Pose Invariant Approach
for Face Recognition at Distance

Eslam Mostafa[1], Asem Ali[1,2], Naif Alajlan[3], and Aly Farag[1]

[1] CVIP Lab, University of Louisville, Louisville, KY, USA
[2] EE Department, Assiut University, Assiut, Egypt
[3] CIS Department, King Saud University, Riyadh, Saudi Arabia
{eslam.mostafa,asem.ali,aly.farag}@louisville.edu, najlan@KSU.EDU.SA
http://www.cvip.uofl.edu

Abstract. We propose an automatic pose invariant approach for Face Recognition At a Distance (FRAD). Since face alignment is a crucial step in face recognition systems, we propose a novel facial features extraction model, which guides extended ASM to accurately align the face. Our main concern is to recognize human faces under uncontrolled environment at far distances accurately and fast. To achieve this goal, we perform an offline stage where 3D faces are reconstructed from stereo pair images. These 3D shapes are used to synthesize virtual 2D views in novel poses. To obtain good synthesized images from the 3D shape, we propose an accurate 3D reconstruction framework, which carefully handles illumination variance, occlusion, and the disparity discontinuity. The online phase is fast where a 2D image with unknown pose is matched with the closest virtual images in sampled poses. Experiments show that our approach outperforms the-state-of-the-art approaches.

1 Introduction

The field of face recognition has been an attractive area of research the past three decades and has made a significant progress. Despite the maturity of the face recognition algorithms, the problem is still challenging. The challenging associated with face recognition can be attributed to many variations such as illumination, occlusion, pose, expression, and age.

In recent surveys of face recognition techniques, pose variation was identified as one of the prominent unsolved problems in the research of face recognition and it gains great interest in the computer vision and pattern recognition research community. The face recognition across pose can be broadly classified into two categories [1] 2D and 3D approaches.

In this paper, we propose an automatic pose invariant Face Recognition At a Distance (FRAD) approach. Our proposed approach consists of two stages: offline and online stages. In the offline stage, 3D faces are reconstructed from stereo pair images. Then, virtual 2D views in novel poses are synthesized. The online phase is a face recognition process where a probe 2D image with unknown pose is matched with the closest virtual images in sampled poses.

1.1   Related Works and Contributions

Active appearance model (AAM) [2] was proposed as a model-based approach for face alignment and pose normalization. Once the model is fitted on an input image, the optimized model parameters can be used for face recognition. Instead of using model parameters for recognition, Guillemaut et.al [3] and Gao et al. [4] warped the texture inside the fitted model shape to the frontal pose. Chai et al. [5] presented a statistical affine transformation for pose normalization. They first estimated the face pose with a pose subspace algorithm, and divided the face region into three rectangles. An affine transformation was performed independently in the three rectangles and the transformation parameters were statistically learned from the correspondence information between the specific pose and the frontal face. Sarfraz et al. [6] go away from the texture wrapping to the face signature by proposing a linear mapping of the face signature of a non-frontal face to the face signature of the corresponding frontal face. They assume that the mapping from non-frontal to frontal feature vectors is linear. However, this assumption seems to be overly restrictive [7].

Castillo and Jacobs [8] proposed to use the cost of stereo matching between a gallery face image and a probe face image to recognize faces. Since the approach is purely image based, it does not consider appearance change due to pose variation.

In the middle between 2D face recognition and 3D face recognition, Zhang et al. [9] proposed an automatic texture synthesis (ATS) approach to synthesise rotated virtual face views from a single frontal view for recognition using a generic face shape model. This face shape was generated by averaging 40 3D face shapes in range data format which were aligned using two eyes locations. Liu and Chen [10] proposed a probabilistic geometry assisted (PGA) face recognition algorithm to handle pose variations. In their algorithm, human heads were approximated as an ellipsoid whose radiuses, locations, and orientations were estimated based on universal mosaic model. Although, these 3D-based recognition approaches achieve better results than most of 2D-based pose invariant recognition approaches, their assumption that all faces have the same 3D geometry is weak, where the generic face model does not capture all 3D faces variances.

Mostafa et al.[11] proposed to use a different weighting for facial feature in recognition. The weight of each facial feature is based on robustness to occlusion. Their method based on using a mean 3D shape.

Facial 3D geometry either can be acquired using 3D sensing devices such as laser scanners [12] or reconstructed from one or more images [13]. Although using 3D sensing devices has proven to be effective in 3D face recognition [14], their high cost and limited availability, appropriate only at close ranges and controlled environment. This has created the need for 3D face reconstruction methods to enable the extraction of 3D information from 2D acquired facial images.

Blanz and Vetter [15] proposed a face recognition system using 3D morphable model based on image-based reconstruction and prior knowledge of human faces. The prior knowledge of face shapes and textures was learned from a set of 3D face scans. Then shape and texture information in the forms of vertices and diffuse reflectance coefficients was spanned into different eigen spaces where principal

component analysis was performed to form a 3D morphable model. However, the identity-related shape and texture coefficients may be affected during cost function minimization [1].

Georghiades et al. [16] proposed a face recognition approach using a photometric stereo technique. From a set of frontal face images under different near frontal illumination conditions, identity face shape and surface reflectance information were reconstructed. Virtual views in novel illumination and viewing conditions were then synthesized and used in face recognition to match the probe image with the closest virtual images in sampled poses and illuminations. This approach requires multiple images under different lighting direction with fixed pose; however, this requirement is impractical for real face recognition systems.

In this work, we are mainly concerned with a fast automatic pose invariant Face Recognition At a Distance (FRAD) approach. Our main contributions are a proposed algorithm for nine facial features detector. These nine features improve the performance of the face alignment algorithm. Also, we propose an accurate face reconstruction provides good synthesized images at far ranges ranges. The proposed 3D reconstruction framework carefully handles three main problems: illumination variance in the stereo pair, occlusion, and finally, the disparity discontinuity is overcome by fitting a grid on the reconstructed face's points.
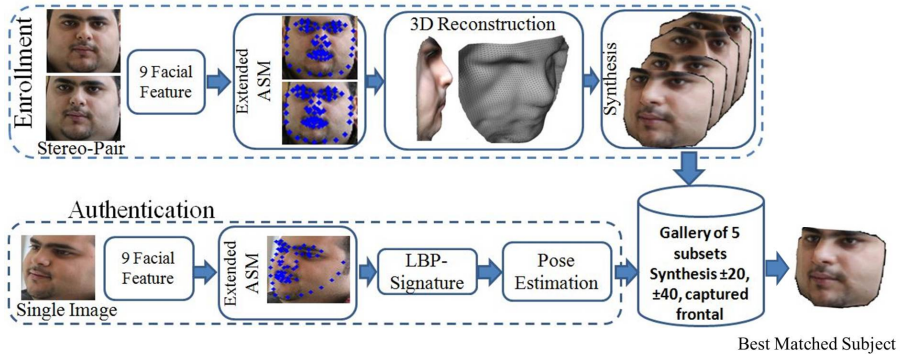
2   The Proposed Approach

As illustrated in Fig. 6, our proposed approach consists mainly of two stages: an offline stage "Enrollment" and an online stage "Authentication". The input to the Enrollment stage is a stereo pair. A 3D face is reconstructed from that stereo pair after a preprocessing step, which includes face detection, face alignment, rectification, and face cropping. 2D images with different poses are synthesized from the reconstructed 3D face and are enrolled in our gallery. A 2D face image with unknown pose is the input to the online stage. The 2D image is preprocessed by detecting, aligning and cropping the face. Finally a 2D recognition approach, which is Local Binary Pattern (LBP) [17], is used to match the probe image with the closest virtual images in sampled poses.

3   Face Alignment

Face alignment is a crucial step in face recognition system. It comes after face detection, which is used for segmentation of the facial region from the background. It aims to locate a precise facial features point such as eyes, nose, mouth, and face boundary.

Active Appearance Models (AAM) and Active Shape Models (ASM) are the most popular algorithms for face alignment. In these algorithms, principle component analysis is applied to a set of labeled faces in order to model the intrinsic variation in shape, texture, or both. This results in a parameterized model that can represent large variation in shape and texture with a small set of

Fig. 1. A block diagram describes the proposed approach

parameters. Fitting a model to a new image is generally accomplished by maximizing the posterior of the model parameters. This posterior has a complex shape and is defined over a high dimensional space, which makes it impossible to find posterior global maximum. Therefore, most algorithms are concerned with the efficient local maximization of the posterior starting from an initial guess. In some applications, the initial guess can be obtained from a face detector [18], but if the face is non-frontal, or a highly precise fit is required then it is necessary to start with a better initialization. One way to specify an accurate initialization is by detecting landmarks correspondences between points in the model domain and points in the image. We propose a novel facial features (nine points) extraction model. These features are wrapped to a mean shape which contains more facial features (68 feature points). Furthermore, the initial boundary points are refined. Then Extended active shape model [19] is used to accurately align the face (i.e., locating the 68 facial features).

3.1   Facial Feature Extraction

Most common algorithms for detecting landmarks based on a sliding window detectors such as [18]. These detectors classify each patch of the image separately as being one of the landmarks or the background. These detectors perform well for some cases such as frontal faces, which have a relatively unique appearance. However, these detectors may fail in case of occluded faces e.g., hair, glasses, hands, etc. The detection of each facial feature is independent and it ignores the relation among these facial feature points. Therefore, researchers filter the output of landmarks detector by a shape model. Most of the existing works model the relation between the facial features as gaussian model. In [20], the facial features' relative positions were modeled by a pairwise reinforcement of feature response instead of Gaussian distribution. Valstar et.al [21] proposed using a Markov Random Field (MRF) as a shape model. These approaches used

a single distribution, which is not suitable for modeling wide range of poses. Everingham and et.al [22] extended the model of the single gaussian distribution into mixture of Gaussian trees.

We propose a probabilistic framework, which models the uncertainty of detection landmarks and the relation between the facial features. This framework has advantages over the existing frameworks that it combines the uncertainty of landmark detection with shape model instead of making them two separate steps. We use a Mixture of Complex Bingham distribution as a shape model. It is a powerful model in statistics shape community since it has a symmetric property [23]. Because of the symmetric property, we do not need to do the least square fitting, which is time consuming and is prone to some errors, to filter the rotation effect.

Formally, let $Z = \{z_1, z_2, ..., z_N\}$ denote the locations of facial features and it is called a shape vector. N is the number of facial features. The feature's location is described in complex domain as $z_i = x_i + jy_i$. The probability distribution function of a shape using complex Bingham is given by

$$P(Z) = c(A)^{-1} exp(\frac{HZ^*}{\parallel HZ \parallel} A \frac{HZ}{\parallel HZ \parallel}). \tag{1}$$

H is a $(N-1) \times N$ Helmert sub-matrix [23]. A is a $(N-1) \times (N-1)$ Hermitian parameter matrix. The spectral decomposition can be written as $A = U\Lambda U^*$. $U = (u^1 u^2 \cdots u^{N-1})$ is a matrix whose columns $u^i, i = 1, .., N - 1$, which correspond to the eigenvectors of A. $\Lambda = diag(\lambda_1, \cdots, \lambda_{N-1})$ is the diagonal matrix of corresponding eigenvalues. $c(A)$ is a normalizing constant and it is given by

$$c(A) = 2\pi^{N-1} \sum_{i=1}^{N-1} a_i \exp(\lambda_i), \tag{2}$$

where $a_i^{-1} = \prod_{m \neq i}(\lambda_i - \lambda_m)$. The eigenvectors and the eigenvalues of A is proved to be $u^i = g_i$, and $\lambda_i = l_i/N$ where $i = 1, .., N - 1$. $G = (g_1, g_2, \cdots g_{N-1})$ denotes the corresponding eigenvector of S, which is a $N - 1 \times N - 1$ matrix that denotes the auto correlation matrix for manually annotated shapes (training shapes). $L = diag(l_1, l_2 \cdots l_{N-1})$ is the diagonal matrix of corresponding eigenvalues of matrix S. Although, complex Bingham distribution can efficiently represent a shape and handle situations such as in-plan rotation, occlusion, and expression, it cannot handle a wide range of poses. The shape deformation, which is due to different poses, is large and cannot be handled by a single distribution [22,24]. Therefore, we divide the training annotated shapes into $M = 5$ classes. Each class contains the shapes that correspondence to 45, -25, 0, 25, 45 degree respectively . Bayes classifier rule is used to estimate the class of the testing shape.

The index of a class is given by

$$\hat{m} = \arg \min_m \frac{Z^*}{\parallel Z \parallel} A_m \frac{HZ}{\parallel Z \parallel} + \log(C_m(\Lambda)). \tag{3}$$

The problem of facial features detection is formulated in Bayesian framework of maximum-a-posteriori. We want to find the vector Z, which maximizes the response probability for the texture model and shape model.

$$\hat{Z} = \arg\max P(I|Z)P(Z). \tag{4}$$

$P(I|Z)$ represents the probability of similarity between the texture of the face to off-line model given the facial feature vector. Since the similarity of the face can be expressed in the similarity of the windows around each facial feature, it can be written as $P(W(Z_i), W(Z_2) \cdots W(Z_N)|Z)$. Where $W(Z_i)$ is the image window around the facial point Z_i. The windows around each facial point can be considered independent from each others. Therefore

$$P(I|Z) = \prod_{i=1}^{N} P(W(Z_i)|Z_i), \tag{5}$$

where $P(W(Z_i)|Z_i)$ can be interpreted as the probability of a pixel being feature based on texture model Based on boosted classifier and Haar-like feature vector the probability can be written as Each detector for each facial feature is trained on MUCT dataset where positive examples are taken at manually annotated point and another sample located randomly inside a circle with radius 3 pixels around annotated point. The negative samples are taken far away for the annotated point at least 1/8 of the distance between two eyes.

$$P(W(Z_i)|Z_i) = \sum_{t=1}^{K} w_{t_i} B_{t_i}, \tag{6}$$

where w_{t_i} is the weight of the weak classifier t for the facial feature i . B_{t_i} is the binary response of the weak classifier t for the facial feature i. Then

$$P(I|Z) = \prod_{i=1}^{N} \sum_{t=1}^{K} w_{t_i} B_{t_i}. \tag{7}$$

Therefore, the maximum-a-posteriori estimate of facial features can be formulated as energy minimization of function $E(Z)$

$$E(Z) = \sum_{i=1}^{N} -\log(\sum_{t=1}^{K} w_{t_i} B_{t_i}) - \frac{HZ^*AHZ}{\| HZ \|^2}. \tag{8}$$

This energy function is non-linear and not amenable to gradient descent-type algorithms. The problem has been solved by simulated annealing algorithm where maximum number of iterations is 100. It takes 0.47sec on Intel core i7 2.93 GHz machine for detection the nine facial features. This problem is equivalent from the computational point of view to the denoising of an image has 9 pixels and each pixel has 10 degree of freedom in intensity level. Examples of the extracted nine features are shown in Fig. 2. Behind the scope of this paper, the proposed approach is also used as essential step for camera steering in camera network application [25]
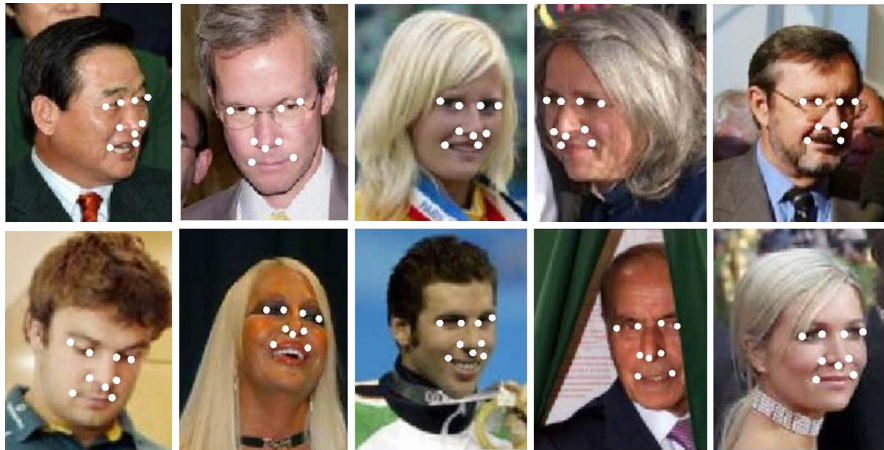
Fig. 2. A sample of our results for detection the nine facial features on LFW database

4   Passive 3D Facial Reconstruction

After the face alignment step, we use the estimated mesh to crop the faces from the background of the stereo pair images as follows. First, the stereo pair is roughly rectified by making the left and right eye centers have the same y coordinates. This shrinks the disparities range in y direction. After that, the faces are cropped using the mesh boundary such that the disparity between left eyes in the cropped stereo pair becomes zero. Example of a cropped stereo pair is illustrated in Fig. 3 (a,b).

| (a) | (b) | (c) | (d) | (e) |
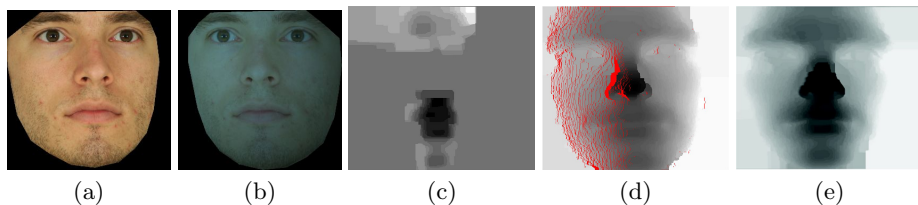
Fig. 3. Example of a stereo pair and its depth map. (a) left image, (b) right image with sever radiometric changes, (c) depth map using [26], (d) depth map using proposed approach (occluded regions colored in red,) (e) depth map after occlusion handling.

stereo matching step aims to find the pair of points p and q in two images that corespondent to the same point in the same point in the space (X, Y, Z). In this work, We follow the most conventional approaches such that stereo problem is formulated as a MAP-MRF framework [27]. We build on Kolmogorov's

approach [26] which treats the two images symmetrically by computing the disparity maps D for both images simultaneously by minimizing the function $E(D) = E_{data}(D) + E_{smooth}(D)$. Smoothness term $E_{\text{smooth}}(D)$ is chosen to be $\sum_{\{p,q\}\in\mathcal{N}} min(|d_p - d_q|, M)$, where $\mathcal{N}$ is the neighborhood system and M is empirically chosen to 1.

stereo matching step aims to find the pair of points p and q in two images that corespondent to the same point in the same point in the space (X, Y, Z). In this work, We follow the most conventional approaches such that stereo problem is formulated as a MAP-MRF framework [27]. We build on Kolmogorov's approach [26] which treats the two images symmetrically by computing the disparity maps D for both images simultaneously by minimizing the function $E(D) = E_{data}(D) + E_{smooth}(D)$. Smoothness term $E_{\text{smooth}}(D)$ is chosen to be $\sum_{\{p,q\}\in\mathcal{N}} min(|d_p - d_q|, M)$, where $\mathcal{N}$ is the neighborhood system and M is empirically chosen to 1.

Most of stereo matching approaches are based on an assumption that corresponding pixels have a similar color value. Thus the data term $E_{\text{data}}(D)$ is usually chosen as a similarity measure (e.g., [28]) that handles slight variation between corresponding pixels' colors. However, colors of the real scene are transformed nonlinearly to another colors in the stereo pair images, which violates color consistency assumption. In this work, the color normalization approach described in [29] is used to convert the transformation between the pixels' colors in two images from non-linear to a liner transformation. Then, the Normalized Cross Correlation (NCC) [30] is used in data term since (NCC) is invariant to linear transformation. We reduce the execution time of the NCC by reducing the calculation of the means of the pixels in the windows. To do this, we compute the integral image $\mathcal{I}$ where the value of $\mathcal{I}$ at pixel p is the sum of the pixels' color above and left of p: $\mathcal{I}_p = \sum_{r \leq p} I_r$. So, we calculate the mean of a window, which its corners from top left and in a clockwise direction are (p, q, r, s), as $(\mathcal{I}_p + \mathcal{I}_r - \mathcal{I}_q - \mathcal{I}_s)/\big((p_x - q_x)(p_y - s_y)\big)$.

To asses the proposed technique, we make a severe synthetic radiometric changes on a stereo pair and the compute the disparity map. As shown in Fig.3, the proposed technique accurately computes the disparity map, however, the output of conventional graph cut approach [26] is not that accurate.

After finding the disparity map $\mathfrak{D}$, we propose to fill the occluded regions (e.g., red pixels in Fig. 3(d)), by interpolating between the correctly reconstructed pixels of each scan line using a cubic Splines interpolation model.

The cloud of 3D points, which are estimated using the disparity map and system geometry, is denser than required for reproducing the amount of actual detail present in the face. So first, we downsample these points. Then to remove some artifacts of the reconstruction, an additional surface fitting step is done. We approximate the reconstructed scattered data in a least squares sense to generate a smoothed surface. Finally, we generate a triangular mesh from the smoothed and downsampled points. More details can be found in [31].

5   Signature Extraction and Similarity Measure

Face recognition using histograms of LBP (Local Binary Pattern) features was originally proposed in [17], and has been used by many groups since then. In the original approach, the face image is divided into small regions from which LBP features are extracted and histograms are used for efficient texture information representation. The signature of the face is constructed by concatenating the regional LBP extracted features. In this work, the LBP features are extracted around the detected facial feature points, eyes, mouth, and nose. This representation is more compact and efficient since most of face part does not have a distinguish characteristic. The region size around each facial feature is empirically estimated for best performance is 21×21 for face image resolution 128*128. The vector length of descriptor is 885 around each facial feature. The recognition is performed using a nearest neighbor classifier in the computed feature space using Chi square distance.

6   Experimental Results

We perform the following experimental results on our own stereo-based human faces database. The motivation behind collecting this dataset is that there is no public dataset, which contains stereo images for human faces at different distances.

Our database is really a challenging dataset, since it is collected under uncontrolled environment, outdoor with different illumination conditions (sunny, cloudy, ... etc.). The stereo pairs are collected at different distances $(50, 80, 100,$ and 150 meter), two expressions (smile and neutral), and different poses. The ranges of poses are, yaw angle $-45°$ to $45°$, pitch angle $-20°$ to $20°$, and roll angle $-20°$ to $20°$. The width of detected faces ranges from 80 to 400 pixels. Database is collected within one year with time laps three months. Each subject (identity), at least, is collected into two time lapse. The database contains 172 subjects with 1820 sessions. Subject's sessions are a set of stereo pairs, which are captured under different conditions.

The 1820 sessions is divided into a gallery and testing data (probe) as follows. The sessions is divided into twenty subsets of sessions according to their poses and captured distances. Each subset is identified by the mean pose of the sessions' poses that it contains and its captured distance. The six mean poses are $\pm45°$, $\pm25°$ and $\pm15°$. The gallery is the two subsets at 50 meter distance and at poses around $\pm15°$. Testing data is the remaining 18 subsets.

In the "Enrollment" stage, for each subject, the frontal neutral expression session at a distance 50 meter is used to reconstruct a 3D face. After reconstructing the 3D face it is a simple matter using ray-tracing technique to render synthetic images under different poses. A gallery entry of each subject consists of five images: the captured left image plus four synthesized images at poses of yaw angles ±20 and ±40. Figure 4 illustrates samples from the gallery. Finally, LBP technique is used to generate five signatures from the five images.

In the online stage, the input is a probe session, which is a 2D image that is captured under unknown environmental condition (i.e., pose, distance, illumination, ... etc.). Firstly, the face is detected [32]. Then our system automatically detects the nine facial features that are used to initialize the extended Active Shape Model (ASM). After that LBP signature is extracted from patches around these facial features. This signature not only depends on the texture but it also depends on the shape since the signature is taken around the facial features. Finally, the probe pose is estimated [7], then its signature is compared with the signatures of the gallery subset, which has the closest pose to prob pose.

It is worth mentioning that it takes around one sec per synthesized image generation. Therefore; if we detect a probe pose first then synthesize images at that pose for all the gallery subjects, for a large gallery, it will be too slow to be a practical recognition system (e.g., it needs around half an hour for 2000 subjects gallery). Therefore our scenario is much faster since the synthesizing process is done offline.

To highlight the need for the synthesizing process, we assume that the gallery consists of the captured images only. Table 1 shows performance, rank one recognition, of identification of the testing dataset against that gallery (captured images only).

For studying the effect of the reconstructed shape in the recognition, we conduct the following experiment. We replace the stereo-based reconstructed 3D with a generic 3D shape. This generic 3D face shape was generated by averaging 80 3D face shapes of USF database. This generic shape is used to synthesis the images at the different poses. We call this approach "generic+synthesized". Table 1 shows the performance, rank one recognition, of identification of the testing dataset against that gallery (generic shape).

The results in Table 1 show how using synthesized images in "generic + synthesized", instead of using captured images only, enhances the recognition rate.Moreover, it shows how our proposed recognition approach, which uses the stereo-based 3D reconstructions to synthesize images, outperforms the "generic + synthesized" approach, which use generic shape to synthesize images.

Table 1. Rank-1 recognition rates (number are percentage) on the test dataset using two approaches: without including the synthesis images in the gallery (left column in each pose) and "generic+synthesized" approach (middle column in each pose) and proposed framework (right column in each pose)

Distance	$-45°$			$-25°$			±15			$+25°$			$+45°$			**Avg**		
50m	63	74	79	71	80	82	—	—	—	86	88	89	74	76	84	74	79	**84**
80m	52	68	77	60	72	74	83	85	85	76	81	84	63	64	69	67	74	**78**
100m	46	53	61	56	64	72	78	78	78	76	79	82	59	62	62	63	67	**71**
150m	31	33	49	41	43	50	49	49	49	33	49	60	26	36	48	36	42	**51**
Avg	48	57	**67**	57	65	**70**	70	71	**71**	68	74	**79**	56	60	**66**			

Fig. 4. Samples from the gallery. Columns from left to right are: the left captured image, the right captured image and the synthesized images at poses $40°$, $20°$, $−20°$, and $−40°$.

Most of the-state-of-the-art approaches for pose invariant face recognition reported their results on non-stereo databases such as CMU-PIE and FERET. To compare with these approach, we have to use the same database. But CMU-PIE and FERET databases don't have stereo information, also we don't have the code of these works to accurately run them on our database. Therefore; we will run the "generic+synthesized" approach, which doesn't need stereo information, on the CMU-PIE and FERET database. For CMU-PIE, the gallery consists of the frontal image (POSE ID c27) with neutral expression and ambient lighting for each of the 68 subjects. The probe set consists of 6 non-frontal poses with yaw angle from $−40°$ to $40°$ and pitch from $−22.5°$ to $22.5°$ also with neutral expression and ambient lighting for each subject. For FERET database, we use all 200 subjects at 7 different poses with yaw angle $−40° to 40°$. The frontal image for each subject is used as gallery and the remaining 6 images per subject are used as probes (1,200 total).

The results in Tables 2 and 3 show the performance of the "generic + synthesized" approach compared to the performances, which are reported for other approaches. It is clear that "generic+synthesized" approach outperforms the-state-of-the-art approaches with rank-1 recognition rate 94.2% for FERET and 99.3% for CMU-PIE. Since our proposed approach outperform the "generic+synthesized" approach, we expect that it will outperform the-state-of-the-art approaches as well if it used in our database.

Table 2. Recognition rates for CMU-PIE

Method	Face Alignment	Trained on PIE	c11 −45°	c29 −22.5°	c07 up22.5°	c09 down22.5°	c05 +22.5°	c37 +45°	Avg
Kanade et al. [33]	Manual	Yes	96.8	100	100	100	100	100	99.5
Zhang et al. [34]	Automatic	No	71.6	87.9	78.8	93.9	86.4	74.6	82.2
Chai et al. [5]	Manual	No	89.8	100	98.7	98.7	98.5	82.6	94.7
Castillo et al. [8]	Manual	No	100	100	90	100	100	99.0	98.2
Sarfraz et al.[6]	Automatic	No	84.0	87.0	−	−	94.0	90.0	88.8
Asthana et al. [7]	Automatic	No	88.1	100	98.5	98.5	95.5	89.4	95.0
generic+synthesized	Automatic	No	95.6	100	100	100	100	100	99.3

Table 3. Recognition rates for FERET

Method	Face Alignment	Trained on FERRET	bh −40°	bg −25°	bf −15°	be +15°	bd +25°	bc +40°	Avg
Zhang et al. [34]	Automatic	No	62.0	91.0	98.0	96.0	84.0	51.0	80.5
Gao et al. [4]	Manual	Yes	78.5	91.5	98.	97.0	93.0	81.5	90.0
Asthana et al. [35]	Manual	Yes	87.0	93.0	98.0	98.5	95.5	74.0	91.0
Sarfraz et al. [6]	Automatic	Yes	92.4	89.7	100	98.6	97.0	89.0	94.5
generic+synthesized	Automatic	No	87.5	97.98	100	98.97	98.47	82.35	94.2

7   Conclusions

In this work, we propose a human face recognition approach. In this approach subjects at far distances with different poses are recognized automatically. To make the recognition process fast, we perform the steps of the approach that consume time offline. Where 3D faces are reconstructed from stereo pair images. Then, virtual 2D views in novel poses are synthesized. In face recognition process, a probe 2D image with unknown pose is matched with the closest virtual images in sampled poses. The proposed work was indirectly compared with the-state-of-the-art approaches, which reported their performance on CMU-PIE and FERET database. This comparison concludes that our approach is promising and can outperform the-state-of-the-art approaches.

References

1. Zhang, X., Gao, Y.: Face recognition across pose: A review. Pattern Recogn. 42 (2009)
2. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 681–685 (2001)
3. Guillemaut, J.-Y., Kittler, J., Sadeghi, M.T., Christmas, W.J.: General Pose Face Recognition Using Frontal Face Model. In: Martínez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 79–88. Springer, Heidelberg (2006)

4. Gao, H., Ekenel, H.K., Stiefelhagen, R.: Pose Normalization for Local Appearance-Based Face Recognition. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 32–41. Springer, Heidelberg (2009)
5. Chai, X., Shan, S., Chen, X., Gao, W.: Locally linear regression for pose-invariant face recognition. IEEE Transactions on Image Processing 16, 1716–1725 (2007)
6. Saquib Sarfraz, M., Hellwich, O.: Probabilistic learning for fully automatic face recognition across pose. Image Vision Comput. 28, 744–753 (2010)
7. Asthana, A., Jones, M., Marks, T., Tieu, K., Goecke, R.: Pose normalization via learned 2D warping for fully automatic face recognition. In: Proceedings of the BMVC, pp. 127.1–127.11 (2011)
8. Castillo, C., Jacobs, D.: Using stereo matching with general epipolar geometry for 2D face recognition across pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 2298–2304 (2009)
9. Zhang, X., Gao, Y., Leung, M.: Automatic texture synthesis for face recognition from single views. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 3, pp. 1151–1154 (2006)
10. Liu, X.: Pose-robust face recognition using geometry assisted probabilistic modeling. In: Proceedings of CVPR, vol. 1, pp. 502–509 (2005)
11. Mostafa, E.A., Farag, A.A.: Dynamic weighting of facial features for automatic pose-invariant face recognition. In: 2012 Ninth Conference on Computer and Robot Vision (CRV), pp. 411–416 (2012)
12. Hesher, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: Proc. of 7th Int. Symp. on Sig. Proc. and Its App., pp. 201–204 (2003)
13. Ahmed, A., Farag, A.: A New Statistical Model Combining Shape and Spherical Harmonics Illumination for Face Reconstruction. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Paragios, N., Tanveer, S.-M., Ju, T., Liu, Z., Coquillart, S., Cruz-Neira, C., Müller, T., Malzbender, T. (eds.) ISVC 2007, Part I. LNCS, vol. 4841, pp. 531–541. Springer, Heidelberg (2007)
14. Park, U., Jain, A.K.: 3D face reconstruction from stereo video. In: Proc. of First International Workshop on Video Processing for Security (2006)
15. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1063–1074 (2003)
16. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 643–660 (2001)
17. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 2037–2041 (2006)
18. Viola, P., Jones, M.: Robust real-time object detection. International Journal of Computer Vision (2001)
19. Milborrow, S., Nicolls, F.: Locating Facial Features with an Extended Active Shape Model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
20. Cristinacce, D., Cootes, T., Scott, I.: A multi-stage approach to facial feature detection. In: 15th BMVC, London, England, pp. 277–286 (2004)
21. Valstar, M., Martinez, B., Binefa, X., Pantic, M.: Facial point detection using boosted regression and graph models. In: CVPR 2010, San Francisco, USA, pp. 2729–2736 (2010)

22. Everingham, M., Sivic, J., Zisserman, A.: "Hello! My name is... Buffy" – automatic naming of characters in TV video. In: Proceedings of the British Machine Vision Conference (2006)
23. Dryden, I., Mardia, K.V.: The statistical analysis of shape. Wiley, London (1998)
24. Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2011)
25. Mostafa, E., El-Melegy, M., Farag, A.A.: Passive single image-based approach for camera steering in face recognition at a distance application. In: Proc. of IEEE 5th International Conference on Biometrics: Theory, Applications and Systems, BTAS (2012)
26. Kolmogorov, V., Zabih, R.: Multi-camera Scene Reconstruction via Graph Cuts. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 82–96. Springer, Heidelberg (2002)
27. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A Comparative Study of Energy Minimization Methods for Markov Random Fields. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part II. LNCS, vol. 3952, pp. 16–29. Springer, Heidelberg (2006)
28. Birchfield, S., Tomasi, C.: A pixel dissimilarity measure that is insensitive to image sampling. IEEE Trans. on PAMI 20, 401–406 (1998)
29. Finlayson, G., Xu, R.: Illuminant and gamma comprehensive normalisation in log RGB space. Pattern Recogn. Lett. 24, 1679–1690 (2003)
30. Heo, Y., Lee, K., Lee, S.: Illumination and camera invariant stereo matching. In: Proc. of CVPR (2008)
31. Abdelrahman, M., Ali, A.M., Elhabian, S., Rara, H., Farag, A.A.: A passive stereo system for 3D human face reconstruction and recognition at a distance. In: WACV (2012)
32. Mostafa, E., El-Barkouky, A., Rara, H., Farag, A.A.: Rejecting Pseudo-Faces using the Likelihood of Facial Features and Skin. In: Proc. of IEEE 5th International Conference on Biometrics: Theory, Applications and Systems, BTAS (2012)
33. Kanade, T., Yamada, A.: Multi-subregion based probabilistic approach toward pose-invariant face recognition. In: Proceedings of the 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation, vol. 2, pp. 954–959 (2003)
34. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In: ICCV 2005, vol. 1 (2005)
35. Asthana, A., Gedeon, T., Goecke, R., Sanderson, C.: Learning-based face synthesis for pose-robust recognition from single image. In: British Machine Vision Conference (2009)