Class-Specific Kernel Fusion of Multiple Descriptors for Face Verification Using Multiscale Binarised Statistical Image Features

Shervin Rahimzadeh Arashloo and Josef Kittler, *Member, IEEE*

***Abstract*— This paper addresses face verification in unconstrained settings. For this purpose, first, a nonlinear binary class-specific kernel discriminant analysis classifier (CS-KDA) based on spectral regression kernel discriminant analysis is proposed. By virtue of the two-class formulation, the proposed CS-KDA approach offers a number of desirable properties such as specificity of the transformation for each subject, computational efficiency, simplicity of training, isolation of the enrolment of each client from others and increased speed in probe testing. Using the proposed CS-KDA approach, a regional discriminative face image representation based on a multiscale variant of the binarized statistical image features is proposed next. The proposed component-based representation when coupled with the dense pixel-wise alignments provided by a symmetric MRF matching model reduces the sensitivity to misalignments and pose variations, gauging the similarity more effectively. Finally, the discriminative representation is combined with two other effective image descriptors, namely the multiscale local binary patterns and the multiscale local phase quantization histograms via a kernel fusion approach to further enhance system accuracy. The experimental evaluation of the proposed methodology on challenging databases demonstrates its advantage over other methods.**

***Index Terms*— Face verification, class-specific kernel discriminant analysis, binarized statistical image features, descriptor fusion.**

I. Introduction

With the saturation of performance of face recognition systems on controlled data, the recent focus of research in this area has been directed more towards recognizing faces in challenging conditions of real life photos. In these situations, imaging conditions previously kept under control in laboratory settings such as pose, illumination, expression, occlusion, low resolution, *etc.* may vary uncontrollably, perturbing image data and subsequently leading to classification errors. From a pattern classification point of view, the problem has been approached in a variety of ways with different levels of success. A line of research in this respect is focused on designing classifiers which can better cope with the nonlinearities of face manifold [1]–[4]. Complementary to the design of more effective classifiers in dealing with image degradations are the attempts to develop or combine robust low level image representations [4]–[7]. The current work follows both avenues by proposing a new kernel space regional face image representation and a class-specific nonlinear classifier which combines multiple representations in a discriminative subspace.

The algorithms based on subspace techniques are the most widely employed methods in face recognition. These methods usually represent facial images by vectors and then try to find a projection function by optimising some criterion over these vectors. PCA [8] and LDA [9] are two prominent examples of the methods in this category. As LDA seeks to find an optimal projection such that the separation between different classes is maximised, it is generally believed that it performs better than its PCA-based counterpart. However, the performances of linear classifiers drop as soon as the data to be classified is highly complex and nonlinear. A suitable alternative in such cases is offered by nonlinear classification techniques such as kernel discriminant analysis. In this framework, the data are implicitly mapped into a very high dimensional space (possibly infinite dimensional) with the hope that they become linearly separable in this new space [10]–[12]. Nevertheless, a drawback of these methods is their high computational complexity which often requires eigen-analysis of dense matrices. Recently, a spectral regression based kernel discriminant analysis (SR-KDA) has been proposed which uses spectral regression instead of costly eigen-analysis computation [13]. The method has been shown to be orders of magnitude faster than the ordinary KDA. Drawing on this approach and motivated by similar works [14], [15], in this paper a client-specific spectral regression-based KDA approach (CS-KDA) is proposed which casts a multi-class classification task into a binary-class problem. Using a spectral regression framework, in contrast to other class-specific KDA projections, the eigen-analysis computation in the proposed approach is avoided. In addition, by virtue of the two-class formulation, the individual patterns mapped onto the feature space are represented as one dimensional data compared to the $C-1$ dimensional vectors obtained in the multi-class scenario (C being the number of classes). This has significant implications on training, enrolment and testing stages of a verification system as follows. First of all, as each client class is represented by a distinct transforma-

Manuscript received January 18, 2014; revised June 24, 2014; accepted August 28, 2014. Date of publication September 19, 2014; date of current version November 10, 2014. This work was carried out as part of EPSRC project Signal Processing in a Networked Battlespace under Contract EP/K014307/1, and the European Union project Beat. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jie Yang.

S. R. Arashloo is with the Department of Electrical Engineering, Faculty of Engineering, Urmia University, Urmia 57135, Iran (e-mail: sh.rahimzadeh@hotmail.co.uk).

J. Kittler is with the Centre for Vision, Speech and Signal Processing, Department of Electronic Engineering, University of Surrey, Surrey GU2 7XH, U.K. (e-mail: j.kittler@surrey.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2014.2359587

tion, the mapping onto the feature space is subject-specific. As a result, personal facial characteristics of each subject are embedded in the transformation function which in turn enhances system accuracy. Moreover, as the enrolment of each subject into the verification system is isolated from the enrolment of others, registering a new user into the system does not alter the transformation functions of the subjects previously enrolled. In addition, the proposed CS-KDA approach is computationally more efficient than the conventional multi-class SR-KDA in various stages of a face recognition system. An appealing characteristic of the proposed approach is the operability in a single sample framework, thus providing a suitable solution for various applications including image pair matching [16], passport check, *etc.*

The CS-KDA method is used to construct a discriminative face image descriptor using the binarised statistical image features (BSIF) [17]. Similar to MLBP [18] and MLPQ [19] representations, the binarised statistical image features encode local micro-structures of image content into a string of binary codes. But unlike these approaches, it employs statistics of images to learn its filters which improves its representational capacity. Moreover, the employed descriptor is be able to capture image content at multiple resolutions using filters of different spatial scales. Motivated by the works in [20]–[23], for reduced sensitivity of the proposed system to misalignments and out-of-plane head rotations, two techniques are employed here. First, the multiscale BSIF filter responses are summarised regionally as histograms. Second, an MRF image matching model is employed at the image level to provide dense pixelwise correspondences between a pair of images. By symmetrising the MRF matching process, the similarity of a pair of images in the kernel space is captured more effectively in two directions. Finally, the MLBP, MLPQ and MBSIF representations are combined via a kernel fusion approach to further increase system accuracy.

In summary, the main contributions of the current work include:

- A class-specific kernel discriminant analysis approach (CS-KDA) based on spectral regression;
- A kernelised regional discriminative face image descriptor (kernel MBSIF);
- Combination of the MBSIF, MLBP and MLPQ representations via the CS-KDA approach;
- Symmetric comparison of a pair of face images in the kernel space.

The rest of the paper is organised as follows: In Section II, we review the literature with an emphasis on KDA-based methods in face recognition. In Section III, after a short overview of the SR-KDA method of [13], the proposed CS-KDA approach is introduced. In Section IV, the proposed discriminative face descriptor based on the CS-KDA approach and the multiscale BSIF features is discussed. The discussion is then followed by a component based approach for kernel fusion of multiple descriptors. An experimental evaluation of the proposed methodology along with a comparison to other approaches is provided in Section V. Finally, the paper is drawn to conclusion in Section VI.

II. Related Work

There are numerous attempts to deal with the complex patterns of faces employing different variants of kernel discriminant analysis. While in [24] a kernel discriminant analysis method is utilised for the classification of faces under difficult facial expression changes, the authors in [25] propose a null space-based KDA method. In this approach, samples are first mapped onto the kernel space through the so-called cosine kernel. Next, a truncated null space KDA is used which requires only a single eigenvalue analysis. In [14], the problem of face verification is cast as a two-class problem and a class-specific transformation providing a multi-dimensional projection is derived. The optimisation criterion is defined so as to maximise the similarities of genuine claims while minimising the similarity of imposter claims to the mean of each claimed class. The eigenproblem solution and the transformation function obtained include more than one kernel fisher face per class.

In [26], the authors propose to use a bagging technique to decrease the computational cost of kernel fisher discriminant analysis in the training phase by dividing the training data into several subsets and training a different classifier for each subset. A multiple kernel construction method for kernel based fisher discriminant analysis is proposed in [27]. The proposed kernel is constructed as a linear combination of several base kernels with a constraint on their weights which is then used for face recognition. In [28], using the minimum squared error criterion a new kernel-based nonlinear discriminant analysis algorithm is proposed. Once the data is mapped onto a higher-dimensional feature space, the minimum squared errors criterion is employed as the discriminant rule and the corresponding transformation is derived. As this solution does not require the scatter matrices to be nonsingular, the proposed method is applicable to the under-sampled multi-class problems. In order to improve the heterogeneous face recognition performance, a coupled discriminant analysis method is proposed in [29]. The upside of the method is that all samples from different modalities are employed to represent the coupled projections. In addition, the locality information in the kernel space is incorporated into the coupled discriminant analysis as a constraint to improve the generalization capability. The authors in [30] proposed a multi-view dynamic face model to extract the shape-and-pose-free facial texture patterns. A kernel discriminant analysis approach is then developed to extract the significant nonlinear features maximising the between-class variance while minimising the within-class variance. In [31], a new kernel fisher discriminant analysis called complete KFD is developed where the problem formulation is divided into two phases as the kernel PCA and Fisher linear discriminant analysis. In [32], authors proposed an algorithm called KDA/QR, which extends the LDA/QR algorithm to deal with nonlinear data by using the kernel operator. Then an efficient approximation of KDA/QR called AKDA/QR is proposed. In [33], a new kernel direct discriminant analysis based on the direct linear discriminant analysis (DLDA) algorithm [34] is proposed. A robust kernel model with statistical local features for face recognition is proposed in [35]. In this approach, first, multi-partition max

pooling is used to enhance the method's invariance to image alignment errors. Next, a kernel based representation is proposed to exploit the discriminative information available in the statistical local features. In [36], a kernel direct discriminant analysis approach is proposed. The motivating idea is that the null space of within class scatter can include discriminant information if the projection of the between-class scatter in that direction is not zero and that no significant information is lost if the null space of between class scatter is removed. In [4], a KDA fusion approach is proposed to combine MLBP and MLPQ histograms for face recognition. The method is reported to achieve good performance in challenging conditions on a number of different databases.

III. Kernel Discriminant Analysis (KDA)

A. Overview

Let us assume that there exist m samples $x_1, x_2, \ldots x_m \in \mathbb{R}^n$, which belong to C classes and $\mathcal{F}$ is a feature space induced by a nonlinear mapping $\phi : \mathbb{R}^n \rightarrow \mathcal{F}$. For a suitably chosen mapping, an inner product $\langle ., . \rangle$ on $\mathcal{F}$ may be represented as $\langle \phi(x_i), \phi(x_j) \rangle = \kappa(x_i, x_j)$, where $\kappa(., .)$ is a positive semi-definite kernel function. Let S_b^ϕ, S_w^ϕ and S_t^ϕ denote respectively the between-class, within-class and total scatter matrices in the feature space $\mathcal{F}$. Then we have

$$S_b^\phi = \sum_{k=1}^{C} m_k (\mu_\phi^k - \mu_\phi)(\mu_\phi^k - \mu_\phi)^\top \tag{1}$$

$$S_w^\phi = \sum_{k=1}^{C} \sum_{i=1}^{m_k} (\phi(x_i^k) - \mu_\phi^k)(\phi(x_i^k) - \mu_\phi^k)^\top \tag{2}$$

$$S_t^\phi = S_b^\phi + S_w^\phi = \sum_{i=1}^{m} \Big(\phi(x_i) - \mu_\phi\Big)\Big(\phi(x_i) - \mu_\phi\Big)^\top \tag{3}$$

where μ_ϕ^k and μ_ϕ are the centroid of the k^{th} class and the global centroid in the feature space, respectively. m_k is the number of samples in the k^{th} class and x_i^k denotes the i^{th} sample in the k^{th} class. KDA seeks to find an optimal projection function V_{opt} in the feature space by solving the following optimisation problem

$$V_{opt} = \arg\max_V \frac{V^\top S_b^\phi V}{V^\top S_w^\phi V} \tag{4}$$

which is equivalent to [37]

$$V_{opt} = \arg\max_V \frac{V^\top S_b^\phi V}{V^\top S_t^\phi V} \tag{5}$$

The columns of V_{opt} (v's) are the generalized eigenvectors satisfying

$$S_b^\phi v = \lambda S_t^\phi v \tag{6}$$

It is known that v's satisfying the preceding problem can be expressed as linear combinations of all samples [11], [12]. Thus, there exists coefficients α_i such that each eigenvector v can be represented as $v = \sum_{i=1}^{m} \alpha_i \phi(x_i)$.

In [11], it is shown that Eq. 5 is equivalent to

$$U_{opt} = \arg\max_U \frac{U^\top K W K U}{U^\top K K U} \tag{7}$$

where K is the kernel matrix ($K_{ij} = \kappa(x_i, x_j)$) and W is a matrix reflecting the number of samples in each class, defined as

$$W_{ij} = \begin{cases} 1/m_k & \text{if } x_k \text{ and } x_j \text{ both belong to the } k^{th} \text{ class;} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In this case, the columns of U_{opt} (α's) are given by the eigenvectors corresponding to the non-zero eigenvalues satisfying

$$K W K \alpha = \lambda K K \alpha \tag{9}$$

The number of α's satisfying Eq. 9 is bounded by $C - 1$ as the rank of S_b^ϕ is at most $C - 1$. Once α's are found, the projection of a new sample (x) onto the feature space using each eigenvector v (hereafter referred to as kernel fisher face) is given by

$$\langle v, \phi(x) \rangle = \sum_{i=1}^{m} \alpha_i \langle \phi(x_i), \phi(x) \rangle = \sum_{i=1}^{m} \alpha_i \kappa(x_i, x) = \alpha^\top K(:, x) \tag{10}$$

where $K(:, x) = [\kappa(x_1, x), \ldots, \kappa(x_m, x)]^\top$.

B. Spectral Regression KDA (SR-KDA)

In [13], an efficient method to solve the eigen-problem in Eq. 9 via spectral regression is proposed which avoids costly eigen-analysis computations. The method uses the following theorem:

Theorem 1: Let y be the eigenvector of the eigen-problem

$$W y = \lambda y \tag{11}$$

with the eigenvalue λ. If $K\alpha = y$, then α is the eigenvector of the eigen-problem in Eq. 9 with the same eigenvalue λ.
See [13] for a proof.

Using Theorem 1, one may, instead of solving the eigen-problem in Eq. 9 directly, use a two step approach to solve for α's as follows:

1) Solve $Wy = \lambda y$ for y;
2) Find α satisfying $K\alpha = y$.

If K is positive-definite, then there exists a unique solution for α. If K is singular, it may be approximated by the positive definite matrix $K + \delta I$ where I is the identity matrix and $\delta > 0$ is a regularisation parameter. In this paper, a Gaussian RBF is used for the kernel function, *i.e.* $K_{ij} = \kappa(x_i, x_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / M)$, resulting in a positive definite kernel matrix [12], [13]. Solving for α may then be performed using the Cholesky factorisation and forward-back substitution as follows. If $K = R^\top R$, then α may be found by first solving $R^\top \theta = y$ for θ using forward substitution and then solving $R\alpha = \theta$ for α using back substitution.

C. Incremental SR-KDA

As the Cholesky decomposition has the largest computational cost in the SR-KDA method, in [13] it is proposed to perform this step more efficiently using an incremental approach. In the incremental scheme, the goal is to find the Cholesky decomposition of an $m \times m$ matrix given the Cholesky decomposition of its $(m-1) \times (m-1)$ submatrix. Hence, given the Cholesky decomposition of the kernel matrix K_{m-1} of $m-1$ samples we want to compute the Cholesky factorisation of the kernel matrix K_m for the augmented training set where a single sample (x_m) is injected into the system. The incremental Cholesky decomposition technique may be applied via the *Sherman's March* algorithm [38] for K_m as

$$K_m = \begin{pmatrix} K_{m-1} & \mathbf{k}_{1m} \\ \mathbf{k}_{1m}^\top & k_{mm} \end{pmatrix} = \begin{pmatrix} R_{m-1}^\top & \mathbf{0} \\ \mathbf{r}_{1m}^\top & r_{mm} \end{pmatrix} \begin{pmatrix} R_{m-1} & \mathbf{r}_{1m} \\ \mathbf{0} & r_{mm} \end{pmatrix} \quad (12)$$

where $\mathbf{k}_{1m}$ is an $(m-1) \times 1$ vector given by $\mathbf{k}_{1m} = [\kappa(x_1, x_m), \ldots, \kappa(x_{m-1}, x_m)]^\top$ and $k_{mm} = \kappa(x_m, x_m)$. Eq. 12 reads

$$\begin{aligned} K_{m-1} &= R_{m-1}^\top R_{m-1} \\ \mathbf{k}_{1m} &= R_{m-1}^\top \mathbf{r}_{1m} \\ r_{mm} &= \sqrt{k_{mm} - \mathbf{r}_{1m}^\top \mathbf{r}_{1m}} \end{aligned} \quad (13)$$

Thus, one first solves $\mathbf{k}_{1m} = R_{m-1}^\top \mathbf{r}_{1m}$ for $\mathbf{r}_{1m}$ using forward substitution and then computes r_{mm}. The employed incremental technique reduces the computational cost of the SR-KDA approach from cubic in number of training samples in the batch mode to quadratic in the incremental mode [13].

D. Class-Specific SR-KDA (CS-KDA)

Class-specific projections have been considered previously for face recognition and shown to provide advantages over their multi-class alternatives. Inspired by similar approaches in LDA [15] and KDA [14], in this section a class-specific kernel discriminant analysis based on the SR-KDA approach is proposed. In the proposed framework, a C-class classification problem is recast into a set of two-class problems, *i.e.* a probe either belongs to a claimed genuine class or to a fixed class represented by imposters. The new CS-KDA approach contrasts with the conventional KDA representation involving multiple shared kernel fisher faces in having only one class-specific kernel fisher face per class. Hence, a distinguishing characteristic of the proposed technique is the specificity of the transformation for each subject. Some other features of the proposed approach are simplicity of training, isolation of the enrolment of each client from others and computational efficiency discussed next.

Let us assume that there is one sample of the client class ω ($x_m \in \omega$) and a fixed set of $m-1$ imposter samples $\{x_i | i = 1, \ldots, m-1\}$ (the extension to the case where more than one instance of each client are available is straightforward). The approach in this two-class problem starts by building the $m \times m$ matrix W as

$$W = \begin{pmatrix} \frac{1}{m-1} & \cdots & \frac{1}{m-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{1}{m-1} & \cdots & \frac{1}{m-1} & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix} \quad (14)$$

Since the imposter set is fixed and the number of training samples for each client is assumed to be one, matrix W would be common to all classes. Moreover, matrix W would have exactly two eigenvectors corresponding to the same eigenvalue of 1 [38] as

$$\begin{aligned} y_1 &= [\overbrace{1, \ldots, 1}^{m-1}, 0]^\top \\ y_2 &= [\overbrace{0, \ldots, 0}^{m-1}, 1]^\top \end{aligned} \quad (15)$$

As 1 is a repeated eigenvalue of W, any linear combination of the corresponding eigenvectors is also an eigenvector of W. A vector of all ones of size m (e) would clearly be an eigenvector. Hence, following [13], we consider e as the first eigenvector and the second eigenvector (y_2) is orthogonalised with respect to e using the Gram-Schmidt process [38] to produce y'. The Gram-Schmidt process to orthogonalise y_2 with respect to e is as follows:

$$y' = y_2 - \frac{e^\top y_2}{e^\top e} e = y_2 - \frac{1}{m} e = [\overbrace{\frac{-1}{m}, \ldots, \frac{-1}{m}}^{m-1}, \frac{m-1}{m}]^\top \quad (16)$$

In order to obtain a unit norm eigenvector, y' is divided by its norm:

$$\begin{aligned} y'' &= \frac{y'}{\sqrt{y'^\top y'}} = \frac{y'}{\sqrt{\frac{m-1}{m}}} \\ &= [\overbrace{\frac{-1}{\sqrt{m(m-1)}}, \ldots, \frac{-1}{\sqrt{m(m-1)}}}^{m-1}, \quad \sqrt{\frac{m-1}{m}}]^\top \end{aligned} \quad (17)$$

Discarding vector e, we are left with only a single eigenvector, *i.e.* y''.

After computing K_m using the new sample $x_m \in \omega$, the next step is to find α_ω satisfying $K_m \alpha_\omega = y''$. A procedure that performs this task can be summarised as follows. Given the kernel matrix K_{m-1} of the fixed imposter set, its Cholesky decomposition is computed offline. For the enrolment of a new sample, after computing the augmented kernel matrix K_m, the Cholesky factorisation of K_m only includes the computation of the vector $\mathbf{r}_{1m}$ and scalar r_{mm}. Once the Cholesky decomposition of K_m is obtained, α_ω can be found using the forward-back substitution. Note that since in the proposed CS-KDA approach there is only one eigenvector associated with the equation $K_m \alpha_\omega = y''$, only a single vector of size m, *i.e.* α_ω, should be computed. This is in contrast to the multi-class scenario where a projection matrix of size $m \times (C-1)$ should be estimated each time a single sample is injected into the system.

E. Decision Strategy for CS-KDA

If a probe subject asserts class ω as his/her identity, the features of that person (z) is mapped onto the subspace represented by ν_ω to form a discriminative representation as

$$p_z = \langle \nu_\omega, \phi(z) \rangle = \sum_{i=1}^{m} \alpha_{\omega i} \langle \phi(x_i),\ \phi(z) \rangle \tag{18}$$

where ν_ω is the kernel fisherface for class ω. For decision making, p_z can be compared against the mean of the imposter set (η) in the same discriminative subspace of ν_ω. The projection of η onto ν_ω is obtained as

$$p_\eta = \langle \nu_\omega, \phi(\eta) \rangle = \sum_{i=1}^{m} \alpha_{\omega i} \langle \phi(x_i),\ \phi(\eta) \rangle \tag{19}$$

In this case, one would expect the projected probe of a genuine claimant to be far from the projected imposter mean, resulting in the following decision criterion:

$$\begin{aligned} |p_z - p_\eta| &\leq t_\omega \quad \text{reject claim} \\ |p_z - p_\eta| &> t_\omega \quad \text{accept claim} \end{aligned}$$

where t_ω is the threshold for class ω to accept/reject a claim. $|p_z - p_\eta|$ can be rewritten as

$$\begin{aligned} |p_z - p_\eta| &= |\langle \nu_\omega, \phi(z) \rangle - \langle \nu_\omega, \phi(\eta) \rangle| \\ &= |\sum_{i=1}^{m} \alpha_{\omega i} \langle \phi(x_i), \phi(z) \rangle - \sum_{i=1}^{m} \alpha_{\omega i} \langle \phi(x_i), \phi(\eta) \rangle| \\ &= |\sum_{i=1}^{m} \alpha_{\omega i} \kappa(x_i, z) - \sum_{i=1}^{m} \alpha_{\omega i} \kappa(x_i, \eta)| \\ &= |\sum_{i=1}^{m} \alpha_{\omega i} [\kappa(x_i, z) - \kappa(x_i, \eta)]| = |\alpha_\omega^\top K_\omega^-| \end{aligned} \tag{20}$$

$K_\omega^- = [\kappa(x_1, z) - \kappa(x_1, \eta), \ldots, \kappa(x_m, z) - \kappa(x_m, \eta)]^\top$. Since, the imposter set and as a result the mean is fixed, the terms $\kappa(x_i, \eta)$ (for $i = 1, \ldots, m$) can be computed during training and enrolment phases and used in the verification phase to speed up probe testing.

F. Discussion

Let us assume that the verification system is in the operation phase (after training and enrolment of $C - 1$ users) and an additional user is to be enrolled into the system. In this case, in the common multi-class SR-KDA approach, matrix W should be updated and its $C - 1$ eigenvectors would have to be found using the Gram-Schmidt method. In contrast, in the CS-KDA approach as W remains constant and common to all classes, the requirement to resolve for its eigenvectors is circumvented. As the main computational cost in this step is the cost of Gram-Schmidt process, the proposed CS-KDA approach saves $mC^2 - \frac{1}{3}C^3$ compound arithmetic operations (flams) each consisting of one multiplication and one addition compared to the multi-class approach. Next, in the multi-class scenario, after computing the Cholesky decomposition of the new augmented kernel matrix (K_m), $C - 1$ linear equations of the form $K_m \alpha = y$ need to be solved to form the new transformation function. In contrast, in the proposed CS-KDA approach only one linear equation corresponding to the new user should be solved and the projection functions of the classes previously enrolled remains unaltered. This results in a computational saving of Cm^2 flams.

During the verification (test) phase, the computational advantages of the proposed CS-KDA approach are as follows. Commonly, in a multi-class scenario, the KDA representation spans a subspace of the order of $C - 1$ dimensions. The employed one dimensional representation in the proposed CS-KDA approach in this case results in a computational saving of mC flams compared with the multidimensional representation often employed.

An interesting property of the proposed CS-KDA approach is the operability in an unseen pair matching paradigm [16]. In this scenario, two images which were not available before are presented to the system and a decision must be made whether they belong to the same subject or not. In this case, one can construct a binary classifier using a fixed imposter set and one of the images and then measure the likelihood of the second image belonging to the first class and not to the class represented by imposters. The general procedures for training (offline), enrolment (online) and testing (online) stages of the CS-KDA approach are summarised in the Algorithms 1, 2 and 3.

Algorithm 1 Training

▷ Offline computations

procedure
  Set $y'' = [\overbrace{\frac{-1}{\sqrt{m(m-1)}}, \ldots, \frac{-1}{\sqrt{m(m-1)}}}^{m-1}, \sqrt{\frac{m-1}{m}}]^\top$
  Calculate the kernel matrix K_{m-1}
  Find R_{m-1} satisfying $K_{m-1} = R_{m-1}^\top R_{m-1}$
  Estimate mean of the imposter set (η)
  Calculate $\kappa(x_i, \eta)$ for $i = 1, \ldots, m-1$
  Save R_{m-1}, $\kappa(x_i, \eta)'$s and y''
end procedure

Algorithm 2 Enrolling Class ω by Sample x_m

▷ Online computations

procedure
  Calculate $\mathbf{k}_{1m} = [\kappa(x_1, x_m), \ldots, \kappa(x_{m-1}, x_m)]^\top$ and $k_{mm} = \kappa(x_m, x_m)$
  Find $\mathbf{r}_{1m}$ satisfying $\mathbf{k}_{1m} = R_{m-1}^\top \mathbf{r}_{1m}$
  Calculate $r_{mm} = \sqrt{k_{mm} - \mathbf{r}_{1m}^\top \mathbf{r}_{1m}}$
  Form the Cholesky decomposition of K_m as

$$K_m = R_m^\top R_m = \begin{pmatrix} R_{m-1}^\top & \mathbf{0} \\ \mathbf{r}_{1m}^\top & r_{mm} \end{pmatrix} \begin{pmatrix} R_{m-1} & \mathbf{r}_{1m} \\ \mathbf{0} & r_{mm} \end{pmatrix}$$

  Solve $R_m^\top \theta = y''$ for θ
  Solve $R_m \alpha_\omega = \theta$ for α_ω
  Calculate $\kappa(x_m, \eta)$
  Save α_ω and $\kappa(x_m, \eta)$
end procedure

Algorithm 3 Testing Probe z for Class ω

▷ Online computations

procedure
  Set $K_\omega^- = [\kappa(x_1, z) - \kappa(x_1, \eta), \ldots, \kappa(x_m, z) - \kappa(x_m, \eta)]^\top$
  if $|\alpha_\omega^\top K_\omega^-| \leq t_\omega$ **then**
    Reject claim
  else
    Accept claim
  end if
end procedure

IV. Face Representation Using Multiscale BSIF in the Kernel Space

A. Multiscale BSIF Image Representation

The underlying principle of the binarised statistical image features in [17] is to model the local micro-structures of image content using a set of linear filters in a neighbourhood of a pixel. The BSIF filters are applied to individual patches of an image centred at a pixel. Given an image patch p of size $d \times d$ pixels and a linear filter f_n of the same size, the filter response r_n is given by

$$r_n = \sum_{j,k} f_n(j,k)p(j,k) = \mathbf{f}_n^\top \mathbf{p} \tag{21}$$

where vectors $\mathbf{f_n}$ and $\mathbf{p}$ include elements of f_n and pixels of p, respectively. The operation of applying N different linear filters to the same patch can be represented by stacking all $\mathbf{f}_n$ filters into a single matrix $\mathbf{F}$ of size $N \times d^2$ which generates the responses by a single matrix multiplication as

$$\mathbf{r} = \mathbf{Fp} \tag{22}$$

In the BSIF representation, the statistical dependencies of r_n's are minimised via independent component analysis (ICA). For this purpose, the filter matrix $\mathbf{F}$ is decomposed into two parts as

$$\mathbf{r} = \mathbf{Fp} = \mathbf{UVp} = \mathbf{Uz} \tag{23}$$

where $\mathbf{z} = \mathbf{Vp}$, and $\mathbf{U}$ is an $N \times N$ square matrix which is estimated using ICA. Matrix $\mathbf{V}$ applies a whitening transformation to the data. The dimensionality of each patch in this step is reduced using $N < d^2$ principal eigenvectors of the covariance matrix of randomly chosen image patches. Next, given the whitened data samples $\mathbf{z}$, the independent component analysis is employed to estimate an orthogonal matrix $\mathbf{U}$. The filter matrix $\mathbf{F}$ is then derived as

$$\mathbf{F} = \mathbf{UV} \tag{24}$$

Finally, the response of each filter is binarised via thresholding at zero to produce a binarised feature b_n

$$b_n = \begin{cases} 1 & r_n > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

The number of BSIF filters in each scale is controlled by the number of the eigenvectors retained after the whitening transform. In this work, the first 8 principal eigenvectors of the whitening transformation are used, giving rise to eight filters in each scale. Moreover, the sizes of the filters can be varied in order to capture image content at multiple resolutions. While larger filters can better handle low frequency content and blurring effects, smaller filters are able to capture high frequency variations of image texture. By varying the filter sizes and combining BSIF descriptors in different scales, a multiresolution representation (MBSIF) is derived. The number of scales we use for the multi-scale BSIF is eight. As such, in total 64 filters are learned. Finally, the responses of filters are summarised regionally via histograms. The construction of MBSIF descriptors may be described as follows. We first apply MBSIF operators at Z scales to each face image after photometric normalisation using the method of [39]. The result is a grey level code for each pixel at each resolution, Fig. 1. After cropping the resultant code images to the same size, they are divided into non-overlapping rectangular regions $G_0, G_1, \ldots, G_{J\times J-1}$. The BSIF pattern histogram for region j in the scale of s, $\mathbf{h}_{j,s}$, is computed as

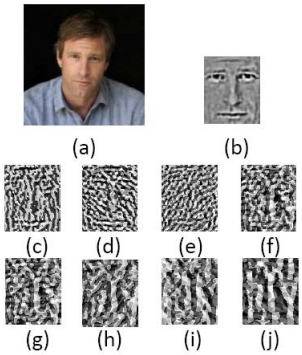

Fig. 1. (a) Original image, (b) normalised and cropped image, (c)-(j) BSIF images at different scales.

$$\begin{aligned} \mathbf{h}_{j,s} &= [h_{j,s}^0, h_{j,s}^1, \ldots, h_{j,s}^{L-1}] \\ h_{j,s}^i &= \sum_{p_c \in G_j} \mathbb{1}\{\text{BSIF}_s(p_c) = i\} \\ j &\in [0, 1, \ldots, J \times J - 1], \\ s &\in [1, 2, \ldots, Z], L = 256 \end{aligned} \tag{26}$$

where $\mathbb{1}\{.\}$ is an indicator function indicating whether its argument is true/false. L represents the number of histogram bins and p_c denotes the centre pixel where the filter is applied. The size of the BSIF filter at scale s is set to $d \times d$ where $d = 2 \times s + 1$. By concatenating all the histograms for each region computed at different scales into a single vector, the multiresolution MBSIF regional descriptor is formed

$$\mathbf{q}_j = [h_{j,1}, h_{j,2}, \ldots, h_{j,Z}]^\top \tag{27}$$

Experimentally it has been found that eight scales ($Z = 8$) are enough to capture a wide range of image frequency content.

B. MBSIF in the Kernel Space

Once the MBSIF histograms are extracted from each region, a regional kernel matrix (K_j) is constructed using a fixed set of imposters and an instance of the subject to be recognised. Following the preceding method, a class-specific regional

kernel fisherface may be estimated and used to project an MBSIF histogram onto a discriminative subspace to form the MBSIF representation in the kernel space. In order to fuse all regional information we follow the approach advocated in [4] and form a combined kernel matrix G as

$$G = \sum_j K_j \tag{28}$$

G can be used to estimate a kernel fisherface for the whole face. In order to boost the efficacy of the system, one may combine the MLBP, MLPQ and MBSIF representations in the kernel space. When multiple image representations are employed, the information from different channels is fused in the CS-KDA approach by adding all regional kernels $K_{j,r}$ corresponding to different regions $j = 0, \ldots, J \times J - 1$ and different image representations $r \in \{\text{MLBP, MLPQ, MBSIF}\}$

$$K^c = \sum_{j,r} K_{j,r} \tag{29}$$

As a result, in order to find the class-specific multiple kernel fisher face for class ω, the equation $K^c \alpha_\omega = y''$ is solved for α_ω with the K^c given by Eq. 29 and y'' by Eq. 17. In the test phase, the vector K_ω^- is formed as $K_\omega^- = \sum_j [\kappa(x_1^j, z^j) - \kappa(x_1^j, \eta^j), \ldots, \kappa(x_m^j, z^j) - \kappa(x_m^j, \eta^j)]^\top$, where x_i^j denotes the j^{th} region of i^{th} training sample and z^j stands for the j^{th} region of the probe.

C. Symmetric Image Matching

In order to reduce the sensitivity of the system to misalignment and out of plane head rotations, we employ the dense MRF matching approach proposed in [20]. In [20], an efficient MRF-based method for dense image registration based on [21], [40], and [41] is proposed. The method matches a block of pixels of the template to a block of pixels in the target image. The correspondences are achieved in a multiresolution framework down to the pixel level. In order to symmetrise the process, following [20], we initially match the template to the target and then exchange the roles of the two images. We repeat the procedure for the horizontally mirrored versions of both images. The regional histograms are finally constructed using the correspondences thus obtained. Once the similarity between each pair of images is computed, the final similarity score is formed by averaging the similarity scores of all pairs of matches. This way, the similarity between a pair of images is gauged symmetrically in the kernel space.

V. Experiments

A. Experiments in Unseen Pair Matching: LFW

The labelled faces in the wild (LFW) dataset is a large database including real world variations of facial images such as pose, illumination, expression, occlusion, low resolution, blur, *etc.* It contains 13,233 images of 5,749 subjects. Evaluation of a method on this dataset is performed by determining whether a pair of images belongs to the same person or not. We evaluate the proposed approach on the "View 2" of the dataset consisting of 3,000 matched and 3,000 mismatched pairs divided into 10 sets. The evaluation is performed in a leave-one-out cross validation scheme on the entire test sets. The overall performance of the method over ten folds is then reported as the mean accuracy and the standard error of the mean. Different evaluation settings on this database are the image "restricted", "unrestricted" and "unsupervised" settings. The restricted setting provides training data for the image pairs as "same" or "not same". The image unrestricted setting in addition provides the identities of the subjects in each pair. In the "unsupervised" setting, no training data in the form of same/not same pairs is provided. We evaluate the proposed approach on the most "restricted" protocol where strictly LFW data is used, *i.e.* without any outside training data. In addition, as we do not use any training data in the form of "same" or "not same", our method is "unsupervised" and is equally comparable with the results in this setting. In an ideal case, the imposter set should not contain images of the subjects in a pair being compared. This might cause the false intuition that the method is operating in a supervised mode. However, if the number of images in the imposter set (m) is sufficiently large then inclusion of a few samples of either one of the subjects in a pair being compared in the imposter set has negligible effects. Similar observations have been made in [15] in the case of linear discriminant analysis. As the inclusion of a few samples of either one of the subjects in a pair in the imposter set does affect the model, the imposter set can be considered as a random collection of face images and as a result the method is unsupervised since neither class labels nor pair labels (same/not same) are used in the construction of the model or its training data. In each of the ten experiments on the LFW data set, one out of ten subsets is used as the test set and the remaining as the training data. We use one of the 9 training subsets as the imposter set. Two separate subsets are used to learn filters for the MBSIF descriptor. Filters are learned in eight scales and in each scale eight filters are learned giving rise to an 8-bit BSIF code for a pixel in each scale. The remaining training subsets are used to set a global acceptance/rejection threshold. We use the funnelled version of the LFW data set and after computing the MLBP, the MLPQ or MBSIF code images, crop the images and keep an area of 80×96 pixels in the centre of the coded image. The number of regions is set to 64 ($J = 8$). The kernel function $\kappa(x_i, x_j)$ is defined as $\kappa(x_i, x_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / M)$. Following [42], M is set to the average squared Euclidean distance between all training samples. In all experiments, once the MRF correspondences are established, the template image is considered as the model and a CS-KDA space is constructed using the model image and the fixed imposter set. Eight systems are evaluated in this experiment:

1) MRF-MLBP-LDA
2) MRF-MLPQ-LDA
3) MRF-MBSIF-LDA
4) MRF-Fusion-LDA
5) MRF-MLBP-CSKDA
6) MRF-MLPQ-CSKDA
7) MRF-MBSIF-CSKDA
8) MRF-Fusion-CSKDA

TABLE I
COMPARISON OF THE PERFORMANCE OF THE PROPOSED APPROACH TO THE STATE-OF-THE-ART METHODS ON THE LFW DATABASE IN THE MOST RESTRICTED SETTING (STRICT LFW, NO OUTSIDE TRAINING DATA USED)

Method	$\mu \pm S_E$
Eigenfaces, original [8]	0.6002 ± 0.0079
Nowak, original [41]	0.7245 ± 0.0040
Nowak, funnelled [41]	0.7393 ± 0.0049
Hybrid descriptor-based, funnelled [42]	0.7847 ± 0.0051
3×3 Multi-region Histograms(1024) [43]	0.7295 ± 0.0055
Pixels/MKL, funnelled [44]	0.6822 ± 0.0041
V1-like/MKL, funnelled [44]	0.7935 ± 0.0055
APEM (fusion), funnelled [45]	0.8408 ± 0.0120
MRF-MLBP, funnelled [20]	0.7908 ± 0.0014
Fisher vector faces [46]	0.8747 ± 0.0149
MRF-MLBP-LDA	0.8212 ± 0.0155
MRF-MLPQ-LDA	0.8326 ± 0.0150
MRF-MBSIF-LDA	0.8446 ± 0.0116
MRF-Fusion-LDA	0.8802 ± 0.0064
MRF-MLBP-CSKDA	0.9068± 0.0132
MRF-MLPQ-CSKDA	0.9242± 0.0103
MRF-MBSIF-CSKDA	0.9363± 0.0127
MRF-Fusion-CSKDA	**0.9589 ± 0.0194**

The results obtained in the most restricted protocol are reported in Table I. We have tested four variants of the CS-KDA approach. These are the CS-KDA approach on the MLBP, MLPQ, MBSIF histograms and a combined kernel using all three descriptors. Reported in the table are also the systems employing a linear discriminant analysis rather than a kernel approach. When using linear discriminant analysis, the similarities of two corresponding regions in a pair of images are measured in the LDA space using cosine similarity and all the regional similarities obtained using different descriptors are combined via a sum rule. A number of observations from the table are in order. First of all, the proposed MBSIF descriptor is shown to perform better than the two other commonly used texture representations, namely the MLBP [18] and MLPQ [19]. This is illustrated both in the kernel and LDA subspaces. The improved representational capacity achieved in the new descriptor can be attributed to different facts. First, the filters employed in the construction of the MBSIF descriptor are based on statistical analysis of facial images in contrast to other ad hoc design schemes such as those used in LBP. Second, the redundancy in the input data is minimised via a whitening transform before applying the filters. Finally, by using an independent component analysis in the BSIF filter design, the filter responses become statistically independent, thus suitable for further processing under independence assumptions. Another observation is that the CS-KDA versions of all four systems consistently perform better than their LDA variants, thanks to the nonlinear nature of the CS-KDA method. In comparison to other approaches, our single descriptor system using MBSIF in the CS-KDA space (MRF-MBSIF-CSKDA) achieves an impressive performance of 93.63%, more than 6% better than the previous best result in this setting. Combination of the three different descriptors results in an improved accuracy both in the kernel and linear subspaces. Comparing our best performing system (MRF-Fusion-CSKDA) against other approaches in this setting, it can be observed that the proposed approach outperforms the previous best result by 8.42% on average. As mentioned earlier, our method is unsupervised and hence it can be compared to other methods under this protocol. The results are provided in Table II. It can be observed that the proposed system improves the best result in this setting by more than 8% margin on average. The proposed system is even comparable to other approaches in the "unrestricted" protocol. The results in this setting are provided in Table III. It can be observed that the MRF-Fusion-CSKDA approach achieves the best performance in this protocol despite following a restricted protocol.

TABLE II
COMPARISON OF THE PERFORMANCE OF THE PROPOSED APPROACH TO THE STATE-OF-THE-ART METHODS ON THE LFW DATABASE IN THE UNSUPERVISED SETTING

Method	AUC
SD-MATCHES, 125×125, aligned [47]	0.5407
H-SX-40, 81×150, aligned [47]	0.7547
GJD-BC-100, 122×225, aligned [47]	0.7392
LARK unsupervised, aligned [6]	0.7830
LHS, aligned [48]	0.8107
Pose Adaptive Filter (PAF) [49]	0.9405
MRF-MLBP [20]	0.8994
MRF-Fusion-CSKDA	0.9894

TABLE III
COMPARISON OF THE PERFORMANCE OF THE MRF-FUSION-CSKDA APPROACH TO THE STATE-OF-THE-ART METHODS ON THE LFW DATABASE IN THE UNRESTRICTED SETTING

Method	$\mu \pm S_E$
LDML-MkNN, funnelled [50]	0.8750 ± 0.0040
Combined multishot, aligned [51]	0.8950 ± 0.0051
LBP multishot, aligned [51]	0.8517 ± 0.0061
LBP PLDA, aligned [52]	0.8733 ± 0.0055
combined PLDA, funnelled/aligned [52]	0.9007 ± 0.0051
combined Joint Bayesian [53]	0.9090 ± 0.0148
high-dim LBP [54]	0.9318 ± 0.0107
Fisher vector faces [46]	0.9303 ± 0.0105
Sub-SML [55]	0.9075 ± 0.0064
VMRS [56]	0.9205 ± 0.0045
MRF-Fusion-CSKDA	**0.9589 ± 0.0194**

B. Experiments in Verification: XM2VTS

Next, we evaluate the proposed system on the rotation shots of the XM2VTS database [58] in a verification scenario. This experiment is designed to explore the capabilities of the proposed approach subject to severe head rotations. The XM2VTS rotation data set [58] is comprised of 295 subjects consisting of 200 clients, 25 evaluation imposters and 70 test imposters. The performance of the system is stated in *Equal Error Rate* (EER) where the False Acceptance and False Rejection rates are equal and the threshold for acceptance or rejection of a claimant is set using the true identities of test subjects. We crop the frontal training images using manually

TABLE IV
COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD TO THE STATE-OF-THE-ART METHODS ON THE XM2VTS DATABASE

Method	EER
3D correc.[59]	7.12
face matching[38]	4.85
MRF-MLBP [20]	4.27
MRF-Fusion-LDA	3.62
MRF-Fusion-CSKDA	**3.0**

annotated eye coordinates to a size of 128×144 pixels so that the distance between the eyes is 70 pixels. The parameter J is set to 8 as before. The test images are detected using the Viola-Jones face detection method [59]. It has been observed that the detection method failed to detect faces in $\sim 2\%$ of the images where there were a severe head rotation. In these cases, we roughly crop out the face area from the image manually. After detection, each face image is scaled in a way that the face area corresponds roughly to an area of 128×144 pixels. As a result, the system is evaluated in the presence of misalignment and moderate scale deviations in addition to pose changes. In this experiment, we compare the proposed method to other pose-invariant approaches on this data set. The results are reported in Table IV. From the table, it can be observed that the MRF-Fusion-CSKDA approach outperforms the MRF-Fusion-LDA and other competitors on this data set. It can be concluded that when dense pixelwise correspondences are available, the combination of the MLBP, MLPQ and MBSIF descriptors via the CS-KDA approach can provide a robust representation for face recognition across pose.

VI. CONCLUSION

Face verification in unconstrained settings was addressed in this paper. In order to cope better with the complex nonlinear face patterns, first, a nonlinear binary class-specific classifier (CS-KDA) based on spectral regression kernel discriminant analysis was proposed. The proposed approach offered a number of desirable characteristics such as specificity of the transformation for each subject, computational efficiency, simplicity of training, isolation of the enrolment of each client from others. Using the proposed nonlinear method, a regional discriminative face image descriptor using the multiscale binarised statistical image features (MBSIF) in the kernel space was proposed next. The component-based representation was coupled with the dense pixelwise alignments provided by an MRF matching model in order to reduce the sensitivity to misalignments and pose variations. Finally, the descriptor was combined with the multiscale local binary patterns (MLBP) and the multiscale local phase quantisation (MLPQ) histograms via a kernel fusion approach which resulted in improved performance. In conclusion, the main contributions of the current work can be summarised as follows:

- a class-specific kernel discriminant analysis approach based on spectral regression was proposed which avoided costly eigen-analysis computation and high dimensional feature projections;
- An effective multiscale regional discriminative face image descriptor (kernel MBSIF) in the kernel space was proposed. The descriptor enjoyed better representational capacity by employing filters learnt via statistical analysis of images and the nonlinear nature of the projection function;
- A class-specific KDA fusion approach for the combination of MBSIF, MLBP and MLPQ representations was proposed. The combination was demonstrated to enhance system performance compared to either one of the single descriptor systems;
- The sensitivity to pose variations was minimised by employing an MRF model at the heart of the system to provide dense symmetric alignment between a pair of face images;
- The similarity between a pair of images was gauged more effectively via a symmetric approach in the kernel space.

REFERENCES

[1] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3499–3506.

[2] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Computer Vision* (Lecture Notes in Computer Science), vol. 6493, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Berlin, Germany: Springer-Verlag, 2010, pp. 709–720.

[3] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE ICCV*, Oct. 2009, pp. 365–372.

[4] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikäinen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1164–1177, May 2013.

[5] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE CVPR*, Jun. 2013, pp. 3025–3032.

[6] H. J. Seo and P. Milanfar, "Face verification using the LARK representation," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1275–1286, Dec. 2011.

[7] Z. Lei, M. Pietikainen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, Feb. 2014.

[8] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1991, pp. 586–591.

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

[10] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," in *Proc. 9th IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process.*, Aug. 1999, pp. 41–48.

[11] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, 2000.

[12] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[13] D. Cai, X. He, and J. Han, "Speed up kernel discriminant analysis," *VLDB J.*, vol. 20, no. 1, pp. 21–33, Feb. 2011.

[14] G. Goudelis, S. Zafeiriou, A. Tefas, and I. Pitas, "Class-specific kernel-discriminant analysis for face verification," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 570–587, Sep. 2007.

[15] J. Kittler, Y. P. Li, and J. Matas, "Face verification using client specific fisher faces," in *The Statistics of Directions, Shapes and Images*, Leeds Annual Statistical Reserach, 2000.

[16] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," School Comput. Sci., Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.

[17] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Tsukuba, Japan, Nov. 2012, pp. 1363–1366.

[18] C.-H. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," in *Proc. Int. Conf. Biometrics*, Aug. 2007, pp. 809–818.

[19] M. A. Tahir, C.-H. Chan, J. Kittler, and A. Bouridane, "Face recognition using multi-scale local phase quantisation and linear regression classifier," in *Proc. 18th IEEE ICIP*, Sep. 2011, pp. 765–768.

[20] S. R. Arashloo and J. Kittler, "Efficient processing of MRFs for unconstrained-pose face recognition," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep./Oct. 2013, pp. 1–8.
[21] S. R. Arashloo, J. Kittler, and W. J. Christmas, "Facial feature localization using graph matching with higher order statistical shape priors and global optimization," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–8.
[22] S. R. Arashloo and J. Kittler, "Pose-invariant face matching using MRF energy minimization framework," in *Energy Minimization Methods in Computer Vision and Pattern Recognition* (Lecture Notes in Computer Science), vol. 5681, D. Cremers, Y. Boykov, A. Blake, and F. Schmidt, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 56–69.
[23] S. R. Arashloo and J. Kittler, "Pose-invariant face recognition by matching on multi-resolution MRFs linked by supercoupling transform," *Comput. Vis. Image Understand.*, vol. 115, no. 7, pp. 1073–1083, 2011.
[24] P. Tsai, T. Jan, and T. Hintz, "Kernel-based subspace analysis for face recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2007, pp. 1127–1132.
[25] W. Liu, Y. Wang, S. Z. Li, and T. Tan, "Null space-based kernel Fisher discriminant analysis for face recognition," in *Proc. 6th Int Conf. Autom. FGR*, May 2004, pp. 369–374.
[26] Y. Li, B. Zhang, S. Shan, X. Chen, and W. Gao, "Bagging based efficient kernel fisher discriminant analysis for face recognition," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3. 2006, pp. 523–526.
[27] X.-Z. Liu and G.-C. Feng, "Multiple kernel learning in fisher discriminant analysis for face recognition," *Int. J. Adv. Robot. Syst.*, vol. 10, pp. 1–7, 2013.
[28] W.-J. Zeng, X.-L. Li, X.-D. Zhang, and E. Cheng, "Kernel-based nonlinear discriminant analysis using minimum squared errors criterion for multiclass and undersampled problems," *Signal Process.*, vol. 90, no. 8, pp. 2333–2343, 2010.
[29] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, "Coupled discriminant analysis for heterogeneous face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1707–1716, Dec. 2012.
[30] Y. Li, S. Gong, and H. Liddell, "Recognising trajectories of facial identities using kernel discriminant analysis," in *Proc. Brit. Mach. Vis. Conf.*, 2001, p. 2003.
[31] J. Yang, A. F. Frangi, and J.-Y. Yang, "A new kernel Fisher discriminant algorithm with application to face recognition," *Neurocomputing*, vol. 56, no. 0, pp. 415–421, 2004.
[32] T. Xiong, J. Ye, Q. Li, R. Janardan, and V. Cherkassky, "Efficient kernel discriminant analysis via QR decomposition," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran & Associates Inc., 2004.
[33] X.-J. Wu, J. Kittler, J.-Y. Yang, K. Messer, and S.-T. Wang, "A new kernel direct discriminant analysis (KDDA) algorithm for face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2004, pp. 1–11.
[34] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data—With application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, 2001.
[35] M. Yang, L. Zhang, S. C.-K. Shiu, and D. Zhang, "Robust kernel representation with statistical local features for face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 900–912, Jun. 2013.
[36] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 117–126, Jan. 2003.
[37] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Computer Science & Scientific Computing), 2nd ed. San Francisco, CA, USA: Academic, Oct. 1990.
[38] G. W. Stewart, *Matrix Algorithms: Basic Decompositions*, vol. 1. Philadelphia, PA, USA: SIAM, 2001.
[39] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. 3rd Int. Workshop AMFG*, 2007, pp. 168–182.
[40] S. R. Arashloo and J. Kittler, "Energy normalization for pose-invariant face recognition based on MRF model image matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1274–1280, Jun. 2011.
[41] A. Shekhovtsov, I. Kovtun, and V. Hlavac, "Efficient MRF deformation model for non-rigid image matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–6.
[42] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, Jun. 2007.
[43] E. Nowak and F. Jurie, "Learning visual similarity measures for comparing never seen objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.
[44] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Proc. Faces Real-Life Images Workshop ECCV*, Oct. 2008.
[45] C. Sanderson and B. C. Lovell, "Multi-region probabilistic histograms for robust and scalable identity inference," in *Advances in Biometrics* (Lecture Notes in Computer Science), vol. 5558, M. Tistarelli and M. S. Nixon, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 199–208.
[46] N. Pinto, J. J. DiCarlo, and D. D. Cox, "How far can you get with a modern face recognition test set using only simple features?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2591–2598.
[47] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3499–3506.
[48] K. Simonyan, O. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2013, pp. 8.1–8.12.
[49] J. Ruiz-del-Solar, R. Verschae, and M. Correa, "Recognition of faces in unconstrained environments: A comparative study," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 1–19, Apr. 2009.
[50] G. Sharma, S. ul Hussain, and F. Jurie, "Local higher-order statistics (LHS) for texture categorization and facial analysis," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 1–12.
[51] D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2013, pp. 3539–3545.
[52] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 498–505. [Online]. Available: http://lear.inrialpes.fr/pubs/2009/GVS09
[53] Y. Taigman, L. Wolf, and T. Hassner, "Multiple one-shots for utilizing class label information," in *Proc. BMVC*, Sep. 2009, pp. 1–12.
[54] P. Li, S. Prince, Y. Fu, U. Mohammed, and J. H. Elder, "Probabilistic models for inference about identity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 144–157, Jan. 2012.
[55] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," in *Computer Vision* (Lecture Notes in Computer Science), vol. 7574, A. W. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer-Verlag, 2012, pp. 566–579.
[56] Q. Cao, Y. Ying, and P. Li, "Similarity metric learning for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013.
[57] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1960–1967.
[58] K. Messer, J. Matas, J. Kittler, and K. Jonsson, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int. Conf. Audio Video-Based Biometric Person Authentication*, 1999, pp. 72–77.
[59] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
[60] J. R. Tena, R. S. Smith, M. Hamouz, J. Kittler, A. Hilton, and J. Illingworth, "2D face pose normalisation using a 3D morphable model," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 1–6.

Shervin Rahimzadeh Arashloo received the Ph.D. degree in computer vision and pattern recognition from the Centre for Vision, Speech and Signal Processing at University of Surrey, Guildford, U.K. His research interests include biometrics, graphical models, and cognitive vision. He is currently an Assistant Professor of Electrical Engineering with Urmia University, Urmia, Iran, and a Visiting Senior Fellow with the University of Surrey.

Josef Kittler is currently a Professor of Machine Intelligence with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, U.K., where he conducts research in biometrics, video and image database retrieval, medical image analysis, and cognitive vision. He has authored a textbook entitled *Pattern Recognition: A Statistical Approach* (Prentice Hall), and over 170 journal papers. He serves on the Editorial Board of several scientific journals in Pattern Recognition and Computer Vision.