Expert Systems With Applications 47 (2016) 23–34

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Fully automatic face normalization and single sample face recognition in unconstrained environments

Mohammad Haghighat[a,*], Mohamed Abdel-Mottaleb[a,b], Wadee Alhalabi[b,c]

[a] *Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146 USA*
[b] *Department of Computer Science, Effat University, Jeddah, Saudi Arabia*
[c] *Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia*

ARTICLE INFO

Keywords:
Face recognition in-the-wild
Pose-invariance
Frontal face synthesizing
Feature-level fusion
Canonical correlation analysis
Active appearance models

ABSTRACT

Single sample face recognition have become an important problem because of the limitations on the availability of gallery images. In many real-world applications such as passport or driver license identification, there is only a single facial image per subject available. The variations between the single gallery face image and the probe face images, captured in unconstrained environments, make the single sample face recognition even more difficult. In this paper, we present a fully automatic face recognition system robust to most common face variations in unconstrained environments. Our proposed system is capable of recognizing faces from non-frontal views and under different illumination conditions using only a single gallery sample for each subject. It normalizes the face images for both in-plane and out-of-plane pose variations using an enhanced technique based on active appearance models (AAMs). We improve the performance of AAM fitting, not only by training it with in-the-wild images and using a powerful optimization technique, but also by initializing the AAM with estimates of the locations of the facial landmarks obtained by a method based on flexible mixture of parts. The proposed initialization technique results in significant improvement of AAM fitting to non-frontal poses and makes the normalization process robust, fast and reliable. Owing to the proper alignment of the face images, made possible by this approach, we can use local feature descriptors, such as Histograms of Oriented Gradients (HOG), for matching. The use of HOG features makes the system robust against illumination variations. In order to improve the discriminating information content of the feature vectors, we also extract Gabor features from the normalized face images and fuse them with HOG features using Canonical Correlation Analysis (CCA). Experimental results performed on various databases outperform the state-of-the-art methods and show the effectiveness of our proposed method in normalization and recognition of face images obtained in unconstrained environments.

1. Introduction

Although face recognition has been a challenging topic in computer vision for the past few decades, most of the attention was focused on recognition based on face images captured in controlled environments. Capturing a face image naturally without controlling the environment, so-called in the wild (Huang, Ramesh, Berg, & Learned-Miller, 2007; Le, 2013), may result in images with different illumination, head pose, facial expressions, and occlusions. The accuracy of most of the current face recognition systems drops significantly in the presence of these variations, specially in the case of pose and illumination variations (Moses, Adini, & Ullman, 1994; Zhao, Chellappa, Phillips, & Rosenfeld, 2003).

Regardless of the face variations in pose, illumination and facial expressions, we humans have an ability to recognize faces and identify persons at a glance. This natural ability does not exist in machines; therefore, we design intelligent and expert systems that can simulate the recognition artificially (Haghighat, Zonouz, & Abdel-Mottaleb, 2015). Building deterministic or stochastic face models is a challenging task due to the face variations. However, normalization can be used in a preprocessing step to reduce the effect of these variations and pave the way for building face models. Pose variations are considered to be one of the most challenging issues in face recognition. Due to the complex non-planar geometry of the face, the 2D visual appearance significantly changes with variations in the viewing angle. These changes are often more significant than the variations of innate characteristics, which distinguish individuals (Zhang & Gao, 2009). In this paper, we propose a fully automatic single sample face

[*] Corresponding author. Tel.: +1 305 284 3291; fax: +1 305 284 4044.
E-mail addresses: haghighat@umiami.edu, haghighat@ieee.org (M. Haghighat), mottaleb@miami.edu (M. Abdel-Mottaleb), walhalabi@effatuniversity.edu.sa (W. Alhalabi).

http://dx.doi.org/10.1016/j.eswa.2015.10.047
0957-4174/© 2015 Elsevier Ltd. All rights reserved.

recognition method that is capable of handling pose variations in unconstrained environments. In the following two sections, we present a literature review of related methods and our contributions in this paper.

1.1. Related work

The Active Appearance Models (AAMs) proposed by (Cootes, Edwards, & Taylor, 1998; 2001) have been used in face modeling for recognition. After fitting the model to a face image, either the model parameters, the location of the landmarks, or the local features extracted at the landmarks are used for face recognition (Edwards, Cootes, & Taylor, 1998; Ghiass, Arandjelovic, Bendada, & Maldague, 2013; Hasan, Abdullaha, & Othman, 2013; Lanitis, Taylor, & Cootes, 1995) or facial expression analysis (Lucey et al., 2010; Martin, Werner, & Gross, 2008; Tang & Deng, 2007; Trutoiu, Hodgins, & Cohn, 2013; Van Kuilenburg, Wiering, & Den Uyl, 2005). For face recognition, (Guillemaut, Kittler, Sadeghi, & Christmas, 2006) and (Heo & Savvides, 2008) proposed using the normalized face images created by warping the face images into the frontal pose.(Gao, Ekenel, & Stiefelhagen, 2009) improved the performance of this technique using a modified piecewise affine warping. None of these methods, however, is fully automatic and they require a manual labeling or manual initialization.

(Chai, Shan, Chen, & Gao, 2007) assumed that there is a linear mapping between a non-frontal face image and the corresponding frontal face image of the same subject under the same illumination. They create a virtual frontal view by first partitioning the face image into many overlapped local patches. Then, a local linear regression (LLR) technique is applied to each patch to predict its corresponding virtual frontal view patch. Finally, the virtual frontal view is generated by integrating the virtual frontal patches. (Li, Shan, Chen, & Gao, 2009) proposed a similar patch-based algorithm; however, they measured the similarities of the local patches by correlations in a subspace constructed by Canonical Correlation Analysis. (Du & Ward, 2009) proposed a similar method based on the facial components. Unlike (Chai et al., 2007) and (Li et al., 2009), where the face image is partitioned into uniform blocks, the method in (Du & Ward, 2009) divides it into the facial components, *i.e.*, two eyes, mouth and nose. The virtual frontal view of each component is estimated separately, and finally the virtual frontal image is generated by integrating the virtual frontal components. The common drawback of these three patch-based approaches, (Chai et al., 2007; Du & Ward, 2009; Li et al., 2009), is that the head pose of the input face image needs to be known. Moreover, these methods require a set of prototype non-frontal face patches, which are in the same pose as the input non-frontal faces; hence, they cannot handle a continuous range of poses and are restricted to a discrete set of predetermined pose angles.

(Blanz & Vetter, 2003) proposed a face recognition technique that can handle variations in pose and illumination. In their method, they derive a morphable face model by transforming the shape and texture of example prototypes into a vector space representation. New faces at any pose and illumination are modeled by forming linear combinations of the prototypes. The morphable model represents shapes and textures of faces as vectors in a high-dimensional space. The knowledge of face shapes and textures is learned from a set of textured 3D head scans. This method requires a set of manually annotated landmarks for initialization and the optimization process often converges to local minima due to a large number of parameters, which need to be tuned. (Breuer, Kim, Kienzle, Scholkopf, & Blanz, 2008) presented an automatic method for fitting the 3D morphable model; however, their method seems to have a high failure rate (Asthana, Marks, Jones, Tieu, & Rohith, 2011).

(Castillo & Jacobs, 2009) used the cost of stereo matching as a measure of similarity between two face images in different poses. This method does not construct a 3D face or a virtual frontal view; however, using stereo matching, it finds the correspondences between pixels in the probe and gallery images. This method requires manual specification of feature points and in case of automatic feature matching, it is fallible in scenarios where an in-plane rotation is present between the image pair.

The method proposed by (Sarfraz & Hellwich, 2010) handles the pose variations for face recognition by learning a linear mapping from the feature vector of a non-frontal face to the feature vector of the corresponding frontal face. However, their assumption of the mapping being linear seems to be overly restrictive (Asthana et al., 2011).

(Asthana et al., 2011) used several AAMs each of which covering a small range of pose variations. All these AAMs are fitted on the query face image and the best fit is selected. The frontal view is then synthesized using the pose-dependent correspondences between 2D landmark points and 3D model vertices. (Mostafa, Ali, Alajlan, & Farag, 2012; Mostafa & Farag, 2012) constructed 3D face shapes from stereo pair images. These 3D shapes are used to synthesize virtual 2D views in different poses, *e.g.*, frontal view. A 2D probe image is matched with the closest synthesized images using the local binary pattern (LBP) features (Ahonen, Hadid, & Pietikäinen, 2006). The drawback of this method is the need for stereo images. In order to solve this problem, the authors developed another method where the 3D shapes are constructed using only a frontal view and a generic 3D shape created by averaging several 3D face shapes.

(Sharma, Al Haj, Choi, Davis, & Jacobs, 2012) proposed the Discriminant Multiple Coupled Latent Subspace method for pose-invariant face recognition. They propose to obtain pose-specific representation schemes so that the projection of face vectors onto the appropriate representation scheme will lead to correspondence in the common projected space, which facilitates direct comparison. They find the sets of projection directions for different poses such that the projected images of the same subject in different poses are maximally correlated in the latent space. They claim that the discriminant analysis with artificially simulated pose errors in the latent space makes it robust to small pose errors due to subjectś incorrect pose estimation.

(De Marsico, Nappi, Riccio, & Wechsler, 2013) proposed a face recognition approach, called “FACE”, in which an unknown face is identified based on the correlation of local regions from the query face and multiple gallery instances, that are normalized with respect to pose and illumination, for each subject. For pose normalization, the facial landmarks are first located by an extension of the active shape model (Milborrow & Nicolls, 2008) and then the in-plane face rotation is normalized using the locations of the eye centers. The rows in the best exposed half of the face are then stretched to a constant length. Then, the other side of the face image is reconstructed by mirroring the first half. The illumination normalization is performed using the Self-Quotient Image (SQI) algorithm (Wang, Li, Wang, & Zhang, 2004), in which the intensity of each pixel is divided by the average intensity of its $k \times k$ square neighborhood.

(Ho & Chellappa, 2013) proposed a patch-based method for synthesizing the frontal view from a given nonfrontal face image. In this method, the face image is divided into several overlapping patches, and a set of possible warps for each patch is obtained by aligning it with frontal faces in the training set. The alignments are performed using an extension of the Lucas–Kanade image registration algorithm (Ashraf, Lucey, & Chen, 2010; Lucas & Kanade, 1981) in the Fourier domain. The best warp is chosen by formulating the optimization problem as a discrete labeling algorithm using a discrete Markov random field and a variant of the belief propagation algorithm (Komodakis & Tziritas, 2007). Each patch is then transformed to the frontal view using its best warp. Finally, all the transformed patches are combined together to create a frontal face image. A shortcoming of this method is that they divide both frontal and non-frontal images into the same regular set of local patches. This division strategy results in the loss of semantic correspondence for some patches when the pose difference is large; therefore, the learnt patch-wise affine warps may lose practical significance.

(Yi, Lei, & Li, 2013) proposed an approach for unconstrained face recognition that is robust against pose variations. A 3D deformable model is generated and a fast 3D model fitting algorithm is proposed to estimate the pose of the face image. Then, a set of Gabor filters is transformed according to the pose and shape of the face image for feature extraction. Finally, Principal Component Analysis (PCA) is applied on the Gabor features to eliminate the redundancies, then, the dot product is used to compute the similarity between the feature vectors.

Most recently, (Guo, Ding, & Xue, 2015) extended the Linear Discriminant Analysis (LDA) approach to multi-view scenarios. Multi-view Linear Discriminant Analysis (MiLDA) is a subspace learning framework for multi-view data analysis based on graph embedding (Yan et al., 2007). The authors introduced a new measure of distance between projected vertex sets of intrinsic graphs to mitigate the effect of the differences between views and preserve the intrinsic graphs. This distance is defined as the weighted sum of squared Euclidean distances between every cross-view data pair in two graph embedding models. Having sets of multi-view data, MiLDA aims to find a common subspace of higher discriminability between classes. The transformed feature vectors in the common subspace are classified using a nearest neighbor classifier.

In a recent publication, (Gao, Zhang, Jia, Lu, & Zhang, 2015) presented a face recognition approach based on deep learning using a single training sample per person. A deep neural network is an artificial neural network with multiple hidden layers between the input and output layers. In (Gao et al., 2015), the authors propose a supervised auto-encoder to build the deep neural network by training a nonlinear feature extractor at each layer. After the layer-wise training of each building block and building a deep architecture, the output of the network is used for face recognition. One of the shortcomings of this method is the manual cropping and alignment of the face images. It is also tested only on near frontal face images. The other well-known deep learning based algorithm, *DeepFace* (Taigman, Yang, Ranzato, & Wolf, 2014), focuses on solving the unconstrained face recognition problem by learning a set of features in the image domain. It uses a nine-layer deep neural network with more than 120 million parameters. The high accuracy of DeepFace owes, to a great extent, to its enormous training database of 4.4 million labeled faces.

1.2. Contributions

In this paper, we propose a fully automated single sample face recognition system suitable for images captured in unconstrained environments. The system is robust to pose and illumination variations, which usually affect images captured in the wild. The system includes a face normalization method based on an enhanced active appearance model approach. We propose a novel initialization technique for AAM, which results in significant improvements in its fitting to non-frontal poses and makes the normalization process robust and fast. Our AAM is trained using face images in-the-wild, which cover a vast range of illumination, pose and expression variations.

In contrast with majority of the algorithms encountered in the literature, our proposed normalization algorithm is fully automatic and handles a continuous range of poses, *i.e.*, it is not restricted to any predetermined pose angles. Moreover, it uses only a single gallery image and does not require additional non-frontal gallery images or stereo images. Relying on the competence of our algorithm in normalizing the face images, we can assume that the face images are properly aligned. This alignment allows us to use corresponding local feature descriptors such as Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005) for feature extraction, which makes the system robust against illumination variations. In addition, we fuse the HOG features with Gabor features using Canonical Correlation Analysis (CCA) to have a more discriminative feature set.

It is worth mentioning that our system is capable of recognizing a face from a non-frontal view and under different illumination conditions using only a single gallery image for each subject. This is important because of its potential applications in many realistic scenarios like passport identification and video surveillance. Experimental results performed on the FERET (Phillips, Moon, Rizvi, & Rauss, 2000), CMU-PIE (Sim, Baker, & Bsat, 2002) and Labeled Faces in the Wild (LFW) (Huang et al., 2007) databases verify the effectiveness of our proposed method, which outperforms the above-mentioned state-of-the-art algorithms.

This paper is organized as follows: Section 2 describes our face normalization technique. Section 3 describes the feature extraction and fusion approaches used in the proposed system. The implementation details and experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

2. Preprocessing for face normalization

As stated in (Moses et al., 1994), “the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity”. Pose variations cause major problems in real-world face recognition systems. In an unconstrained environment, there are usually in-plane and out-of-plane face rotations. In order to achieve better recognition results, we preprocess the facial images to handle these variations.

In this section, we present a pose normalization technique based on piece-wise affine warping, which can normalize both in-plane and out-of-plane pose changes. The warping is applied on triangular pieces determined by enhanced active appearance models described below. The overall process is illustrated in Fig. 1. In the following sections, we describe the fitting and warping process of the active appearance models and present a novel initialization technique for AAMs, which results in significant improvement in the fitting accuracy.

2.1. Active appearance models and piece-wise affine warping

Active appearance models have been widely used in pattern recognition research (Cootes et al., 1998). Face modeling has been the most ubiquitous application of AAMs. Given the model parameters, AAMs reconstruct a specific face via statistical models of shape and appearance. The model parameters are obtained by maximizing the match between the model instance and the face by fitting the AAM to the input face image.

The shape, S, of an AAM, is defined by the coordinates of a set of landmarks on the face. Learning the shape model requires annotating these landmarks on a training set of face images, then, applying principal component analysis (PCA) to these shapes. The shape model of a specific face is expressed as a base shape, s_0, plus a linear combination of the n shape eigenvectors, s_i, $i = 1 , \dots , n$,that correspond to the n largest eigenvalues:

$$S = s_0 + \sum_{i=1}^{n} p_i s_i \, , \tag{1}$$

where p_is are the shape parameters.

The appearance of an AAM is defined within the base shape, s_0, which means that learning the appearance model requires removing the shape variations. The appearance of an AAM is an image A($\mathbf{x}$), where $\mathbf{x}$ is the set of pixels inside the base mesh s_0 ($\mathbf{x} \in s_0$). In order to obtain the appearance model, PCA is applied on these shape-free images. The appearance model of a specific face is expressed as a base

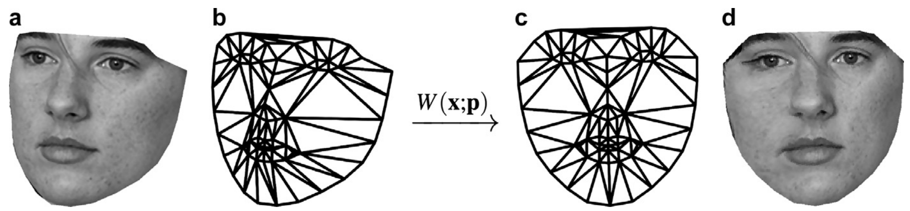

Fig. 1. Warping the face image into the base (frontal) mesh. (a) Rotated face image. (b) Fitting mesh corresponding to the rotated face image. (c) Triangulated base (frontal) mesh, s_0. (d) Face image warped into the base mesh.

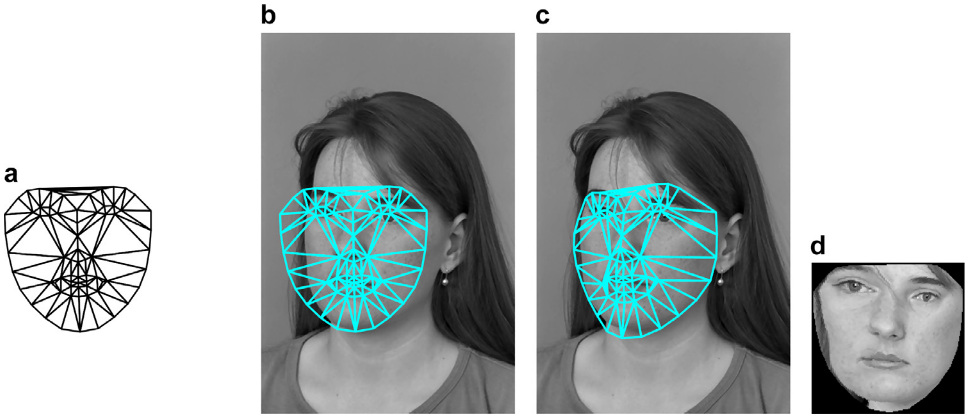

Fig. 2. Initialization problem in AAM fitting. (a) Initial shape used in POIC and SIC algorithms $\mathbf{p} = 0$. (b) Initialization of the base mesh on the target face image. (c) Fitting result of the Fast-SIC method after 100 iterations. (d) Result of the piecewise affine warping into the base mesh.

appearance, a_0, plus a linear combination of $m$appearance eigenvectors, $a_i, i = 1, \ldots, m$corresponding to the m largest eigenvalues:

$$A(\mathbf{x}) = a_0(\mathbf{x}) + \sum_{i=1}^{m} q_i a_i(\mathbf{x}) , \tag{2}$$

where q_is are the appearance parameters.

The shape and appearance parameters for a given face image are obtained in the process of AAM fitting. Project-Out Inverse Compositional (POIC) algorithm (Matthews & Baker, 2004) and Simultaneous Inverse Compositional (SIC) algorithm (Gross, Matthews, & Baker, 2005) are two well-known algorithms for AAM fitting. SIC performs significantly better than POIC on images of subjects that are not included in the training. However, the computational cost of SIC is very high (Baker, Gross, & Matthews, 2003). Recently, (Tzimiropoulos & Pantic, 2013) proposed Fast-SIC, which reduces the computational complexity of SIC. In our experiments, we use the Fast-SIC optimatization technique for fitting the AAM.

Let $\mathbf{p} = \{p_1, p_2, \ldots, p_n\}$be the set of shape parameters obtained from AAM fitting. As shown in Fig. 1, a piecewise affine warp, $W(\mathbf{x}; \mathbf{p})$,transfers a face instance into the base shape. After fitting the AAM, each triangle in the AAM mesh has a corresponding triangle in the base (frontal) mesh. Using the coordinates of the vertices in the AAM mesh, the coordinates of the corresponding triangle in the base mesh are computed from the current shape parameters $\mathbf{p}$using Eq. (1). Using the coordinates of the vertices in corresponding triangles, we compute an affine transformation for each triangle, such that the vertices of the first triangle map to the vertices of the second triangle (Matthews & Baker, 2004). For every pixel inside the target triangle in the frontal mesh, the corresponding location in the AAM mesh is calculated. Then, the value of this pixel is obtained based on a nearest neighbor interpolation in the calculated location. This process is applied to all the triangles and the synthesized frontal face is created in the base mesh s_0. In our approach, we use the warped face within the base shape as the normalized face image. This step results in a shape-free facial appearance ($\mathbf{p} = 0$), which allows face identification to be performed in the coordinates of the base shape.

2.2. Proposed AAM initialization

Despite the popularity of the AAMs, there is no guarantee for obtaining correct fitting, specially when the images are not in near-frontal pose. As mentioned before, both POIC and SIC algorithms use the base mesh s_0, when $\mathbf{p} = 0$,as the initial shape model. The base mesh represents the mean shape of all the training samples, which is usually in frontal pose as shown in Fig. 2(a). Typical fitting methods use a face detection algorithm to find the face and then scale the base mesh to the size of the detected face and use it as the initial shape model. However, in semi-profile poses, this initialization sometimes falls out of face region and if the algorithm starts with this mesh, it may not converge to the actual shape. Fig. 2(b) shows the initialization of the base mesh on a sample face image. The result of the AAM fitting using Fast-SIC method after 100 iterations is shown in Fig. 2(c). Fig. 2(d) shows the result of the piecewise affine warping into the base mesh, which is supposed to represent the normalized face image.

For better initialization, in this paper, we use the flexible mixture of parts proposed in (Yang & Ramanan, 2011) to automatically initialize the locations of the landmarks. Every facial landmark with its predefined neighborhood patch is defined as a *part*. The landmarks on a face define a mixture of these parts, which are used to build a tree graph to represent the spatial structure of the landmarks. Due to the topological changes caused by pose variations, (Zhu & Ramanan, 2012) proposed a model based on mixture of trees with a shared pool of parts for face detection, pose estimation, and landmark localization. We modified this approach to initialize the landmark locations for our AAM.

Let I denote the facial image, in which $l_i = (x_i, y_i)$is a landmark location in part i. For each viewpoint t, we define a tree graph $G_t = (V_t, E_t)$,where $V_t \subseteq V$, and V is the shared pool of parts. A configuration of parts $L = \{l_i : i \in V\}$is scored as:

$$S(I, L, t) = \sum_{i \in V_t} \omega_i^{t_i} . \phi(I, l_i) + \sum_{i,j \in E_t} \lambda_{i,j}^{t_i,t_j} . \psi(l_i, l_j) + \alpha^t . \tag{3}$$

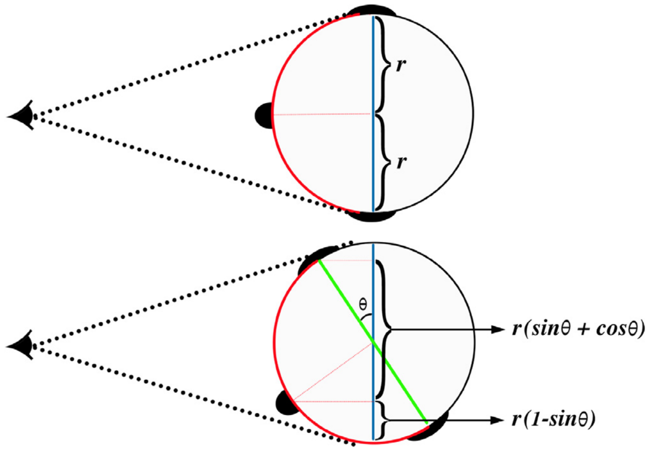

Fig. 3. Top view perspective of a human head in frontal and rotated poses.

The first term in Eq. (3) is an appearance evaluation function, indicating how likely a landmark is in an aligned position. $\phi(I, l_i)$ is a feature vector extracted from a neighborhood centered at l_i, where in our experiments, we use HOG features (Dalal & Triggs, 2005); and $\omega_i^{t_i}$is a template for part i tuned for the mixture for viewpoint t_i. The second term is the shape deformation cost, *i.e.*, computes the cost associated with the relative positions of neighboring landmarks. $\lambda_{i,j}^{t_i,t_j}$is used to encode parameters of rest location and rigidity, controlling the shape displacement of part i relative to part j defined as $\psi(l_i, l_j) = [dx\ dx^2\ dy\ dy^2]^T$,where $dx = x_i - x_j$and $dy = y_i - y_j$. Finally, the last term $\alpha^t$is a scalar bias associated with the mixture for viewpoint t.

We seek to maximize $S(I, L, t)$ over the landmark locations, L, and viewpoint, t, and find the best configuration of parts. Since each mixture is a tree-structured graph, maximization can be efficiently done with dynamic programming (Felzenszwalb & Huttenlocher, 2005) to find the global optimum solution.

Learning: To learn the model, a fully supervised scenario using labeled positive and negative samples is used. Assume that $\{I_n, L_n, t_n\}$ and $\{I_n\}$ denote the nth positive and negative samples, respectively. The scoring function, Eq. (3), is linear in its parameters. Concatenating the parameters, we can write $S(I, k) = \mu.\Phi(I, k)$,where $\mu = (\omega, \alpha)$and $k_n = (l_n, t_n)$. Now, learning the model can be formulated as:

$$\arg\min_{\mu,\xi_n\geq 0} \ \frac{1}{2} \parallel \mu \parallel + C\sum_n \xi_n \tag{4}$$

$$\begin{array}{ll} s.t.\ \forall n \in pos & \mu.\Phi(I_n, k_n) \geq 1 - \xi_n \\ \forall n \in neg, \forall k & \mu.\Phi(I_n, k) \leq -1 + \xi_n\,. \end{array}$$

(Zhu & Ramanan, 2012) trained their model in 13 viewpoints spanning 180° with sampling every 15°. They used images from CMU Multi-PIE face database (Gross, Matthews, Cohn, Kanade, & Baker, 2010) with 68 facial landmarks in poses between $-45^\circ$and $+45^\circ$,and 39 facial landmarks in poses $\pm$ 60°, $\pm$ 75° and $\pm$ 90°. In order to cover the whole range of pose variations, we used the model in (Zhu & Ramanan, 2012), which uses 900 positive samples from Multi-PIE, and 1218 negative samples from INRIA Person database (Dalal & Triggs, 2005), including outdoor scenes with no people in them.

AAM Initialization: In the testing stage, since we use the landmarks for the initialization of our AAM, in cases of detecting a mixture with 39 vertices (landmarks), we estimate the location of the remaining 29 landmarks based on the information obtained from the topology of the facial landmarks in the viewpoint corresponding to the detected mixture. Without loss of generality, if we assume that the top view of a human head is a circle with radius r, Fig. 3 shows the visible area of the left and right sides of the face in frontal and rotated poses. As illustrated, the ratio between the visible areas in two sides of the face is

$$\gamma = \frac{1 - sin(\theta)}{sin(\theta) + cos(\theta)}\,, \tag{5}$$

θ being the pose angle.

In cases where the landmark localization stage selects a mixture of 39 vertices, these landmarks are fitted on the best exposed half of the face. The selected mixture provides an estimation of the pose angle, θ. γ, obtained from Eq. (5), is used as a scaling factor to roughly calculate the location of the landmarks on the other half of the face by relatively mirroring the current landmarks across the face mid-line.

The landmark localization algorithm based on the flexible mixture of parts works very well in finding the contour of the face but it is not accurate enough in the more detailed regions such as the eyes or the mouth. Fig. 4(a) shows the result of this method on a sample face image.

In this paper, instead of using the base mesh, s_0, we create the initial shape model for AAM using the estimated landmarks obtained from the flexible mixture of parts model. Fig. 4(b) shows the triangularized initial mesh using these landmarks. The result of the AAM fitting using Fast-SIC method after only five iterations is shown in Fig. 4(c). It is clear from Fig. 4(c) that, using this initialization, the fitting is much more accurate. Fig. 4(d) shows the result of the piecewise affine warping into the base mesh, which in comparison with Fig. 2(d), provides a better representation of the face. In the rest of this paper, we use these warped images as the normalized face images.

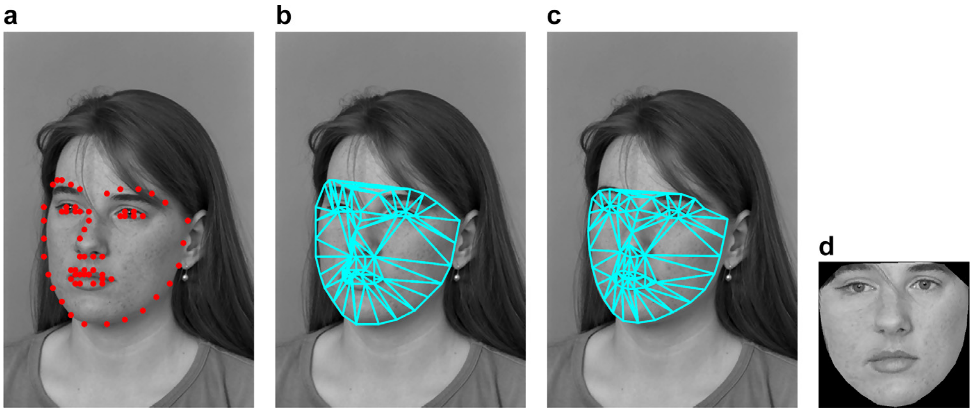

Fig. 4. Our proposed initialization method for AAM fitting. (a) Estimated landmarks using the flexible mixture of trees. (b) Triangularization of the initial mesh created by the estimated landmarks. (c) Fitting result of the Fast-SIC method after only 5 iterations. (d) Result of the piecewise affine warping into the base mesh.

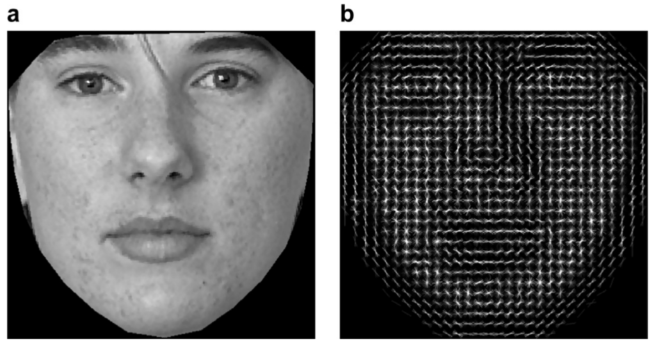

Fig. 5. Histogram of Oriented Gradients (HOG) features in 4 × 4 cells.

3. Feature extraction and fusion

The face images of an individual subject are similar to each other and different from the face images of other subjects. However, face images of an individual are not exactly the same either. The question is how these changes are different from the changes between different subjects. The proper alignment of the face images made possible by the proposed normalization technique reduces the variations between feature vectors of the samples of the same subject, which facilitates building a more accurate face model. In this section we describe the feature extraction techniques as well as the feature fusion method employed in our approach.

3.1. Feature extraction

In our experiments, the normalized face images are resized to 120 × 120 pixels. We use two different techniques to extract features from the normalized images. These techniques include Gabor wavelet features (Haghighat, Zonouz, & Abdel-Mottaleb, 2013; Liu & Wechsler, 2002) and Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005).

Since the face images are aligned, we can make use of local descriptors such as the histograms of oriented gradients (HOG) (Dalal & Triggs, 2005) for feature extraction. Here, we extract the HOG features in 4 × 4 cells for nine orientations. We use the UOCTTI variant for the HOG presented in (Felzenszwalb, Girshick, McAllester, & Ramanan, 2010). UOCTTI variant computes both directed and undirected gradients as well as a four dimensional texture-energy feature, but projects the result down to 31 dimensions, 27 dimensions corresponding to different orientation channels and 4 dimensions capturing the overall gradient energy in square blocks of four adjacent cells. Fig. 5(b) shows the HOG features extracted from a sample face image in Fig. 5(a)[1].

On the other hand, we employ forty Gabor filters in five scales and eight orientations. The most important advantage of Gabor filters is their invariance to rotation, scale, and translation. Furthermore, they are robust against photometric disturbances, such as illumination change and image noise (Haghighat et al., 2015; Kämäräinen, Kyrki, & Kälviäinen, 2006). Since the adjacent pixels in an image are usually correlated, the information redundancy can be reduced by downsampling the feature images that result from Gabor filters (Liu & Wechsler, 2002). In our experiments, the feature images are downsampled by a factor of five. Fig. 6 shows the Gabor features for the normalized face image in Fig. 5(a). The dimensionality of both Gabor and HOG feature vectors are reduced using principal component analysis (PCA) (Turk & Pentland, 1991).

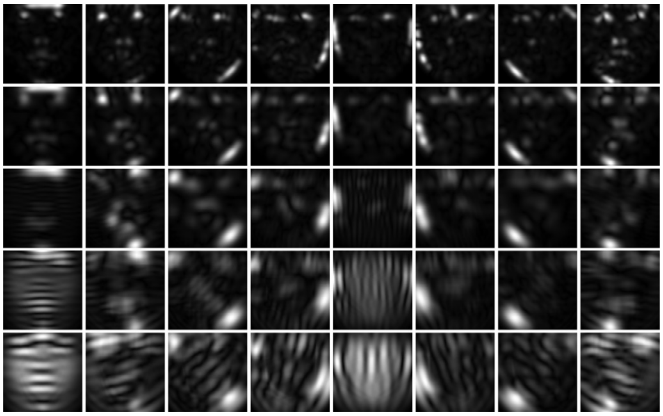

Fig. 6. Gabor features in five scales and eight orientations.

3.2. Feature fusion using canonical correlation analysis

We combine the two feature vectors to obtain a single feature vector, which is more discriminative than any of the input feature vectors. This is achieved by using a feature fusion technique based on Canonical Correlation Analysis (CCA) (Sun, Zeng, Liu, Heng, & Xia, 2005).

Canonical correlation analysis has been widely used to analyze associations between two sets of variables. Suppose that $X \in \mathbb{R}^{p\times n}$and $Y \in \mathbb{R}^{q\times n}$are two matrices, each contains n training feature vectors from two different modalities. In other words, there are n samples for each of which $(p+q)$features have been extracted. Let $S_{xx} \in \mathbb{R}^{p\times p}$and $S_{yy} \in \mathbb{R}^{q\times q}$denote the within-sets covariance matrices of X and Y and $S_{xy} \in \mathbb{R}^{p\times q}$denote the between-set covariance matrix (note that $S_{yx} = S_{xy}^T$). The overall $(p+q)\times(p+q)$covariance matrix, S, contains all the information on associations between pairs of features:

$$S = \begin{pmatrix} cov(x) & cov(x,y) \\ cov(y,x) & cov(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}. \tag{6}$$

However, the correlation between these two sets of feature vectors may not follow a consistent pattern, and thus, understanding the relationships between these two sets of feature vectors from this matrix is difficult (Krzanowski, 1988). CCA aims to find the linear combinations, $X^* = W_x^T X$and $Y^* = W_y^T Y$,that maximize the pair-wise correlations across the two data sets:

$$corr(X^*, Y^*) = \frac{cov(X^*, Y^*)}{var(X^*).var(Y^*)}, \tag{7}$$

where $cov(X^*, Y^*) = W_x^T S_{xy} W_y$,$var(X^*) = W_x^T S_{xx} W_x$and $var(Y^*) = W_y^T S_{yy} W_y$. Maximization is performed using Lagrange multipliers by maximizing the covariance between $X^*$and Y^* subject to the constraints $var(X^*) = var(Y^*) = 1$. The transformation matrices, $W_x$and W_y, are then found by solving the eigenvalue equations (Krzanowski, 1988):

$$\begin{cases} S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx} \hat{W}_x = \Lambda^2 \hat{W}_x \\ S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy} \hat{W}_y = \Lambda^2 \hat{W}_y \end{cases}, \tag{8}$$

where $\hat{W}_x$and $\hat{W}_y$are the eigenvectors and Λ^2 is the diagonal matrix of eigenvalues or squares of the *canonical correlations*. The number of non-zero eigenvalues in each equation is $d = rank(S_{xy}) \leq min(n, p, q)$,which will be sorted in decreasing order, $\lambda_1 \geq \lambda_1 \geq \cdots \geq \lambda_d$. The transformation matrices, W_x and W_y, consist of the sorted eigenvectors corresponding to the non-zero eigenvalues. $X^*, Y^* \in \mathbb{R}^{d\times n}$are known as canonical variates. For the transformed data, the

[1] VLFeat open source library is used to extract and visualize the HOG features (Vedaldi & Fulkerson, 2008).

sample covariance matrix defined in Eq. (6) will be of the form:

$$S^* = \left(\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & \lambda_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & \lambda_d \\ \lambda_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & \lambda_d & 0 & 0 & \dots & 1 \end{array}\right).$$

The above matrix shows that the canonical variates have nonzero correlation only on their corresponding indices. The identity matrices in the upper left and lower right corners show that the canonical variates are uncorrelated within each data set.

As defined in (Sun et al., 2005), feature-level fusion is performed either by concatenation or summation of the transformed feature vectors:

$$Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix} = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{9}$$

or

$$Z_2 = X^* + Y^* = W_x^T X + W_y^T Y = \begin{pmatrix} W_x \\ W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{10}$$

where Z_1 and Z_2 are called the Canonical Correlation Discriminant Features (CCDFs). In this paper, we use the concatenation method defined in Eq. (9). The fused feature vectors (Z) are used to build face models following using the face modeling approach presented in (Haghighat, Abdel-Mottaleb, & Alhalabi, 2014). The query sample is then classified as the nearest neighbor based on the Euclidean distance between the query's model and the models in the gallery.

4. Experimental setup and results

4.1. Experimental setup: AAM training

In our experiments, we trained the AAMs using in-the-wild databases. For this purpose, we use three of the training sets provided for *300 Faces in-the-Wild Challenge* (Sagonas, Tzimiropoulos, Zafeiriou, & Pantic, 2013). These images contain large variations in pose, expression, illumination and occlusion. These databases are Labeled Face Parts in-the-Wild (LFPW) (Belhumeur, Jacobs, Kriegman, & Kumar, 2011), Helen (Le, Brandt, Lin, Bourdev, & Huang, 2012), and a database collected by Intelligent Behavior Understanding Group (IBUG) (Sagonas et al., 2013). LFPW database consists of 1, 035 annotated images collected from *Yahoo, Google*, and *Flickr*. HELEN database contains 2, 330 annotated faces downloaded from *Flickr*. Most of the expressions in these two databases are neutral and smile. Therefore, IBUG database, which contains 135 highly expressive face images, is added to include a larger variety of facial expressions. In total, 3500 in-the-wild face images are used to train the AAM. Note that these databases are only used for training the AAM and since they are not labeled, they are not employed in evaluating the recognition accuracy of our system.

4.2. Normalization performance

Here we discuss the self-occlusion problem in case of large pose variations. Fig. 7 shows a semi-profile face image with a large pose angle, where only a small fraction of the right side of the face is visible. According to Eq. (5), for instance in the case of a 60° pose angle, the visible area of the *occluded* side of the face shrinks by a factor of $1 - sin(60^\circ) = 0.13$,while for the other side of the face, the visible area stretches by a factor of $sin(60^\circ) + cos(60^\circ) = 1.36$. The ratio between these two areas is less than 10%.

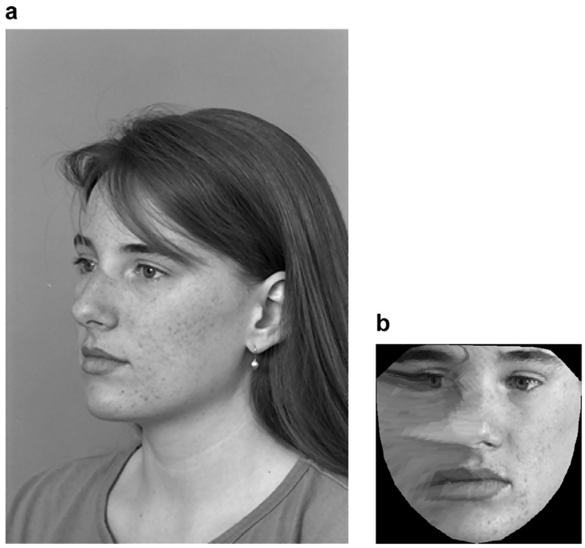

Fig. 7. (a) Self-occluded face image with 60° rotation. (b) Normalized face image with a stretched half face.

In the proposed normalization technique, after fitting the AAMs, the face image is warped into the base frontal mesh. Since the areas of the left and right halves of the base mesh have the same size, the occluded side of the face will be over-sampled (stretched) in the process of piecewise warping. In this case, a small misalignment in the AAM fitting may cause a large error in the warped face image, which will result in a distorted half-face. Even if a semi-profile face is perfectly fitted, the warped frontal view will still be distorted due to the stretching (Gao et al., 2009). This phenomenon is clearly seen in Fig. 7, which has a 60° of face rotation. In the normalization process, the right half of the face, *i.e.*, the occluded half, is stretched, which results in a distorted half face. This distortion will have negative effect on the recognition accuracy. Therefore, in these cases, we only use half of the face that corresponds to the visible side and ignore the distorted half.

In order to automatically distinguish between the well-normalized and the distorted half-faces in semi-profile images, we trained a two class minimum distance classifier using Discrete Cosine Transform (DCT) features. This classifier is trained using 400 well-normalized half-faces generated from frontal faces in the *ba* set of FERET database (Phillips et al., 2000), and 400 distorted half-faces randomly chosen from the *hl* and *hr* sets of FERET database, which include poses at −67.5°and +67.5°rotations. After face normalization, this classifier uses the DCT features extracted from each half of the face to determine whether it is well-normalized or distorted. Based on the outcome, we either use only the well-normalized side or the whole face for identification. The complexity of this step is negligible not only because DCT features are very simple to calculate, but also because the decision is made based on the Euclidean distances from the centroids of only two classes.

In this following, we present several sets of experiments to demonstrate the performance of our proposed face normalization and recognition system. We conduct three sets of experiments, on three databases: Facial Recognition Technology (FERET) (Phillips et al., 2000), CMU-PIE (Sim et al., 2002) and Labeled Faces in the Wild (LFW) (Huang et al., 2007).

4.3. Experiments on FERET database

The first set of experiments was performed on the FERET b-series database (Phillips et al., 2000). It contains 2, 200 face images for 200 subjects, *i.e.*, eleven images per subject. Three of the images include frontal faces with different facial expressions

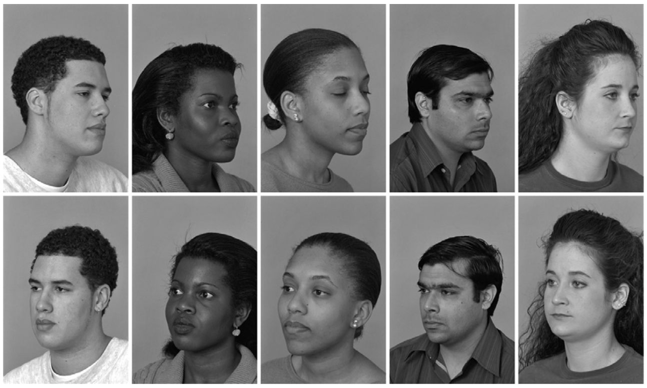

Fig. 8. Symmetry issue in FERET database. The upper row includes the sample images at +60° (bb) and the lower row shows the corresponding images at −60° (bi).

and illuminations. These images are letter coded as *ba, bj*, and *bk*. The other eight images are faces in different poses with +60°, +40°, +25°, +15°, −15°, −25°, −40°, and −60° degrees of rotation. These images are letter coded as *bb, bc, bd, be, bf, bg, bh*, and *bi*, respectively. Fig. 8 shows these images for a sample subject along with the results of the proposed normalization approach. Note that, our proposed normalization approach is fully automatic and no manual adjustments were needed in any of the 2200 samples.

In our experiments, only a single image, *i.e.*, the frontal face image with neutral expression labeled *ba*, is used for enrollment and the remaining ten images with different poses, expressions, and illumination conditions are used for testing. Table 1 shows the accuracy of our proposed method for each set in comparison with previous methods in the literature. Note that, the proposed method is evaluated with all the pose angles presented in FERET database. However, only five of the previous methods used the images from all the pose angles (Asthana, Sanderson, Gedeon, & Goecke, 2009; Gao et al., 2009; Sarfraz & Hellwich, 2010; Sharma et al., 2012; Yi et al., 2013), and the other studies (Asthana et al., 2011; Ho & Chellappa, 2013; Mostafa et al., 2012; Sagonas, Panagakis, Zafeiriou, & Pantic, 2015; Zhang, Shan, Gao, Chen, & Zhang, 2005) only used a subset of the pose angles.

The recognition rates for +60°and +45°poses (*bb* & *bc*) are less than those for −60°and −45°poses (*bi* & *bh*). The reason goes back to the setup of the FERET database in which the positive rotations are slightly more than the negative ones. Fig. 9 shows examples of this difference. The upper row shows the sample images at +60°(*bb*) and the lower row shows the corresponding images at −60°(*bi*) for the same subjects.

As seen in Table 1, our proposed algorithm outperforms the previous algorithms (Asthana et al., 2011; Asthana et al., 2009; Gao et al., 2009; Ho & Chellappa, 2013; Mostafa et al., 2012; Sagonas et al., 2015; Sarfraz & Hellwich, 2010; Sharma et al., 2012; Yi et al., 2013; Zhang et al., 2005) in most of the pose angles. In the case of high rotations (± 60°), the recognition rates are comparable with the best method PAF (Yi et al., 2013). It is worth mentioning that some of the methods in Table 1 are not fully automatic and they require manual intervention, some of these methods also use the same database (FERET) in training their normalization approach. However, our approach is fully automatic and does not use FERET database in training the normalization technique.

Note that in (Ho & Chellappa, 2013) and (Asthana et al., 2011), if the face and both eyes are not detected using the cascade classifiers, a Failure to Acquire (FTA) is reported and the image is not included in the test set. However, we tested the recognition rate on all the 200 images of each set and no images were excluded in the evaluation process (no FTA is considered).

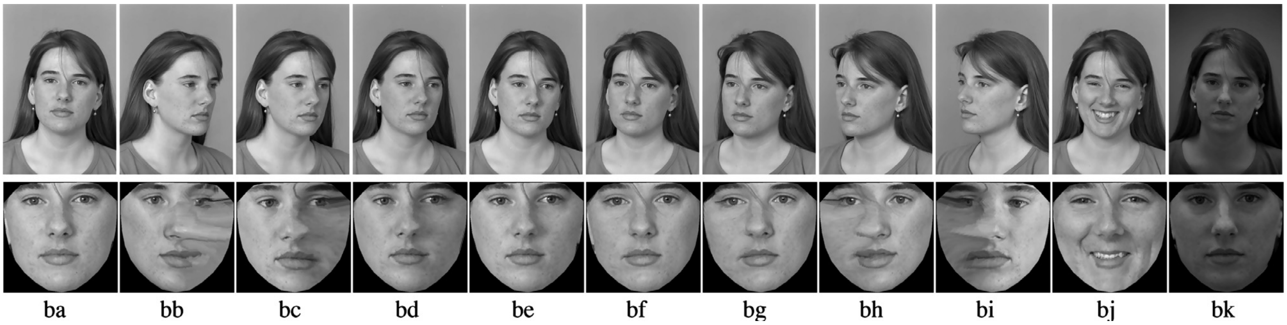

Fig. 9. Face images of a sample subject from FERET b-series database (upper row), and their normalized faces (lower row).

Table 1
Face recognition rates of different approaches in confrontation with different face distortions on the FERET database. The frontal face images with neutral expression, labeled *ba*, are used for training.

Method	Face Alignment	Trained on FERET	bb +60°	bc +45°	bd +25°	be +15°	bf −15°	bg −25°	bh −45°	bi −60°	bj expr.	bk illum.
LGBP (Zhang et al., 2005)	Automatic	No	–	51.0	84.0	96.0	98.0	91.0	62.0	–	–	–
PAN (Gao et al., 2009)	Manual	Yes	44.0	81.5	93.0	97.0	98.5	91.5	78.5	52.5	–	–
Asthana (Asthana et al., 2009)	Manual	Yes	32.5	74.0	95.5	98.5	98.0	93.0	87.0	48.0	–	–
Sarfraz (Sarfraz & Hellwich, 2010)	Automatic	Yes	78.0	89.0	97.0	98.6	100	89.7	92.4	84.0	–	–
3DPN (Asthana et al., 2011)	Automatic	No	–	91.9	97.0	97.5	98.5	98.0	90.5	–	–	–
CLS (Sharma et al., 2012)	Manual	Yes	70.0	82.0	90.0	95.0	96.0	94.0	85.0	79.0	–	–
FRAD (Mostafa et al., 2012)	Automatic	No	–	82.35	98.47	98.97	100	97.98	87.5	–	–	–
PIMRF (Ho & Chellappa, 2013)	Automatic	No	–	91.5	96.5	98.5	98.0	97.3	91.0	–	–	–
PAF (Yi et al., 2013)	Automatic	No	93.75	98.0	98.5	99.25	99.25	98.5	98.0	93.75	–	–
FAR (Sagonas et al., 2015)	Automatic	No	–	96.0	100	100	100	99.0	96.5	–	–	–
Proposed Method	Automatic	No	91.5	96.0	100	100	100	100	99.0	93.0	99	100

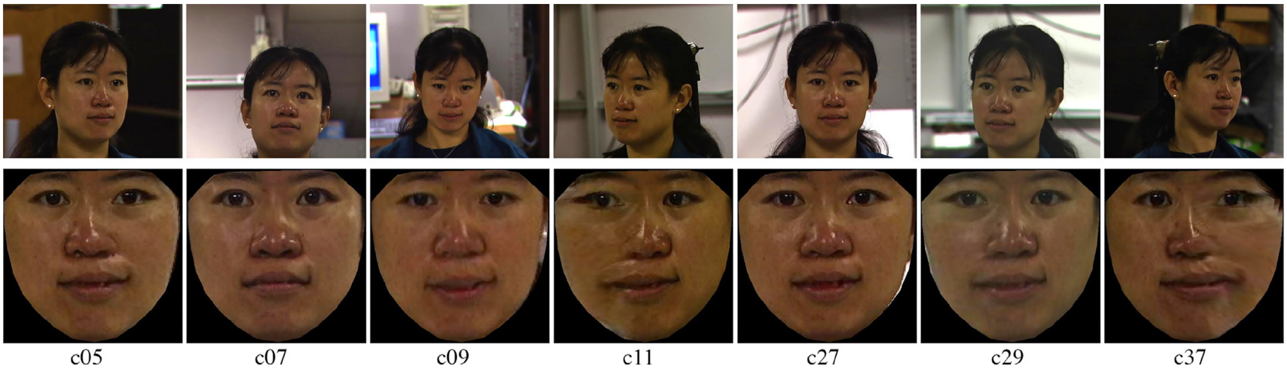

Fig. 10. Face images of a sample subject from CMU-PIE database (upper row), and their normalized faces (lower row).

Table 2
Face recognition rates of different approaches in confrontation with different pose changes on the CMU-PIE database. The frontal faces captured by camera *c27* is used for training.

Method	Face Alignment	Trained on PIE	Gallery Size	c11 -45°	c29 -22.5°	c07 22.5°*up*	c09 22.5°*down*	c05 $+22.5^\circ$	c37 $+45^\circ$
LGBP (Zhang et al., 2005)	Automatic	No	67	71.6	87.9	78.8	93.9	86.4	75.8
LLR (Chai et al., 2007)	Manual	No	34	89.7	100	98.5	98.5	98.5	82.4
3ptSMD (Castillo & Jacobs, 2009)	Manual	No	34	97.0	100	100	100	100	100
Sarfraz (Sarfraz & Hellwich, 2010)	Automatic	No	68	83.8	86.8	–	–	94.1	89.7
3DPN (Asthana et al., 2011)	Automatic	No	67	98.5	100	98.5	100	100	97.0
CLS (Sharma et al., 2012)	Manual	Yes	34	100	100	100	100	100	100
FRAD (Mostafa et al., 2012)	Automatic	No	68	95.6	100	100	100	100	100
PIMRF (Ho & Chellappa, 2013)	Automatic	No	67	97.0	100	98.5	100	100	97.0
PAF (Yi et al., 2013)	Automatic	No	68	100	100	100	100	100	100
MiLDA (Guo et al., 2015)	Automatic	No	68	90.30	99.58	–	–	98.73	92.55
SSAE (Gao et al., 2015)	Manual	Yes	48	–	68.06	71.45	71.96	67.52	–
Proposed Method	Automatic	No	68	100	100	100	100	100	100

4.4. Experiments on CMU-PIE database

The second set of experiments were performed on CMU-PIE database (Sim et al., 2002). This database consists of face images taken from sixty eight subjects under thirteen different poses. Similar to the previous methods (Asthana et al., 2011; Castillo & Jacobs, 2009; Chai et al., 2007; Ho & Chellappa, 2013; Mostafa et al., 2012; Sarfraz & Hellwich, 2010; Zhang et al., 2005), seven poses are used in our experiments. The frontal pose, labeled *c27*, is used as the gallery image. The probe set consists of six non-frontal poses labeled as *c37* and *c11* (the yawn angle about ± 45°), *c05* and *c29* (the yawn angle about ± 22.5°), and *c07* and *c09* (the pitch angle about ± 22.5°). Fig. 10 shows these images for a sample subject along with the results of applying the proposed normalization method to them.

The performance of the proposed system is compared with the state-of-the-art approaches in (Asthana et al., 2011; Castillo & Jacobs, 2009; Chai et al., 2007; Gao et al., 2015; Guo et al., 2015; Ho & Chellappa, 2013; Mostafa et al., 2012; Sarfraz & Hellwich, 2010; Sharma et al., 2012; Yi et al., 2013; Zhang et al., 2005). Table 2 shows the outstanding accuracy of our proposed method for each pose in comparison with these methods. We obtain 100% accuracy in all sets. In our experiment, all the 68 subjects were employed for the evaluations; however, in some of the previous methods, *e.g.*, (Asthana et al., 2011; Ho & Chellappa, 2013; Zhang et al., 2005), the probe size is 67, because when their algorithm fails to normalize an image, they do not consider it as a recognition error and exclude that image from the test set. Some methods, in Table 2, only used 34 subjects out of the 68, *e.g.*, (Castillo & Jacobs, 2009; Chai et al., 2007; Sharma et al., 2012). (Gao et al., 2015) used 20 subjects for training their proposed deep neural network and the remaining 48 subjects were for evaluation. It is important to note that the deep learning based face recognition algorithm presented in (Gao et al., 2015) is not robust to pose variations and it is only tested in near frontal poses[2].

4.5. Experiments on LFW database

Our last experiment is on the Labeled Faces in the Wild (LFW) (Huang et al., 2007) database. LFW is one of the most challenging databases for evaluating the performance of face verification systems in unconstrained environments. This database contains 13, 233 face images of 5, 749 subjects labeled by their identities. 1, 680 of these subjects have more than one face images. The images are collected from *Yahoo! News* in 2002-2003, and have a wide variety of variations in pose, illumination, expression, scale, background, color saturation, focus, etc. Fig. 11 shows some sample images from this database and Fig. 12 shows the results of the proposed normalization method on these images. It is obvious from the figure that even with the changes in pose, expression, illumination and occlusion, the normalization results are impressive as the faces are precisely detected and aligned.

In order to compare with a wide range of methods, we evaluated our proposed algorithm in two different experiments. The first experiment follows the directions used in (Cox & Pinto, 2011; Hussain, Napoléon, & Jurie, 2012; Yi et al., 2013). As in (Cox & Pinto, 2011), the LFW dataset is organized into two disjoint sets: 'View 1' is used as gallery whereas 'View 2' is used for probe. Although (Cox & Pinto, 2011; Hussain et al., 2012; Yi et al., 2013) use the aligned version of

[2] We do not compare the proposed algorithm with the other well-known deep learning based algorithm, *DeepFace* (Taigman et al., 2014), because we only use a single gallery image, while DeepFace is trained using a large number of gallery images per subject. Moreover, the code of DeepFace and the training dataset are not available and we do not have the resources to handle such data in laboratory environment.

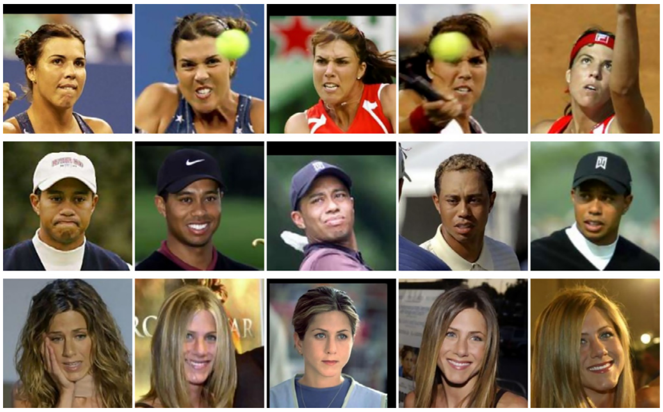

Fig. 11. Sample images of three subjects from LFW database.

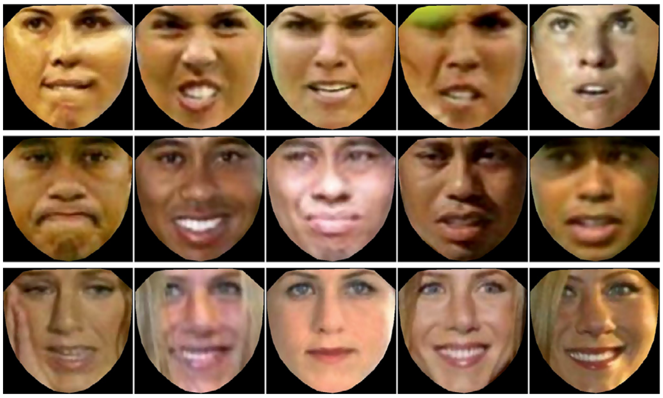

Fig. 12. Normalized face images corresponding to the ones shown in Fig. 11.

Table 3
Mean classification accuracy of different approaches following the first experiment on LFW database.

BIF (Cox & Pinto, 2011)	I-LQP (Hussain et al., 2012)	PAF (Yi et al., 2013)	Proposed Method
88.13	86.20	87.77	91.46

the faces provided by (Wolf, Hassner, & Taigman, 2010), we use the original version of the LFW database and all face images are aligned using our normalization technique described in Section 2. The mean classification accuracies of the proposed method and the methods following the same protocol are shown in Table 3.

Although LFW is basically designed for metric learning for face verification, (De Marsico et al., 2013) evaluated some of the most popular face recognition algorithms as well as their own method on a subset of this database. This subset is made from the first fifty subjects who have at least eight images. Five of the images are used as gallery images and three as probes. We used the same setting to evaluate the performance of our proposed method. Table 4 shows the performance of our proposed system in comparison with the Eigenface approach (Turk & Pentland, 1991), which is based on PCA, Independent Component Analysis (ICA) method proposed in (Bartlett, Movellan, & Sejnowski, 2002), Incremental Linear Discriminant Analysis (ILDA) approach (Kim, Wong, Stenger, Kittler, & Cipolla, 2007), a method using Support Vector Machines (SVM) (Guo, Li, & Chan, 2000), a recent approach based on Hierarchical Multiscale LBP (HMLBP) (Guo, Zhang, & Mou, 2010), and the method called “FACE” proposed in (De Marsico et al., 2013), which is the most recent method evaluated on this dataset.

Table 4 shows that our proposed system outperforms all the above-mentioned methods including the recent method proposed in (De Marsico et al., 2013) with an impressive margin of 26% in the recognition rate. Note that the experiments are performed using the original, not the aligned, version of the LFW database.

5. Conclusions and future work

In this paper, we proposed a single sample face recognition system for real-world applications in unconstrained environments. The potential application of this system is in many realistic scenarios like passport identification and video surveillance. The proposed system is fully automatic and robust to pose and illumination variations in face images. The system synthesizes the frontal views using a piece-wise affine warping. The warping is applied to the triangles of a mesh determined by an enhanced AAM. In order to enhance the fitting accuracy, we initialize the AAM using estimates of the facial landmark locations obtained by a method based on flexible mixture of parts. The fitting accuracy is further improved by training the AAM with in-the-wild images and using a powerful optimization technique. Experimental results demonstrated the efficacy of our proposed fitting approach. HOG and Gabor wavelet features are extracted from the synthesized frontal views. We use CCA to fuse these two feature sets into a single but more discriminative feature vector.

In contrast with other state-of-the-art methods, our approach uses only a single gallery image and does not require additional non-frontal gallery images or stereo images. It is also fully automatic and does not require any manual intervention. Moreover, it handles a wide and continuous range of poses, *i.e.*, it is not restricted to any predetermined pose angles. Experimental results performed on FERET, CMU-PIE and LFW databases demonstrated the effectiveness of our proposed method, which outperforms the state-of-the-art algorithms.

Our algorithm works very well in normalizing the near-frontal poses; however, its main weakness is in normalizing facial images with large pose variations. In semi-profile poses, half of the face is usually occluded, which results in a distorted normalized face. This distortion has a negative impact on the recognition accuracy. Although we use the other well-normalized half of the face for recognition, the accuracy in these cases is still low. Another limitation of the proposed method is that it does not handle the normalization of facial expressions.

In the future, we will investigate the possibility of synthesizing frontal faces with neutral expression to make the system invariant to facial expressions. We will also investigate the use of features that are less invariant to aging variations. This will make the system more reliable in recognizing people from images that have been taken with large time gaps. Moreover, we plan to design an intelligent system that can integrate multiple sources of biometric information, *e.g.*, frontal face, profile face and ear, to obtain a more

Table 4
Face recognition rates of different approaches following the second experiment on LFW database.

PCA (Turk & Pentland, 1991)	ICA (Bartlett et al., 2002)	ILDA (Kim et al., 2007)	SVM (Guo et al., 2000)	HMLBP (Guo et al., 2010)	FACE (De Marsico et al., 2013)	Proposed Method
37	41	48	45	49	61	87.3

reliable recognition. Fusion of multiple biometric modalities can be applied at different levels of a recognition system, *i.e.*, at feature level, matching-score level, or decision level. We plan to find a method that not only increases the accuracy of the system but also is computationally efficient.

References

Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(12), 2037–2041.

Ashraf, A. B., Lucey, S., & Chen, T. (2010). Fast image alignment in the fourier domain. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2480–2487).

Asthana, A., Marks, T. K., Jones, M. J., Tieu, K. H., & Rohith, M. (2011). Fully automatic pose-invariant face recognition via 3D pose normalization. In *Proceedings of the IEEE international conference on computer vision (ICCV)* (pp. 937–944).

Asthana, A., Sanderson, C., Gedeon, T. D., & Goecke, R. (2009). Learning-based face synthesis for pose-robust recognition from single image. In *Proceedings of the BMVC* (pp. 1–10).

Baker, S., Gross, R., & Matthews, I. (2003). *Lucas-Kanade 20 Years On: A Unifying Framework: Part 3. Technical Report, CMU-RI-TR-03-35*. Pittsburgh, PA: Robotics Institute.

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks, 13*(6), 1450–1464.

Belhumeur, P. N., Jacobs, D. W., Kriegman, D., & Kumar, N. (2011). Localizing parts of faces using a consensus of exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 545–552).

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(9), 1063–1074.

Breuer, P., Kim, K.-I., Kienzle, W., Scholkopf, B., & Blanz, V. (2008). Automatic 3D face reconstruction from single images or video. In *Proceedings of the 8th IEEE international conference on automatic face & gesture recognition* (pp. 1–8).

Castillo, C. D., & Jacobs, D. W. (2009). Using stereo matching with general epipolar geometry for 2D face recognition across pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(12), 2298–2304.

Chai, X., Shan, S., Chen, X., & Gao, W. (2007). Locally linear regression for pose-invariant face recognition. *IEEE Transactions on Image Processing, 16*(7), 1716–1725.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1998). Active appearance models. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 484–498). Springer.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(6), 681–685.

Cox, D., & Pinto, N. (2011). Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *Proceedings of the IEEE international conference on automatic face & gesture recognition and workshops (FG)* (pp. 8–15).

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR): 1* (pp. 886–893).

De Marsico, M., Nappi, M., Riccio, D., & Wechsler, H. (2013). Robust face recognition for uncontrolled pose and illumination changes. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43*(1), 149–163.

Du, S., & Ward, R. (2009). Component-wise pose normalization for pose-invariant face recognition. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 873–876).

Edwards, G. J., Cootes, T. F., & Taylor, C. J. (1998). Face recognition using active appearance models. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 581–595). Springer.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(9), 1627–1645.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2005). Pictorial structures for object recognition. *International Journal of Computer Vision, 61*(1), 55–79.

Gao, H., Ekenel, H. K., & Stiefelhagen, R. (2009). Pose normalization for local appearance-based face recognition. In *Advances in biometrics* (pp. 32–41). Springer.

Gao, S., Zhang, Y., Jia, K., Lu, J., & Zhang, Y. (2015). Single sample face recognition via learning deep supervised autoencoders. *IEEE Transactions on Information Forensics and Security, 10*(10), 2108–2118.

Ghiass, R. S., Arandjelovic, O., Bendada, H., & Maldague, X. (2013). Vesselness features and the inverse compositional aam for robust face recognition using thermal ir. In *Proceedings of the twenty-seventh AAAI conference on artificial intelligence*.

Gross, R., Matthews, I., & Baker, S. (2005). Generic vs. person specific active appearance models. *Image and Vision Computing, 23*(12), 1080–1093.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2010). Multi-PIE. *Image and Vision Computing, 28*(5), 807–813.

Guillemaut, J.-Y., Kittler, J., Sadeghi, M. T., & Christmas, W. J. (2006). General pose face recognition using frontal face model. In *Proceedings of the progress in pattern recognition, image analysis and applications* (pp. 79–88). Springer.

Guo, G., Li, S. Z., & Chan, K. L. (2000). Face recognition by support vector machines. In *Proceedings of the fourth IEEE international conference on automatic face and gesture recognition* (pp. 196–201).

Guo, Y., Ding, X., & Xue, J.-H. (2015). MiLDA: A graph embedding approach to multiview face recognition. *Neurocomputing, 151*, 1255–1261.

Guo, Z., Zhang, D., & Mou, X. (2010). Hierarchical multiscale LBP for face and palmprint recognition. In *Proceedings of the IEEE international conference on image processing (ICIP)* (pp. 4521–4524).

Haghighat, M., Abdel-Mottaleb, M., & Alhalabi, W. (2014). Computationally efficient statistical face model in the feature space. In *Proceedings of the IEEE symposium on computational intelligence in biometrics and identity management (CIBIM)* (pp. 126–131).

Haghighat, M., Zonouz, S., & Abdel-Mottaleb, M. (2013). Identification using encrypted biometrics. In *Proceedings of the computer analysis of images and patterns (CAIP)* (pp. 440–448). Springer.

Haghighat, M., Zonouz, S., & Abdel-Mottaleb, M. (2015). CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Systems with Applications, 42*(21), 7905–7916.

Hasan, M., Abdullaha, S. N. H. S., & Othman, Z. A. (2013). Efficient face recognition technique with aid of active appearance model. In *Intelligent robotics systems: Inspiring the next* (pp. 101–110). Springer.

Heo, J., & Savvides, M. (2008). Face recognition across pose using view based active appearance models (VBAAMs) on CMU Multi-PIE dataset. In *Computer vision systems* (pp. 527–535). Springer.

Ho, H. T., & Chellappa, R. (2013). Pose-invariant face recognition using markov random fields. *IEEE Transactions on Image Processing, 22*(4), 1573–1584.

Huang, G. B., 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.

Hussain, S. U., Napoléon, T., & Jurie, F. (2012). Face recognition using local quantized patterns. In *Proceedings of the British machine vision conference* (pp. 11–pages).

Kämäräinen, J.-K., Kyrki, V., & Kälviäinen, H. (2006). Invariance properties of gabor filter-based features-overview and applications. *IEEE Transactions on Image Processing, 15*(5), 1088–1099.

Kim, T.-K., Wong, K.-Y. K., Stenger, B., Kittler, J., & Cipolla, R. (2007). Incremental linear discriminant analysis using sufficient spanning set approximations. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1–8).

Komodakis, N., & Tziritas, G. (2007). Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Transactions on Image Processing, 16*(11), 2649–2661.

Krzanowski, W. J. (1988). *Principles of multivariate analysis: a user's perspective*. Oxford University Press, Inc.

Lanitis, A., Taylor, C. J., & Cootes, T. F. (1995). A unified approach to coding and interpreting face images. In *Proceedings of the Fifth international conference on computer vision* (pp. 368–373). IEEE.

Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 8595–8598).

Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive facial feature localization. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 679–692). Springer.

Li, A., Shan, S., Chen, X., & Gao, W. (2009). Maximizing intra-individual correlations for face recognition across pose differences. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 605–611).

Liu, C., & Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image processing, 11*(4), 467–476.

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision.. In *Proceedings of the IJCAI: 81* (pp. 674–679).

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 94–101).

Martin, C., Werner, U., & Gross, H.-M. (2008). A real-time facial expression recognition system based on active appearance models using gray images and edge images. In *Proceedings of the 8th IEEE international conference on automatic face and gesture recognition (FG)* (pp. 1–6).

Matthews, I., & Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision, 60*(2), 135–164.

Milborrow, S., & Nicolls, F. (2008). Locating facial features with an extended active shape model. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 504–513). Springer.

Moses, Y., Adini, Y., & Ullman, S. (1994). Face recognition: The problem of compensating for changes in illumination direction. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 286–296). Springer.

Mostafa, E., Ali, A., Alajlan, N., & Farag, A. (2012). Pose invariant approach for face recognition at distance. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 15–28). Springer.

Mostafa, E. A., & Farag, A. A. (2012). Dynamic weighting of facial features for automatic pose-invariant face recognition. In *Proceedings of the 9th conference on computer and robot vision (CRV)* (pp. 411–416). IEEE.

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(10), 1090–1104.

Sagonas, C., Panagakis, Y., Zafeiriou, S., & Pantic, M. (2015). Face frontalizationfor alignment and recognition. *arXiv preprint arXiv:1502.00852*.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE international conference on computer vision workshops (ICCVW)* (pp. 397–403).

Sarfraz, M. S., & Hellwich, O. (2010). Probabilistic learning for fully automatic face recognition across pose. *Image and Vision Computing, 28*(5), 744–753.

Sharma, A., Al Haj, M., Choi, J., Davis, L. S., & Jacobs, D. W. (2012). Robust pose invariant face recognition using coupled latent space discriminant analysis. *Computer Vision and Image Understanding, 116*(11), 1095–1110.

Sim, T., Baker, S., & Bsat, M. (2002). The CMU pose, illumination, and expression (PIE) database. In *Proceedings of the 5th IEEE international conference on automatic face and gesture recognition, 2002. proceedings* (pp. 46–51).

Sun, Q.-S., Zeng, S.-G., Liu, Y., Heng, P.-A., & Xia, D.-S. (2005). A new method of feature fusion and its application in image recognition. *Pattern Recognition, 38*(12), 2437–2448.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1701–1708). IEEE.

Tang, F., & Deng, B. (2007). Facial expression recognition using aam and local facial features. In *Proceedings of the 3rd international conference on natural computation (ICNC): 2* (pp. 632–635). IEEE.

Trutoiu, L. C., Hodgins, J. K., & Cohn, J. F. (2013). The temporal connection between smiles and blinks. In *Proceedings of the 10th IEEE international conference on automatic face and gesture recognition (FG)* (pp. 1–6).

Turk, M. A., & Pentland, A. P. (1991). Face recognition using eigenfaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 586–591).

Tzimiropoulos, G., & Pantic, M. (2013). Optimization problems for fast AAM fitting in-the-wild. In *Proceedings of the IEEE international conference on computer vision (ICCV)* (pp. 593–600).

Van Kuilenburg, H., Wiering, M., & Den Uyl, M. (2005). A model based method for automatic facial expression recognition. In *Proceedings of the 16th european conference on machine learning*. In *ECML'05* (pp. 194–205). Springer.

Vedaldi, A. Fulkerson, B. (2008). VLFeat: An open and portable library of computer visionalgorithms. http://www.vlfeat.org/. Accessed 17.03.15.

Wang, H., Li, S. Z., Wang, Y., & Zhang, J. (2004). Self quotient image for face recognition. In *Proceedings of the IEEE international conference on image processing (ICIP): 2* (pp. 1397–1400).

Wolf, L., Hassner, T., & Taigman, Y. (2010). Similarity scores based on background samples. In *Proceedings of the Computer vision-ACCV 2009* (pp. 88–97). Springer.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., & Lin, S. (2007). Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(1), 40–51.

Yang, Y., & Ramanan, D. (2011). Articulated pose estimation with flexible mixtures-of-parts. In *Proceedings of the IEEE conference on computer vision and pattern recognition (cvpr)* (pp. 1385–1392).

Yi, D., Lei, Z., & Li, S. Z. (2013). Towards pose robust face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3539–3545).

Zhang, W., Shan, S., Gao, W., Chen, X., & Zhang, H. (2005). Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In *Proceedings of the IEEE international conference on computer vision (ICCV): 1* (pp. 786–791).

Zhang, X., & Gao, Y. (2009). Face recognition across pose: A review. *Pattern Recognition, 42*(11), 2876–2896.

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR), 35*(4), 399–458.

Zhu, X., & Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2879–2886).