Pattern Anal Applic (2017) 20:113–126
DOI 10.1007/s10044-015-0475-1

THEORETICAL ADVANCES

Multiscale binarised statistical image features for symmetric face matching using multiple descriptor fusion based on class-specific LDA

Shervin Rahimzadeh Arashloo[1]

Received: 22 March 2014 / Accepted: 16 April 2015 / Published online: 5 May 2015

Abstract Local binary image coding for face image representation is established as a successful methodology mostly popularized by the well-known local binary pattern operator (LBP) and its variants. In this paper, an alternative learning-based binary image coding scheme is introduced which operates by projecting local image patches linearly onto a subspace using learnt filters. Most importantly, independent binarisation of filter responses is justified theoretically using independent component analysis in the filter learning stage. The extension of the method to a multiscale framework makes the feature capable to capture image content at multiple resolutions, improving its expressive power. Taking a local feature-based approach, the coded images are summarised regionally by histograms exploiting dense correspondences between images. A discriminative face image descriptor is constructed next by projecting the regional multiscale histograms onto a class-specific LDA space. The proposed discriminative descriptor can be learnt in an unsupervised fashion and hence perfectly suited for face recognition in unconstrained settings, including the unseen face pair matching task. Finally, the proposed MBSIF descriptor is combined with two state-of-the-art face image representations, namely the multiscale LBP and local phase quantisation features to further enhance the accuracy. The proposed approach has been evaluated extensively on the extended Yale B, LFW, FERET and the XM2VTS databases in various scenarios and shown to perform very favourably compared to the state-of-the-art methods.

Keywords Unconstrained face recognition · Face image representation · Binarised statistical image features · Multiscale representation

1 Introduction

Motivated by its widespread range of practical applications in surveillance, identification systems, access control, social networks, etc. face recognition has been an active research topic in pattern recognition over decades. Promoted by the face recognition grand challenge, recognition rates under well controlled settings have almost saturated [47]. With this achievement, the recent focus of research has been directed towards recognizing faces in the presence of undesired perturbations in imaging conditions such as variability in lighting conditions, subject pose and expression, misalignment, occlusion, low resolution, etc [31]. The challenges in this case are caused by the large variability in appearance of the same subject and small sample size compared to the dimensionality of the data.

Although not fundamental to the operational logic of a system, quality of the feature representation adopted in an algorithm may impose serious limitations on performance. Consequently, much effort has recently been focused on designing new low level image descriptors and/or combining multiple features to surpass standard representations such as SIFT [41], Gabor [40], HOG [22], LBP [61], LPQ [50], etc. Moreover, many of the current descriptors in image analysis such as LBP [61], SIFT [41], etc. have a hand-crafted design, not benefiting much from statistical learning which limits their representation capacity. An alternative is to develop new features via statistical learning [17, 32]. In this paper, a new face image representation based on binarised statistical image features (BSIF) [24] is

✉ Shervin Rahimzadeh Arashloo
sh.rahimzadeh@hotmail.co.uk

[1] Department of Electrical Engineering, Faculty of Engineering, Urmia university, Urmia, Iran

introduced and then extended into a multiscale framework (MBSIF). Similar to some other common representations in face image analysis such as LBP [61] and LPQ [50], the new descriptor converts the local micro-structures of a face image into a set of discrete codes. This is realised using a number of different filters and projecting an image/sub-images linearly onto a subspace, the basis vectors of which are estimated via unsupervised learning. In other words, the MBSIF descriptor benefits from a learning stage in contrast to ad hoc design schemes used in some other alternatives.

The binary string code generation in many descriptors including LBP and LPQ is achieved by independently binarising each element. A fundamental prerequisite for independent binarisation of code elements is their statistical independence. While this condition is only approximately met in LBP or LPQ descriptors, in the MBSIF descriptor the justification for independent processing is provided using independent component analysis (ICA) in the filter design procedure. Extending the BSIF descriptor into a multiscale framework increases its representation capability, enabling the feature to capture image content at multiple resolutions. It is shown that the extension of the BSIF representation to a multiscale scheme is fundamentally beneficial, rendering the representation to perform on par or better than widely employed descriptors in the field. By stacking the frequency of occurrences of the MBSIF binary codes into a histogram, one may characterise the statistical distributions of filter responses at different scales.

In practice prior to extracting features, an alignment step is performed on the images. The alignment is usually imposed via an affine or similarity transformation using detected facial landmark points. However, such 2D holistic alignments will be insufficient in presence of out-of-plane head rotations. Even in frontal poses, an error in the localisation of a landmark will result in misalignment of the whole face. To address the problem, two approaches are pursued in the present work. First, a Markov random field (MRF) image matching model is embedded at the pixel level to provide dense alignment between a pair of images [5, 6]. The benefits of employing such an approach are two fold. First, it provides dense pixelwise alignment between a pair of images which is quite useful for face recognition in unconstrained settings [3]. Second, the matching is discriminative in the sense that two images of the same subject would most probably provide a good match while images of different subjects are less likely to be matched accurately. As a result, the method acts as a discriminative pre-processing step for the subsequent stages of a recognition pipeline. The MBSIF histogram is then constructed locally taking into account the correspondences and then mapped into an LDA space for comparison. Finally, the regional MBSIF descriptor similarities are summed up to produce the final similarity score.

An appealing characteristic of the proposed approach is the capability to perform unseen face pair matching. That is, given a pair of face images which were not available to the system before, the system should decide whether they belong to the same subject or to different individuals. The decision in this case can be made using a class-specific fisher discriminant analysis (CSLDA) [34]. The employed class-specific LDA transformation is used to construct discriminative subspaces for the features extracted from each image in a pair using a single sample and a fixed set of training data (imposter set). As will be described, the CSLDA transformation can be constructed in an unsupervised fashion making it a suitable candidate for the unseen face matching task. A further characteristic of the proposed technique is the symmetric face comparison. To this end, the method computes the similarity between a pair of face images by symmetrising the MRF matching process and as a result the LDA space feature construction and matching. This is in contrast to previous widely employed asymmetric methods where the similarity is measured only in one direction, compromising performance. The similarity score of the proposed MBSIF + CSLDA descriptor is finally combined with those of the MLBP [18] and MLPQ [62] representations via a sum rule to further increase the accuracy. As will be illustrated, the proposed method provides better discrimination and robustness than many of the existing state-of-the-art approaches in the most challenging situations of real life photos.

In summary the main contributions of the present work can be summarised as follows.

- A novel discriminative multiscale image descriptor (MBSIF + CSLDA) using statistical learning based on a variant of linear discriminant analysis is proposed. The discriminative descriptor can be learnt in an unsupervised fashion, suitable for unseen image pair matching tasks.
- In order to gauge the similarity of a pair of images, the face pair matching task is symmetrised. For this purpose, the discriminative LDA subspace learning is performed symmetrically, improving recognition performance.
- The proposed descriptor is combined with the MLBP and MLPQ features in a score level fusion scheme in an LDA space to further enhance the recognition accuracy.
- Last but not least, a dense pixelwise image pair matching method embedded at the pixel level makes the proposed method applicable to the problem of pose robust recognition of faces.

The rest of the paper is organised as follows. In Sect. 2, we briefly review the literature. Section 3 presents the details of our proposed multiscale local descriptor. In Sect. 4, the symmetric face matching approach is introduced. An evaluation of the proposed method including a comparison to the state-of-the-art methods is presented in Sect. 5 following which the conclusions are drawn in Sect. 6.

2 Related work

A great extent of the early efforts at face recognition made extensive use of features extracted globally from an image and mapped onto a lower dimensional space called subspace. Two prominent examples in this group are eigenfaces [67] and fisherfaces [12]. However, as local feature-based approaches demonstrated a higher degree of robustness against image perturbation, presently the majority of the best performing methods widely exploit local features for the characterisation of face image data. As an example, the authors in [17] use vector quantized local pixels to extract discriminative information from different face regions. While references [2, 64] use histogram of local pattern features (such as LBP, LTP etc.), reference [49] uses spatially localised Gabor filters in a multi-layer framework for face verification. In [44], the authors propose to use histogram of local binary patterns extracted from orientation images, achieving good performance using a single training sample per subject. A more recent approach to boosting the performance under unconstrained settings is to jointly use multiple local descriptors [17, 36, 75], where in the combination is applied via a wide range of methods from combination at the decision level to multiple kernel learning (MKL). Some other recent methods adopt metric-learning approaches for improved similarity comparison [23, 28, 43]. In [72, 73], the authors propose a two-level classifier, training a small number of one-shot and two-shot classifiers for each test pair employing one or both test images as positive samples and an additional set of negative samples. Employing a set of attribute (race, gender, hair colour, etc.) classifiers, the authors in [35, 36] also make use of this two-level classifier. Recently, a blur tolerant image descriptor called local phase quantization (LPQ) operator is introduced by Rahtu et al. [50]. LPQ has been shown to perform better than the local binary pattern (LBP) operator in face recognition and texture classification. In [76], global and local Gabor phase pattern histograms are proposed for face recognition. Graph-based approaches constitute a major category in part-based local face matching. In this framework [5, 7, 70, 71], different subregions of a face are processed independently of other non-neighbouring regions. Such a processing model is helpful in dealing with local geometrical distortions and handling occlusions and cluttered background. In addition, under this framework, good performance may be achieved even using only one training image per class. The current work uses a graph-based method for dense symmetric pixelwise alignment of faces. After establishing dense correspondences, regional multi-resolution features are employed for decision making in an LDA space.

3 Face representation via multiscale binarised statistical image features (MBSIF)

3.1 BSIF image coding

The binarised statistical image features (BSIF) is a generative model based on the independent component analysis (ICA) [32]. ICA represents the data as a linear transformation of some latent independent components. Let $\mathbf{p}$ denote the pixel grey values in an image patch concatenated into a vector. Using ICA, $\mathbf{p}$ can be represented using a feature matrix $\mathscr{F}$ as

$$\mathbf{p} = \mathscr{F}\mathbf{r} \tag{1}$$

where the elements of the vector $\mathbf{r}$ are some unknown random variables which differ from one patch to another. Conversely, the elements of $\mathscr{F}$ are constant and the same for all different image patches. A fundamental assumption regarding this linear generative model is that the elements of $\mathbf{r}$ are statistically independent. In this case, one may, using a large enough number of training samples, recover a reasonable approximation to $\mathscr{F}$ up to a multiplicative constant without explicitly knowing the latent vector $\mathbf{r}$ [32]. Estimation of $\mathscr{F}$ is equivalent to determining the matrix $\mathbf{F}$ which produces $\mathbf{r}$ as the output of a number of linear filters as

$$\mathbf{r} = \mathbf{F}\mathbf{p} \tag{2}$$

where each row of $\mathbf{F}$ represents a filter to be applied on the pixels of patch $\mathbf{p}$.

In practice, the statistical models are applied on pre-processed data. Suppose that the pixels of a single patch after pre-processing are collected into the vector $\mathbf{z} = (z_1, \ldots, z_N)$. Commonly, for pre-processing a linear transformation is used. In this case, z_i's would be linear transformations of the independent components r_i's. This can be readily observed by multiplying both sides of Eq. 1 by the matrix performing the pre-processing and obtain

$$\mathbf{z} = \mathscr{U}\mathbf{r} \tag{3}$$

where matrix $\mathscr{U}$ is obtained by multiplying matrix $\mathscr{F}$ by the pre-processing transformation matrix, $\mathbf{V}$. In practice, a whitening transformation is used as the pre-processing step as it is found to be instrumental in contrast gain and luminance control [32]. In this case, for matrix $\mathscr{U}$ to be invertible, the number of independent components should be chosen in a way that it equals the number of variables produced after the whitening transformation. Under this condition, the system in Eq. 3 would be invertible in a unique way, producing the vector **r** as a linear function of **z** as

$$\mathbf{r} = \mathbf{U}\mathbf{z} \tag{4}$$

where matrix **U** is obtained by inverting matrix $\mathcal{U}$. The filter matrix **F** in Eq. 2 can then be obtained by multiplying the linear transformations given by **U** and **V**, i.e.

$$\mathbf{F} = \mathbf{U}\mathbf{V} \tag{5}$$

As a result, the independent components r_i's of vector **r** are obtained as

$$\mathbf{r} = \mathbf{U}\mathbf{V}\mathbf{p} \tag{6}$$

In summary, given an image **p** of size $d \times d$ pixels, one applies N filters on the pixels of **p** using the filter matrix $\mathbf{F}^{N \times d^2}$ and obtains N responses which are stacked into the vector **r**. As the filter responses r_i's are independent, they can be processed independently. A useful post-processing step is binarising r_i's by thresholding at zero to produce the binarised features b_i's as

$$b_i = \begin{cases} 1 & r_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The binarised features of b_i's can then be summarised using aggregate statistics such as histograms.

3.1.1 *Training for BSIF filters*

The training procedure for filter matrix **F** can be summarised as follows. Using a training set of image patches randomly taken from images, their covariance matrix is estimated and eigen-decomposed. The dimensionality of each patch is then reduced using N (number of the filters used) principal eigenvectors of the covariance matrix divided by their standard deviations. At the end of this step, whitened data samples **z** are obtained. In more detail, if the eigen decomposition of the covariance matrix **C** is $\mathbf{C} = \mathbf{Y}\mathbf{D}\mathbf{Y}^\top$, where **D** is the diagonal matrix of eigen values of **C** in a descending order and the columns of **Y** are the corresponding eigen vectors of **C**, then matrix **V** which is used for whitening and dimensionality reduction is given by

$$\mathbf{V} = \left[\mathbf{D}^{-1/2}\mathbf{Y}\right]_{1:N} \tag{8}$$

where $[.]_{1:N}$ denotes the first N rows of a matrix. Next, given the whitened data samples **z**, the independent component analysis is employed to estimate an orthogonal matrix **U**. Having estimated the matrices of **U** and **V**, the final filter matrix is obtained as **UV**. Some sample filters learnt are depicted in Fig. 1. In the figure, eight BSIF filters corresponding to a of size 17×17 are depicted. By applying the filters, eight filter responses are obtained which are then binarised to form an 8-bit binary code for each pixel.

An essential prerequisite in the binarisation is the independency of filter responses [32, 46]. As ICA is used for filter design, the dependencies between filter responses in the binarised statistical image features approach are minimised. This is in contrast to some commonly employed techniques such as local binary patterns where the independency holds only approximately.

3.2 Multiscale analysis

Suppose the size of each individual BSIF filter is fixed at $d \times d$. In this case, using a larger number of filters (increasing N) would include more high frequency components into the descriptor. This is because the N eigenvectors of the covariance matrix of the training data are sorted in a descending order with respect to their corresponding eigenvalues and increasing N would include more eigenvectors corresponding to smaller eigenvalues into the whitening transformation. Conversely, using a fixed number of filters (N), by increasing the size of each filter, the variations of the signal over a larger support region are taken into account. In others words, the descriptor now captures large scale image content. It has been observed that using eight filters ($N = 8$) results in an acceptable frequency response, able to capture a wide range of frequency content of images [24]. Hence, the number of filters in all experiments in this work is fixed to 8, producing an 8 bit binary code for each pixel. As noted earlier, the other parameter controlling the frequency content of the feature is the filter size. While smaller filters capture small scale variations of texture, larger filters are better suited to deal with blurring effects and low frequency contents. In this work, the compromise brought about by this trade-off is moderated via a simple yet powerful texture representation, called multiscale binarised statistical image feature.

The proposed multi-resolution representation is derived by varying the filter size, and combining the BSIF descriptors in different scales. However, in this case the common problem of high dimensionality and small sample size may result in instability of the representation in the presence of image noise. The problem, however, can be minimised using histograms as aggregate statistics which can capture the most fundamental statistical properties of the feature. The benefits of employing histograms of the code words are three fold. First of all, using a histogram reduces the feature dimension from the image size to that of the histogram. Moreover, by optimising the dimensionality of histogram and projection onto other spaces, the

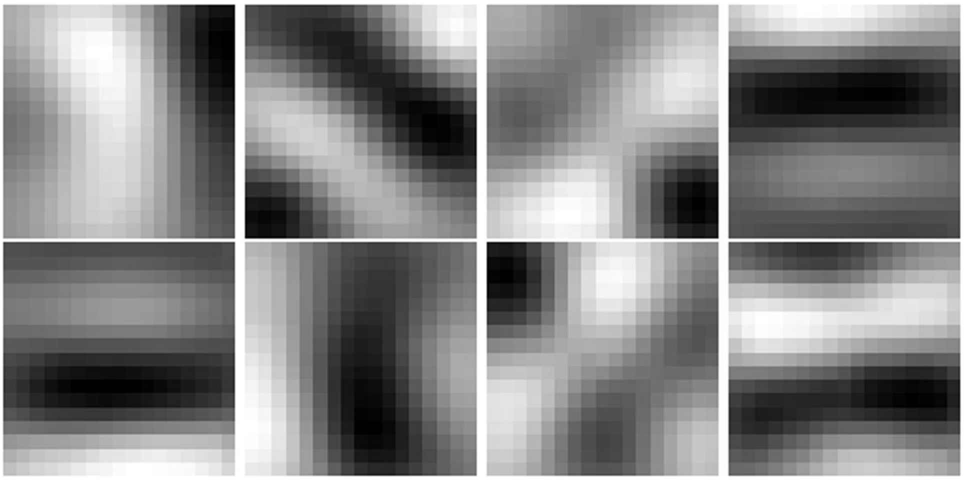

Fig. 1 Sample 17 × 17 BSIF filters ($N = 8$)

effects of the image noise on the feature can be regulated. Finally, a histogram summary is more robust to spatial image transformations such as rotation and translation and hence the sensitivity to misalignment is decreased [39].

3.3 MBSIF face descriptor

In the proposed approach to multi-resolution analysis, BSIF operators at Z scales are first applied to a face image after photometric normalisation [64]. A grey level code for each pixel at each resolution is thus obtained, Fig. 2. The c–j coded images are obtained by applying eight BSIF filters each. The coded image of (c) in the figure corresponds to the finest scale, i.e. the result of applying 3 × 3 filters while the coded image of (j) represents the output of applying BSIF filters at the coarsest scale, i.e. using filters of size 17 × 17. The resulting BSIF code images are divided into non-overlapping rectangular regions $G_0, G_1, \ldots, G_{J\times J-1}$ after cropping to the same size. The BSIF pattern histogram for region j in the scale of s, $\mathbf{h}_{j,s}$, is computed by

$$\begin{aligned} &\mathbf{h}_{j,s} = \left[h^0_{j,s}, h^1_{j,s}, \ldots, h^{L-1}_{j,s}\right] \\ &h^i_{j,s} = \sum_{m\in G_j} \mathbb{1}\{\mathrm{BSIF}_s(m) = i\} \\ &j \in [0, 1, \ldots, J \times J - 1], \\ &s \in [1, 2, \ldots, Z], L = 256 \end{aligned} \tag{9}$$

where $\mathbb{1}\{.\}$ is the indicator function equal to one when its argument is true and zero otherwise. L is the number of histogram bins (determined by the number of filters used) and the size of the BSIF filter at scale s is $d \times d$ where $d = 2 \times s + 1$. By concatenating all the histograms computed at different scales for each region into a single vector, the final multi-resolution regional face descriptor is obtained

$$\mathbf{q}_j = \left[h_{j,1}, h_{j,2}, \ldots, h_{j,Z}\right]^\top \tag{10}$$

3.4 Single sample model construction using class-specific LDA

In order to obtain a discriminative regional descriptor, we use a client-specific linear discriminant analysis (CSLDA) [34] to project the multi-resolution features onto a discriminative subject-specific subspace. The client-specific LDA operates in a two-class framework. That is, when comparing a pair of images, one of them is assumed to be the model (f) and the likelihood of the second image (f') belonging to the first one and not to a class of imposters is measured. The two-class linear discriminant transformation for region j taking f as the model, $\mathbf{a}^f_j$, is given by

$$\mathbf{a}^f_j = S^{-1}_j\left(\mu^f_j - \mu_j\right) \tag{11}$$

where S^{-1}_j denotes inverse of the within-class scatter matrix for region j while μ^f_j and μ_j are the mean histograms of the model image f and training data for the same region, respectively. In [34], it has been shown if the number of training samples excluding those belonging to the subject f are large enough, the inverse of the within-class scatter matrix can be approximated as

$$S^{-1}_j \approx \Psi_j \Phi^{-1}_j \Psi^\top_j \tag{12}$$

where Ψ_j is the matrix of leading eigenvectors of the mixture covariance matrix and Φ_j is the diagonal matrix of corresponding eigenvalues for region j. The reasons supporting the use of client-specific LDA are its perfect adaptability to the unseen face pair matching, computational efficiency, ease of training and lower error rates in classification [34]. Once a regional linear discriminant transformation is estimated, the similarity of two

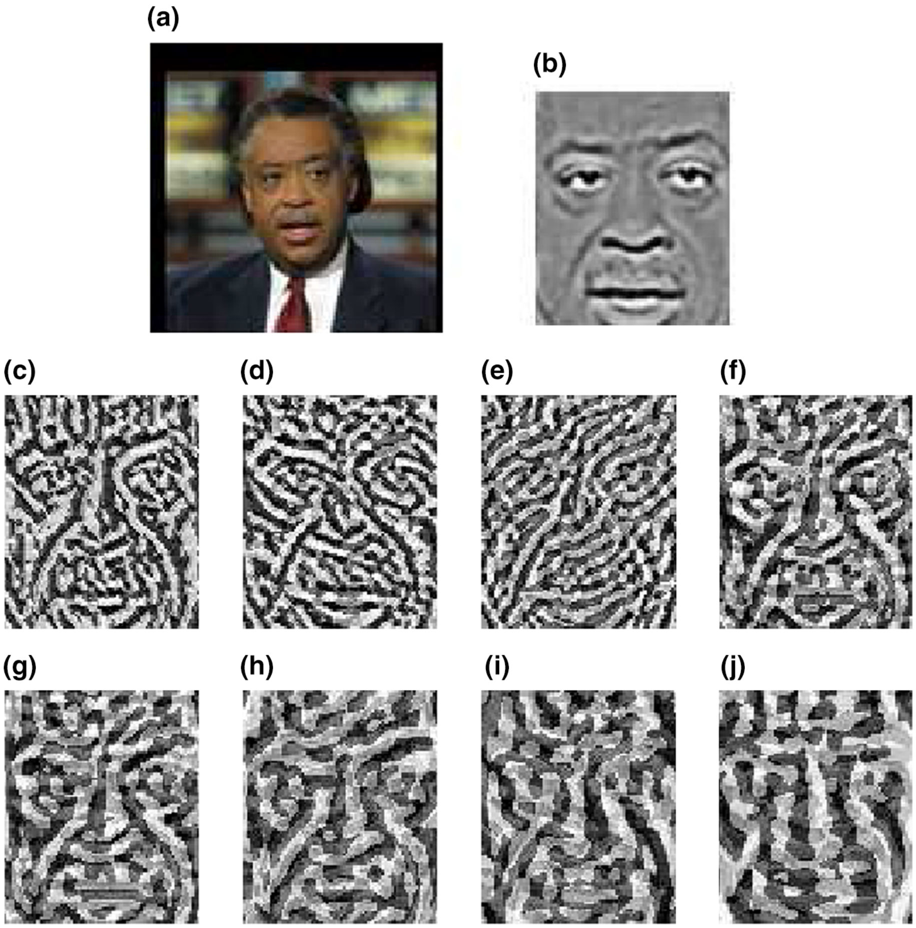

Fig. 2 **a** Original image, **b** normalised and cropped image, **c–j** BSIF coded images at different scales

corresponding regions is measured as the cosine similarity measure $\frac{(\mathbf{a}_j^f)^\top \mathbf{q}_j^{f'}}{\|\mathbf{a}_j^f\|\|\mathbf{q}_j^{f'}\|}$ and the final similarity between a pair of images, $\mathrm{Sim}(f,f')$, is measured as the sum of regional similarities, *i.e.*

$$\mathrm{Sim}(f,f') = \sum_j \frac{\left(\mathbf{a}_j^f\right)^\top \mathbf{q}_j^{f'}}{\left\|\mathbf{a}_j^f\right\| \|\mathbf{q}_j^{f'}\|} \tag{13}$$

3.4.1 Discussion

The rationale for using CSLDA is to obtain a discriminative compact descriptor for face representation and matching. However, the common fisher discriminant analysis is a supervised technique requiring class labels of training examples. Thus, at the first glance it might seem that for a pairwise face matching task where the goal is to gauge the similarity of a pair of images, labelled training data of images belonging to both subjects is required. This is a rather restrictive assumption in practical applications where the two images are never seen before [31]. However, the problem is easily circumvented using the CSLDA approach as follows. Assume there is a set of random training face images. We call this set the imposter set. There is no restriction on this set except that if by any chance a number of images belonging to either one of the subjects to be compared exists in the training set, the number of such samples should be small compared to the total number of training images. This requirement can be easily fulfilled by choosing a large number of training images of random subjects in the imposter set. This condition is studied in [34] and using it the approximation to the within-class scatter matrix in Eq. 12 is derived. Once the imposter set is selected, the within-class scatter matrix for the class-specific LDA can be approximated using Eq. 12. Note that the approximation in Eq. 12 does not require any labels as it only entails an eigen decomposition of the features extracted from the imposter set. Next, we construct a class-specific LDA transformation using Eq. 11, taking μ^f to be the feature extracted from the first image and μ the mean over the imposter set. That is, the transformation for the CSLDA can be constructed using only a single model sample. In this case, the second image would either belong to the imposter set or to the class represented by the first image. The probability of the second image belonging to the first image and not the class represented by imposters is then measured by Eq. 13. Exchanging the roles of the two images, we construct a second CSLDA transformation using the second image as the model and measure the probability of the first image belonging to the second image and not to the class represented by the imposter set. Finally, the similarity of the two images is taken as the average of the two similarity scores thus obtained. In practice, we also make use of the mirrored versions of both images in a pair to reduce the effect of self-occlusion in inconsistent poses. As we use both images and their horizontally flipped versions as model images, four CSLDA transformations would be required. In addition, a pair of images and their horizontally mirrored versions can be matched in eight different ways by exchanging the roles of the model and the test images in each pair. As a result, four CSLDA subspaces and eight image pair comparisons are performed for each pair of images.

Note that the preceding approach for comparing a pair of images is completely unsupervised as no class labels are utilised in obtaining the CSLDA transformation, thanks to the approximation given by Eq. 12. This is extremely advantageous and different from most commonly employed approaches based on linear discriminant analysis in comparing a pair of face images.

4 Dense image alignment

Alignment prior to recognition has a fundamental impact on performance. This has fuelled the research leading to a growing number of methods for object alignment [4, 10, 13, 16, 20, 25, 51, 53, 54, 59, 66, 68, 77]. However, obviously aligning a non-planar object using a 2D transformation such as similarity or affine can only partly correct for the misalignment existing objects. This shortcoming is successfully approached via 2D or 3D methods such as the well-known active appearance model (AAM) [20] or the 3D morphable model (3DMM) [14]. An alternative to these methods is the dense image matching approaches using Markov random fields which estimate pixelwise alignment between a pair of images. For dense image alignment we adopt the method proposed in [4, 6, 7]. The reasons supporting such a choice are as follows. First of all, it provides dense pixelwise alignments between a pair of images. This has been found to be quite advantageous in pose-invariant and also frontal pose face recognition. Second, unlike most MRF-based methods which are rather slow due to high computational complexity of the optimisation problem involved, the method in [4, 7] uses a variety of different techniques including multi-resolution analysis and GPU acceleration to perform matching much faster than many other alternatives. Next, the matching is performed in a discriminative way. That is, unlike other 2D or 3D approaches such as AAMs [20] or 3DMMs [14] which try to fit a generic model to an image, the method in [4, 7] tries to find the best alignment between a pair of images without using a pre-learnt generic model. As a result, one expects to have good alignment (smooth deformation maps) when the two images belong to the same subject and poor alignment when the images are from different subjects. This in effect is likely to lead to high similarity scores in the subsequent stages of a recognition system between images of the same subject and low similarity between images of different individuals. Last but not least, the procedure can be modified to compare a pair of face images symmetrically. Some matching results of this method are depicted in Fig. 3. In this work, we symmetrise the process of matching two images as follows. Initially, the template is matched to the target and then the roles of the two images are exchanged. The procedure is also repeated for the horizontally mirrored versions of both images. As a result, for each pair of images we perform eight matchings. The MBSIF histograms are then computed taking into account the correspondences thus obtained. Once the similarity between each pair of images is computed, the final score is obtained by averaging the similarity scores of all the eight pairs of matches. As will be illustrated in the experiments, the symmetric matching serves to improve the performance by a great extent.

5 Experiments

5.1 Implementation details

In the following experiments, after geometrically normalising the images (the details will be given separately) before extracting features, the cropped face images are preprocessed using an effective photometric normalisation scheme [64]. The applied method is designed to decrease

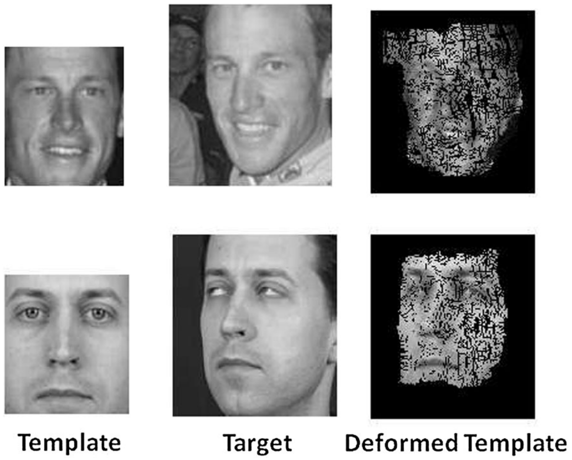

Fig. 3 Some results of dense image-to-image matching using the method of [6, 7]

the effects of changes in illumination conditions, highlights and local shadowing, while keeping the fundamental visual information. In the multi-resolution analysis, the numbers of scales for the multiscale local binary pattern (MLBP) and the multiscale local phase quantisation (MLPQ) operators are set to 10 and 7 respectively, as advocated in [19]. For the BSIF descriptor, while applying the BSIF operator in a small number of scales does not provide sufficient discriminatory information for face representation, an operator with a larger filter size captures lower frequency components which tend to be influenced by the illumination conditions more severely. Here, the number of scales is optimised empirically and is set to 8. The other parameter to tune is the number of local regions ($J \times J$) from which the histograms are extracted. While using fewer regions provides robustness against misalignment, in the case of dense correspondences, using a larger number of regions a larger amount of spatial information becomes available for classification. We investigate the effect of varying J on system performance. Finally, for the construction of within scatter matrices, the dimensionality of the Ψ_j's and Φ_j's is chosen in a way that 95 % of the variation in the training data is preserved.

5.2 Comparison of different descriptors: combined Yale database B and the extended Yale face database B

In this section, a face identification experiment is performed on the combined Yale database B [27] and the extended Yale database B [37] under varying illumination conditions to compare the single scale BSIF descriptor to the proposed discriminative multiscale representation and the multiscale local binary pattern and the multiscale local phase quantisation histograms. The data set consists of 2432 images of 38 subjects under 64 different illumination conditions. For each of the 38 individuals in the database, a single image corresponding to the normal illumination condition is selected as the gallery and all the remaining images are considered as the test samples. Each image in this data set is cropped to a of size 192 × 160 (rows × columns) and then divided into 16 × 16 non-overlapping rectangular regions. For the construction of the imposter set for the class-specific LDA, frontal images of the XM2VTS database [42] are used. The BSIF filters used in this experiment are learnt using an external set of natural images, provided by the authors of [24]. As a result, the generalisation capability of the method is also evaluated. A number of investigations are made in this experiment. The 8-bit single scale BSIF descriptor with varying filter sizes using a χ^2 distance measure is examined. The multiscale BSIF descriptor using the χ^2 distance measure is also evaluated and compared to the single scale BSIF descriptors. In addition, we have also evaluated the multiscale LBP and the multiscale LPQ descriptors for comparison using a χ^2 distance measure. Finally, we have also included the results obtained using the client-specific LDA using the multiscale LBP, multiscale LPQ and the multiscale BSIF descriptor. For the client-specific LDA, for each probe-gallery pair, four client-specific LDA subspaces, two corresponding to the probe and the mirrored probe image and two for the gallery and the mirrored gallery image are learnt. As noted earlier, for each pair of images eight scores are obtained which are averaged to produce the final score. The results obtained are reported in Table 1. A number of observations from the table are in order. First, the proposed multiscale descriptor using a χ^2 distance measure consistently performs better than the single scale versions using the χ^2 distance measure by a large margin. Second, the MBSIF descriptor with a χ^2 measure outperforms both the MLBP and MLPQ representations using the same distance metric. Third, all the three multiscale descriptors using a client-specific LDA perform better than the χ^2 distance measure. The proposed MBSIF + CSLDA approach while performing much better than the MBLP + CSLDA also outperforms the MLPQ + CSLA approach by nearly 6 percent. The improved representational capacity achieved in the new descriptor can be analysed from different viewpoints. First, the filters used in constructing the MBSIF descriptor are estimated using statistical analysis of image properties 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 using PCA in the filter learning procedure. And finally, by using an independent

Table 1 Comparison of the performance of different descriptors on the combined Yale database B and the extended Yale face database B

Method	Accuracy
$\mathrm{BSIF}_3 + \chi^2$	88.34
$\mathrm{BSIF}_5 + \chi^2$	87.46
$\mathrm{BSIF}_7 + \chi^2$	86.55
$\mathrm{BSIF}_9 + \chi^2$	88.13
$\mathrm{BSIF}_{11} + \chi^2$	89.22
$\mathrm{BSIF}_{13} + \chi^2$	89.05
$\mathrm{BSIF}_{15} + \chi^2$	88.17
$\mathrm{BSIF}_{17} + \chi^2$	88.55
$\mathrm{MBSIF} + \chi^2$	93.19
$\mathrm{MLPQ} + \chi^2$	89.05
$\mathrm{MLBP} + \chi^2$	81.99
MBSIF + CSLDA	97.07
MLPQ + CSLDA	91.10
MLBP + CSLDA	84.58

BSIF_d denotes a single scale BSIF filter of size $d \times d$

component analysis in the filter design, the codes generated become statistically independent, thus suitable for further processing under independence assumptions. It can be observed that the proposed discriminative multiscale regional descriptor (MBSIF + CSLDA) improves the performance of the single scale BSIF descriptor to a large extent making it comparable to other alternatives, also emphasised by the following experiments.

5.3 Experiment in unseen pair matching: LFW

Recently with the development of the LFW data set [31] studying the performance of face recognition methods in unconstrained settings has been facilitated. The LFW data set includes real-world variations in facial images such as pose, illumination, expression, occlusion, low resolution, blur etc. It contains 13,233 images of 5749 subjects. The task is to determine whether a pair of images belongs to the same person or not. We evaluate the proposed approach on the "View 2" of the data set consisting of 3000 matched and 3000 mismatched pairs divided into 10 sets. The evaluation is performed in a leave-one-out cross-validation scheme on the entire test sets. The aggregate performance of the method over tenfolds is reported as the mean accuracy and the standard error on the mean. There are different evaluation settings on this database: the image restricted setting and the unrestricted setting. 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. There is also the unsupervised setting where no training data in the form of same/not same pairs are provided. We evaluate the proposed approach on the most restricted protocol where strictly LFW data are used, without any outside training data. In addition, as our method is unsupervised (both the MBSIF filter learning and the CSLDA approach are unsupervised), it is equally comparable with the results in this setting. 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 nine as the training data. We use one of the nine training subsets to learn the projection matrix of the class-specific LDA. Two separate subsets of the remaining eight subsets are used to learn filters for the BSIF descriptor. Filter learning is performed using 20,000 randomly sampled image patches. Filters are learned in eight scales, i.e. $m = \{3, 5, \ldots, 17\}$ and in each scale, eight filters are learned ($N = 8$) giving rise to an 8-bit BSIF code. The remaining training subsets are used to set the acceptance/rejection threshold. We use the funnelled and aligned versions of the LFW data set and after computing the LBP, LPQ or BSIF code images, crop the images and keep an area of 80 × 96 pixels in the centre of the code image. In the experiments on the LFW, a number of investigations are made. First of all, the proposed MBSIF descriptor is compared against two other commonly used texture representations for face recognition, namely the MLBP [18] and MLPQ [62] against a varying J. The results are obtained using the proposed method described in earlier sections, i.e. using the symmetric matching and the client-specific LDA approach on the MBSIF histograms. The results are shown in Fig. 4. A number of observations can be made from the figure. First, it can be seen that the proposed MBSIF descriptor outperforms both MLPQ and MLBP representations. Second, by increasing J and as a result the number of regions, the performance of all three descriptors is improved. This is due to the fact that the underlying MRF matching method provides good pixelwise alignment and by increasing J more spatial information becomes available for recognition. The boost in performance with increasing J is better observed from $J = 2$ to $J = 8$ than from $J = 8$ to $J = 16$ with the performance being almost saturated around $J = 16$.

Next, we study the effect of symmetric MRF matching on recognition performance. We compare the mean accuracies obtained using each descriptor with the proposed symmetric matching method versus the non-symmetric approach. The results are illustrated in Fig. 5. It is observed that irrespective of the value of J, the proposed symmetric face matching method consistently performs better than the

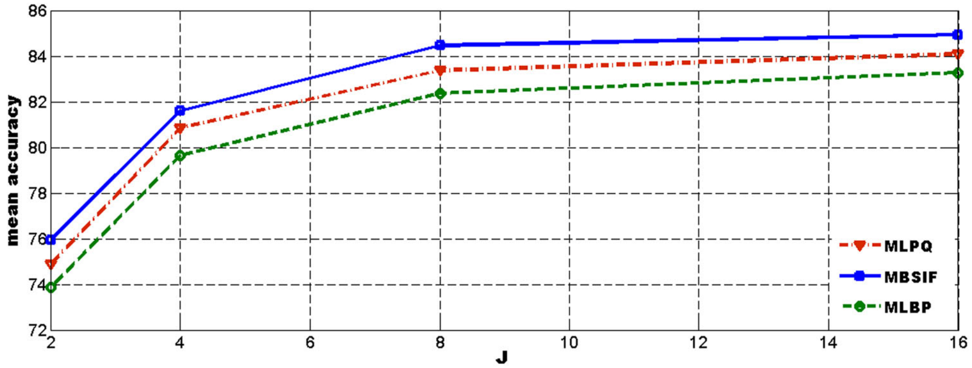

Fig. 4 Comparison of mean recognition accuracies between MBSIF, MLPQ and MLBP descriptors on the LFW data set against varying J

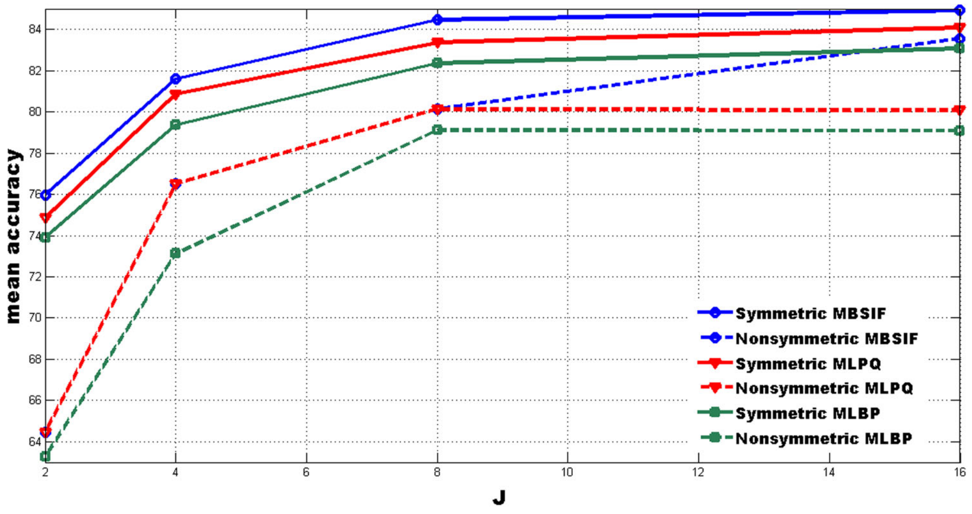

Fig. 5 Effect of symmetric MRF matching on mean recognition accuracy using different descriptors on the LFW data set against varying J

Table 2 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 [67]	0.6002 ± 0.0079
Nowak, original [45]	0.7245 ± 0.0040
Nowak, funnelled [45]	0.7393 ± 0.0049
Hybrid descriptor-based, funnelled [72]	0.7847 ± 0.0051
3 × 3 multi-region histograms (1024) [52]	0.7295 ± 0.0055
Pixels/MKL, funnelled [49]	0.6822 ± 0.0041
V1-like/MKL, funnelled [49]	0.7935 ± 0.0055
APEM (fusion), funnelled [29]	0.8408 ± 0.0120
MRF-MLBP, funnelled [3]	0.7908 ± 0.0014
Fisher vector faces [33]	0.8747 ± 0.0149
MRF-MLBP-CSLDA, funnelled	0.8196 ± 0.0145
MRF-MLPQ-CSLDA, funnelled	0.8336 ± 0.0170
MRF-MBSIF-CSLDA, funnelled	0.8493 ± 0.0126
MRF-fusion, funnelled	0.8819 ± 0.0079

Table 3 Comparison of the performance of the proposed approach to the state-of-the-art methods on the LFW database in the unsupervised setting

Method	$\mu \pm S_E$
SD-MATCHES, 125 × 125, aligned [60]	0.6410 ± 0.0062
H-SX-40, 81 × 150, aligned [60]	0.6945 ± 0.0048
GJD-BC-100, 122 × 225, aligned [60]	0.6847 ± 0.0065
LARK unsupervised, aligend [56]	0.7223 ± 0.0049
LHS, aligned [58]	0.7340 ± 0.0040
I-LPQ*, aligned [1]	0.8620 ± 0.0046
Pose adaptive filter (PAF) [21]	0.8777 ± 0.0051
MRF-MLBP, aligned [3]	0.8008 ± 0.0013
DFD [38]	0.8402 ± 0.0044
VMRS [11]	0.8857 ± 0.0037
MRF-fusion, aligned	0.8935 ± 0.0079

conventional non-symmetric approach. The improvement is more pronounced with a fewer number of subregions yet with the largest number of regions used ($J = 16$), the improvements for MLBP, MLPQ and MBSIF compared to the non-symmetric approach are more than 3.5 , 4 and 1.4 %, respectively.

Next, as the MLBP, MLPQ and MBSIF descriptors provide different representations, it is expected that the information they provide would be complementary to each other and that the recognition performance can be boosted by combining them. For combination, a sum rule over scores obtained in different regions using different descriptors is employed. The result of fusion along with other state-of-the-art results on the LFW data set ($J = 16$) in the most restricted protocol is presented in Table 2. From the table, it is observed that by using only the proposed MBSIF descriptor one can achieve a comparable performance to the previous best results under this setting. Fusing the three MLBP, MLPQ and MBSIF descriptors together we achieve an impressive mean performance of 88.19 %, ranking the proposed approach first under this setting. As noted earlier, our method is unsupervised and can be compared to other approaches under this protocol. In this case, we ran the experiment on the aligned version of the LFW data set [63]. The results of this comparison are provided in Table 3. It can be observed that the proposed approach achieves the best result in this setting.

5.4 Experiment in identification: FERET

In real-world scenarios in-depth rotation of faces is commonly present in face images. In this experiment, we evaluate the proposed method on the rotation shots of the

Table 4 Comparison of the performance of the proposed approach to the state-of-the-art methods on the FERET database

Pose	bi	bh	bg	bf	be	bd	bc	bb
Horizontal deviation angle	−60°	−40°	−25°	−15°	+15°	+25°	+40°	+60°
MRF [30]	na	91.0	97.3	98.0	98.5	96.5	91.5	na
PAF [74]	93.75	98.0	98.50	99.25	99.25	98.50	98.0	93.75
Sarfraz [55]	79.2	92.4	89.7	100	98.6	97.0	89.0	82.5
CLS [57]	79.0	85.0	94.0	96.0	95.0	90.0	82.0	70.0
PAN [26]	52.0	78.5	91.5	98.5	97.0	93.0	81.5	44.0
3D Morph. model [15]	90.7	95.4	96.4	97.4	99.5	96.9	95.4	94.8
Prob. stack flow [8]	∼43	∼65	∼89	∼95	∼93	∼82	∼57	∼34
3D pose norm. [9]	na	90.5	98	98.5	97.5	97.0	91.5	na
MRF-MLBP [4]	92.0	98.5	99.5	100.0	99.5	99.0	99.5	91.0
MRF-MLBP-CSLDA	92.5	98.5	99.0	100.0	100.0	99.0	99.5	92.0
MRF-MLPQ-CSLDA	93.0	98.0	99.5	100.0	100.0	99.0	99.0	93.0
MRF-MBSIF-CSLDA	93.0	98.5	99.5	100.0	100.0	99.5	99.5	93.5
MRF-fusion	94.0	99.0	99.5	100.0	100.0	99.5	99.5	93.5

FERET database [48] i.e. the ***b*** series in an identification scenario. For this experiment, frontal images of 200 clients of the XM2VTS [42] data set are used as the imposter set. This experiment is designed particularly to explore the capabilities of the proposed methodology for recognition in varying pose conditions. This part of the database consists of 200 subjects captured under 9 different yaw angles ranging from nearly −60° to +60°. We use the *ba* image of each subject (almost frontal) as the gallery image and all the rest as test images. Frontal gallery images are cropped using manually annotated eye coordinates to a size of 128 × 144 pixels with an interocular distance of 70 pixels. The test/evaluation images are detected using the Viola and Jones [69] method and scaled so that the face area roughly corresponds to an area of 128 × 144 pixels. Hence, the method is evaluated subject to misalignments and moderate scale deviations. Region parameter J is set to 16. Table 4 reports the correct identification rates obtained on this data. The results of some other methods are also included for comparison. From the table, it can be observed that the proposed approach outperforms all alternative methods in most poses, except the ***bb*** pose (corresponding to an extreme pose deviation of +60° from frontal) in which losing only by approximately 1 %.

5.5 Experiments in verification: XM2VTS

We also evaluate our method on the rotation shots of the XM2VTS database [42]. In the XM2VTS rotation data set the evaluation protocol is based on 295 subjects consisting of 200 clients, 25 evaluation imposters and 70 test imposters. The performance of a verification system is often stated in equal error rate (EER) in which 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. In this experiment, frontal training images are cropped using manually annotated eye coordinates to a size of 128 × 144 pixels so that the distance between the eyes is 70 pixels. As in the FERET experiment, the test/evaluation images are detected and cropped using the Viola and Jones [69] method. After face detection, each image is scaled so that the face area roughly corresponds to an area of 128 × 144 pixels. Parameter J is set to 16. This experiment enables one to compare the proposed method to other similar pose-invariant approaches in a verification scenario subject to challenging settings of face misalignment and pose variation. The rest of the procedure is as described in Sect. 5.1. As in the previous experiment on the FERET database, the imposter set is chosen to be the frontal images of the 200 clients of the XM2VTS database. The best results obtained on this data set are listed in Table 5. It can be observed from the table that the proposed approach obtains the lowest error rate on the rotation shots of the XM2VTS [42] database. In addition to the multi-resolution nature of the descriptors employed, the achieved high performance is attributed to the dense pairwise matching provided by the symmetric matching process and the functionality of the client-specific LDA transformation employed.

Table 5 Comparison of the performance of the proposed method to the state-of-the-art methods on the XM2VTS database

Method	EER
3D correc. [65]	7.12
Face matching [5]	4.85
MRF-MLBP [3]	4.27
MRF-MLBP-CSLDA	3.87
MRF-MLPQ-CSLDA	3.62
MRF-MBSIF-CSLDA	3.37
MRF-fusion	3.12

6 Conclusion

The paper presented a novel discriminative multiscale image descriptor (MBSIF + CSLDA) using statistical learning based on a variant of linear discriminant analysis. The discriminative descriptor which can be learnt in an unsupervised fashion, was shown to be a suitable solution for the unseen image pair matching tasks. Next, in order to gauge the similarity of a pair of images more effectively, the face pair matching task was symmetrised. For this purpose, the discriminative LDA subspace learning was performed symmetrically, improving recognition performance. A dense pixelwise image pair matching method embedded at the pixel level made the proposed technique applicable to pose robust recognition of faces. Finally, the proposed descriptor was combined with the MLBP and MLPQ features in a score level fusion scheme in an LDA space to further enhance the recognition accuracy.

References

1. Hussain SU, Napoléon T, Jurie F (2012) Face recognition using local quantized patterns. In: British machive vision xonference, Guildford, United Kingdom, p.11
2. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns:application to face recognition. PAMI 28(12):2037–2041
3. Arashloo S, Kittler J (2013) Efficient processing of mrfs for unconstrained-pose face recognition. In: Biometrics: theory, applications and systems (BTAS), 2013 IEEE sixth international conference, pp 1–8. doi:10.1109/BTAS.2013.6712721
4. Arashloo S, Kittler J, Christmas W (2010) Facial feature localization using graph matching with higher order statistical shape priors and global optimization. In: Biometrics: theory applications and systems (BTAS), 2010 fourth IEEE international conference, pp 1–8
5. Arashloo SR, Kittler J (2011) Energy normalization for pose-invariant face recognition based on mrf model image matching. IEEE Trans Pattern Anal Mach Intell 33(6):1274–1280
6. Arashloo SR, Kittler J (2014) Fast pose invariant face recognition using super coupled multiresolution markov random fields on a GPU. Pattern Recognit Lett 48(0):49–59 (celebrating the life and work of Maria Petrou)
7. Arashloo SR, Kittler J, Christmas WJ (2011) Pose-invariant face recognition by matching on multi-resolution mrfs linked by super-coupling transform. Comput Vis Image Underst 115(7):1073–1083
8. Ashraf A, Lucey S, Chen T (2008) Learning patch correspondences for improved viewpoint invariant face recognition. CVPR 2008:1–8
9. Asthana A, Marks TK, Jones MJ, Tieu KH, Rohith M (2011) Fully automatic pose-invariant face recognition via 3d pose normalization. In: Computer vision, IEEE international conference, vol 0, pp 937–944. doi:10.1109/ICCV.2011.6126336
10. Baker S, Matthews I (2004) Lucas-Kanade 20 years on: a unifying framework. Int J Comput Vis 56(3):221–255
11. Barkan O, Weill J, Wolf L, Aronowitz H (2013) Fast high dimensional vector multiplication face recognition. In: ICCV, pp 1960–1967
12. Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. Pattern Anal Mach Intell IEEE Trans 19(7):711–720. doi:10.1109/34.598228
13. Belhumeur PN, Jacobs DW, Kriegman DJ, Kumar N (2011) Localizing parts of faces using a consensus of exemplars. In: CVPR. IEEE, pp 545–552
14. Blanz V, Vetter T (2003) Face recognition based on fitting a 3d morphable model. IEEE Trans Pattern Anal Mach Intell 25(9):1063–1074
15. Blanz V, Vetter T (2003) Face recognition based on fitting a 3d morphable model. IEEE Trans Pattern Anal Mach Intell 25:2003
16. Cao X, Wei Y, Wen F, Sun J (2012) Face alignment by explicit shape regression. In: CVPR. IEEE, pp 2887–2894
17. Cao Z, Yin Q, Tang X, Sun J (2010) Face recognition with learning-based descriptor. In: Computer vision and pattern recognition (CVPR), 2010 IEEE conference, pp 2707–2714. doi:10.1109/CVPR.2010.5539992
18. Chan CH, Kittler J, Messer K (2007) Multi-scale local binary pattern histograms for face recognition. In: Proceedings of international conference on biometrics. Springer, pp 809–818
19. Chan CH, Tahir MA, Kittler J, Pietikainen M (2013) Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. IEEE Trans Pattern Anal Mach Intell 35(5):1164–1177. doi:10.1109/TPAMI.2012.199
20. Cootes T, Edwards G, Taylor C (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
21. Dong Y, Zhen L, Stan L (2013) Towards pose robust face recognition. In: IEEE Computer vision and pattern recognition
22. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: CVPR, pp 886–893
23. Davis JV, Kulis B, Jain P, Sra S, Dhillon IS (2007) Information-theoretic metric learning. In: Proceedings of the 24th international conference on machine learning, ICML '07ACM, New York, pp 209–216
24. Kannala J, Esa R (2012) Bsif: binarized statistical image features. In: Proceedings of 21st international conference on pattern recognition (ICPR 2012), Tsukuba, Japan, pp 1363–1366
25. Friedman JH (2000) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232
26. Gao H, Ekenel HK, Stiefelhagen R (2009) Pose normalization for local appearance-based face recognition. In: Proceedings of the third international conference on advances in biometrics, ICB '09. Springer, Berlin, Heidelberg, pp 32–41
27. Georghiades AS, Belhumeur PN, Kriegman DJ (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 23(6):643–660
28. Guillaumin M, Verbeek JJ, Schmid C (2009) Is that you? metric learning approaches for face identification. In: ICCV. IEEE, pp 498–505
29. Li H, Hua G, Lin Z, Brandt L, Yang J (2013) Probabilistic elastic matching for pose variant face verification. In: IEEE computer vision and pattern recognition
30. Ho HT, Chellappa R (2013) Pose-invariant face recognition using markov random fields. Image Process IEEE Trans 22(4):1573–1584
31. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report 07–49, University of Massachusetts, Amherst
32. Hyvrinen A, Hurri J, Hoyer P (2009) Natural image statistics a probabilistic approach to early computational vision. Springer, New York

33. Simonyan K, Parkhi OM, Vedaldi A, Andrew Z (2013) Fisher vector faces in the wild. In: British machine vision conference (BMVC)
34. Kittler J, Li YP, Matas J (2000) Face verification using client specific fisher faces. In: The statistics of directions, shapes and images
35. Kumar N, Berg A, Belhumeur PN, Nayar S (2011) Describable visual attributes for face verification and image search. IEEE Trans Pattern Anal Mach Intell 33(10):1962–1977
36. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009 Attribute and simile classifiers for face verification. In. In IEEE international conference on computer vision (ICCV)
37. Lee K, Ho J, Kriegman D (2005) Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans Pattern Anal Mach Intell 27(5):684–698
38. Lei Z, Pietikainen M, Li SZ (2014) Learning discriminant face descriptor. IEEE Trans Pattern Anal Mach Intell 36(2):289–302. doi:10.1109/TPAMI.2013.112
39. Li,SZ, Jain AK (eds) (2011) Handbook of face recognition, 2nd edn. Springer, Berlin. doi:10.1007/978-0-85729-932-1
40. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans Image Process 11:467–476
41. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110
42. Messer K, Matas J, Kittler J, Jonsson K (1999) Xm2vtsdb: the extended m2vts database. In: Second international conference on audio and video-based biometric person authentication, pp 72–77
43. Mignon A, Jurie F (2012) PCCA: a new approach for distance learning from sparse pairwise constraints. In: IEEE conference on computer vision and pattern recognition, France, pp 2666–2672
44. Nguyen H, Bai L, Shen L (2009) Local gabor binary pattern whitened PCA: a novel approach for face recognition from single image per person. In: Tistarelli M, Nixon M (eds) Advances in biometrics, Lecture notes in computer science, vol 5558. Springer, Berlin, Heidelberg, pp 269–278
45. Nowak E, Jurie F (2007) Learning visual similarity measures for comparing never seen objects. In: Computer vision and pattern recognition, 2007. CVPR '07. IEEE conference, pp 1–8. doi:10.1109/cvpr.2007.382969
46. Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381(6583):607–609
47. Phillips PJ, Flynn PJ, Scruggs WT, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek WJ (2005) Overview of the face recognition grand challenge. In: CVPR (1). IEEE Computer Society, pp 947–954
48. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The feret evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104
49. Pinto N, DiCarlo JJ, Cox DD (2009) How far can you get with a modern face recognition test set using only simple features? In: IEEE computer vision and pattern recognition
50. Rahtu E, Heikkilä J, Ojansivu V, Ahonen T (2012) Local phase quantization for blur-insensitive image analysis. Image Vis Comput 30(8):501–512
51. Rivera S, Martnez AM (2012) Learning deformable shape manifolds. Pattern Recognit 45(4):1792–1801
52. Sanderson C, Lovell BC (2009) Multi-region probabilistic histograms for robust and scalable identity inference. In: Tistarelli M, Nixon MS (eds) ICB, Lecture notes in computer acience, vol 5558. Springer, pp 199–208
53. Saragih J, Gcke R (2007) A nonlinear discriminative approach to aam fitting. In: ICCV. IEEE, pp. 1–8
54. Saragih JM, Lucey S, Cohn JF (2009) Face alignment through subspace constrained mean-shifts. In: ICCV. IEEE, pp 1034–1041
55. Sarfraz MS, Hellwich O (2010) Probabilistic learning for fully automatic face recognition across pose. Image Vis Comput 28(5):744–753
56. Seo HJ, Milanfar P (2011) Face verification using the lark representation. IEEE Trans Inf Forensics Secur 6(4):1275–1286
57. Sharma A, Haj MA, Choi J, Davis LS, Jacobs DW (2012) Robust pose invariant face recognition using coupled latent space discriminant analysis. Comput Vis Image Underst 116(11): 1095–1110
58. Sharma G, ul Hussain S, Jurie F (2012) Local higher-order statistics (lhs) for texture categorization and facial analysis. In: Proceedings of the 12th European conference on computer vision—volume part VII, ECCV'12. Springer, Berlin, Heidelberg, pp 1–12
59. Snchez-Lozano E, la Torre FD, Gonzlez-Jimnez D (2012) Continuous regression for non-rigid image alignment. In: Fitzgibbon AW, Lazebnik S, Perona P, Sato Y, Schmid C (eds) ECCV (7), Lecture notes in computer science, vol 7578. Springer, pp 250–263
60. del Solar JR, Verschae R, Correa M (2009) Recognition of faces in unconstrained environments: a comparative study. EURASIP J Adv Signal Process 2009:1–19
61. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach 24(7):971–987
62. Tahir MA, Chan CH, Kittler J, Bouridane A (2011) Face recognition using multi-scale local phase quantisation and linear regression classifier. In: Macq B, Schelkens P (eds) ICIP. IEEE, pp 765–768
63. Taigman Y, Wolf L, Hassner T (2009) Multiple one-shots for utilizing class label information. In: BMVC, British Machine Vision Association
64. Tan X, Triggs B (2007) Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: AMFG, pp 168–182
65. Tena J, Smith R, Hamouz M, Kittler J, Hilton A, Illingworth J (2007) 2d face pose normalisation using a 3d morphable model. In: International conference on video and signal based surveillance, pp 1–6
66. Tresadern PA, Sauer P, Cootes TF (2010) Additive update predictors in active appearance models. In: Labrosse F, Zwiggelaar R, Liu Y, Tiddeman B (eds) BMVC. British Machine Vision Association, pp 1–12
67. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: Proceedings. 1991 IEEE Computer Society conference on computer vision and pattern recognition. IEEE Computer Society Press, pp 586–591
68. Tzimiropoulos G, Zafeiriou S, Pantic M (2011) Robust and efficient parametric face alignment. In: Metaxas DN, Quan L, Sanfeliu A, Gool LJV (eds) ICCV. IEEE, pp 1847–1854
69. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154. doi:10.1023/B:VISI.0000013087.49260.fb
70. Wang R, Lei Z, Ao M, Li S (2009) Bayesian face recognition based on markov random field modeling. In: ICB, pp 42–51
71. Wiskott L, Fellous J, Kuiger N, von der Malsburg C (1997) Face recognition by elastic bunch graph matching. PAMI 19(7):775–779
72. Wolf L, Hassner T, Taigman Y (2008) Descriptor based methods in the wild. In: Faces in real-life images workshop in ECCV [(b) similarity scores based on background samples]
73. Wolf L, Hassner T, Taigman Y (2009) Similarity scores based on background samples. In: Asian conference on computer vision (ACCV)
74. Yi D, Lei Z, Li S (2013) Towards pose robust face recognition. In: Computer vision and pattern recognition (CVPR), 2013 IEEE conference, pp 3539–3545

75. Ying Y, Li P (2012) Distance metric learning with eigenvalue optimization. J Mach Learn Res 13(1). http://jmlr.csail.mit.edu/papers/v13/ying12a.html

76. Zhang B, Shan S, Chen X, Gao W (2007) Histogram of gabor phase patterns (HGPP): a novel object representation approach for face recognition. IEEE Trans Image Process 16(1):57–68

77. Zhu X, Ramanan D (2012) Face detection, pose estimation, and landmark localization in the wild. In: CVPR. IEEE, pp 2879–2886