Multimodal Biometrics Fusion Using Correlation Filter Bank*

Yan Yan and Yu-Jin Zhang
Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
yan-y04@mails.tsinghua.edu.cn, zhang-yj@mail.tsinghua.edu.cn

Abstract

In this paper, a novel class-dependence feature analysis method based on Correlation Filter Bank (CFB) technique for effective multimodal biometrics fusion at the feature level is developed. In CFB, the unconstrained correlation filter trained for a specific modality is designed by optimizing the overall original correlation outputs. Therefore, the differences between modalities have been taken into account and useful information in various modalities is fully exploited. Preliminary experimental results on the fusion of face and palmprint biometrics show the superiority of the novel method.

1. Introduction

Recently, multimodal biometrics fusion techniques have attracted much attention as the supplementary information between different modalities could improve the recognition performance. Many works have concentrated on this area [1-4]. In general, they can be classified into three categories: fusion at the feature level, fusion at the match level and fusion at the decision level [1]. In this paper, the focus is on fusion at the feature level which is believed to be very promising as feature sets can provide more information about the input biometrics than other levels [1, 2].

On the other hand, subspace learning methods, *e.g.* Principal Component Analysis (PCA) [5], Linear Discriminant Analysis (LDA) [6], Locality Preserving Projections (LPP) [7] and Class-dependence Feature Analysis (CFA) [8], which select low-dimensional features to represent raw data, have been widely used in the biometric researches. Unlike traditional subspace learning methods [5-7], the projection axis obtained by CFA which is based on the design of correlation filter (CF) technique tries to discriminate one specific class from all other classes. According to different criterions, different correlation filters can be designed.

Subspace learning methods which perform at the feature level for multimodal biometrics is usually implemented by concatenating two or more original or low-dimensional features to form a long vector [3]. Usually, a normalization step is carried out prior to the concatenation of different features. Unfortunately, this method may not work well when the features of various modalities are compatible [1]. Furthermore, the combined feature vector can become high-dimensional if more biometric modalities are added.

In this paper, we propose a novel Correlation Filter Bank (CFB) based CFA method for the multimodal biometrics fusion application. It is worthwhile to highlight several aspects of the novel method here:

- A new filter technique called Correlation Filter Bank (CFB) is developed for multimodal biometrics feature vector extraction. In CFB, the unconstrained correlation filter trained for a specific modality is designed by optimizing the overall origin correlation outputs. Hence, CFB takes full advantage of the information in different modalities.
- Compared with traditional fusion methods at the feature level which concatenate features to form a long vector, the dimensionality of the features obtained by the novel method is equal to the number of classes in the training set no matter how many modalities are used.
- This paper mainly discusses the fusion of face and palmprint biometrics. However, the algorithm and analysis presented here can also be applied to other multimodal biometrics fusion applications.

* This work has been supported by Grant NNSF-60573148.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

2. Technical Approach

The general framework is first described, then the design of correlation filter is detailed.

2.1. General framework

After original feature extraction, the face or palmprint biometric is usually represented as a high-dimensional vector. Then, PCA is performed to reduce the data dimensionality. Finally, in the low-dimensional subspace, CFB is designed by using the 1-D Fourier transforms of the low-dimensional features (face features and palmprint features). The detailed derivation of CFB is given in Section 2.2.

Once CFBs are designed for each class in the training set, the multimodal biometrics feature vector can be extracted as shown in Figure 1. To be specific, each component of the feature vector is derived by the addition of the inner products of low-dimensional features and designed 1-D correlation filters (CF) in the CFB that are represented in terms of space domain.

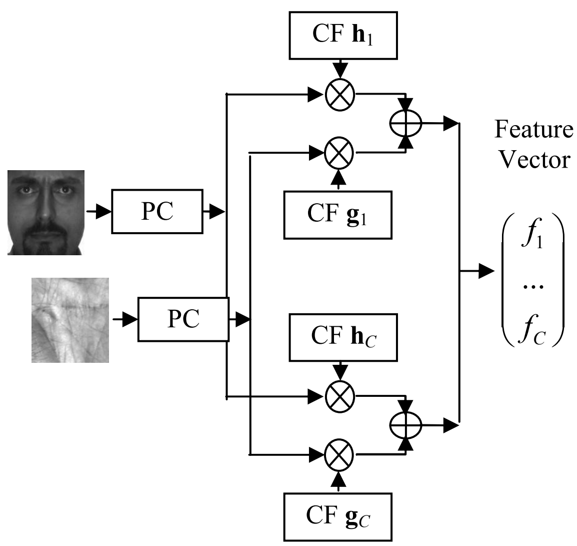

Figure 1. Feature vector extraction based on correlation filter bank.

2.2. Correlation Filter Bank (CFB)

From Figure 1, it needs to design a CFB for each class in the training set. Assume that two correlation filters (CF) in CFB designed for the l-th class are $\boldsymbol{h}$ and $\boldsymbol{g}$, respectively. In CFB, two correlation filters are jointly optimized to provide the best performance. The architecture of the novel CFB is shown in Figure 2.

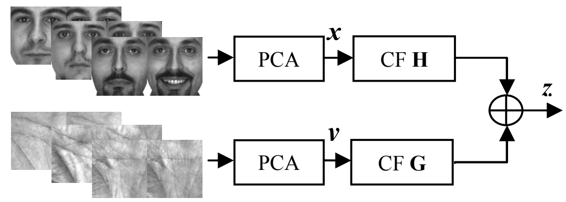

Figure 2. Architecture of correlation filter bank.

Suppose there are N training images and C classes for each modality in the training set. N_l is the number of training images in the l-th class for each modality. p is the dimensionality of low-dimensional feature obtained by PCA. According to Figure 2, the output of CFB is:

$$\boldsymbol{z}_i(n) = \boldsymbol{x}_i(n) \odot \boldsymbol{h}(n) + \boldsymbol{y}_i(n) \odot \boldsymbol{g}(n) \qquad (1)$$

where '$\odot$' stands for the correlation function of two signals. $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are the low-dimensional features of the i-th face training image and i-th palmprint training image in the reduced PCA subspace, respectively.

The output can also be expressed using the frequency domain representations, that is:

$$\boldsymbol{z}_i(n) = \sum_{k=0}^{p-1} \boldsymbol{X}_i(k)^* . \boldsymbol{H}(k) e^{\frac{j2\pi kn}{p}} + \sum_{k=0}^{p-1} \boldsymbol{Y}_i(k)^* . \boldsymbol{G}(k) e^{\frac{j2\pi kn}{p}} \qquad (2)$$

where $\boldsymbol{X}_i(k)$, $\boldsymbol{Y}_i(k)$, $\boldsymbol{H}(k)$ and $\boldsymbol{G}(k)$ are the 1-D Fourier transforms of $\boldsymbol{x}_i$, $\boldsymbol{y}_i$, $\boldsymbol{h}$ and $\boldsymbol{g}$, respectively. '*' denotes the conjugate operation. Note that the point $\boldsymbol{z}_i(0)$ which is the sum of the inner products of inputs and correlation filters is often referred to the overall origin correlation output or overall origin peak.

In this paper, CFB is designed by optimizing the overall origin correlation outputs. More specifically, the design objective of CFB is to minimize the overall origin correlation output energy for extra-class samples while maximizing the overall average origin peak for intra-class samples in the l-th class.

According to Eq. (2), the overall origin correlation output energy for extra-class samples can be derived as:

$$\begin{aligned}
&\text{Orgin Output Energy} = \frac{1}{N-N_l} \sum_{i=1}^{N-N_l} \left| \boldsymbol{z}_i^{\boldsymbol{E}}(0) \right|^2 \\
&= \frac{1}{N-N_l} \sum_{i=1}^{N-N_l} \left| \sum_{k=0}^{p-1} \boldsymbol{X}_i^{\boldsymbol{E}}(k)^* . \boldsymbol{H}(k) + \sum_{k=0}^{p-1} \boldsymbol{Y}_i^{\boldsymbol{E}}(k)^* . \boldsymbol{G}(k) \right|^2 \\
&= \frac{1}{N-N_l} \sum_{i=1}^{N-N_l} \left| \boldsymbol{X}_i^{\boldsymbol{E}+} \boldsymbol{H} + \boldsymbol{Y}_i^{\boldsymbol{E}+} \boldsymbol{G} \right|^2 \qquad (3) \\
&= \frac{1}{N-N_l} \sum_{i=1}^{N-N_l} (\boldsymbol{H}^+ \boldsymbol{X}_i^{\boldsymbol{E}} \boldsymbol{X}_i^{\boldsymbol{E}+} \boldsymbol{H} + \boldsymbol{G}^+ \boldsymbol{Y}_i^{\boldsymbol{E}} \boldsymbol{Y}_i^{\boldsymbol{E}+} \boldsymbol{G} \\
&\qquad + \boldsymbol{H}^+ \boldsymbol{X}_i^{\boldsymbol{E}} \boldsymbol{Y}_i^{\boldsymbol{E}+} \boldsymbol{G} + \boldsymbol{G}^+ \boldsymbol{Y}_i^{\boldsymbol{E}} \boldsymbol{X}_i^{\boldsymbol{E}+} \boldsymbol{H}) \\
&= \boldsymbol{H}^+ \boldsymbol{R}_{xx} \boldsymbol{H} + \boldsymbol{G}^+ \boldsymbol{R}_{yy} \boldsymbol{G} + \boldsymbol{H}^+ \boldsymbol{R}_{xy} \boldsymbol{G} + \boldsymbol{G}^+ \boldsymbol{R}_{yx} \boldsymbol{H}
\end{aligned}$$

where $\boldsymbol{z}_i^{\boldsymbol{E}}(0)$ is the overall origin correlation output for the i-th extra-class sample for the l-th class.

$$R_{xx} = \frac{1}{N-N_l}\sum_{i=1}^{N-N_l} X_i^E X_i^{E+}, \quad R_{yy} = \frac{1}{N-N_l}\sum_{i=1}^{N-N_l} Y_i^E Y_i^{E+}$$

$$R_{xy} = \frac{1}{N-N_l}\sum_{i=1}^{N-N_l} X_i^E Y_i^{E+}, \; R_{yx} = \frac{1}{N-N_l}\sum_{i=1}^{N-N_l} Y_i^E X_i^{E+}$$

X_i^E and Y_i^E, $i = 1,...,N-N_l$ are the 1-D Fourier transforms of extra-class low-dimensional face features and palmprint features for the l-th class. '+' denotes conjugate transpose.

The overall average origin correlation output (origin peak) for the intra-class samples in the l-th class is:

$$\begin{aligned} &\text{Average Orgin Peak} = \frac{1}{N_l}\sum_{i=1}^{N_l} z_i^I(0) \\ &= \frac{1}{N_l}\sum_{i=1}^{N_l}\sum_{k=0}^{p-1}[X_i^I(k)^* .H(k) + Y_i^I(k)^* .G(k)] \\ &= \frac{1}{N_l}\sum_{i=1}^{N_l}(X_i^{I+}H + Y_i^{I+}G) \\ &= X_m^{I+}H + Y_m^{I+}G \end{aligned} \quad (4)$$

where X_m^I and Y_m^I are the mean vectors of the Fourier transforms of intra-class low-dimensional face features and palmprint features, respectively.

Therefore, by maximizing Eq. (4) and minimizing Eq. (3), we can obtain the following optimization criterion through simple matrix operations:

$$J(C) = \arg\max_C \frac{|m^+ C|^2}{C^+ RC} \quad (5)$$

where $C = \begin{pmatrix} H \\ G \end{pmatrix}$, $R = \begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix}$ and $m = \begin{pmatrix} X_m^I \\ Y_m^I \end{pmatrix}$.

The well-known solution of Eq. (5) is given by $C = R^{-1}m$. Once C is computed, H and G can be obtained accordingly.

After feature extraction for both the training set and test set, it needs to design a classifier. In this paper, the simple nearest neighbor classifier is applied. For subspace learning methods, the whitened cosine distance is observed to give better performance than the Euclidean distance [8]. Therefore, we use whitened cosine distance in the paper. The computation of whitened cosine distance is shown in (6):

$$S(x, y) = \frac{-(x \bullet y)}{\| x \| \| y \|} \quad (6)$$

2.3. Discussions

CFB is designed by jointly optimizing two correlation filters (one for face and the other for palmprint) at the overall origin correlation outputs. Therefore, the differences between modalities are taken into account and useful information in various modalities is fully exploited in CFB. As a result, CFB is effective for the extraction of feature vector.

It's worth noting that CFB becomes an unconstrained correlation filter if only one input is considered. Compared with constrained correlation filter, such as OTF (Optimal Tradeoff Filter) [8], the generalization performance of the unconstrained correlation filter is better since the hard constraints on the correlation outputs are removed.

3. Experimental Results

In this section, the recognition performance of the novel method is evaluated on two face databases (AR and FRGC) and a palmprint database. We compare the novel method with other subspace learning based multimodal biometrics fusion methods (including PCA [5], LDA [6], LPP [7] and OTF based CFA [8]) which is fused at the feature level (combine face image and palmprint image at the pixel level) [3].

The AR face database [9] contains over 4,000 face images of 126 people, including frontal view of faces with different facial expressions, lighting conditions and occlusions. Fourteen face images from 100 individuals are selected and used in our experiment. For FRGC face database [10], we select 2000 images for 100 individuals (each one has 20 images). The face images were captured in both controlled and uncontrolled conditions with harsh illumination and expression variations. For PolyU palmprint database [11], we select 2000 images of 100 different palms. Twenty samples from each of these palms were collected in two different sessions, where the first 10 were captured in the first session and the other 10 in the second session. The size of all images is 64×64.

We take the sample subsets of the same size from AR and $PolyU_1$ databases. In other words, $PolyU_1$ database is composed of 14 images with 100 palmprints from PolyU palmprint database. For FRGC and $PolyU_2$ databases, all 2000 images are used.

For all databases, 2 or 3 images per class are randomly chosen to form the training set. And the rest of the images are used for testing. Totally, twenty experiments are performed. The final result is the average recognition rate over 20 random training sets.

Table 1 shows the top mean recognition rate by each method and the corresponding dimensionality of reduced subspace on the face databases. Table 2 shows the top mean recognition rate by each method and the corresponding dimensionality of reduced subspace on the palmprint databases. The final fusion results are given in Table 3.

Table 1 Top recognition rate and the corresponding dimensionality based on the face biometric.

Method	AR (2 train)	AR (3 train)	FRGC (2 train)	FRGC (3 train)
PCA	0.7466 (191)	0.8369 (299)	0.5691 (199)	0.5982 (298)
LDA	0.6794 (38)	0.7871 (56)	0.4304 (34)	0.5609 (49)
LPP	0.6236 (92)	0.7670 (92)	0.4341 (97)	0.5597 (94)
CFA (OTF)	0.8601 (100)	0.9023 (100)	0.6167 (100)	0.7246 (100)
Proposed method	0.8831 (100)	0.9248 (100)	0.6617 (100)	0.7720 (100)

Table 2 Top recognition rate and the corresponding dimensionality based on the palmprint biometric.

Method	$PolyU_1$ (2 train)	$PolyU_1$ (3 train)	$PolyU_2$ (2 train)	$PolyU_2$ (3 train)
PCA	0.9443 (38)	0.9717 (38)	0.9266 (199)	0.9667 (43)
LDA	0.9220 (74)	0.9602 (92)	0.9029 (70)	0.9572 (97)
LPP	0.9252 (101)	0.9606 (101)	0.9100 (100)	0.9586 (100)
CFA (OTF)	0.9395 (100)	0.9593 (100)	0.9280 (100)	0.9549 (100)
Proposed method	0.9564 (100)	0.9715 (100)	0.9439 (100)	0.9654 (100)

Table 3 Top recognition rate and the corresponding dimensionality based on the face and palmprint biometric fusion.

Method	AR+$PolyU_1$ (2 train)	AR+$PolyU_1$ (3 train)	FRGC+ $PolyU_2$ (2 train)	FRGC+ $PolyU_2$ (3 train)
PCA	0.9178 (191)	0.9504 (272)	0.8143 (196)	0.8909 (298)
LDA	0.9599 (47)	0.9790 (74)	0.8758 (46)	0.9459 (52)
LPP	0.9534 (92)	0.9781 (92)	0.8618 (97)	0.9428 (97)
CFA (OTF)	0.9714 (100)	0.9749 (100)	0.9529 (100)	0.9664 (100)
Proposed method	0.9923 (100)	0.9964 (100)	0.9761 (100)	0.9895 (100)

From Table 1, Table 2 and Table 3, it is obvious that the proposed method outperforms other subspace learning based fusion methods when the training sample size per class is small. Furthermore, the proposed method obtains a better recognition results than other methods when only one modality (face or palmprint) is used. In a word, experimental results demonstrate the superiority of the novel method in the multimodal biometrics fusion application.

4. Conclusions

In this paper, we propose an effective multimodal biometrics fusion method based on the design of Correlation Filter Bank (CFB) technique. By focusing on the overall original correlation outputs, CFB can extract the discriminant features of multimodal biometrics effectively. Preliminary experimental results show the effectiveness and robustness of the novel method. However, in this paper, experiments are conducted on the non-real multimodal biometrics data for convenience. In the future work, more experimental tests will be performed on the real multimodal biometrics data.

References

[1] A.K. Jain, K. Nandakumar, and A. Ross. "Score normalization in multimodal biometric systems", *Pattern Recognition*, 38(12): 2270-2285, 2005.

[2] R. Wang and B. Bhanu. "Performance prediction for multimodal biometrics", In *Proc. of the IEEE International Conference on Pattern Recognition*, pages 586-589, 2006.

[3] X. Jing, Y. Yao, D. Zhang, J. Yang, and M. Li. "Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometric recognition", *Pattern Recognition*, 40(11): 3209-3224, 2007.

[4] T. Zhang, X. Li, D. Tao, and J. Yang. "Multi-modal biometrics using geometry preserving projections", *Pattern Recognition*, 41(3): 805-813, 2008.

[5] M. Turk and A. Pentland. "Eigenfaces for recognition", *J. Cognitive Neuroscience*, 3(1): 71-86, 1991.

[6] P.N. Belhumeur, J.P. Hepanda, and D.J. Kriegman. "Eigenface vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7): 711-720, 1997.

[7] X. He, S. Yan, Y. Hu, and H.-J. Zhang. "Face recognition using Laplacianfaces", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(3): 328-340, 2005.

[8] B.V.K.V. Kumar, M. Savvides, and C. Xie, "Correlation pattern recognition for face recognition", *Proc. IEEE*. 94(11): 1963-1976, 2006.

[9] A.R. Martinez and R. Benavente, "The AR face database", *CVC Technical Report #24*, 2003.

[10] PolyU palmprint database <http://www.comp.polyu.edu.hk/~biometrics/>, 2006.

[11] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge", In *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 947-954, 2006.