Research Article

Directional Correlation Filter Bank for Robust Head Pose Estimation and Face Recognition

Si Chen,¹ Dong Yan ,² and Yan Yan ²

¹School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China

²Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Yan Yan; yanyan@xmu.edu.cn

Received 24 July 2018; Accepted 25 September 2018; Published 21 October 2018

Academic Editor: Krzysztof Okarma

Copyright © 2018 Si Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During the past few decades, face recognition has been an active research area in pattern recognition and computer vision due to its wide range of applications. However, one of the most challenging problems encountered by face recognition is the difficulty of handling large head pose variations. Therefore, the efficient and effective head pose estimation is a critical step of face recognition. In this paper, a novel feature extraction framework, called Directional Correlation Filter Bank (DCFB), is presented for head pose estimation. Specifically, in the proposed framework, the 1-Dimensional Optimal Tradeoff Filters (1D-OTF) corresponding to different head poses are simultaneously and jointly designed in the low-dimensional linear subspace. Different from the traditional methods that heavily rely on the precise localization of the key facial feature points, our proposed framework exploits the frequency domain of the face images, which effectively captures the high-order statistics of faces. As a result, the obtained features are compact and discriminative. Experimental results on public face databases with large head pose variations show the superior performance obtained by the proposed framework on the tasks of both head pose estimation and face recognition.

1. Introduction

Face recognition has attracted significant attention in computer vision and pattern recognition. It has a wide range of practical engineering applications, including access control, video surveillance, and human computer interaction. Many works have been developed towards robust face recognition systems [1–4], and they usually encounter several challenging issues, such as occlusions, illumination changes, and pose variations [1]. Among these issues, face recognition under large head pose variations is one of the most challenging problems [5–10], mainly due to the fact that the face appearance dramatically changes under different head poses.

Generally speaking, the image can be described by using its two-dimensional Fourier transform in the spatial frequency domain, where significant advantages can be exploited (such as shift-invariance, graceful degradation, and closed-form solution [10]). However, most of the traditional

methods work in the image domain, which may lose much important information for image recognition. The main objective of this paper is to investigate the benefits of using the spatial frequency domain representation for effective head pose estimation and face recognition.

In this paper, we propose a novel feature extraction framework, called Directional Correlation Filter Bank (DCFB), for robust head pose estimation and face recognition by taking advantage of the correlation filter technique. Specifically, in the proposed framework, the 1-Dimensional Optimal Tradeoff Filters (1D-OTF) corresponding to different head poses are jointly designed in the low-dimensional subspace obtained by Principal Component Analysis (PCA). Then, the origin correlation outputs corresponding to different 1D-OTFs constitute a compact and discriminative feature vector, which can be used for the final head pose classification. Based on the shift-invariance property of the correlation filter technique, one distinguishing advantage of the propose

framework is that it does not need the precise localization of the key facial feature points used in the traditional methods. Experimental results on public face databases show that the obtained compact feature vector is very discriminative for head pose estimation and face recognition.

In summary, the main contributions of this paper are given as follows. Firstly, we develop a novel feature extraction framework for head pose estimation. The extracted directional feature vector (a three-dimensional feature vector) effectively encodes the information for pose estimation. Secondly, an effective correlation filter is designed. Compared with the traditional correlation filter designed in the 2D image space, the proposed filter is designed in the 1D low-dimensional PCA subspace, which not only reduces the computational complexity but also improves the final performance. Thirdly, experimental results show the superiority of the proposed framework for head pose estimation and face recognition. Moreover, we show that the proposed framework can also be extended to occlusion estimation.

A preliminary version of this work was reported in [11]. However, we have made several significant improvements. More specifically, we reformulate the original method and offer more mathematical details and motivations of the proposed method. We conduct extensive experiments to demonstrate the superiority of the proposed method for head pose estimation and face recognition. Moreover, we successfully extend the proposed method to the application of occlusion estimation, which validates the generalization ability of the proposed method for different tasks.

The remainder of the paper is organized as follows: Section 2 describes related work, where head pose estimation methods and correlation filter methods are, respectively, discussed. Section 3 explains the methodology of the proposed framework in detail. Section 4 shows extensive experimental results on the tasks of head pose estimation and face recognition. Finally, we make the conclusions in Section 5.

2. Related Work

Our work mainly focuses on the head pose estimation for robust face recognition [5]. Existing head pose estimation methods can be roughly classified into two categories: model-based methods [12–14] and appearance-based methods [7, 8, 15–26].

On one hand, model-based methods aim to construct a 3D model of human head, where a large number of 3D samples are usually required during the training stage. On the other hand, appearance-based methods receive much attention recently. These methods depend on pose-invariant local features or the localization of facial feature points [7], which include appearance template methods [8, 17], detector array methods [15], nonlinear regression methods [7, 16–20], manifold embedding methods [21–23], and Convolutional Neural Network (CNN) based methods [24–26].

For instance, Kim et al. [17] estimated the head pose based on a part-based face matching algorithm. But the performance of head pose estimation heavily relies on the accuracy of the localization of facial features. Ho and Chellappa [8] proposed using the dense SIFT features [9] for head pose estimation. Jones and Viola [15] built different face detectors

for different head poses based on the decision tree, where the face detectors are trained separately. Dantone et al. [7] demonstrated the benefits of Conditional Regression Forests (CRF) by modeling the appearance and the location of facial feature points that are conditionally dependent to the head pose. Guo et al. [18] proposed to combine regression and classification for head pose estimation. Lu et al. [22] proposed an ordinary preserving manifold analysis method for head pose estimation. Wei et al. [23] developed a robust face pose estimation method by using the geometry-preserving visual phase.

Recently, Convolutional Neural Network (CNN) has made significant progress for head pose estimation. Xu et al. [24] proposed a framework to jointly perform head pose estimation and face alignment based on the global and local CNN features. Later, Patacchiola and Cangelosi [25] developed a head pose estimation method using CNN and adaptive gradient methods to obtain the promising performance. Ahn et al. [26] used the multitask deep CNN for real-time head pose estimation. However, most of the CNN based methods suffer from high computational complexity.

The correlation filter technique has been shown to be effective for the task of object recognition [10], due to its desirable properties, such as graceful degradation, shift-invariance, and closed-form solutions. Graceful degradation means that although some pixels in a face image are occluded or contaminated, the face can be recognized since only the peak is decreased on the correlation plane (but can still be discernible). Shift-invariance shows that any shift operation in an input image will result in the correlation output being shifted by the same distance. Therefore, the correlation filter based method does not need to align the face images, while the traditional feature-based head pose estimation methods are required to align the face images before feature extraction. Finally, instead of the iterative operations used in many filtering methods (e.g., [27]), correlation filters usually have closed-form solutions.

Among decades of development, many types of correlation filters have been proposed. For example, Mahalanobis et al. [28] proposed the Minimum Average Correlation Energy (MACE) filter. Kumar [29] proposed the Minimum Variance Synthetic Discriminant Function (MVSDF) filter. The Optimal Tradeoff Filter (OTF) in [30] combines the MACE filter and the MVSDF filter to produce the sharp correlation peaks and suppress noises. Note that all these correlation filters are designed based on the 2D image space.

Correlation filter is not only employed in face recognition [10], but also applied in other pattern recognition tasks. Venkataramani et al. [31] showed the application of correlation filter in fingerprint verification for access control. Henning et al. [32] illustrated the application of correlation filter in palmprint identification and verification. Henriques et al. [33] proposed the popular Kernel Correlation Filter (KCF) for robust object tracking. In the paper, we will extend the application of correlation filter towards robust head pose estimation and face recognition.

3. Methodology

In Section 3.1, we present a novel feature extraction framework, called DCFB, for head pose estimation. In this framework,

FIGURE 1: The feature extraction framework of DCFB.

the 1-Dimensional Optimal Tradeoff Filter (1D-OTF) is designed for each directional correlation filter in Section 3.2.

3.1. Directional Correlation Filter Bank. The proposed feature extraction framework, i.e., the Directional Correlation Filter Bank (DCFb) is illustrated in Figure 1. Firstly, a high-dimensional feature vector is extracted for each input image. After that, Principal Component Analysis (PCA) is used to perform dimension reduction so that the prominent information of the face is preserved while the noise is reduced. Then, DCFb consists of the three correlation filters corresponding to left-profile pose, frontal pose, and right-profile pose, respectively, which are separately used to correlate with the low-dimensional PCA features to generate the final direction feature vector.

In Figure 1, the 1-Dimensional Optimal Tradeoff Filter (called 1D-OTF) is proposed for each correlation filter in the 1D frequency domain. Note that the Optimal Tradeoff Filter (OTF) has been successfully applied in face recognition [10], but its computational complexity is very high. Besides, OTF is relatively sensitive to variations in illumination and facial expressions due to the fact that the filter is designed in the original 2D image space. Different from the traditional OTF which is based on the 2D image space, we propose to design 1D-OTF in the low-dimensional PCA subspace. A directional correlation filter bank consisting of three 1D-OTF correlation filters is obtained. In other words, the DCFb only concerns the pose information and ignores the person identity information. More specifically, each correlation filter in DCFb tries to discriminate one specific pose from all the other poses. Finally, a direction feature vector is obtained for head pose estimation. In this paper, the pose of a head can be easily estimated by using the simple nearest neighbor classifier, which is based on the Euclidean distance of the direction feature vectors.

3.2. 1-Dimensional Optimal Tradeoff Filter. In the proposed DCFb, the PCA is firstly used to preserve the dominant information and remove the noises in the face image. Moreover, we design the 1D-OTF in the PCA feature subspace. The advantages of 1D-OTF are that the computational cost can be reduced significantly and the variations caused by illumination and pose can be effectively alleviated.

The PCA is a classic and popular dimension reduction method [34]. Formally, if we have the N face images denoted

as $\{x_1, x_2, \dots, x_N\}$, where $x_i \in R^{H \times 1}$ and each image belongs to one class. The total scatter matrix S_T is defined as

$$S_T = \sum_{k=1}^N (x_k - \mu)(x_k - \mu)^T, \quad (1)$$

where N is the number of all the training images and μ is the mean of all the training images.

Projection matrix $W_{pca} \in R^{H \times K}$ is defined as

$$W_{pca} = \arg \max_W |W^T S_T W| = [w_1, w_2, \dots, w_K], \quad (2)$$

where $[w_i \mid i = 1, 2, \dots, K]$ is the set of the eigenvectors of S_T corresponding to the K largest eigen-values.

The new feature vector $y_i \in R^{m \times 1}$ ($i = 1, 2, \dots, N$) is obtained by using the following transformation:

$$y_i = W^T x_i. \quad (3)$$

After performing PCA, we obtain a 1D low-dimensional feature vector for each face image. Then, a variant of OTF is designed based on the frequency domain of the obtained feature vector. Unlike the traditional OTF which is designed in the 2D image space, the proposed 1D-OTF is designed in the low-dimensional PCA subspace. Then, the three different 1D-OTF correlation filters are obtained, where each filter corresponds to a specific pose thus distinguishing from the other two poses.

Specifically, the 1D-OTF is derived by combining the 1D-MACE (Minimal Average Correlation Energy) filter and the 1D-MVSDF (Minimal Variance Synthetic Discriminant Function) filter. The objective of the 1D-MACE filter is to minimize the average correlation energy (ACE), which can be formulated as

$$\begin{aligned} \min \frac{1}{N} \sum_{t=1}^N \sum_{p=0}^{m-1} |g_t(p)|^2 &= \min \frac{1}{Nm} \sum_{t=1}^N \sum_{p=0}^m |G_t(p)|^2 \\ &= \min \frac{1}{Nm} \sum_{t=1}^N \sum_{p=0}^m |F(p)|^2 |Y_t(p)|^2 = \min_F \frac{1}{m} F^+ Q F, \end{aligned} \quad (4)$$

where $G_t(p)$, $F(p)$, and $Y_t(p)$ are the 1D Fourier transforms of the output g_t , the correlation filter f and the input y_t , respectively; F is the vector of $F(p)$; “+” means the conjugate transpose; Q is a diagonal matrix whose diagonal entries are the average power spectrum of all N features.

FIGURE 2: Some examples from the PIE and HPI databases.

The origin value of the correlation output is $g_t(0) = \mathbf{Y}_t^+ \mathbf{F}$. The constraints of the 1D-MACE filter are that the values of the outputs at the origin are equal to 1 for the authentic images (corresponding to the images of a specific pose) and 0 for the imposter images (corresponding to the images of the other poses), expressed as

$$\mathbf{Y}^+ \mathbf{F} = \mathbf{c}, \quad (5)$$

where $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N]$ is the 1D Fourier transform of the low-dimensional features obtained by PCA; $\mathbf{c} = [c_1, c_2, \dots, c_N]$ is a $N \times 1$ vector and c_t denotes the correlation peak amplitude of the t th training image; c_t is equal to 1 for the authentic images and 0 for the imposter images.

Therefore, the objective of the 1D-MACE filter is

$$\begin{aligned} \min_{\mathbf{F}} \quad & \mathbf{F}^+ \mathbf{Q} \mathbf{F}; \\ \text{s.t.} \quad & \mathbf{Y}^+ \mathbf{F} = \mathbf{c}. \end{aligned} \quad (6)$$

As a result, the solution of the above objective function is obtained by using the method of Lagrange multipliers. The optimum solution of 1D-MACE is

$$\mathbf{F}_{\text{1D-MACE}} = \mathbf{Q}^{-1} \mathbf{Y} (\mathbf{Y}^+ \mathbf{Q}^{-1} \mathbf{Y})^{-1} \mathbf{c}. \quad (7)$$

The solution of the 1D-MVSDF filter is derived in the same way as the 1D-MACE filter. The optimum solution is

$$\mathbf{F}_{\text{1D-MVSDF}} = \mathbf{G}^{-1} \mathbf{Y} (\mathbf{Y}^+ \mathbf{G}^{-1} \mathbf{Y})^{-1} \mathbf{c}, \quad (8)$$

where $\mathbf{G}$ is an identity matrix if the input noise is modelled as the white noise.

Based on the combination of (7) and (8), 1D-OTF is written as

$$\mathbf{F}_{\text{1D-OTF}} = \mathbf{T}^{-1} \mathbf{Y} (\mathbf{Y}^+ \mathbf{T}^{-1} \mathbf{Y})^{-1} \mathbf{c}. \quad (9)$$

Here, $\mathbf{T} = \alpha \mathbf{Q} + (1 - \alpha) \mathbf{G}$ where α ($0 \leq \alpha \leq 1$) is a trade-off parameter ($\alpha = 0$ leads to the 1D-MVSDF filter and $\alpha = 1$ leads to the 1D-MACE filter).

4. Experimental Results and Analysis

In this section, we firstly introduce the face databases used in the experiments in Section 4.1. Then, we give the experimental results obtained by the different methods for the task of head pose estimation in Section 4.2. Next, the generalization ability of the proposed framework is shown in Section 4.3. To further validate the effectiveness of the proposed framework, we extend our framework for the task of occlusion estimation in Section 4.4. Finally, we describe the experimental results for the task of face recognition in Section 4.5.

4.1. Face Databases. In this paper, three popular face databases with large pose variations, including PIE [35], HPI [36], and UMIST [37], are used to demonstrate the performance of the proposed method for head pose estimation and face recognition. Figure 2 shows some examples from the PIE and HPI databases. Besides, the AR database [38] is used to evaluate the performance for occlusion estimation, which contains occlusions in face images.

The PIE face database contains 41,368 images of 68 different persons with variations in pose, illumination, and expression. We choose 612 images with three different poses, that is, left-profile ($[-45^\circ, -15^\circ]$), frontal ($[-15^\circ, 15^\circ]$), and right-profile ($[15^\circ, 45^\circ]$). And each pose has nine images for each person. The HPI face database contains 15 persons with various poses. We choose ten images for both the left-profile and right-profile poses and six images for the frontal pose for each person. The UMIST face database consists of 564 different pose images of 19 persons. We, respectively, choose six images for each pose in our experiments. The AR face database contains over 4000 face images of 126 people, including the frontal view of faces with different facial expressions and occlusions (including sun-glasses and scarfs). The images of 120 individuals are taken in two sessions (separated by two weeks) and each session contains 13 color images.

All the faces in the images are cropped and resized to the size of 64×64 . For all the databases, we randomly choose 30% images as the training set and the rest are used as the test set. We also compare our method with several other competing methods (see the following subsections for details). The experiments are repeated 30 times. We report

FIGURE 3: Performance of head pose estimation under different block sizes.

the average pose/face classification rates obtained by all the competing methods.

4.2. Results on Head Pose Estimation. In this section, we show the results of head pose estimation obtained by the competing methods (including PCA [34], LDA [39], QRLDA [40], and OTF [41]) and our proposed DCFB method. To validate the effectiveness of the proposed method, we use three different face representations, including gray, Gabor [42], and HOG [43] features.

The HOG features use the sliding block to record the gradient information of images, where the block size will greatly affect the performance of head pose estimation. The large block catches the global information while the small block records the details. The head pose estimation usually needs the outline information rather than the identity characteristic, which indicates that we should use the large block size. To verify our assumption, we set ten different kinds of sizes of blocks used in the HOG features for head pose estimation. From Figure 3, we can see that the proposed method with the larger block size have better head pose estimation performance than the smaller size. However, if we use one block to cover an image, we cannot get the best performance, because not only the outline gradient but also the gradient information of the mouth and the eye will be recorded, thus leading to the performance decrease. The best results can be obtained by using four blocks and each block has nine bins. Therefore, we only need 36-dimensional features to estimate the head pose. We will use the 36-dimensional HOG features for the other following experiments.

Table 1 shows the results obtained by all the competing methods for head pose estimation based on the three types of features. From Table 1, we can see that the proposed method obtains the best performance by employing the three different face representations for the most cases, which shows the feasibility and robustness of DCFB. In particular, the DCFB achieves the 100% accuracy on the PIE and UMIST databases.

TABLE 1: The results obtained by the different competing methods for head pose estimation based on three feature representations.

Method	Features	Face Database		
		PIE	HPI	UMIST
PCA	Gray	57.56%	78.04%	52.34%
	Gabor	43.73%	75.71%	32.74%
	HOG	89.33%	89.68%	86.55%
LDA	Gray	62.20%	82.34%	64.09%
	Gabor	62.20%	75.60%	61.02%
	HOG	93.54%	90.64%	87.28%
QRLDA	Gray	77.05%	76.00%	48.48%
	Gabor	60.62%	70.75%	16.04%
	HOG	94.98%	78.00%	60.91%
CFA	Gray	93.23%	64.36%	46.74%
	Gabor	73.63%	56.09%	6.47%
	HOG	93.94%	87.18%	84.54%
DCFB	Gray	94.11%	76.55%	70.56%
	Gabor	75.08%	87.09%	79.37%
	HOG	100.00%	99.61%	100.00%

Most of methods using the HOG features can achieve the higher performance for pose estimation than those using the gray and Gabor features. This is because that the HOG features use the histogram of gradients to describe the shape of a face image, where the information of face directions is included. Therefore, the HOG features are more effective for head pose estimation. In contrast, the performance of the Gabor features is worse than the HOG features. This is because the Gabor features can extract features that are insensitive to the head pose variations.

To demonstrate the superiority of the proposed DCFB for head pose estimation, we further show the distance distributions between the test images and the template images for all the competing methods on the PIE database (see Figure 4). We can see that PCA is not suitable for head pose

FIGURE 4: Distance distributions between the test images and the templates. X axis represents different persons (i.e., class ID) while Y axis means the distance between a test image and the templates. The blue points (“+”) indicate the distances between the test images and the left-profile template; the green points (“o”) show the distances between the test images and the frontal template; and the red points (“*”) represent the distances between the test images and the right-profile template. The test images are left-profile images (left column), frontal images (middle column), and right-profile images (right column), respectively.

estimation since the distances between different pose images are not big enough to be distinguished. Although PCA can effectively reduce the noises, the extracted features are not distinguishable for head pose estimation. LDA achieves better results than PCA, since LDA considers the class information. LDA and QRLDA achieve the similar performance because these two methods try to differentiate all the classes. In contrast, OTF gives the better distance distributions than LDA and QRLDA. However, compared with OTF, DCFB shows the excellent capability to separate the distance distributions for different templates, since a specific filter is designed to classify between one pose and the other two poses in the PCA feature subspace. This makes the head pose estimation more effective and robust. The effectiveness of DCFB can be attributed to the superiority of the correlation filter in dealing with frequency domain representations and the advantage of the filter bank design in learning head pose features.

4.3. Generalization Ability. In this section, we show the generalization ability of the proposed framework. That is, we use one database for training while applying the other database for testing. In this way, the generalization ability of the proposed method across different databases can be evaluated.

More specifically, we randomly choose 30% images of one database to train the proposed DCFB and all the images in the other database for testing. The results are given in Tables 2–4. We can see that the proposed DCFB using the HOG features obtains the excellent performance in all the experiments, which demonstrates the effectiveness of the correlation filter for cross database validation. Note that the DCFB using the gray features does not work well in the experiments. This is mainly because the gray features that only preserve the appearance information. The DCFB using the HOG features achieve the better results than the DCFB using the Gabor features. There are three main reasons for the good generalization ability of DCFB. Firstly, the correlation filter bank is designed in the low-dimensional PCA subspace, which not only significantly suppresses the noises in the facial images but also focuses on extracting intrinsic features for head pose estimation. Secondly, the proposed framework exploits the frequency domain representation, which inherits the advantages of graceful degradation and shift-invariance in the correlation filter. Therefore, the precise localization of the key facial feature points for different individuals is not required. Thirdly, the HOG features are used to compute the gradient histograms so as to effectively reduce the appearance differences between individuals. Therefore, one advantage of using the proposed DCFB is that it provides the robustness against overfitting (i.e., we can train on one database and test on the other database with completely different individuals). In summary, the proposed framework using the HOG features can effectively extract the discriminative information, and thus the superiority generalization ability is obtained for head pose estimation.

4.4. Occlusion Estimation. In this section, we show the results of the proposed framework for the task of occlusion estimation. Occlusion estimation is to determine whether

TABLE 2: Generalization ability of the proposed framework (PIE for training, and HPI and UMIST for testing).

Training database	Test database	
PIE	HPI	UMIST
Gray	87.23%	61.64%
Gabor	74.20%	56.51%
HOG	100%	100%

TABLE 3: Generalization ability of the proposed framework (HPI for training, and PIE and UMIST for testing).

Training database	Test database	
HPI	PIE	UMIST
Gray	62.93%	63.57%
Gabor	72.50%	72.72%
HOG	99.80%	99.72%

TABLE 4: Generalization ability of the proposed framework (UMIST for training, and PIE and HPI for testing).

Training database	Test database	
UMIST	PIE	HPI
Gray	35.62%	36.36%
Gabor	58.62%	61.61%
HOG	98.91%	99.20%

the given input image is occluded or not. Therefore, there are only two 1D-OTFs (one for occluded images and the other for non-occluded images) designed in DCFB. The AR database is used for evaluation. The images of the AR database are separated into two classes, i.e., occluded images and nonoccluded images. The gray, HOG and Gabor features are extracted from the AR database for all the images. Figure 5 shows the distance distribution of occlusion estimation of the proposed framework using the gray, HOG, and Gabor features. As demonstrated in Figure 5, the occluded images can be better classified than the nonoccluded images. This is mainly because that there is a distinguish gap between two classes. The proposed framework using the gray features does not work well for non-occluded images (it only obtains 67% classification rate). The Gabor features can classify images correctly, but the gap between the occluded images and nonoccluded images is not distinct. The gray features are sensitive to illumination changes, while the Gabor features mainly depict the textures of facial images. In contrast, the HOG features effectively characterize the edge variations, which are of great importance to occlusion estimation. In a word, the proposed framework using the HOG features achieves the steady and excellent performance in occlusion estimation.

4.5. Face Recognition with Head Pose Estimation. In this section, we show the results on face recognition by taking advantage of head pose estimation. Note that if a face is not frontal, then its appearance is usually surrounded by the background. The information provided by the background is

FIGURE 5: Distance distributions between the test images and the templates. X axis represents different persons (i.e., class ID) while Y axis means the distance between a test image and the templates. The blue points (“+”) indicate the distances between the test images and the non-occluded template; green points (“o”) show the distances between the test images and the occluded template; The test images are nonoccluded images (first row) and occluded images (second row). The proposed DCFB framework using the gray, HOG and Gabor features are given in the left column, middle column, and right column, respectively.

not effective for face recognition. To overcome such a problem, we use the partial information (i.e., the nonbackground face region) to perform the face recognition process. The CFA method [10] (without relying on head pose estimation) is used as the baseline face recognition method. For head pose estimation, we use the OTF and DCFB for comparisons. Table 5 shows the results of the recognition accuracy obtained by all the competing methods. For each method, three different face representations (i.e., gray, Gabor, and HOG) are used for face recognition. The CFA method directly performs face recognition without head pose estimation. The method (A+B) denotes the combination of the head pose estimation method A (OTF or DCFB) and the face recognition method B (CFA in our paper). Note that we use the HOG features for both DCFB and OTF.

From Table 5, we can see that DCFB+CFA with the Gabor features obtains the best recognition performance among three feature representations. By using the head pose estimation, the performance of face recognition (i.e., DCFB+CFA) can be significantly improved. Specifically, DCFB+CFA with the Gabor features can improve at least 40% than CFA with the Gabor features. DCFB+CFA can obtain the higher

TABLE 5: The recognition accuracy obtained by different methods for face recognition based on the three feature representations.

Method	Features	Face Database		
		PIE	HPI	UMIST
CFA	Gray	13.52%	24.17%	41.54%
	HOG	50.44%	25.78%	66.61%
	Gabor	31.99%	28.10%	34.68%
OTF+CFA	Gray	55.75%	50.00%	61.05%
	HOG	60.23%	53.73%	84.23%
	Gabor	67.62%	51.16%	41.07%
DCFB+CFA	Gray	27.80%	61.14%	61.83%
	HOG	60.72%	57.86%	92.73%
	Gabor	86.08%	80.73%	97.54%

recognition accuracy than OTF+CFA because DCFB has the better head pose estimation ability than OTF. In addition, DCFB+CFA with the Gabor features achieves the highest recognition accuracy among all the competing methods. CFA achieves the worst performance since head pose estimation

is not used. In this way, the intraclass variations caused by pose differences are much larger than the interclass variations caused by individual differences. As a result, the recognition rates greatly drop. Note that different from head pose estimation where the HOG features show the best performance for DCFB, Gabor is much more effective than Gray and HOG for DCFB in face recognition, due to the fact that Gabor is less sensitive to pose variations during the filtering steps. From the above results, the robust head pose estimation is an essential step for face recognition.

5. Conclusions

In this paper, a novel feature extraction framework DCFB has been proposed for robust head pose estimation and face recognition. An effective 1-Dimensional optimal tradeoff filter, called 1D-OTF, is designed in DCFB by using the frequency representations of 1D features on the linear subspace obtained by principal component analysis. Experimental results have demonstrated the effectiveness of DCFB for head pose estimation, face recognition, and occlusion estimation. Moreover, the superiority of the generalization capability of DCFB has been shown.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

An earlier version of this paper was presented as a conference paper at the 4th International Conference on Intelligent Science and Big Data Engineering, Beijing, China, 2013.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the National Key R&D Program of China (no. 2017YFB1302400), by the National Nature Science Foundation of China (nos. 61503315, 61571379) and the National Nature Science Foundation of Fujian (nos. 2018J01576, 2017J01127). The authors are indebted to Professor Hanzhi Wang for helpful criticism of an earlier version of this paper.

References

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2017, pp. 6738–6746, USA, July 2017.
- [3] F. Z. Chelali and A. Djeradi, "Face Recognition Using MLP and RBF Neural Network with Gabor and Discrete Wavelet Transform Characterization: A Comparative Study," *Mathematical Problems in Engineering*, vol. 2015, 2015.
- [4] X. Wu, B. Fang, Y. Tang, X. Zeng, and C. Xing, "Reconstructed error and linear representation coefficients restricted by ℓ_1 -minimization for face recognition under different illumination and occlusion," *Mathematical Problems in Engineering*, vol. 2017, no. 5, Article ID 1458412, pp. 1–16, 2017.
- [5] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, 2009.
- [6] T. F. Cootes, K. Walker, and C. J. Taylor, "View-based active appearance models," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, FG 2000, pp. 227–232, France, March 2000.
- [7] M. Dantone, J. Gall, G. Fanelli, and L. Van Gool, "Real-time facial feature detection using conditional regression forests," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2578–2585, IEEE, Providence, RI, USA, June 2012.
- [8] H. T. Ho and R. Chellappa, "Automatic head pose estimation using randomly projected dense SIFT descriptors," in *Proceedings of the 2012 19th IEEE International Conference on Image Processing*, ICIP 2012, pp. 153–156, USA, October 2012.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] B. V. K. Vijaya Kumar, M. Savvides, and C. Xie, "Correlation pattern recognition for face recognition," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1963–1975, 2006.
- [11] D. Yan, Y. Yan, and H. Wang, "Robust Head Pose Estimation with a New Principal Optimal Tradeoff Filter," in *Intelligence Science and Big Data Engineering*, vol. 8261 of *Lecture Notes in Computer Science*, pp. 320–327, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [12] . Beymer, "Face recognition under varying pose," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 756–761, Seattle, WA, USA, June 1993.
- [13] J. Sherrah, S. Gong, and E. J. Ong, "Face distributions in similarity space under varying head pose," *Image and Vision Computing*, vol. 19, no. 12, pp. 807–819, 2001.
- [14] Y. Yu, K. A. F. Mora, and J.-M. Odobez, "Robust and Accurate 3D Head Pose Estimation through 3DMM and Online Head Model Reconstruction," in *Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition*, FG 2017, pp. 711–718, USA, June 2017.
- [15] M. Jones and P. Viola, "Fast multi-view face detection," *Proceedings of Computer Vision and Pattern Recognition*, pp. 276–286, 2003.
- [16] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi, "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems*, ITSC 2007, pp. 709–714, USA, October 2007.
- [17] K. H. Kim, C. Zhang, Z. Zhang et al., "Robust part-based face matching with multiple templates," in *Proceedings of the 2013 10th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2013)*, pp. 1–7, Shanghai, China, April 2013.
- [18] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Head pose estimation: Classification or regression?" in *Proceedings of the 2008*

- 19th International Conference on Pattern Recognition (ICPR)*, pp. 1–4, Tampa, FL, USA, December 2008.
- [19] M. A. Haj, J. Gonzalez, and L. S. Davis, "On partial least squares in head pose estimation: how to simultaneously deal with misalignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2602–2609, IEEE, Providence, RI, USA, June 2012.
 - [20] C. Gou, Y. Wu, F. Wang, and Q. Ji, "Coupled cascade regression for simultaneous facial landmark detection and head pose estimation," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2906–2910, Beijing, September 2017.
 - [21] S. Yan, Z. Zhang, Y. Fu et al., "Learning a person-independent representation for precise 3D pose estimation," in *Proceedings of the International Workshop Classification of Events Activities and Relationships*, pp. 297–306, 2008.
 - [22] J. Lu and Y.-P. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 2, pp. 249–258, 2013.
 - [23] W. Wei, C. Tian, and Y. Zhang, "Robust face pose classification method based on geometry-preserving visual phrase," pp. 3342–3346.
 - [24] X. Xu and I. A. Kakadiaris, "Joint Head Pose Estimation and Face Alignment Framework Using Global and Local CNN Features," in *Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017*, pp. 642–649, USA, June 2017.
 - [25] M. Patacchiola and A. Cangelosi, "Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods," *Pattern Recognition*, vol. 71, pp. 132–143, 2017.
 - [26] B. Ahn, D.-G. Choi, J. Park, and I. S. Kweon, "Real-time head pose estimation using multi-task deep neural network," *Robotics and Autonomous Systems*, vol. 103, pp. 1–12, 2018.
 - [27] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–123, 2013.
 - [28] A. Mahalanobis, B. V. K. V. Kuma, and D. Casasent, "Minimum average correlation energy filters," *Applied Optics*, vol. 26, no. 17, pp. 3633–3640, 1987.
 - [29] B. V. K. V. Vijaya Kumar, "Minimum-variance synthetic discriminant functions," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 3, no. 10, pp. 1579–1584, 1986.
 - [30] P. Refregier, "Filter design for optical pattern recognition: Multicriteria optimization approach," *Optics Express*, vol. 15, no. 15, pp. 854–856, 1990.
 - [31] K. Venkataramani and B. V. K. V. Kumar, "Performance of composite correlation filters in fingerprint verification," *Optical Engineering*, vol. 43, no. 8, pp. 1820–1827, 2004.
 - [32] P. Hennings and B. V. K. V. Kumar, "Palmprint recognition using correlation filter classifiers," in *Proceedings of the Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, pp. 567–571, USA, November 2004.
 - [33] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.
 - [34] M. Turk and A. Pentland, "Eigenfaces for recognition," *Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
 - [35] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.
 - [36] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial features," *Fg Net Workshop on Visual Observation of Deictic Gestures*, 2004.
 - [37] J. Woodall, M. Agúndez, A. J. Markwick-Kemper, and T. J. Millar, "The UMIST database for astrochemistry 2006," *Astronomy & Astrophysics*, vol. 466, no. 3, pp. 1197–1204, 2007.
 - [38] A. Martinez and R. Benavente, "The AR face database," Tech. Rep. CVC 24, 1998.
 - [39] J. Yang, A. F. Frangi, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.
 - [40] J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 929–941, 2005.
 - [41] Y. Yan, H. Wang, C. Li, C. Yang, and B. Zhong, "A Novel Unconstrained Correlation Filter and Its Application in Face Recognition," in *Intelligent Science and Intelligent Data Engineering*, vol. 7751 of *Lecture Notes in Computer Science*, pp. 32–39, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
 - [42] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of gabor magnitude and phase for face recognition," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1349–1361, 2010.
 - [43] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, June 2005.

Submit your manuscripts at
www.hindawi.com