point B_i' and P_j , Q_s , M_o can be written as

$$\begin{aligned}
|B_i' Q_s| &\geq |B_i' M_o| - |Q_s M_o| \geq |B_i' M_o| - r_o = th_i; \\
|B_i' M_o| &> |B_i' M_o| - r_o = th_i; \\
|B_i' P_j|^2 &= |B_i' M_o|^2 + |P_j M_o|^2 - 2 \cos \theta_j \cdot |B_i' M_o| \cdot |P_j M_o| \\
&< |B_i' M_o|^2 + |P_j M_o|^2 - |P_j M_o|^2 - 2r_o |B_i' M_o| + r_o^2 \\
&= |B_i' M_o|^2 - 2r_o |B_i' M_o| + r_o^2 = th_i'^2.
\end{aligned} \tag{5}$$

That means $|B_i' P_j| \leq th_i'$, $|B_i' M_o|, |B_i' Q_s| > th_i'$. So, using B_i' as a distinguish point, the point P_j , Q_s and M_o are then transformed into bit $b_i(P_j) = 0$, $b_i(M_o) = 1$, $b_i(Q_s) = 1$. That means, data points within the class Ω_o will be transformed into same bit value while data points outside the class Ω_o will be transformed into opposite value.

So far, we have shown that under the above two scenarios, the transformed bit of any feature vector Q in class Ω_o will be the same with M_o , and the transformed

bit of any feature vector V in group G_i will be different from M_o . Using the same argument as above, consider all k pairs $\{B_1, th_1\}, \{B_2, th_2\} \dots \{B_k, th_k\}$, the transformed bit string of any feature vector Q in class Ω_o will be the same with M_o . And there is at least one bit in the transformed bit string of any feature vector V out of class Ω_o which is different from M_o . (Because V should belong to one group G_i and the corresponding pair $\{B_i, th_i\}$ will contribute a different bit from M_o .) Therefore, we can prove that

$$\begin{aligned} \text{If } V \text{ belongs to } \Omega_o, \quad d_2(T(V), T(M_o)) &= 0; \\ \text{If } V \text{ dose'nt belong to } \Omega_o, d_2(T(V), T(M_o)) &> 0. \end{aligned}$$

The function d_2 can be Hamming distance between two bit strings.

One important thing on $\cos \theta_j$ we would like to point out. In scenarios 1 and 2, $\cos \theta_j$ needs to satisfy the following two conditions

$$\begin{aligned} \cos \theta_j &> \frac{r_o}{|P_j M_o|} \\ \cos \theta_j &> \frac{|P_j M_o|}{2|B'_i M_o|} + \frac{r_o}{|P_j M_o|} - \frac{r_o^2}{2|B'_i M_o| \cdot |P_j M_o|} \end{aligned} \quad (6)$$

And hence,

$$\begin{aligned} \frac{r_o}{|P_j M_o|} &< 1 \\ \frac{|P_j M_o|}{2|B'_i M_o|} + \frac{r_o}{|P_j M_o|} - \frac{r_o^2}{2|B'_i M_o| \cdot |P_j M_o|} &< 1 \end{aligned} \quad (7)$$

Because P_j is out of class Ω_o , i.e., $|P_j M_o| > r_o$, thus $r_o/|P_j M_o| < 1$. With the given condition that $|B'_i M_o| \geq 2|P_j M_o|$ in scenario 2, we have

$$\frac{|P_j M_o|}{2|B'_i M_o|} + \frac{r_o}{|P_j M_o|} - \frac{r_o^2}{2|B'_i M_o| \cdot |P_j M_o|} < \frac{1}{4} + \frac{r_o}{|P_j M_o|} - \frac{1}{4} \left(\frac{r_o}{|P_j M_o|} \right)^2 < 1. \quad (8)$$

(Because $r_o < |P_j M_o|$.)

To satisfy the requirement $|B'_i M_o| \geq 2|P_j M_o|$ in scenario 2, we give the definition of $a_i = \pm(2 + e_i)d$ in Section 3.2. With it, $|B'_i M_o| = |a_i| = (2 + e_i)d \geq 2d \geq 2|P_j M_o|$. The parameter e_i is used to increase the randomness of the generated pairs of distinguish points and thresholds because they do contain some information of the class Ω_o .

4 Experiment Results

Two experiments on ORL and FERET face databases are performed to evaluate our proposed scheme. The ORL database consists of 40 individuals, 10 images

per individual while FERET database consists of 250 individuals, 4 images per individual. The experiments are conducted based on the block diagram in Fig. 2 in which the Fisherface [12] algorithm (feature extractor) is used to construct the original template while the matching is performed in hashed domain to ensure that the template protection algorithm is non-invertible.

Since the database size is relatively small, leave-one-out method is employed to conduct the experiments. The average error/accuracy is recorded. The results are plotted in Fig. 4 and 5 with different number of distinguish points (k). Fig. 4 and 5 show the ROC curves (with label "CDP") on ORL and FERET databases respectively. It can be shown that the performance increases when the number of distinguish points increase. The performance of the Fisherface algorithm with nearest neighbor classifier are also recorded and plotted in Fig. 4 and 5 (label as "original"). It can be shown that the proposed scheme outperforms original Fisherface algorithm while the biometric template is protected.

Fig. 4. Experiment results on ORL database

To compare with the existing algorithm, the bio-hashing scheme proposed by Ngo et al. [10] is selected because this algorithm also follows the biometric hardening approach. Same experiments setting are used for bio-hashing algorithm and the results are plotted in Fig. 4 and 5 with label "bioh". Parameter r here means the variation σ related to transform of ϕ in the bio-hashing algorithm. It can be seen that our proposed outperforms bio-hashing algorithm as well.

The equal error rates (ERR) of each curve are recorded and tabulated in Tab. 1 and 2. Comparing with the original template, the proposed method improves the performance at 1.1% and 4% on ORL and FERET databases respectively, while it is about 4.6% and 1.3% in comparing with the bio-hashing scheme.

Fig. 5. Experiment results on FERET database

Table 1. The ERR on ORL database

	Original	CDP $k = 50$	CDP $k = 70$	CDP $k = 90$	Bioh $r = 0$	Bioh $r = 1$	Bioh $r = 2$	Bioh $r = 3$
ERR(%)	3.43	2.30	1.90	1.89	6.59	7.92	7.72	6.86

Table 2. The ERR on FERET database

	Original	CDP $k = 50$	CDP $k = 70$	CDP $k = 90$	Bioh $r = 0$	Bioh $r = 1$	Bioh $r = 2$	Bioh $r = 3$
ERR(%)	16.03	12.00	10.94	10.39	11.7	11.87	11.87	11.87

5 Conclusions

A new scheme for the class preserving transform has been developed and reported in this paper. The scheme is designed for the authentication system and determines the distinguish point set automatically so that after transformation, the inter-class and intra-class data distribution are preserved. Two face databases, namely ORL and FERET, are selected for evaluation and LDA is used for creating the original template. Two face databases, namely ORL and FERET, are selected for evaluation and LDA is used for creating the original template. Experimental results show that by integrating the proposed scheme into the LDA (original) algorithm, the system performance can be further improved by 1.1% and 4%, in terms of equal error rate, on ORL and FERET databases respectively.

The results show that the proposed scheme not only can preserve the original template discriminant power, but also improve the performance if the original template is not fully optimized.

Acknowledgments. This project is partially supported by the Faculty Research Grant of Hong Kong Baptist University.

References

1. Uludag, U., Pankanti, S., Prabhakar, S., Jain, A.K.: Biometric cryptosystems: issues and challenges. *Proceedings of the IEEE* 92(6), 948–960 (2004)
2. Ratha, N., Connell, J., Bolle, R.: Enhancing security and privacy in biometric-based authentication systems. *IBM Systems Journal* 40(3), 614–634 (2001)
3. Ratha, N., Connell, J., Bolle, R., Chikkerur, S.: Cancelable biometrics: A case study in Fingerprints. In: *Proceedings of International Conference on Pattern Recognition* (2006)
4. Jain, A.K., Ross, A., Pankanti, S.: Biometrics: A Tool for Information Security. *IEEE Transactions on Information Forensics and Security* 1(2), 125–143 (2006)
5. Adler, A.: Images can be regenerated from quantized biometric match score data. In: *Proceedings of Canadian conference of Electrical and Computer Engineering*, pp. 469–472 (2004)
6. Boulton, T.: Robust distance measures for face-recognition supporting revocable biometric tokens. In: *Proceeding of International Conference on Automatic Face and Gesture recognition* (2006)
7. Juels, A., Wattenberg, M.: A fuzzy commitment scheme. In: *Sixth ACM Conf. on Comp. and Comm. Security*, pp. 28–36 (1999)
8. Juels, A., Sudan, M.: A Fuzzy Vault Scheme. In: *IEEE International Symposium on Information Theory*. IEEE Computer Society Press, Los Alamitos (2002)
9. Clancy, T.C., Kiyavash, N., Lin, D.J.: Secure smartcard-based fingerprint authentication. In: *Proc. ACM SIGMM 2003 Multimedia, Biometrics Methods and Applications Workshop*, pp. 45–52 (2003)
10. Ngo, D.C.L., Teoh, A.B.J., Goh, A.: Biometric Hash: High-Confidence Face Recognition. *IEEE transactions on circuits and systems for video technology* 16(6) (2006)
11. Feng, Y.C., Yuen, P.C.: Class-Distribution Preserving Transform for Face Biometric Data Security. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2007* (in press)
12. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI* 19(7), 711–720 (1997)
13. Goh, A., Ngo, D.C.L.: Computation of cryptographic keys from face biometrics. In: *Proc. 7th IFIP TC6/TC11 Conf. Commun. Multimedia Security*, vol. 22, pp. 1–13 (2003)