Feature Correlation Filter for Face Recognition

Xiangxin Zhu, Shengcai Liao, Zhen Lei, Rong Liu, and Stan Z. Li

Center for Biometrics and Security Research & National Laboratory of Pattern
Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun East Road, 100080 Beijing, China
{xxzhu,scliao,zlei,rliu,szli}@nlpr.ia.ac.cn
http://www.cbsr.ia.ac.cn

Abstract. The correlation filters for pattern recognition, have been
extensively studied in the areas of automatic target recognition(ATR)
and biometrics. Whereas the conventional correlation filters perform di-
rectly on image pixels, in this paper, we propose a novel method, called
"feature correlation filter (FCF)", by extending the concept of correla-
tion filter to feature spaces. The FCF preserves the benefits of conven-
tional correlation filters, i.e., shift-invariant, occlusion-insensitive, and
closed-form solution, and also inherits virtues of the feature representa-
tions. Moreover, since the size of feature is often much smaller than the
size of image, the FCF method can significantly reduce the storage re-
quirement in recognition system. The comparative results on CMU-PIE
and the FRGC2.0 database show that the proposed FCFs can achieve
noteworthy performance improvement compared with their conventional
counterpart.

Keywords: Feature correlation filters, Correlation filters, MACE, Face
recognition.

1 Introduction

Face recognition has received much attention due to its potential values for appli-
cations as well as theoretical challenges. However, despite the research advances
over these years, face recognition is still a highly difficult task in practice due
to the large variability of the facial images. The variations between images of
a same face can generally be divided into two categories: *the appearance vari-
ations*, and *the man-made variations*. The appearance variations include facial
expression, pose, aging, illumination changes, etc. And the man-made variations
are mainly due to the imperfections of the capture devices and image processing
technologies, e.g. the noises from the cameras and the face registration error re-
sulting from imperfect face detections [1]. The performances of many recognition
algorithms degrade significantly in these cases.

A common approach to overcome the effect of the appearance variations is to
use face representations that are relatively insensitive to those variations. To deal
with the man-made variations, using correlation to characterize the similarity

between faces is a natural choice, because of its shift-invariant property and the optimality in the presence of additive white Gaussian noise [2].

As far as correlation is mentioned, we often think of the Matched Filter (MF), which is essentially a template image designed to match with a specific pattern. The correlation peak value provides the likelihood measure, and a shift of the test image simply produces a shift of the correlation peak. However the MFs are very sensitive to even slight deformations [3]. Hence, to capture all possible image variations, a huge number of MFs are needed, which is definitely computational and storage impractical [4]. To solve these problems, more sophisticated correlation filter design techniques are proposed in various literatures [5][6][7]. The advanced correlation filters construct a single filter or template from a set of training images and can be computed analytically and effectively using frequency domain techniques. Among such advanced correlation filters, Minimum Average Correlation Energy(MACE) filter [8] is one of the most well known. Savvides *et al.* showed that the MACE is, to some extent, robust to illumination variations [4]. However, since the correlation filters (including MACE) operate directly on the raw image pixel data (sometimes with illumination normalization [9]) without feature extraction, it may not achieve stability with respect to deformations caused by expression and pose variations, thus may not gain good generalization ability. We think this might be the Achilles heel of this approach [10].

The motivation of this work is to explore potentials of MF types of matchers by combining the advantages of the correlation method and the feature representations of faces. This leads to the idea of the proposed feature correlation filter (FCF) method. To overcome the limitation of the correlation filters, we extend the concept of correlation filter to the feature space; we do the correlation using face image features rather than directly using image pixel values. "The most common motivation for using features rather than the pixels directly is that features can act to encode ad-hoc domain knowledge that is difficult to learn using a finite quantity of training data"(Viola and Jones [11]). Moreover, the FCF method could significantly reduce the size of templates stored in the system, because the size of feature is often much smaller than the size of image.

Many representation approaches have been introduced, such as Eigenface (PCA) [12], Fisherface (LDA) [3], ICA [13], Haar features [11], Gabor wavelet features [14], etc. Among these, we choose to use the PCA features and Gabor wavelet features for constructing FCFs and to demonstrate the effectiveness of the method.

By combining the advantages of the correlation filter and feature representation of faces, we expect to gain improved generalization and discriminant capability. First, we formulate the feature extraction procedure as inner products, then perform the correlation on features rather than raw image pixel values. The FCF is constructed by solving an optimization problem in frequency domain. Finally a closed-form solution is obtained.

While the focus of this work will be on the MACE filter criterion, it should be stated that all of the results presented here are equally applicable to many

other kinds of the advanced correlation filters with appropriate changes to the respective optimization criteria.

The rest of this paper is organized as follows. Section 2 reviews the conventional MACE filter briefly. In Section 3, we introduce the basic concept of the FCF, and explain in detail how to construct the FCF. In Section 4, the experiments on PIE and FRGC2.0 database are presented. Finally, we conclude this paper and point out some further research in Section 5.

2   Minimum Average Correlation Energy (MACE) Filter

Notation: In this and the following section, we denote matrices by light face characters and vectors by bold characters. Uppercase symbols refer to the frequency domain terms, while lowercase symbols represent quantities in the space domain.

Suppose we have N facial images from a certain person. We consider each 2-dimensional image as a $d \times 1$ column vector $\mathbf{x}_i$ $(i = 1, 2, \ldots, N)$ by lexicographically reordering the image, where d is the number of pixels. The discrete Fourier transform (DFT) of $\mathbf{x}_i$ is denoted by $\mathbf{X}_i$, and we define the training image data matrix in frequency domain as $X = [\mathbf{X_1} \ \mathbf{X_2} \ \ldots \ \mathbf{X_N}]$. X is a $d \times N$ matrix.

Let the vector $\mathbf{h}$ be the correlation filter(correlation template) in the space domain and $\mathbf{H}$ be its Fourier transform. The correlation result of the ith image and the filter could be written as

$$c_i(m,n) = \mathbf{h}(m,n) \circ \mathbf{x}_i(m,n)$$
$$= \langle \mathbf{h}, \mathbf{x}_i^{m,n} \rangle \tag{1}$$

where $\circ$ denotes *correlation*, and $\langle \ , \ \rangle$ denotes *inner product* of two vectors. Here $\mathbf{x}_i^{m,n}$ is a vector obtained by circularly shifting the ith training image by m pixels horizontally and n pixels vertically, and reorder it to a 1-dimensional vector. Keep in mind that $\mathbf{h}$ and $\mathbf{x}_i$ are both 1-dimensional vectors obtained by reordering a 2-dimensional array. Since the correlation actually operates on a 2-dimensional plane, here we use two indices, m and n, to indicate the elements in these vectors. From Eq. (1), we can see that each value in the correlation output plane is simply a inner product of the filter and the shifted input image.

By Parseval's theorem, the correlation energy of $c_i(m,n)$ can be rewritten as follows using its frequency domain representation $C_i(u,v)$:

$$E_i = \sum_m \sum_n |c_i(m,n)|^2 = \frac{1}{d} \sum_u \sum_v |C_i(u,v)|^2$$
$$= \frac{1}{d} \sum_u \sum_v |\mathbf{H}(u,v)|^2 |\mathbf{X}_i(u,v)|^2 \tag{2}$$
$$= \frac{1}{d} \mathbf{h}^\dagger D_i \mathbf{h}$$

where D_i is a $d \times d$ diagonal matrix containing the power spectrum of training image $\mathbf{X}_i$ along its diagonal. The superscripts $\dagger$ denote *conjugate transpose*.

The objective of the MACE filter is to minimize the average correlation energy over the image class while simultaneously satisfying an linear constraint that the correlation values at the origin due to training images take on pre-specified values stored in vector $\mathbf{u}$, where $\mathbf{u} = [u_1 \ u_2 \ \cdots \ u_N]^T$. i.e.

$$c_i(0,0) = \mathbf{X}_i{}^{\dagger}\mathbf{H} = u_i \tag{3}$$

The average correlation energy over all training images is

$$E_{avg} = \mathbf{H}^{\dagger}D\mathbf{H} \quad \text{where} \quad D = \frac{1}{N}\sum_{i=1}^{N}D_i \tag{4}$$

Minimize E_{avg} subject to the constraint, $X^{\dagger}\mathbf{H} = \mathbf{u}$. The solution can be obtained using Lagrange multipliers [8]:

$$\mathbf{H} = D^{-1}X(X^{\dagger}D^{-1}X)^{-1}\mathbf{u} \tag{5}$$

3   Feature Correlation Filter for Face Recognition

3.1   MACE in Feature Space

All the linear feature extraction method can be expressed as follows:

$$\mathbf{y}_i = v^{\dagger}\mathbf{x}_i \tag{6}$$

where $v = [\mathbf{v_1} \ \mathbf{v_2} \ \ldots \ \mathbf{v_M}]$ is a $d \times M$ feature extraction matrix. $\mathbf{y}_i \in \mathbb{R}^M$ is the feature vector of length M. M depends on what feature to use. The discrete Fourier transform (DFT) of $\mathbf{v}_j$ is denoted by $\mathbf{V}_j$, and define $V = [\mathbf{V_1} \ \mathbf{V_2} \ \ldots \ \mathbf{V_M}]$.

Note that, if we use PCA feature, the columns of v in Eq. (6) should be the eigenfaces. Extracting Gabor features could also be formulated using Eq. (6), since applying Gabor wavelet kernel to the input image is simply an inner product.

As described in Eq. (1), the correlation plane is a collection of correlation values, each one is obtained by performing an inner product of the image and the template. In our FCF method, each correlation value is obtained by performing inner product of two feature vectors. The feature correlation output of the input image $\mathbf{x}_i$ $(i = 1, 2, \ldots, N)$ and the template $\mathbf{h}$ in spatial domain can be formulated as

$$\begin{aligned} c_i(m,n) &= \langle \mathbf{y}_h, \mathbf{y}_i \rangle = \langle v^{\dagger}\mathbf{h}, v^{\dagger}\mathbf{x_i}^{m,n} \rangle \\ &= \sum_{j=1}^{M} \mathbf{v_j}^{\dagger}h \cdot \mathbf{v_j}^{\dagger}x_i^{m,n} \end{aligned} \tag{7}$$

Here $\cdot$ denotes simply *scalar product*. $\mathbf{y}_h$ is the feature vector extracted from the template $\mathbf{h}$. Comparing Eq. (7) with Eq. (1), we can see the difference in definition between our FCF method and the conventional correlation method.

The frequency representation of $c_i(m,n)$ is,

$$C_i(u,v) = \frac{1}{d}\sum_{j=1}^{M}\mathbf{V_j^\dagger H} \cdot \mathbf{V}_j(u,v) \cdot \mathbf{X}_i{}^*(u,v) \tag{8}$$

where the superscript $*$ denotes *complex conjugation.*

With Eq. (8), we can obtain the expression of the correlation plane energy :

$$\begin{aligned}E_i &= \frac{1}{d}\sum_u \sum_v |C_i(u,v)|^2 \\ &= \frac{1}{d^3}\mathbf{H}^\dagger V V^\dagger D_i V V^\dagger \mathbf{H}\end{aligned} \tag{9}$$

The average correlation plane energy is

$$E_{avg} = \mathbf{H}^\dagger V V^\dagger D V V^\dagger \mathbf{H} \tag{10}$$

And the constraint at the origin is

$$X^\dagger V V^\dagger \mathbf{H} = \mathbf{u} \tag{11}$$

We let $\mathbf{P} = V^\dagger \mathbf{H}$, the objective function and constraint function can be rewritten as

$$E_{avg} = \mathbf{P}^\dagger V^\dagger D V \mathbf{P} \tag{12}$$

subject to

$$X^\dagger V \mathbf{P} = \mathbf{u} \tag{13}$$

Minimize E_{avg} in Eq. (12), subject to the constraint in Eq. (13), we get the closed-form solution:

$$\mathbf{P} = (V^\dagger D V)^{-1} V^\dagger X (X^\dagger V (V^\dagger D V)^{-1} V^\dagger X)^{-1} \mathbf{u} \tag{14}$$

Then, we obtain the *feature correlation filter* $\mathbf{P}$, which is a $M \times 1$ complex vector. Remind that M is the number of the features used.

Let $\tilde{D} = V^\dagger D V$ and $\tilde{X} = V^\dagger X$, Eq. (14) can be rewritten as:

$$\mathbf{P} = \tilde{D}^{-1}\tilde{X}(\tilde{X}^\dagger \tilde{D}^{-1}\tilde{X})^{-1}\mathbf{u} \tag{15}$$

Comparing Eq. (15) with Eq. (5), we can see that the FCF and MACE share the same formulation.

Note that, for each person we only need to store $\mathbf{P}$ in the recognition system, while the conventional correlation filter based system stores a template(filter) $\mathbf{H}$. Typically the size of $\mathbf{P}$ is much smaller than the size of $\mathbf{H}$. Thus the FCF can significantly reduce the size of the templates stored in the system. However, in FCF system, the $d \times M$ matrix V is also needed to be stored for verification. So this advantage on storage may not show up until the number of people enrolled in the system grows larger than M.

3.2   Face Verification Using FCF

Given a test image $\mathbf{x}_t$, the output correlation plane can be obtained using Eq. (16),

$$
\begin{aligned}
c_t(m,n) &= \mathfrak{F}^{-1}(C_t(u,v)) \\
&= \mathfrak{F}^{-1}(\sum_{i=1}^{M} \mathbf{V}_i(u,v) \cdot \mathbf{P}(i) \cdot \mathbf{X}_t^{*}(u,v))
\end{aligned}
\tag{16}
$$

where The superscript T denotes *nonconjugate transpose* of a complex vector. $\mathfrak{F}^{-1}(\cdot)$ is *inverse Fourier transform*. $\mathbf{X}_t$ is the DFT of $\mathbf{x}_t$.

We compute the peak-to-sidelobe ratio (PSR) of $c_t(m,n)$ to measure the peak sharpness.

$$
PSR = \frac{peak - mean}{\sigma}
\tag{17}
$$

where the peak is the largest value in the correlation output, the mean and σ are the average value and the standard deviation of all the correlation outputs, excluding a 5×5 region centered in the peak.

4   Experiments

To demonstrate the effectiveness of our proposed FCF method, several experiments are carried out on the illumination subset of the CMU-PIE database [15] and two subsets of FRGC2.0 data [16]. The experiments on the CMU-PIE is to show that the FCFs preserve the advantages of the conventional MACE, including shift-invariant and insensitive to partially occlusion. The experiments on the FRGC2.0 database, which has much larger size and is more difficult, will show the improved performance of our FCF methods. In the following experiments, the centers of the eyes of an image are manually detected firstly, then rotation and scaling transformations align the centers of the eyes to predefined locations. Finally, the face image is cropped to the size of 100×100 to extract the facial region, which is further normalized to zero mean and unit variance.

We choose the most representative holistic and local features , i.e. Eigenface(PCA) and Gabor wavelets, to construct the FCFs. For PCA, we used the eigenfaces that retain 98% of the total variation, i.e., totally 134 eigenfaces for experiments on CMU-PIE database, and 400 for experiments on FRGC2.0. The three most significant eigenfaces are discarded, since they capture the variation due to lighting [3]. The first 5 eigenfaces used for the CMU-PIE database are shown in Fig. 1. We used Gabor kernels at 5 scales and 4 orientations. In order to obtain a reasonably small-size feature vector, we extract the Gabor features at 49 predetermined positions, shown in Fig. 1. The total number of the Gabor features we used is 980.

The two kinds of FCFs are termed as PCA-FCF and Gabor-FCF, respectively.

The correlation outputs of MACE, PCA-FCF and Gabor-FCF of typical genuine and imposter images are shown in Fig. 2. The storage costs are listed in Table 1.

Fig. 1. Left: the first 5 Eigenfaces used for PIE database. Right: The 49 predetermined positions for Gabor feature extraction.
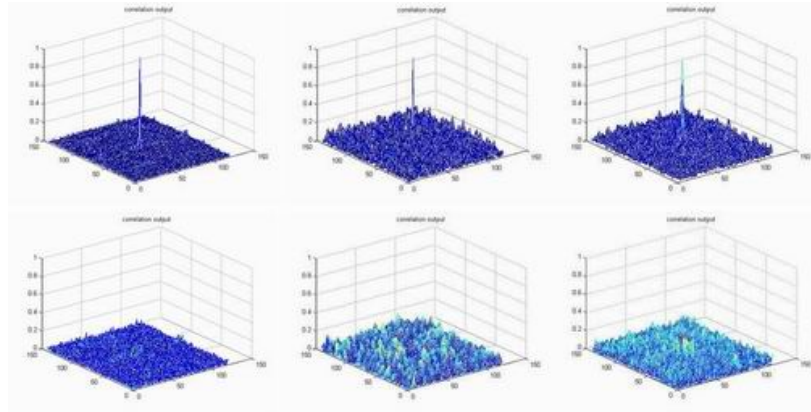

Fig. 2. The correlation output planes of MACE, PCA-FCF and Gabor-FCF, respectively(from left to right). Upper row: The correlation outputs of a typical genuine image. Lower row: The correlation outputs of a typical imposter image.

Table 1. The Storage Cost of Each Template (In KBytes)

MACE	PCA-FCF	Gabor-FCF
80.0	3.2	7.8

4.1   CMU-PIE Database

The illumination subset of the CMU-PIE database we used consists of 67 people with 21 illumination variation (shown in Fig. 3). We use 3 images per person to construct the MACE filter and the FCFs. The other 18 images are used for testing.

The test is carried out in three scenarios (shown in Fig. 3). (1) Using full size facial images for testing. (2) The testing images are partially occluded(10 pixels from top, and 15 pixels from bottom). (3) The testing images are shifted(10 pixels to the left) to simulate the situation of registration error. All these three cases are often encountered in real applications. Note that we train the MACE and the FCFs using full size images. The results are shown in Table 2.

The performance of PCA method degrades significantly in occlusion and shifting scenarios, while the MACE and the FCF methods still perform well. The

Fig. 3. Left: The 21 images of different illuminations from Person 2 of the PIE database. Right: the full size image, partially occluded image and shifted image.

Table 2. The Rank-1 recognition rates on the CMU-PIE database

Methods	Full Size	Partial	Shifted
MACE	100.00%	99.59%	100.00%
PCA	98.76%	24.30%	10.78%
PCA-FCF	100.00%	96.60%	100.00%
Gabor-FCF	100.00%	100.00%	100.00%

results demonstrate that the FCFs preserve the MACE's advantages, i.e., insensitive to occlusion and shift-invariant, which are desired in real applications.

Since there are only 67 people in the CMU-PIE database, and only illumination variation is involved(the 21 images are taken almost simultaneously, thus the expression and pose in the images are exactly the same), both MACE and FCF obtain nearly perfect recognition rates.

4.2 FRGC2.0 Database

To further evaluate the performance of the FCF-based method and show its improved recognition ability, we use the FRGC2.0 database [16], which is much larger and more difficult. Two subsets are used: the controlled subset and uncontrolled one. The controlled images were taken in a studio setting, are full frontal facial images taken under two lighting conditions and with two facial expressions. The uncontrolled images were taken in varying illumination conditions; e.g., hallways, atria, or outdoors. Each set of uncontrolled images contains two expressions. The images are taken in different semesters, and some images are blurred. Some sample images are shown in Fig. 4.

From each subset, we randomly select 222 people and 20 images per person. 5 images of each person are used for training, and the other 15 for testing. The test results are summarized in the Table 3.

The results show that the FCF methods significantly outperform the MACE. This is mainly derived from the combination of the advanced correlation filters and the feature representation.

Fig. 4. Some example image for FRGC database. Top row, controlled images. Bottom row, uncontrolled images.

Table 3. The Rank-1 recognition rates on the FRGC2.0 database

Methods	Controlled	Uncontrolled
MACE	80.99%	34.29%
PCA	76.73%	51.11%
PCA-FCF	**91.92%**	**64.14%**
Gabor-FCF	**93.84%**	**67.57%**

5   Conclusions

This work aims to bring a promising perspective for using correlation filters in biometric recognition. The major contribution is the idea of incorporating the feature representations of faces into correlation filter, to take the advantages of advanced correlation filter and feature representation. We formulated the feature extraction as inner products, and obtained a closed-form solution by solving an optimization problem in frequency domain. The experimental results show that the proposed FCF methods significantly outperform their conventional counterpart. Working with fewer features, the FCF method can significantly reduce the size of the template stored in the system.

While the FCFs in this paper is in essence a linear method, in future work, we would consider the nonlinear extension. We would also consider other types of features to used, and selection of the best features and combinations of different feature types for correlation.

References

1. Rentzeperis, E., Stergiou, A., Pnevmatikakis, A., Polymenakos, L.: Impact of Face Registration Errors on Recognition, presented at Artificial Intelligence Applications and Innovations (AIAI06) (2006)
2. Vanderlugt, A.: Signal detection by complex spatial filtering. IEEE Trans. Inf. Theory 10, 139–145 (1964)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions 19, 711–720 (1997)
4. Savvides, M., Kumar, B.V.K.V., Khosla, P.K.: Corefaces - Robust shift invariant PCA based correlation filter for illumination tolerant face recognition. IEEE Comp. Vision and Patt. Rec (CVPR) (2004)
5. Kumar, B.V.K.V., Mahalanobis, A., Juday, R.D.: Correlation Pattern Recognition. Cambridge Univ. Press, Cambridge, U.K. (2005)
6. Vijara Kumar, B.V.K.: Minimum variance synthetic discriminant functions. J. Opt. Soc. Amer. A 3, 1579C1584 (1986)
7. Mahalanobis, A., Vijaya Kumar, B.V.K., Song, S., Sims, S.R.F., Epperson, J.: Unconstrained correlation filters. Appl. Opt. 33, 3751C3759 (1994)
8. Mahalanobis, A., Kumar, B.V.K.V., Casasent, D.: Minimum average correlation energy filters. Appl. Opt. 26, 3630–3633 (1987)
9. Savvides, M., Kumar, B.V.K.V.: Illumination normalization using logarithm transforms for face authentication. In: 4th Int. Conf. AVBPA (2003)
10. Vijaya Kumar, B.V.K., Savvides, M., Xie, C.: Correlation Pattern Recognition for Face Recognition. Proceedings of the IEEE 94, 1963–1976 (2006)
11. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition (CVPR) (2001)
12. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cognitive Neuroscience 3, 72–86 (1991)
13. Bartlett, M.S., Lades, H.M., Sejnowski, T.J.: Independent component representations for face recognition. In: Proceedings of the SPIE, Conference on Human Vision and Electronic Imaging (1998)
14. Wiskott, L., Fellous, J., Kruger, N., malsburg, C.V.: Face recognition by elastic bunch graph matching. Pattern Analysis and Machine Intelligence, IEEE Transactions 19, 775–779 (1997)
15. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces, Robotics Institute, Carnegie Mellon University CMU-RI-TR-01-02 (2001)
16. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Jin, C., Hoffman, K., Marques, J., Jaesik, M., Worek, W.: Overview of the face recognition grand challenge. In: Proc. IEEE Conf. Comp. Vision Pattern Rec. (CVPR) (2005)