INVITED PAPER

Correlation Pattern Recognition for Face Recognition

Correlation techniques for improving the accuracy of face recognition systems and for reducing the computational complexity of those systems are discussed.

By Bhagavatula V. K. Vijaya Kumar, *Senior Member IEEE*, Marios Savvides, *Member IEEE*, and Chunyan Xie, *Student Member IEEE*

ABSTRACT | Two-dimensional (2-D) face recognition (FR) is of interest in many verification (1 : 1 matching) and identification (1 : *N* matching) applications because of its nonintrusive nature and because digital cameras are becoming ubiquitous. However, the performance of 2-D FR systems can be degraded by natural factors such as expressions, illuminations, pose, and aging. Several FR algorithms have been proposed to deal with the resulting appearance variability. However, most of these methods employ features derived in the image or the space domain whereas there are benefits to working in the spatial frequency domain (i.e., the 2-D Fourier transforms of the images). These benefits include shift-invariance, graceful degradation, and closed-form solutions. We discuss the use of spatial frequency domain methods (also known as correlation filters or correlation pattern recognition) for FR and illustrate the advantages. However, correlation filters can be computationally demanding due to the need for computing 2-D Fourier transforms and may not match well for large-scale FR problems such as in the Face Recognition Grand Challenge (FRGC) phase-II experiments that require the computation of millions of similarity metrics. We will discuss a new method [called the class-dependence feature analysis (CFA)] that reduces the computational complexity of correlation pattern recognition and show the results of applying CFA to the FRGC phase-II data.

KEYWORDS | Class-dependence feature analysis (CFA); correlation filter; face recognition (FR); face recognition grand challenge (FRGC); face verification

Manuscript received March 13, 2006; revised June 13, 2006. This work was supported in part by the U.S. Government's Technical Support Working Group (TSWG) and in part by Carnegie Mellon University's CyLab.
The authors are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15217 USA (e-mail: kumar@ece.cmu.edu; marioss@andrew.cmu.edu; chunyanx@andrew.cmu.edu).
Digital Object Identifier: 10.1109/JPROC.2006.884094

I. INTRODUCTION

Machine recognition of human faces from still images and video frames is an active research area due to the increasing demand for authentication (1 : 1 matching) and identification (1 : *N* matching) in commercial and government applications. Despite the research advances over the years, face recognition (FR) is still a highly challenging task in practice due to the large variability in the facial appearance due to expressions, pose, illumination variations, and aging. Many well-known FR algorithms [1] such as Eigenfaces [2] and Fisherfaces [3] work in the image domain. However, any image is also completely described by its two-dimensional (2-D) Fourier transform (FT) and there are advantages (mainly shift-invariance, graceful degradation, and closed form solutions) to working in the spatial frequency domain. The main objective of this paper is to illustrate the benefits of using spatial frequency domain representation for FR while indicating its potential limitations.

The idea of spatial frequency domain object recognition is not new—much research exists in the use of spatial frequency domain methods for automatic target recognition (ATR) [4], [5], where the goal is to locate and classify various targets in a scene. The targets can exhibit significant variability (due to range variability, rotations, occlusions, etc.), and spatial frequency domain methods (also known as correlation filters or correlation pattern recognition) have been used successfully to deal with the resulting appearance variability. A natural question is whether such correlation filters can be used to handle the appearance variability in FR. In this paper, we show that correlation filters are useful for FR.

While the recognition rates of correlation filters are attractive, these techniques can be computationally demanding because of the need to carry out 2-D FTs and thus may not be easily applicable to large-scale FR problems

such as in the Face Recognition Grand Challenge (FRGC) experiments [6] that require the computation of millions of image similarity metrics. We will discuss a new correlation filter-motivated approach [called the class-dependence feature analysis (CFA)] that reduces the computational complexity. We will illustrate the benefits of CFA by showing its application to the FRGC data.

The rest of this paper is organized as follows. In Section II, we provide a brief overview of correlation filters. This is followed by a discussion, in Section III, of the application of correlation filters to FR, with the goal of illustrating the advantages of shift-invariance, graceful degradation, and closed form solutions. However, conventional correlation filter methods are too computationally demanding for FRGC experiments and the CFA technique is introduced in Section IV as a method to reduce the computational complexity. Section V presents the results of applying CFA to FRGC data and Section VI provides concluding remarks.

II. OVERVIEW OF CORRELATION FILTERS

Correlation is a natural metric for characterizing the similarity between a reference pattern $r(x,y)$ and a test pattern $t(x,y)$, and not surprisingly, it has been used often in pattern recognition applications. Often, the two patterns being compared exhibit relative shifts and it makes sense to compute the cross-correlation $c(\tau_x,\tau_y)$ between the two patterns for various possible shifts τ_x and τ_y as in (1); then, it makes sense to select its maximum as a metric of the similarity between the two patterns and the location of the correlation peak as the estimated shift of one pattern with respect to the other

$$c(\tau_x,\tau_y) = \iint t(x,y)r(x-\tau_x,y-\tau_y)dxdy \qquad (1)$$

where the limits of integration are based on the support of $t(x,y)$. The correlation operation in (1) can be equivalently expressed as

$$\begin{aligned} c(\tau_x,\tau_y) &= \iint T(f_x,f_y)R^*(f_x,f_y)e^{j2\pi(f_x\tau_x+f_y\tau_y)}df_xdf_y \\ &= \mathrm{FT}^{-1}\{T(f_x,f_y)R^*(f_x,f_y)\} \end{aligned} \qquad (2)$$

where $T(f_x,f_y)$ and $R(f_x,f_y)$ are the 2-D FTs of $t(x,y)$ and $r(x,y)$, respectively, with f_x and f_y denoting the spatial frequencies. Equation (2) can be interpreted as the test pattern $t(x,y)$ being filtered by a filter with frequency response $H(f_x,f_y)=R^*(f_x,f_y)$ to produce the output $c(\tau_x,\tau_y)$ and hence the terminology "correlation filtering" for this operation. However, unlike in simple low-pass and high-pass filters, the phase of the correlation filter $R^*(f_x,f_y)$ is very important for pattern matching.

Matched Filter: The correlation filter in (2) is known as the *matched filter* (MF), i.e., the filter $H(f_x,f_y)=R^*(f_x,f_y)$ is simply the complex conjugate of the 2-D FT of the reference pattern $r(x,y)$. It can be shown [7] that this is optimal for detecting a reference pattern corrupted by additive white noise. But the optimality of the MF holds only if the reference pattern and the input pattern are identical except for the additive white noise and translation. In reality, the test pattern will differ from the reference pattern in many ways, e.g., rotations, scale changes, and the MF does not perform well.

An interesting aside is that MFs were implemented using coherent optical processors in 1960s [8] and much effort has gone into making optical correlators feasible. However, the devices needed for optical correlators did not advance sufficiently fast and most of the current correlation filter implementations are digital.

In FR, the test face image from a subject is bound to differ from the reference face image of the same subject due to normal variations induced by expression changes, illumination differences, pose variations, and aging. In such a case, a theoretically optimum solution is to use one MF for each possible appearance of the face image. Clearly, this is computationally impractical because of the combinatorial explosion in the number of filters when we consider all possible factors that cause face appearance to change.

Synthetic Discriminant Function Filters: Towards the goal of handling pattern variability using fewer correlation filters, Hester and Casasent [9] introduced the concept of the synthetic discriminant function (SDF) filter. This approach assumes the availability of a representative set of training images and the SDF filter is a weighted sum of MFs where the weights are chosen so that the correlation outputs corresponding to the training images would yield prespecified values at the origin. For example, the correlation values (at the origin) corresponding to the training face images of authentic subjects can be set to one, and the origin values due to the impostor training images can be set to zero. The expectation is that the resulting correlation filter would yield correlation peak values close to one for nontraining face images from the authentic class and correlation peak values close to zero for images from the impostor class.

As shown schematically in Fig. 1, FR is performed by filtering the input face image with a synthesized correlation filter and processing the resulting correlation output. The correlation output is searched for a peak, and the height of that peak (relative to the background) is used to determine whether the test face image matches the training face images or not. The location of the correlation peak indicates the shift of the test face image relative to the training images.

Minimum Average Correlation Energy Filter: Although the original SDF filter produces prespecified correlation peak values, it only controls the output at the origin for centered training images. Since the test patterns are not necessarily centered, it is nearly impossible to know where these controlled values in the output are, unless we can control the rest of the correlation plane to take on smaller values. The minimum average correlation energy (MACE) filter [10] aims to make the controlled values (ones at the origin for centered authentic training face images) the largest values in the output plane by making the average energy of the correlation outputs as small as possible. Correlation outputs from well-designed MACE filters typically exhibit sharp peaks for authentic input images, making the peak detection and location relatively easy and robust.

We will now briefly explain the MACE filter design. Suppose we have N training images, each of size $d \times d$. First, the 2-D FTs of these training images are computed and resulting complex arrays are vectorized into columns of a $d^2 \times N$ complex valued matrix $\mathbf{X}$. We also use a $d^2 \times d^2$ diagonal matrix $\mathbf{D}$ whose diagonal entries are the average power spectrum of the N training images. Since $\mathbf{D}$ is diagonal, we need to store only its diagonal entries and not the complete matrix. The filter is represented by a column vector $\mathbf{h}$ with d^2 elements. Finally, the filter $\mathbf{h}$ is required to produce prespecified values u_i at the correlation origin in response to the training images $i = 1, 2, \ldots, N$ and these constraints can be expressed as follows:

$$\mathbf{X}^{+}\mathbf{h} = \mathbf{u} \tag{3}$$

where $\mathbf{u} = [u_1\ u_2\ \cdots\ u_N]^T$ superscripts T and $+$ denote transpose and conjugate transpose, respectively. The average correlation plane energy can be expressed as the quadratic term $\mathbf{h}^{+}\mathbf{D}\mathbf{h}$ and minimizing this quadratic while satisfying the linear constraints in (3) leads to the following closed form solution for the MACE filter:

$$\mathbf{h} = \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^{+}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}. \tag{4}$$

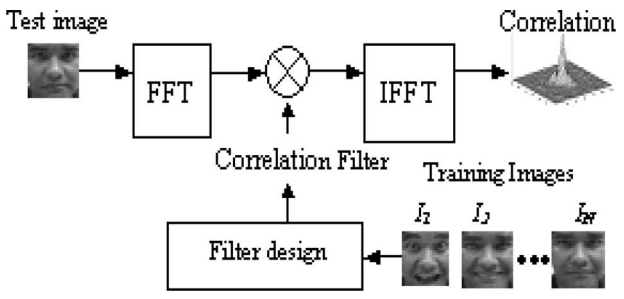


Fig. 1. *Block diagram of correlation process.*

Since $\mathbf{D}$ is a diagonal matrix, determining $\mathbf{h}$ requires, in effect, the inversion of the $N \times N$ matrix $(\mathbf{X}^{+}\mathbf{D}^{-1}\mathbf{X})$ where N is the (usually small) number of training images per class.

Optimal Tradeoff Filters (OTF): In order to produce very sharp correlation peaks for authentic training images, MACE filters emphasize high spatial frequencies (since narrow functions in correlation output should correspond to broad support in spatial frequency domain) and this makes MACE filters very susceptible to input noise and other deviations from the training images. The variance of the noise in the correlation output can be shown [11] to be given by $\mathbf{h}^{+}\mathbf{C}\mathbf{h}$ where $\mathbf{C}$ is a $d^2 \times d^2$ diagonal matrix containing the input noise power spectral density values as its diagonal entries. Often, the input noise is modeled as white leading to $\mathbf{C} = \mathbf{I}$, the identity matrix. Not surprisingly, minimizing output variance $\mathbf{h}^{+}\mathbf{C}\mathbf{h}$ results in a filter that emphasizes low spatial frequencies, whereas minimizing the average correlation energy $\mathbf{h}^{+}\mathbf{D}\mathbf{h}$ leads to a filter that emphasizes high spatial frequencies. Optimally, trading off between $\mathbf{h}^{+}\mathbf{C}\mathbf{h}$ and $\mathbf{h}^{+}\mathbf{D}\mathbf{h}$ while satisfying the linear constraints in (3) results [12] in the following filter known as the *optimal tradeoff filter* (OTF):

$$\mathbf{h} = \mathbf{T}^{-1}\mathbf{X}(\mathbf{X}^{+}\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{u} \tag{5}$$

where $\mathbf{T} = (\alpha\mathbf{D} + \sqrt{1-\alpha^2\mathbf{C}})$, and $0 \le \alpha \le 1$ is a parameter that controls the tradeoff. $\alpha = 0$ leads to the maximally noise-tolerant filter, whereas $\alpha = 1$ leads to the MACE filter that produces very sharp correlation peaks.

Advanced Correlation Filters: Since the introduction of the MACE filters and the OTFs in late 1980's and early 1990's, correlation filter designs have evolved significantly. Due to space constraints, we will not go into the details [5], [13], but will highlight the main ideas.

1) Early correlation filter designs required that the filter vector $\mathbf{h}$ satisfy the constraints in (3) which force the correlation output at the origin to take on prespecified values for centered training images. While this makes sense, it does not ensure that we get the same values (or even close to the same values) for nontraining images. Thus, with the goal of improving the generalization capability of these filters, unconstrained correlation filters [14] are designed to maximize the average correlation height instead of satisfying the hard constraints of (3). Resulting correlation filters have performed well in automatic target recognition applications [14].
2) Most correlation filter designs focus on the correlation output at the origin for centered training images, but there is useful discriminatory information in the rest of the correlation plane.

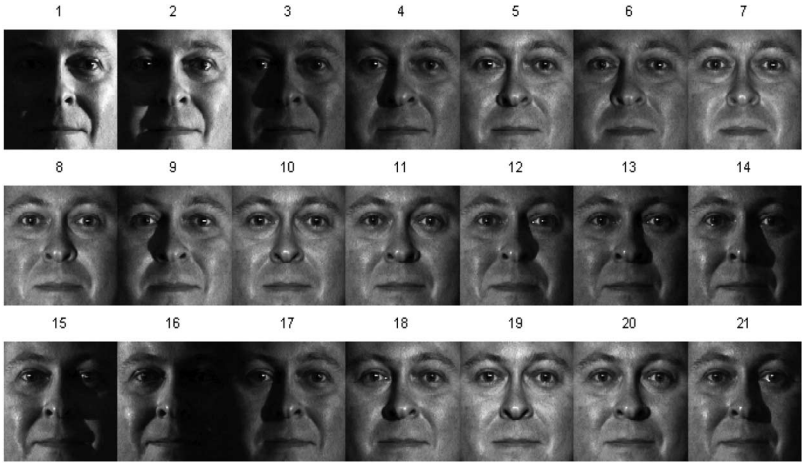


Fig. 2. *Images of one face under different illuminations in the CMU PIE database [21].*

Another correlation filter approach known as the distance classifier correlation filter (DCCF) [15] is aimed at determining the filter $H(f_x, f_y)$ that maps the training images from different classes into well-separated compact clusters.

3) Polynomial correlation filters [16] use as inputs the original input image as well as various nonlinear versions of it (obtained by applying a point nonlinearity to each pixel). Correlation filters for each nonlinear version are designed (in a jointly optimal manner) and the resulting correlation outputs are easily fused since they all should produce peaks at the same location.
4) One of the potential problems of correlation filters is their sensitivity to image rotations. However, rotating an image in-plane around its origin results in the rotation of its 2-D FT in the same way around the origin in the frequency domain. Using this fact as well as the 2π periodicity in angular frequency, Kumar *et al.* [17] developed optimal tradeoff circular harmonic function (OTCHF) filters that produce correlation outputs in a prescribed way (e.g., large correlation peaks for the target rotated between $-\pi/4$ and $+\pi/4$ and small peaks for other rotations). The design methodology borrows ideas from finite impulse response (FIR) filter designs.
5) More recently, Kerekes and Kumar [18] have extended the correlation filter designs using Mellin radial harmonic functions [19] to produce prescribed response to in-plane scale. In principle, such filters can produce nearly constant correlation outputs (at origin) over a scale range (e.g., 90% to 110% of nominal scale) while maintaining sharp correlation peaks.

In feature-based FR methods, the designer must make somewhat arbitrary decisions about which features to use and in that sense may discard some useful information right upfront, although experience is a great teacher in helping us select good features. In correlation filter methods, no information is *a priori* lost as all image pixels are used, and in that sense correlation filters capture all the information present in the training data. Of course, this dependency on training data can be this approach's Achilles heel, when training data are limited. Thus, in practice, we may want to employ both traditional feature-based methods as well as correlation filter methods.

We have already seen [e.g., (4)] that correlation filters have convenient closed-form expressions. In addition, correlation filter methods offer other advantages such as shift-invariance and graceful degradation, which we will illustrate in the next section for FR.

III. FACE RECOGNITION WITH CORRELATION FILTERS

Appearance of a face image varies with changes in expression, pose, and illumination. Savvides *et al.* [20] demonstrated the superior face verification performance of MACE filters in the presence of expression changes. Pose-tolerant FR using correlation filters has not advanced as much; one practical way of achieving pose-tolerant FR appears to design and use multiple correlation filters for multiple (possibly overlapping) sectors of pose angles. In this section, we illustrate the use of correlation filters for FR by focusing on their performance under strong illumination variations. To this purpose, we will use the illumination subset of the images from the CMU pose, illuminations, and expressions (PIE) face database [21]. The PIE database has 65 subjects and 21 images of one subject's face under different illuminations are shown in Fig. 2. All face images have been manually cropped to yield images shown in Fig. 2.

For each subject in this illumination subset, a MACE filter was designed based on images numbered 3 (left half of face in shadow), 7 (frontal illumination), and 16 (right half of the face in shadow) and the resulting correlation filter was tested against all 21 images of each of the 65 subjects in the database. We expect the correlation output to exhibit sharp, high peaks for the authentic and no such peaks for the impostors. We quantify the peak sharpness by the peak-to-sidelobe ratio (PSR) defined as

$$\text{PSR} = \frac{(\text{peak-mean})}{\text{std}} \tag{6}$$

where peak is the largest value in the correlation output and mean and std are the average value and the standard deviation of the correlation outputs in an annular region (of size 20×20 for the PIE database images) centered on the peak, but excluding the peak region (a 5×5 region). PSR is designed to measure the relative height of the correlation peak to the background and is observed to be not too sensitive to the sizes of these regions. One benefit of the PSR definition in (6) is that it is unaffected by constant illumination changes in the input image. For well-designed MACE filters, PSR should be large for authentic images and small for impostor images.

As described earlier, a MACE filter for each subject was designed using only images numbered 3, 7, and 16 of that subject. In Fig. 3, we show two correlation outputs produced by one such correlation filter in response to different input images. Fig. 3 (left) shows the correlation output in response to image 10 from the same subject. The correlation output exhibits a sharp peak and the corresponding PSR value is 40.95. Fig. 3 (right) shows the correlation output in response to the face image of a different subject (i.e., an impostor) in this database. This correlation output has no discernible peak and leads to a PSR of only 4.77.

Fig. 4 shows the PSR values of person 1's correlation filter when tested against all 21 illumination images of that person (solid curve) and when tested against 21 images from each of the other 64 subjects (dashed curves at the bottom) in this database. It is seen that the filter yields the highest PSR values for the three images (3, 7, and 16) used for the training. Although PSR values for nontraining authentic images are lower, they are still higher than any impostor PSR values, indicating that this filter can discriminate Person 1 from 64 others in the CMU PIE database. The discrimination of filters designed for other 64 subjects was equally good in that no verification errors were observed [22] from MACE filters designed from images numbered 3, 7, and 16. In contrast, individual PCA (IPCA) methods trained on the same three images yield nearly 34% equal error rate.

Graceful Degradation: One of the benefits of correlation filters is the graceful degradation they offer because of the integrative nature of the matching operation as seen in (1). If some of the pixels are occluded, they simply do not contribute to the correlation peak, thus decreasing the overall peak. However, no single pixel in the image domain is critical in that recognition can be still carried out successfully. This is illustrated in Fig. 5. The left half of the figure shows an authentic nontraining image, with part of the face image completely blocked, and the right half shows the resulting correlation output that exhibits a very discernible peak. Note that the filter was designed from full images. It is clear that the correlation peak is degraded

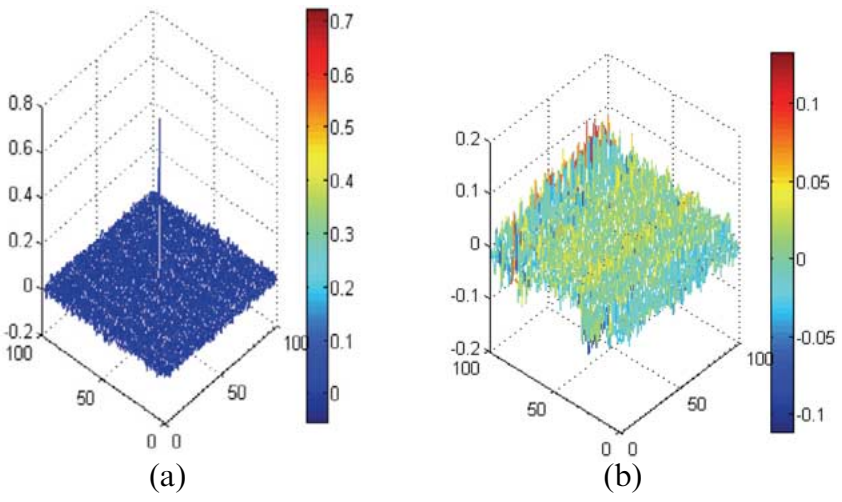


Fig. 3. ***Correlation outputs for (a) authentic input and (b) impostor input.***

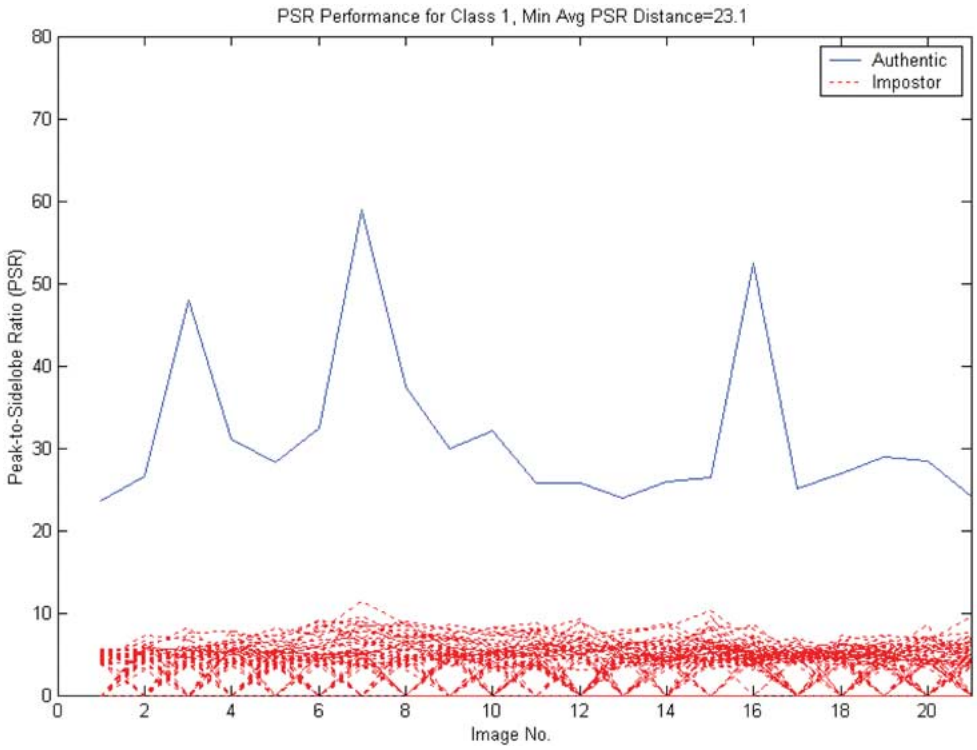


Fig. 4. ***PSR values from Person 1's correlation filter.***

in that the resulting PSR is only 30.6 but still high enough to verify the subject's identity. This tolerance to image occlusions is in contrast to many image-domain (also called space-domain) methods that first locate the two eyes and then center and normalize the input image, prior to computing features.

Shift-Invariance: Another important benefit from the use of correlation filters is the resulting automatic shift invariance. Since correlation filters are linear shift-invariant filters, any translation in the test input image will result in correlation output being shifted by exactly the same amount. Since the first thing we do in processing correlation outputs is to locate the correlation peak and since filters are designed to yield centered correlation peaks for centered training images, we do not need to explicitly center the test image. Instead, we implicitly center the test image by locating the correlation peak and

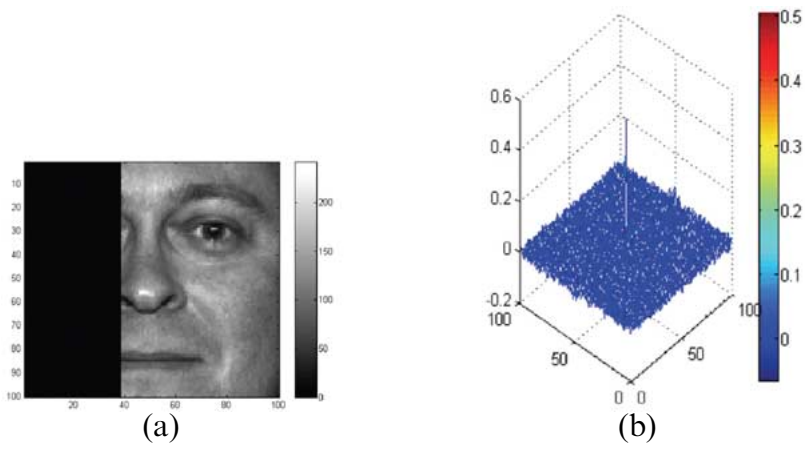


Fig. 5. ***(a) Partially occluded input image and (b) corresponding output from a correlation filter designed from full training images.***

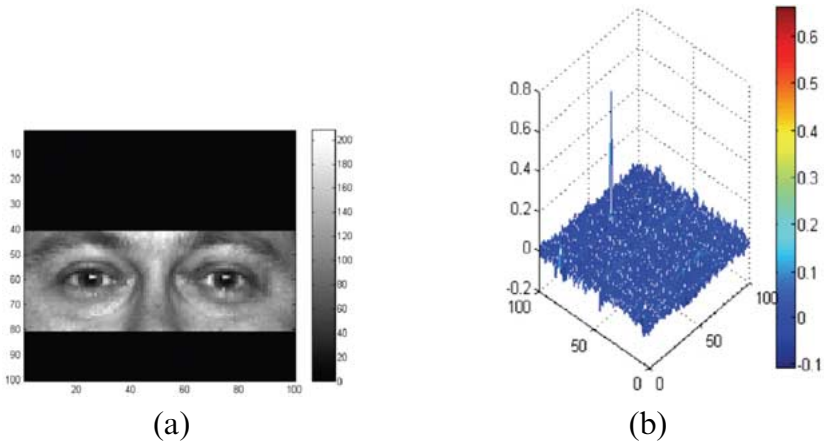


Fig. 6. ***(a) Partially occluded and uncentered image and (b) corresponding output from correlation filters designed from full and centered training images.***

computing the PSR centered at this peak location. In many feature-based FR methods, centering of a test image prior to computing the feature values is critical since they are often sensitive to image centering. We illustrate the shift-invariance property of the correlation filters in Fig. 6. The left half of this figure shows an occluded and off-centered test image and the right half shows the resulting correlation output. The peak is still very visible except that it has moved to a new location. The shift-invariance is not just a theoretical curiosity. In practice, even with our best algorithms, centering test images is only approximate and can be time consuming. Thus, methods that avoid the need for explicit test image centering are attractive.

Identification: While our discussion so far focused on verification (1 : 1 matching), correlation filters can be used for identification (1 : N matching). In identification, the input face image is correlated with each of the N correlation filters, one designed for each of the N subjects in the database. For each correlation output, a PSR is determined and the input face image is assigned to the class of the filter yielding the largest PSR; or, if this largest PSR is below a preselected threshold, it is decided that the test image subject is not represented in the database. Identification requires N times as much work as in verification. In Table 1, we show [23] the recognition rates for three different methods (namely IPCA, Fisherfaces, and MACE filters) trained on different combinations of training images from the CMU PIE illuminations database. It is seen that MACE filters outperform the other two methods.

IV. CLASS-DEPENDENCE FEATURE ANALYSIS (CFA)

While the correlation filters have attractive properties such as shift-invariance, graceful degradation, and closed-form solutions, we must be concerned about their

Table 1 Identification Rates (For Different Training Sets) of IPCA, Fisherfaces, and MACE Filters on CMU PIE Illuminations

	IPCA		Fisherfaces		MACE Filters	
Training Images	# Errors	% Rec. Rate	# Errors	% Rec. Rate	# Errors	% Rec. Rate
3,7,16	55	95.9%	1	99.9%	0	100%
1,10,16	126	90.7%	1	99.9%	0	100%
2,7,16	156	88.57%	0	100%	0	100%
4,7,13	117	91.4%	12	99.1%	0	100%
3,10,16	113	91.7%	1	99.9%	0	100%
3,16	760	44.3%	684	49.9%	1	99.9%

computational complexity, particularly when dealing with large class FR as in the case of the FRGC [6]. Even when the correlation filters are computed beforehand, producing one correlation output requires two 2-D FFTs and this can take too long when millions of similarity computations are needed, as in FRGC. Towards the goal of reducing this computational complexity, we have recently [23] introduced class-dependence feature analysis (CFA) method, and in this section, we will explain the main ideas of CFA.

Face Recognition Grand Challenge (FRGC): The FRGC [6] provides several challenging experiments for the FR research community. In this paper, we focus on the hardest 2-D FR problem in the FRGC, which is Experiment 4. This experiment tests the ability of FR algorithms to match face images captured in controlled conditions against face images captured in uncontrolled indoor and outdoor conditions with harsh lighting conditions as shown in Fig. 7.

The FRGC database consists of three components: a "generic" dataset which is typically used to build a global face model (e.g., a global PCA subspace), a "gallery/target" set which consists of known people that we want to find, and a "probe/query" set which are the unknown images captured that we need to match against the gallery set. The generic training set consists of a total of 12 776 face images representing 222 subjects, with the number of images per subject ranging from 36 to 64. The gallery set consists of 16 028 controlled images representing 466 subjects and the query set has 8014 uncontrolled images representing the same 466 people. The number of images per subject varied substantially, from 4 to 88. There are 153 subjects common between the generic training set and probe and gallery sets, i.e., there are 313 subjects in the probe and gallery set that were not in the generic training set. The face images in the dataset were captured in multiple sessions spanning a period of one year. Experiment 4 is difficult as the gallery images are controlled images with controlled lighting conditions, whereas the probe images are captured in uncontrolled indoor and outdoor settings. The FRGC protocol requires a matrix of similarity scores between all pairs of gallery and probe images. The resulting similarity matrix for FRGC Experiment 4 is of size 16 028 $\times$ 8014. From this similarity matrix, we can extract the similarity scores for all authentic pairs and similarity scores for all imposter pairs. For a similarity score threshold, we can compute the false acceptance rate (FAR, i.e., the fraction of impostor pair similarity scores exceeding the threshold) and the verification rate (VR, i.e., the fraction of authentic pair similarity scores exceeding the same threshold). By varying the threshold, we get a receiver operating characteristic (ROC) curve. The FRGC protocol calls for the VR at a threshold yielding 0.1% FAR. The principal component analysis (PCA) baseline algorithm yields a verification rate of 12% at 0.1% FAR [6]. The research goal in FRGC is to increase the VR at a FAR of 0.1%.

Fig. 7. ***Example query images from FRGC-Phase II dataset [6] showing harsh uncontrolled illumination.***

Difficulty in Applying Correlation Filters to FRGC: Our goal is to apply correlation filters for FRGC. However, several challenges make this difficult. First, the normal procedure for designing a correlation filter for a gallery subject is to use several training images of that subject (perhaps with illumination variations) and construct one filter (e.g., MACE filter) that produces sharp correlation peaks for all these training images. However, the FRGC protocol requires the computation of similarity scores between every gallery image and every probe image. This means that only one gallery image is available for correlation filter design and we do not get to use multiple training images to represent one subject. The second problem is that the standard correlation filter designs are not set up to take advantage of the generic training set; whereas, PCA-type methods use the generic training set to determine eigenfaces. The third difficulty is that Experiment 4 of FRGC requires the computation of 128 million similarity values (between 16 028 gallery images and 8014 query images) and carrying out that many cross correlations (each requiring three 2-D FFTs of size 128 $\times$ 128) is computationally prohibitive. Finally, the face images provided in the FRGC are well centered, rendering the shift-invariance advantages of correlation filters mostly irrelevant. Thus, we may as well be computing inner products rather than full correlations and thus save significant computational expense. Keeping in mind all these factors, we introduced the new CFA approach [24].

CFA Approach: In this approach, one filter (in our method a MACE filter, but one could use other filters) is designed for each of the 222 subjects in the generic

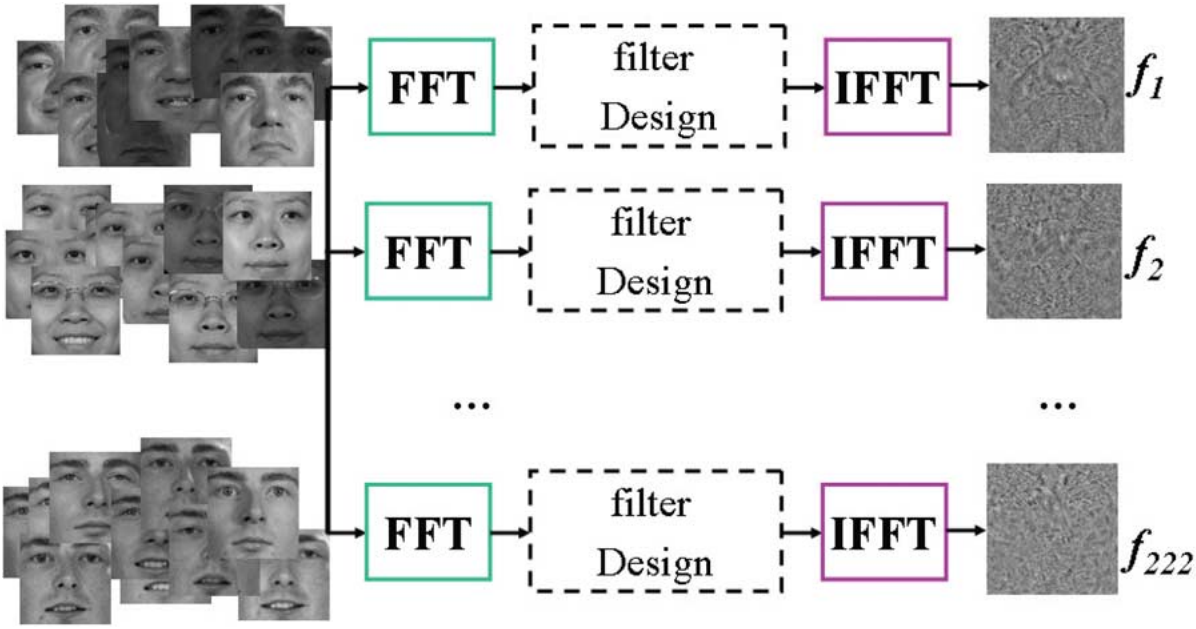


Fig. 8. *Training CFA correlation filter bases using generic training data.*

training set. The design of each of these 222 filters uses all the 12 776 available generic training images in one class versus the rest of the configuration (i.e., correlation output is 1 at the origin for all authentic images and correlation output is 0 for all other images from the remaining 221 people), as illustrated in Fig. 8. Once the N (in this case $N = 222$) filters are designed, they are used (just as in the case of eigenfaces) as shown in Fig. 9 to produce the output feature vector $\mathbf{c}$ for any image represented by $\mathbf{y}$, i.e.,

$$\mathbf{c} = \mathbf{H}^{\mathrm{T}}\mathbf{y} = [\mathbf{h}_{\mathrm{mace}-1}\mathbf{h}_{\mathrm{mace}-2}\ldots\mathbf{h}_{\mathrm{mace}-N}]^{T}\mathbf{y} \qquad (7)$$

where $\mathbf{h}_{\mathrm{mace}-i}$ is a filter is trained to give a small correlation output (close to zero) for all classes except for class i. In this example, the number of filters generated from the FRGC generic training set is 222 since it contains 222 subjects. Then, each input image $\mathbf{y}$ (can be a gallery image or a probe image) is projected onto those non-orthogonal basis vectors to yield a 222-dimensional correlation feature vector $\mathbf{c}$. The similarity between the probe image and the gallery image is computed as the similarity between the corresponding 222-dimensional feature vectors. By limiting the computations to just inner products as in (7), we avoid the need for FFTs and the necessary computations can be carried out in the image domain. The CFA feature space dimensionality is a function of the number of classes in the generic training set and not the total number of images.

Kernel CFA: Due to the nonlinear distortions in human face's appearance, linear subspace methods have not performed well in FR studies, especially when lacking representative training data. As a result, common algorithms such as PCA and LDA have been extended [25] to represent nonlinear features. Since nonlinear mappings increase the dimensionality, kernel tricks are used for computational efficiency as they enable us to obtain the necessary inner products in the higher dimensional feature space without actually having to compute the higher dimensional feature mapping.

Let $\Phi(x)$ denote the nonlinear mapping of interest. Then, the Kernel functions defined by $K(x,y) = \langle\Phi(x), \Phi(y)\rangle$ can be used without having to explicitly map the features to a higher dimensional space as long as kernels form an inner product and satisfy Mercer' theorem [26]. Examples of kernel functions are as follows.

Polynomial kernel:

$$(K(\mathbf{a},\mathbf{b}) = (\langle\mathbf{a},\mathbf{b}\rangle + 1)^{p}). \qquad (8)$$

Radial Basis Function kernel:

$$\big(K(\mathbf{a},\mathbf{b}) = \exp\big(-\|\mathbf{a}-\mathbf{b}\|^{2}/2\sigma^{2}\big)\big). \qquad (9)$$

Sigmoidal Kernel:

$$(K(\mathbf{a},\mathbf{b}) = \tanh(k\langle\mathbf{a},\mathbf{b}\rangle - \delta). \qquad (10)$$

Kernel Correlation Filters: The kernel correlation filters [27] can be extended from the linear correlation filters using the kernel trick. However, we need to perform a prewhitening step before we can derive the closed form solution for the kernel correlation filter. The correlation output of a MACE filter $\mathbf{h}$ and an input image $\mathbf{y}$ can be expressed as

$$\begin{aligned}\mathbf{y}^{+}\mathbf{h} &= \mathbf{y}^{+}\left[\mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^{+}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}\right] \\ &= (\mathbf{D}^{-0.5}\mathbf{y})^{+}(\mathbf{D}^{-0.5}\mathbf{X})\left((\mathbf{D}^{-0.5}\mathbf{X})^{+}\cdot\mathbf{D}^{-0.5}\mathbf{X}\right)^{-1}\mathbf{u} \\ &= \left((\mathbf{y}')^{+}\mathbf{X}'\right)\left((\mathbf{X}')^{+}\mathbf{X}'\right)^{-1}\mathbf{u} \qquad (11)\end{aligned}$$

where $\mathbf{X}' = \mathbf{D}^{-0.5}\mathbf{X}$ indicates prewhitened version of $\mathbf{X}$. From now on, we assume that $\mathbf{X}$ is already prewhitened. Now we can apply the kernel trick to yield the kernel correlation filter as follows:

$$\begin{aligned}\Phi(\mathbf{y})\cdot\Phi(\mathbf{h}) &= (\Phi(\mathbf{y})\cdot\Phi(\mathbf{X}))(\Phi(\mathbf{X})\cdot\Phi(\mathbf{X}))^{-1}\mathbf{u} \\ &= K(\mathbf{y},\mathbf{x}_i)K(\mathbf{x}_i,\mathbf{x}_j)^{-1}\mathbf{u}. \qquad (12)\end{aligned}$$

We use toy examples to illustrate the advantage of using kernel-based nonlinear classifiers. We show in the top half of Fig. 10 three linearly separable classes, ten samples per class. We design the linear CFA, the polynomial CFA, and the Gaussian radial basis function (RBF) CFA. Resulting decision boundaries are shown in Fig. 10(a), (b), and (c). All three schemes are able to separate the three classes. In the bottom half of Fig. 10, we show two linearly nonseparable cases and we show the decision boundaries obtained using two different RBF kernels in Fig. 10(d) and (e). The nonlinear RBF kernels are able to come up with the nonlinear decision boundaries needed to separate the three classes.

To take advantage of the potential benefits from the nonlinear mappings, we use (12) to compute the inner products needed in the CFA framework. This leads to a 222-dimensional feature vector representation in the FRGC experiment and we refer to this approach as the kernel CFA (KCFA).

Distance Metrics: In CFA and KCFA approaches, a face image in the FRGC database will be represented by a 222-dimensional feature vector $\mathbf{c}$, and evaluating the

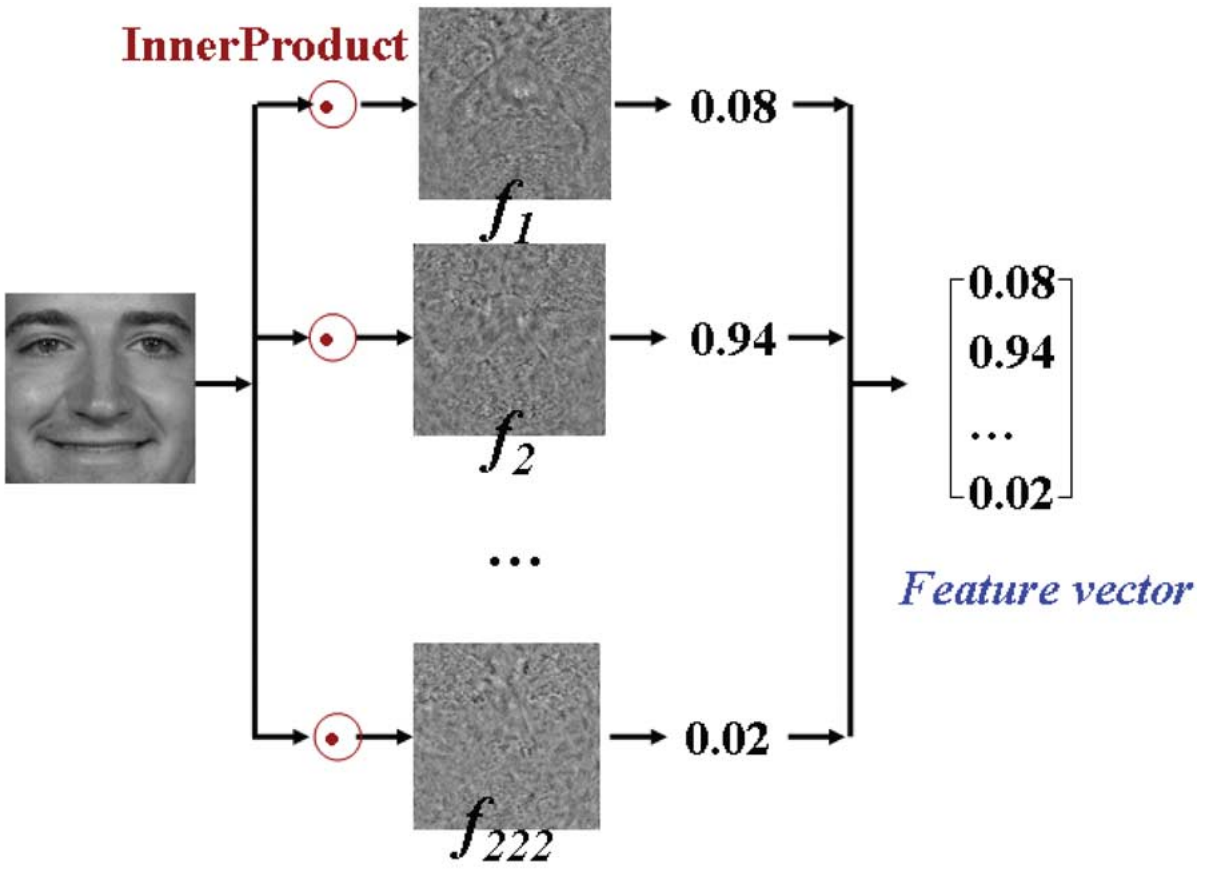


Fig. 9. *Feature vector formation using trained CFA bases.*

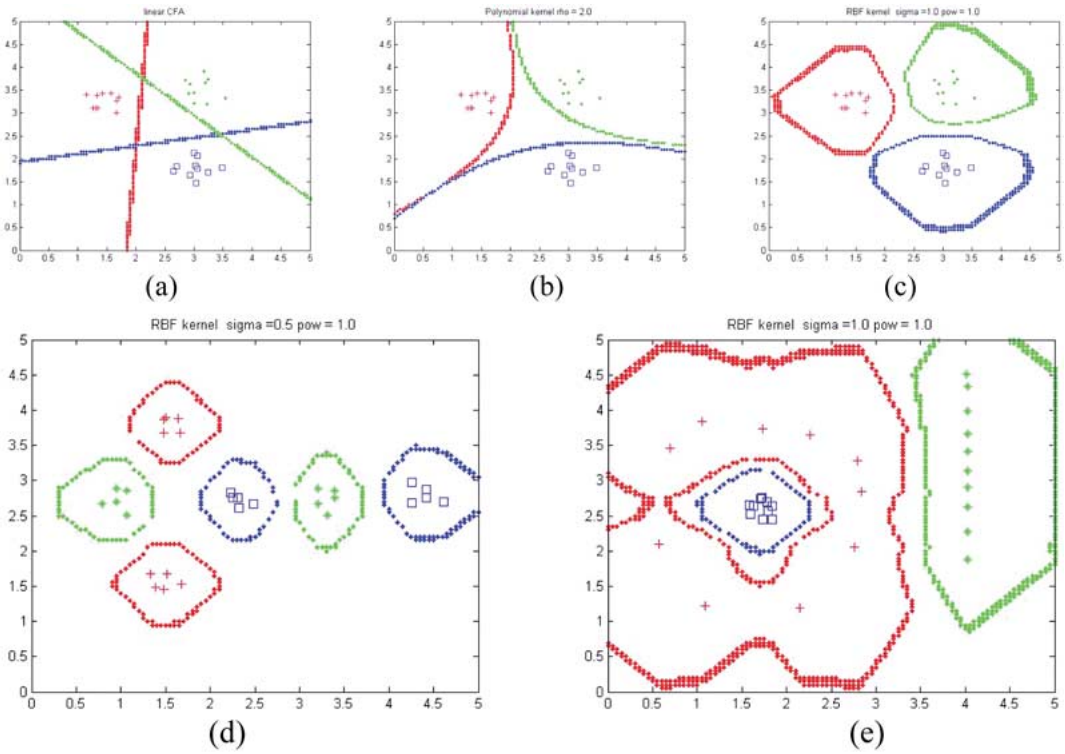


Fig. 10. ***Decision boundaries for (a) linear correlation filter, (b) polynomial correlation filter, (c) Gaussian RBF classifier, (d) RBF case 2, and (e) RBF case 3.***

similarity between these feature vectors comes down to the selection of an appropriate distance or similarity metric. Commonly used distance measures are L1-norm, L2-norm [24], the Mahalanobis distance, and the normalized cosine distance (given below) which exhibited the best results

$$d(\mathbf{x},\mathbf{y}) = \frac{-\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}. \tag{13}$$

V. FRGC NUMERICAL RESULTS

In this section, we show the verification results from the application of CFA and KCFA to FRGC Experiment 4 data (16 028 $\times$ 8014 image pairs or equivalently approximately 128 million similarity values). In order to reduce the image variability, we applied the illumination-compensation preprocessing algorithm developed by Gross and Brajovic [28]. This algorithm first estimates the illumination field and then compensates for it to mostly recover the scene reflectance. The algorithm does not require any training steps, knowledge of 3-D face models, or reflective surface models. We show in Fig. 11 some examples of original and preprocessed FRGC images.

Fig. 12 shows the VR (for a FAR of 0.1%) using PCA, Gram-Schmit Linear Discriminant Analysis (GSLDA) [29], CFA, and KCFA. The performance of PCA is the benchmark 12% verification rate provided by NIST [6]. KCFA offers the best VR of about 57%.

Support Vector Machine Training: Since both CFA and KCFA lead to 222-dimensional feature vectors, support vector machine (SVM) training [30], [31] can be used to separate the feature vector of one class from the others in the SVM space. This approach may offer the flexibility of varying thresholds and thus better classification performance. We designed 466 SVMs (in a one-against all framework) using the gallery set (but not any probe images) of the FRGC data. The probe images are then projected on the class-specific SVMs to obtain a projection

Fig. 11. ***Examples of original (first and third) and preprocessed [28] (second and fourth) images.***

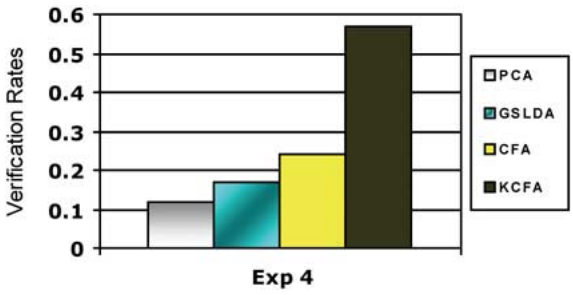


Fig. 12. ***Verification rates for FRGC Experiment 4 at 0.1% FAR. Kernel CFA shows best results (nearly threefold) over all linear methods.***

similarity score. Verification rates (at FAR of 0.1%) for the various algorithm using the normalized cosine distance and using the SVM-based distance are shown in Fig. 13. While this SVM approach boosts performance of all algorithms as shown in Fig. 13, it enhances the KCFA method the most yielding nearly 87.5% verification rate at a FAR of 0.1%. As shown in Fig. 13, the SVM-based distance measure produces better results than using normalized cosine distance. We also compare in Fig. 13 different kernel approaches such as KPCA and kernel discriminant analysis (KDA) with both distance measures.

There are some publicly available FRGC Experiment 4 numerical results. In [32], Liu proposes a method that integrates Gabor image representation, a multiclass Kernel Fisher Analysis (KFA) method, and fractional power polynomial models for FR. Their January 2005 result for Experiment 4 reported 76% VR at 0.1% FAR. Phillips *et al.* [33] provide a summary of FRGC Experiment 4 results, which show a median VR of about 30% (at 0.1% FAR) and the best VR is about 76%. More recently, Phillips [34] has also presented a set of VR results (after anonymizing the algorithm sources) for FRGC Phase II Experiment 4. Of the 16 sets of verification results presented, the best one appears to be about 90% VR, the second best one is about 75% VR, and all others are below 70% VR with the worst less than 10% VR.

VI. CONCLUSION

Face recognition is of interest in many human verification and identification applications because of the relatively passive nature of face image acquisition and because of the ubiquitous nature of digital cameras. However, 2-D FR continues to be challenging because of the appearance variability due to pose, illumination, expression, and aging. Particularly challenging FR tasks involve matching face images acquired under controlled conditions with those collected in less controlled conditions (e.g., outdoors). Most FR methods are based on computing features in the image domain and deriving similarity metrics between feature vectors corresponding to the probe and gallery images. These image-domain methods require careful centering of test images, and in some sense are only as good as the features selected by the designer. In contrast, spatial frequency domain approaches (often called correlation filters) offer a way to extract the recognition information available in the training data. In this paper, we showed how correlation filters can be designed for FR. Well-designed correlation filters should produce sharp correlation peaks in the output for authentic images and no such discernible peaks for the impostor images. We have also illustrated two major advantages of correlation filters. First is that because of the built-in shift-invariance of the correlation filters, there is no need to explicitly center the test face image; instead, by locating the correlation peak and calculating a peak-to-sidelobe ratio metric centered around that peak, we implicitly center the test image. Thus, careful positioning of camera with respect to test faces is not needed. The second advantage is that the integrative nature of correlation operation provides graceful degradation in that no particular pixels become critical for matching. For example, correlation filters work reasonably well even when the eye regions are occluded; whereas, many image-domain approaches must first locate the eye centers to center and normalize the face image.

However, standard correlation filter methods are not well matched to FRGC protocol and a new approach called the CFA is introduced to follow the FRGC protocol. Both CFA and its nonlinear counterpart (KCFA) perform significantly better than PCA, LDA, and their variants.

Correlation filters can be computationally demanding because of the need to carry out multiple 2-D FFTs. This is particularly problematic for the face recognition grand challenge (FRGC) experiments that call for the computation of millions of similarity metrics. Also, the standard correlation filter formulations do not take advantage of the generic training set provided in FRGC experiments or the fact that the face images are well centered, essentially eliminating the benefits of the built-in shift-invariance. Because of these considerations, we introduced the new

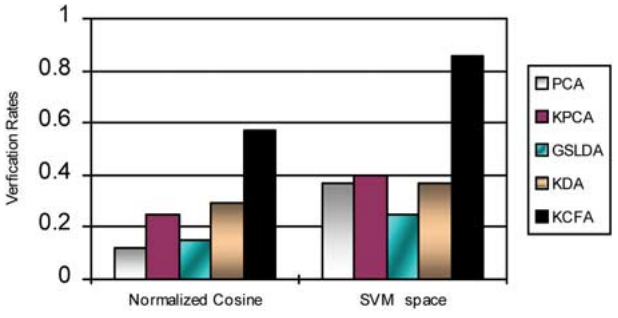


Fig. 13. ***Verification rates (at FAR of 0.1%) for different methods using normalized cosine distance and using SVM-based distance.***

class-dependence feature analysis (CFA) method that computes only inner products and uses the generic datasets. We showed that the basic CFA method as well as its nonlinear counterparts (KCFA) yield very good verification rates on the challenging FRGC phase II Experiment 4 data.

While correlation filters and their derivatives (such as CFA and KCFA) have much to offer for FR, they are only as good as the training data available during their design. In situations where the training data is sparse, these techniques may not work well. Also unsatisfactory is the fact that correlation filter methods do not take advantage of domain-specific knowledge about face images. Thus, we believe that even better FR performance can be achieved by cleverly fusing both image-domain feature-based approaches and frequency-domain correlation filter methods and future research should focus on methods that combine image-domain and spatial frequency-domain representations. ■

REFERENCES

[1] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–741, 1995.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, pp. 72–86, 1991.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[4] B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.*, vol. 31, pp. 4773–4801, 1992.

[5] B. V. K. Vijaya Kumar, A. Mahalanobis, and R. D. Juday, *Correlation Pattern Recognition*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[6] 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. IEEE Conf. Comp. Vision Pattern Rec. (CVPR)*, 2005.

[7] D. O. North, "An analysis of the factors which determine signal/noise discriminations in pulsed carrier systems," *Proc. IEEE*, vol. 51, no. 7, pp. 1016–1027, Jul. 1963.

[8] A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, pp. 139–145, Apr. 1964.

[9] C. F. Hester and D. Casasant, "Multivariant technique for multiclass pattern recognition," *Appl. Opt.*, vol. 21, pp. 4016–4019, 1982.

[10] A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.*, vol. 26, pp. 3630–3633, 1987.

[11] B. V. K. Vijara Kumar, "Minimum variance synthetic discriminant functions," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1579–1584, 1986.

[12] P. Refregier, "Filter design for optical pattern recognition: Multi-criteria optimization approach," *Opt. Lett.*, vol. 15, pp. 854–856, 1990.

[13] B. V. K. Vijaya Kumar and A. Mahalanobis, "Recent advances in composite correlation filter designs," *Asian J. Phys.*, vol. 8, pp. 407–420, 1999.

[14] A. Mahalanobis, B. V. K. Vijaya Kumar, S. Song, S. R. F. Sims, and J. Epperson, "Unconstrained correlation filters," *Appl. Opt.*, vol. 33, pp. 3751–3759, 1994.

[15] A. Mahalanobis, B. V. K. Vijaya Kumar, and S. R. F. Sims, "Distance classifier correlation filters for multi-class target recognition," *Appl. Opt.*, vol. 35, pp. 3127–3133, 1996.

[16] A. Mahalanobis and B. V. K. Vijaya Kumar, "Polynomial filters for higher-order correlation and multi-input information fusion," *Proc. SPIE, Euro-American Optoelectronic Information Processing Workshop*, pp. 221–231, 1997.

[17] B. V. K. Vijaya Kumar, A. Mahalanobis, and A. Takessian, "Optimal tradeoff circular harmonic function (OTCHF) correlation filter methods providing controlled in-plane rotation response," *IEEE Trans. Image Process*, vol. 9, pp. 1025–1034, 2000.

[18] R. Kerekes and B. V. K. Vijaya Kumar, "Correlation filters with controlled scale response," *IEEE Trans. Image Process*, vol. 15, Jul. 2006.

[19] J. Rosen and J. Shamir, "Scale-invariant pattern recognition with logarithmic radial harmonic filters," *Appl. Opt.*, vol. 28, pp. 240–244, 1989.

[20] M. Savvides, B. V. K. Vijaya Kumar, and P. K. Khosla, "Face verification using correlation filters," in *Proc. 3rd IEEE Automatic Identification Advanced Technologies*, Tarrytown, NY, Mar. 2002, pp. 56–61.

[21] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces," Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-01-02, 2001.

[22] M. Savvides and B. V. K. Vijaya Kumar, "Quad-phase minimum average correlation energy filters for reduced-memory illumination-tolerant face authentication," in *Proc. 4th Intl. Conf. Audio- and Video-Based Biometric Person Authentication* (AVBPA), vol. LCNS 2688, pp. 19–26: Springer-Verlag, 2003.

[23] M. Savvides, B. V. K. Vijaya Kumar, and P. K. Khosla, "Corefaces—Robust shift invariant PCA based correlation filter for illumination tolerant face recognition," *IEEE Comp. Vision and Patt. Rec. (CVPR)*, vol. II, pp. 834–841, 2004.

[24] C. Xie, M. Savvides, and B. V. K. Vijaya Kumar, "Redundant class-dependence feature analysis based on correlation filters using FRGC2.0 data," in *Proc. IEEE Workshop on Face Recognition Grand Challenge Experiments in Conjunction With CVPR 2005*, San Diego, CA, 2005.

[25] M. H. Yang, "Kernel Eigenfaces versus Kernel Fisherfaces: Face recognition using Kernel methods," in *Proc. IEEE Conf. Auto. Face Gesture Recog.*, 2002, pp. 215–220.

[26] ——, "Face recognition using kernel methods," in *Advances in Neural Information Processing Systems*, T. Diederich, S. Becker, and Z. Ghahramani, Eds., 2002, vol. 14.

[27] C. Xie, M. Savvides, and B. V. K. Vijaya Kumar, "Kernel Correlation Filter Based Redundant Class-Dependence Feature Analysis (KCFA) on FRGC2.0 data," in *Proc. 2nd Int. Workshop Analysis Modeling of Faces Gestures (AMFG 2005) Held in Conjunction With ICCV 2005*, Beijing, 2005.

[28] R. Gross and V. Brajovic, "An image pre-processing algorithm for illumination invariant face recognition," in *Proc. 4th Int. Conf. AVBPA*, 2003, pp. 10–18.

[29] W. Zheng, C. Zou, and L. Zhao, "Real-time face recognition using Gram-Schmidt orthogonalization for LDA," in *Proc. ICPR'04*, Cambridge, U.K., 2004, pp. 403–406.

[30] B. Scholkopf, *Support Vector Learning*. Munich, Germany: Oldenbourg-Verlag, 1997.

[31] P. J. Phillips, "Support vector machines applied to face recognition," in *Advances in Neural Information Processing Systems 11*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., 1998.

[32] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 725–737, May 2006.

[33] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, and W. Worek, "Preliminary face recognition grand challenge results," in *Proc. 7th Int. Conf. Automatic Face and Gesture Recognition (FGR'06)*, 2006.

[34] P. J. Phillips, "FRGC and ICE workshop," in *Presentation at the FRGC Workshop*, Arlington, VA, Mar. 22–23, 2006. [Online]. Available: http://face.nist.gov/frgc/presentations.htm

ABOUT THE AUTHORS

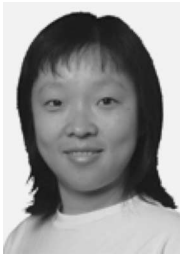
Bhagavatula V. K. Vijaya Kumar (Senior Member, IEEE) is a Professor in the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA. His research interests include automatic target recognition algorithms, biometric recognition methods, and coding and signal processing for data storage systems. His publications include the book entitled *Correlation Pattern Recognition* (coauthored with Dr. A. Mahalanobis and Dr. R. Juday, Cambridge University Press, 2005), eight book chapters, and about 400 technical papers.

Dr. Kumar served as the Pattern Recognition Topical Editor for the Information Processing division of *Applied Optics* and is currently serving as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He has served on many conference program committees and was a Cogeneral Chair of the 2004 Optical Data Storage conference and a Cogeneral Chair of the 2005 *IEEE AutoID Workshop*. He is a Fellow of SPIE—The International Society of Optical Engineering, a Fellow of the International Association of Pattern Recognition (IAPR), and a Fellow of the Optical Society of America (OSA). In 2003, he received the Eta Kappa Nu award for Excellence in Teaching from the ECE Department at CMU and the Carnegie Institute of Technology's Dowd Fellowship for educational contributions.

Marios Savvides (Member, IEEE) received the B.Eng. (hons) degree in microelectronics systems engineering from the University of Manchester Institute of Science and Technology (UMIST), U.K., in 1997, the M.Sc. degree in robotics from the Robotics Institute, Carnegie Mellon University (CMU), Pittsburgh, PA, in 2000, and the Ph.D. degree from the Electrical and Computer Engineering Department also at CMU in May 2004, where he focused on biometrics identification technology for his thesis "Reduced Complexity Face Recognition using Advanced Correlation Filters and Fourier Subspace Methods for Biometric Applications."

He is currently a Research Assistant Professor at the Electrical and Computer Engineering (ECE) Department at CMU with a joint appointment at Carnegie Mellon CyLab. Prior to his current position, he was a Systems Scientist in electrical and computer engineering and CyLab at CMU from May 2004 until January 2006. He was also the Technical Lead in CMU's face recognition and iris efforts in FRGC+FRVT2006 and ICE1.0+ICE2006. His research interests are in biometric recognition of face, iris, fingerprint and palmprint modalities. He has authored and coauthored over 75 conference and journal articles in the biometrics area, has filed two patent applications, and authored one book chapter in this field.

Dr. Savvides is a member of SPIE and serves on the Program Committee of several biometric conferences including IEEE AutoID, SPIE Defense and Security Biometric Identification Technologies and International Conference on Image Analysis and Recognition (ICIAR), and the Biometrics Symposium of the annual Biometrics Consortium (BC 2006). He also served on the Biometrics Panel in SPIE's Defense and Security Symposium 2006 and is also listed in 2005 Edition of *Marquis' Who's Who in America*.



Chunyan Xie (Student Member, IEEE) received the B.S. degree in automation, in 1997, and the M.S. degree in detection technology and control systems, in 2000, both from the Department of Automation, Tsinghua University, Beijing, China. She received the Ph.D. degree in electrical and computer engineering, in 2006, from Carnegie Mellon University, Pittsburgh, PA.

Her research interests include pattern recognition, computer vision and image processing, and focus on face recognition. To date, she has authored or coauthored 18 conference and journal publications.