Pose Varying Face Recognition: Review

Shahnas S[1], Sreeletha S H[2]

[1]Shahnas S, Department of Computer Science and Engineering, LBSITW Trivandrum, India
[2]Sreeletha S H, Department of Computer Science and Engineering, LBSITW Trivandrum, India

---

Abstract - *For biometric authentication, face recognition (FR) has been preferred due to its passive nature. Processing face images are accompanied by a series of complexities, like variation of pose, light, face expression, and make up. Although all aspects are important, the one that impacts the most face-related computer vision applications is pose. Every time people pose differently when they take pictures, so it will be difficult to distinguish and recognize the faces from images with changing poses. Also, pose varying face recognition has great advantages in many applications that deals with uncooperative subjects, in which being a passive biometric technique, face recognition can be implemented. Recently several efforts have been put into the research towards pose varying face recognition and many techniques have been proposed. However, several issues regarding this are still there such as lack of understanding about subspaces of pose varying images, pose-robust feature extraction, complex face synthesis methods etc. This paper, discuss about the difficulties in face recognition of varied poses and presents a comprehensive review of existing techniques. Their advantages/disadvantages, performances and strategies are compared. By generalising different approaches in handling pose variations in face and evaluating their performances, several promising features for future research have been suggested.*

Key Words: pose variation, face recognition, face recognition algorithms, 2D techniques,3D techniques, face synthesis

1. INTRODUCTION

Face recognition (FR) has been the passive biometric technique for identity authentication. Face recognition has clear advantages of being natural and passive over other biometric techniques requiring cooperative subjects such as fingerprint, iris, and retina recognition and it has the potential to recognize uncooperative subjects in a nonintrusive manner. Therefore, it can be applied to many areas, such as military, finance, surveillance security, border control, forensics, digital entertainment, etc. In recent studies of face recognition techniques, one that identified as an unsolved problem is pose variation and it gains great potential in the computer vision and pattern recognition research applications. In this survey, pose varying face recognition techniques are classified into three classes, general algorithms, 2D techniques and 3D techniques. General algorithms mean face recognition algorithms that didn't consider pose variations. They equally handle all image variations in face like illumination variations, expression variations, age variations etc during face recognition. In each class, further classifications were created and the categorisation is given in Table 1. Generally, two trends are used to develop face recognition techniques,

1) Rising the universality and potential of face recognition algorithms to tolerate image variation.

2) Significantly planning methods that can eliminate the difficulties brought by image variations according to its own characteristics, like through 2D transformations or 3D reconstructions.

The problem of pose varying face recognition is given in Section two with discussions of evaluations and challenges. Section three represents a review on general face recognition algorithms. A review based on 2D techniques or 3D techniques that handle pose variations in face recognition is provided on section four, five and six. Finally, conclusion and future work is given in Section 7 and 8 respectively.

2. PROBLEMS & CHALLENGES OF POSE VARYING FACE RECOGNITION

Pose varying face recognition refers to the problem of identifying or authorizing individuals with face images captured under arbitrary poses. Various works in face recognition are completed and great progress has been achieved, from successfully identifying criminal suspects from police investigation. Suppose, face recognition is used in airport security systems to recognize attackers and keep them from boarding plane. Ideally, the faces of attackers are collected and stored in the database to cross check the travellers' faces. The face of every person going through a security checkpoint will be scanned. Once a match is found, cameras are turned on to follow the attacker with a live video, so that the authorities can stop the attacker whose face matches with the one in the database. A solution for this task is to cluster multiple gallery images in all poses possible to find the pose variations in the captured images. But gathering multiple gallery images in different poses is difficult. Pose varying face recognition refers to recognizing face images whose poses are different from the known images. Hence, pose invariance is a key ability for face recognition to realize its advantages of being non-intrusive over other biometric techniques which requires ooperative subjects like iris recognition and fingerprint recognition.

Facial appearance change caused by pose variation make several challenges to face recognition system. Fig 1 and Fig 2 presents the following challenges.

•The rigid rotation of the head ends up in self-occlusion, which means there is loss of information for recognition.

•The position of facial texture varies nonlinearly when pose varies, that indicates the loss of semantic correspondence in 2D images.

•The shape of facial texture is warped nonlinearly along with the pose variation, that causes serious confusion with the inter-personal texture difference.

•The pose variation is usually combined with other factors to simultaneously affect face appearance.ie, subjects being captured at a long distance tend to exhibit large pose variations, as they are unaware of the cameras. Therefore, low resolution as well as illumination variations happens together with large pose variations. For these reasons, the appearance change caused by pose variation usually surpasses the intrinsic differences between people. So, it is not possible to directly compare two images under different poses, as in general face recognition algorithms. Specific strategies are needed to bridge the cross-pose gap.

Table 1. Categorisation Of Pose Varying Face Recognition

Class	Approach
General Algorithms	
Holistic approaches	Principal component analysis, Fisher discriminant analysis, Artificial neural network (Convolutional Networks), Line edge maps.
Local approaches	Template matching, Modular PCA. Local Binary Pattern (LBP).
2D Techniques	
Real view-based matching	Beymer's method, Panoramic view.
Pose transformation in image space	Active appearance models, linear shape model, Eigen light-field.
Pose transformation in feature space	Kernel methods (kernel PCA, kernel FDA), Expert fusion, Local linear regression.
3D Techniques	
Generic shape-based methods	Cylindrical 3D pose recovery, Probabilistic geometry assisted face recognition, Automatic texture synthesis.
Feature-based 3D reconstruction	Composite deformable model, Jiang's method, multi-level quadratic variation minimisation.
Image-based 3D reconstruction	Morphable model, Illumination cone model, Stereo matching.

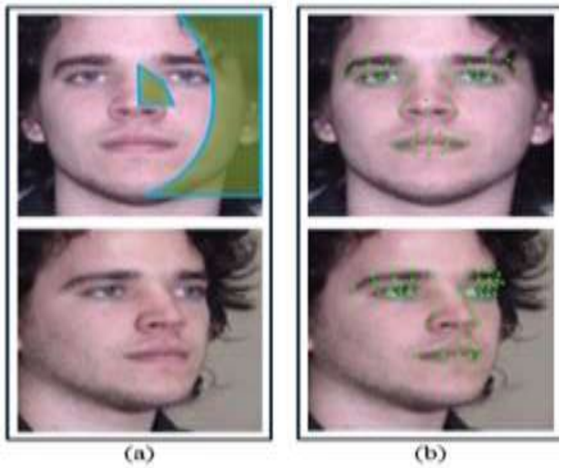

Fig 1: The challenges for face recognition caused by pose variation. (a) self-occlusion: the marked space within the frontal face is invisible in the non-frontal face (b) loss of semantic correspondence: the position of facial textures varies nonlinearly following the pose variation.

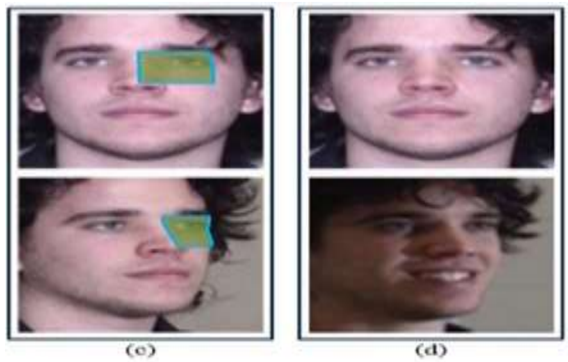

Fig 2: The challenges for face recognition caused by pose variation. (c) nonlinear deformation of facial textures (d) accompanied variations in resolution, illumination, and expression.

3. GENERAL ALGORITHMS FOR FACE RECOGNITION TECHNIQUES

A typical face recognition problem is to identify a person whose image is given as the input image by analyzing face. A number of face recognition methods have been proposed, among which Principal Component Analysis, Fisher Discriminant Analysis, Self-Organizing Map and Convolutional Network, Template matching, Modular PCA, Elastic Bunch Graph Matching, Line Edge Maps and Local Binary Patterns are some of the approaches. All of these methods try to extract features or classification patterns from face image and based on these patterns against the known face images in the database, input image is recognized.

3.1. Holistic Approaches

To represent face images by a small number of coefficients Kirby and Sirovich [1] used principal component analysis (PCA) which are corresponding to the most significant eigen values. This results in an extension of the data and imposes even and odd symmetry on the eigenfunctions of the covariance matrix, without increasing the complexity of the calculation. The resulting approximation of faces projected from outside of the data set onto this optimal basis is improved on average. Turk and Pentland [2,3] used Eigenfaces for face detection and identification. Through principal component analysis, a set of eigen vectors and eigen values were first calculated to form the eigen space of human faces from the training set of face image. The gallery and probe pictures were projected to this eigen space and their eigen values are compared in the recognition time. The Eigenfaces approach seems to be a fast, simple, and practical method, that becomes the most used face recognition technique. However, it does not offer better invariance over changes in poses and scales.

Fisher faces approach (or Fisher discriminant analysis, FDA) [4] was applied to provide the discrimination among classes, when multiple training data per class are available. n. This projection method produces well separated classes in a low-dimensional subspace, even under severe variation in lighting and facial expressions. The Eigenface technique, another method based on linearly projecting the image space to a low dimensional subspace, has similar computational requirements. Yet, extensive experimental results demonstrate that the proposed "Fisherface" method has error rates that are lower than those of the Eigenface technique for tests on the Harvard and Yale Face Databases. To beat the problem of within-class scatter matrix being singular, the face images were initially projected using PCA to reduce the dimensionality to a lower level that FDA can handle. In this case, it needs multiple gallery images per class or FDA will be identical to PCA. As holistic face recognition approaches, both FDA and PCA are very sensitive to pose [5], because in-depth rotations of 3D human faces almost always cause misalignment of image pixels that are the only classification clues for these holistic approaches. The attractiveness of using artificial neural network (ANN) can be because of its nonlinearity within the network. One of the first artificial neural network techniques used for face recognition is that the single layer network WISARD [6], which contains a separate network for every stored individual. Every node is assigned with a set of n weights where n is the dimension of the sub-image. In training, the Best Matching Unit (BMU) to each training sub image is selected as the closest match. The planes of the final layer has only one element, which indicates the classification results. In general, however, neural network approaches encounter problems when the number of classes increases. For pose varying face recognition, one individual may require several classes in different poses. Edge data of faces can also be used for face recognition. A Line Edge Map (LEM) [7] approach was proposed, which provide a distance measurement between two-line edge maps of faces and performs face matching based on those measures. The LEM of a face image is generated by consecutive 1) Extracting edges, 2) Thinning, and 3) Polygonal line fitting. LEM is sensitive to pose variations, because in depth rotations always cause distortions of image edge maps which will affect the performances of the methods using image edges as classification patterns.

3.2. Local Approaches

For holistic approach, the face recognition decisions are made considering the entire face images. In contrast, local approaches mainly consider a set of isolated points or regions on the face images and classification patterns are extracted from a limited region in the face image.

Template matching recognize faces by considering local regions represented in templates, which compares input images pixel-wisely against a template using a suitable metric such as the Euclidean distance. Bruneli and Poggio [8] automatically selected a set of 4 feature template, i.e., the eyes, nose, mouth and the whole face, for all of the available faces. Within each template, the input image region is compared with each database image in the same region through normalized cross correlation. The recognition decision was created using summed matching scores. One drawback of template matching lies in the description of these templates. Since the recognition system need to be tolerant to certain discrepancies between gallery and probe images, this tolerance would possibly average out the differences that make individual faces unique.

Pentland et al. [9] extended PCA to Modular PCA (MPCA) to improve robustness of face recognition. Instead of building a holistic eigen space for the entire images, MPCA establishes multiple eigen spaces around facial components such as eyes, nose, and mouth to form "Eigenfeatures" (Fig 3). Multiple fixed-size sub-regions are first located through facial component detection to the facial components (e.g., eyes) and only image pixels in the subregions are considered in the Eigenfeatures process in training and recognition. Eigen values of a face image are calculated separately in different sub-regions which are then concatenated for classification. The pose tolerance is achieved by eliminating the impact of facial feature misalignment under pose variations, at the price of neglecting some useful image patterns like freckles, birthmarks, and wrinkles which can be considered in holistic approaches. As MPCA depends on the predefined facial features, the feature detection is crucial to this approach similar to other feature-based face recognition methods. In experiments, it didn't offer any test on face recognition across pose due to the difficulty of automatically detecting facial parts under rotated face images.

Ahonen et al. [10] applied Local Binary Patterns (LBP) [11], a successful texture descriptor, used for face recognition. The local pattern is extracted by binarizing the gradients of center point to its 8 neighbouring points pixel-wisely and this binary pattern is used as image features for classification. Then the face image is split into several patches and inside each patch, the histogram of the local pixel-wise patterns is calculated. Comparing two images, the histograms are compared through calculating weighted Chi square distance, whose weights are trained by separate recognition process on a single patch. Though the LBP primarily focuses on pixel-wise local patterns, the holistic information is also considered by concatenating the regional histograms into a single description over the entire image. Compared to holistic approaches, LBP is more robust to pose changes because it doesn't require exact locations of patterns but relies only on histogram of the pattern in a region. It is found that LBP will tolerate small pose variations and achieve excellent recognition rates when the rotations are less than 15 degrees. Once the rotation becomes larger, however, the dividing face pictures into regions becomes problematic. These approaches are not entirely robust to pose variations, because in local regions, image distortions brought by pose variations still exist.

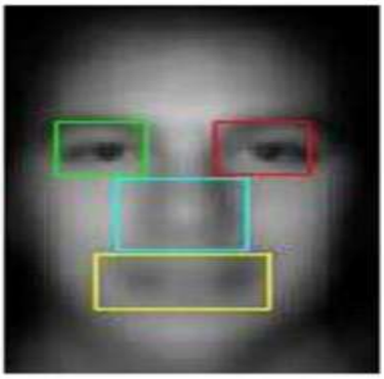

Fig 3: The Modular PCA builds multiple eigenfeature in the regions of facial components (e.g., eyes, nose, and mouth) to achieve pose tolerances.

4.2D TECHNIQUES FOR POSE-VARYING FACE RECOGNITION

Most of the general face recognition approaches are sensitive to pose variations, so a number of approaches have been proposed to handle pose variations. 2D techniques and 3D methods were used to handle or predict the appearance variations of human faces brought by changing poses. In this section, 2D techniques are classified as three groups, i.e, 1) pose-tolerant feature extraction 2) real view-based matching, and 3) 2D pose transformation.

4.1. Real View-Based Matching

The natural way to realize a face recognition system against pose variations in the direction is to prepare multiple real view templates for every known individual. Because general face recognition algorithms which are previously reviewed are able to tolerate small pose variations (ie., 15degree rotation), the number of required real gallery images can be significantly reduced by quantization on the in-depth rotations. Beymer [12] designed a real view-based (RVB) face recognition system employing a template matching of image-based single-view illustration. Every input view was geometrically registered to the known person's templates by using locations of eyes and nose, that were automatically located by that system. The recognizer acquires fifteen gallery face images to cover a range of pose variations with approximately ±40° in yaw and ±20° in tilt. The recognition process is a typical template matching algorithm with templates around eyes, nose and mouth, while the only difference is that it matches an off-centered probe face image with gallery face images in similar poses.

Singh et al. [13] proposed a mosaicing scheme (MS) to form a wide view as shown in Fig 3 from multiple gallery images to cover the possible appearances under all horizontal in-depth rotations. The panoramic view is generated from a frontal view and rotated views in three steps, i.e., 1) view alignment, 2) image segmentation, and 3) image stitching. In the first step, views in several poses were aligned by coarse affine alignment and fine mutual information based general alignment. The boundary blocks of 8 by 8 pixels for the segmentation were detected using phase correlation, that were used as the connection regions of the two views to stitch. A multi-resolution splining was applied to straddle the connecting boundary of the images and also the splined images were expanded and summed along to form the final composite face mosaic. In recognition, the synthesized face mosaics were used as gallery and single normal face images in arbitrary poses were matched using a face recognition algorithm combining log Gabor transform, C2 feature extraction, and 2ν-support vector machine. The clear advantage of using face mosaics over virtual view synthesis is the save of storage spaces, because only a single image per person is required to cover all possible poses. In general, face recognition methods of real-view based matching need multiple real views of every person as gallery. Either the raw gallery pitures or some transformations of them are considered in recognition to cover possible pose variations.

4.2. Pose Transformation in image space

As it is generally impractical to collect multiple images in different poses for real view-based matching, a feasible alternate is to synthesize virtual views to substitute the demand of real views from a restricted number of known views (even from a single view). The virtual view synthesis can be done in 2D space as pose transformation or in 3D space as 3D face reconstruction and projection.

Active Shape Model (ASM) originally proposed by Cootes et al. [14] is one of the most thriving approaches in automatic face image representation, structure locating in medical Images and face recognition. In ASM, Principal Component Analysis (PCA) was applied on the locations of facial components such as facial contours, eyes, lips, etc. which are presented as connected point distributions from a variety of manually labelled images, containing various image variations such as pose, illumination, and expression variations. The distributions of the eigen model parameters obtained by projecting face shapes represented as point distributions, onto this eigen space are then used to exclude invalid shapes, suppose a face shape wherever the mouth location is between those of eyes and nose is excluded. To automatically modify the point distribution to the new face image, a local searching strategy is applied on each point. Initially, a gradient-based local profile on the point is extract along the local line segment perpendicular to the boundary of the point. Based on the training set, an average profile is calculated that captures the local texture variations around the point. This profile, in adjustment step, is employed to seek out the location of that point in the new image whose local profile best fits this reference profile. As an extension of ASM, Active Appearance Models (AAM) [15] has been planned simultaneously to model the variations of shape represented by point distributions and textures represented by pixel intensities. The shape variations were obtained in the same manner as ASM, using PCA on a training set of point distributions. For texture variations, every image in the training set was warped to a uniform shape and the pixel intensities were then analyzed using PCA. With both shape and texture eigen spaces, a new face was represented by a vector of model parameters controlling face variations based on the 2 eigen spaces, a vector of similarity transformation parameters controlling shape transformations and a vector of scaling and offset transformation parameters controlling texture transformations. Because the AAM is based on 2D image transformation, the in-depth rotation cannot be decoupled from the identity changes.

Vetter [16] extended the concept of AAM by substituting the sparse point distributions with a pixel-wise correspondence between two images in different poses using optical flow. A linear shape model of dense point distributions in 3D space was built using PCA on a set of 3D training face shape. Then it was projected to different poses to generate different linear shape models in 2D image space, where a single set of model parameters can describe the 2D projections in these poses of the same 3D shape. To align the linear shape model to new image, optical flow was applied to establish a dense correspondence between the projected model shape in the same pose and the input image, followed by estimating model parameters by projecting the shape distribution onto the eigen space of the linear 2D shape model. The same model parameters were then used on the linear 2D shape model in the target pose to synthesize new shape in that pose.

Gross et al. [17] proposed Eigen Light-Field (ELF) method to increase the capability of Vetter's method of handling multiple gallery images in different poses per face in face recognition across pose. They clustered all possible appearances of faces in different poses within a framework of light field, which is in a 4D space.

Ying Tai and Jian Yang et al. [18] introduce the orthogonal Procrustes problem (OPP) as a model to handle pose variations existed in 2D face images. OPP seeks an optimal linear transformation between two images with different poses so as to make the transformed image best fits the other one. They integrate OPP into the regression model and propose the orthogonal Procrustes regression (OPR) model. To deal with the problem that the linear transformation is not appropriate for handling extreme non-linear pose variation, they further adopt a progressive strategy and propose the stacked OPR. As a sensible framework, OPR can handle face alignment, pose correction, and face representation simultaneously.

4.3. Pose Transformation in Feature Space

Pose tolerance can also be achieved in feature space rather than the image space, where the feature-space transformed data cannot be visually displayed as face images as in the image space. One possible feature space transformation for face recognition is kernel tricks that nonlinearly map face images into a better higher dimensional non-linear feature space, so that the previously nonseparable distributions caused by pose variations could be linearly separable.

Variety of kernel-based face recognizers were proposed to perform face recognition or other pattern recognition tasks, such as various Kernel PCAs [19,20] and Kernel FDA [21,22]. In [20], Schölkopf et al. proposed a framework of performing a non-linear PCA with kernel functions in high-dimensional feature space transformed from the input image space. Liu [19] pre-processed the facial images with Gabor wavelets and extended kernel polynomial functions to have fractional powers in Kernel PCA. Huang et al. [21] proposed to automatically tune to find optimal parameters of a Gaussian radial basis function in their Kernel Fisher Discriminant Analysis (K-FDA) using an Eigenvalue-Stability-Bounded Margin Maximization (ESBMM) algorithm. The kernel tricks, sometimes with Gabor filtering to extract local texture information, improved PCA's or FDA's capability in handling pose variations. However, this improvement is restricted due to the fact that the actual nonlinear transformation forms caused by pose variations are unknown.

Kim and Kittler [23] proposed a hybrid approach, expert fusion, fusing four completely different systems to tolerate pose variations in face images for recognition. The first system is based on a linear pose transformation on PCA features which are then classified using linear discriminant analysis (LDA). The second system simultaneously trains linear transformation matrix and the LDA system and uses raw image data without the previous PCA feature extraction. The third system applies generalized discriminant analysis (GDA), which uses non-linear radial basis function as pose transformation functions. The fourth system applies a pose transformation lookup table generated by rotating generic 3D face shape.

Chai et al. [24] proposed to generate virtual frontal views from single horizontally rotated views through Local Linear Regression (LLR). In training stage, the face image was first divided into 10-30 evenly distributed patches in terms of an average cylindrical face model. In each patch, linear regression was performed to minimize the sum-square of image differences between frontal and non-frontal face images under a linear transformation. Then in testing stage, the input non-frontal image was also divided into patches in the same manner and each patch was transformed using the trained linear transformation matrix to form the appearance in the frontal view. Finally, all reconstructed patches were combined with an intensity averaging of overlapped pixels to form holistic frontal virtual views for recognition.

Zhenyao Zhu et al. [25] have described that face recognition with large pose and illumination variations is a challenging problem in computer vision. Their paper addresses this challenge by proposing a new learning-based face representation, the face identity-preserving (FIP) features. In contrast to conventional face descriptors, the FIP features can significantly reduce intra-identity variances, whereas maintaining discriminativeness between identities. Moreover, the FIP features extracted from a picture under any pose and illumination can be used to reconstruct its face image in the canonical view. This property makes it attainable to enhance the performance of traditional descriptors, such as LBP and Gabor, that can be extracted from our reconstructed images in the canonical view to eliminate variations.

5. 3D TECHNIQUES FOR POSE-VARYING FACE RECOGNITION

Recently, face recognition with the help of 3D models becomes one of the successful approaches, especially when dealing with pose. The success of 3D model-based approaches in handling pose variations is due to the fact that human heads are 3D objects with fine structures and changes in viewpoints all take places in the 3D spaces.

5.1. Generic Shape-Based Approaches

A simple and efficient pose recovery methodology based on a generic cylindrical face shape was proposed [26] to handle face images in small in-depth pose variations. The face images in arbitrary horizontal poses were mapped onto the generic cylindrical face shape and the frontal virtual views can be recovered.

Liu and Chen [27] proposed a Probabilistic Geometry Assisted (PGA) face recognition algorithm to handle pose variations. In their algorithm, human heads were approximated as an ellipsoid whose locations, orientations and radiuses were estimated based on universal mosaic model. Then the facial textures of the image were warped onto the surface of the ellipsoid which became free from pose variations. Due to occlusion, the visible regions of images in totally different poses were different, in order that a normalized pixel-wise Euclidean distance was used for recognition that solely considers the overlapped region of 2 texture maps on the ellipsoid.

Zhang et. al [28] proposed an Automatic Texture Synthesis (ATS) approach to generate rotated virtual face views from a single frontal view image 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.

Techniques using 3D generic shapes are very similar to 2D pose transformation. The only difference is that the transformation space is no longer the image space, but a nonlinear space specified by the 3D generic shape. Despite of their simplicity and efficiency, techniques using 3D generic shapes suffer from the incapability to preserve inter-personal shape difference, which is an important feature for face classification.

5.2. Feature-Based 3D Face Reconstructions

3D reconstruction is a vigorous research area in computer vision, which inversely estimates 3D shape information from 2D images. Generalized 3D reconstruction considers all of the shape modelling, the surface reflectivity descriptions and the estimation of environmental parameters.

Lee and Ranganath [29] presented a composite 3D deformable face model for pose estimation and face synthesis based on a template deformation which maintained connectedness and smoothness. Three sub-models of edge model, colour region model and a wire frame model were deformed correspondingly in minimizing a cost function consisting of edge fitting errors, colour region displacements and deformation energy. Using five images of the same person with different poses, a complete 3D face model for the person can be generated. The model was transformed to novel poses and scales by rigid 3D rotation and the virtual textures were synthesized by estimating an optimal set of coefficients on a linear texture space spanned by training images to best approximate the input image. The recognition was then performed by comparing the synthesized image with the probe real image pixel-wisely in Euclidean distance.

Jiang et al. [30] used facial features to efficiently reconstruct personalized 3D face models from a single frontal face image for recognition. Their technique is based on the automatic detection of facial features on the frontal views using Bayesian shape localization. A set of 100 3D face scans was used as prior knowledge of human faces. Facial features on both input images and 3D scans were used to find principal components of face shapes on the shape spaces spanned by the training 3D shapes. Personalized 3D face shapes were reconstructed and the facial textures were directly mapped onto the face shape to synthesize virtual views in novel conditions.

Zhang et al. [31] proposed to reconstruct the personalized 3D face shape by using Multi-level Quadratic Variation Minimisation (MQVM). From a 3D feature point set manually specified on the frontal view and side view of an input face, the 3D face shape was reconstructed from scratch by reducing a cost function of quadratic variations of 3D surfaces which ensures a second order smoothness. This process was performed in a hierarchical manner to overcome the sparseness of the facial feature points on facial images. MQVM started in a coarse resolution which reached the convergence quickly and provided a good initial shape for the next resolution level. Finally, a pixel-wise 3D surface model was reconstructed in the finest resolution level. After shape reconstruction, this method analyzed facial textures by fitting the pixel intensities in Phong reflection model in considerations of face shape and lighting directions known a prior. Then virtual face views in different poses were synthesized and local binary patterns (LBP) were used for recognition in a view-based manner.

Yaniv Taigman and Ming Yang [32] have discussed that in modern face recognition, the traditional pipeline consists of four stages, detect, align, represent and classify. They get back both the alignment step and also the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and develop a face representation from a nine-layer deep neural network. This deep network includes more than 120 million parameters using several locally connected layers while not weight sharing, instead of the standard convolutional layers. Thus, they trained it on the highest facial dataset to-date, an identity labeled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the more accurate model-based alignment with the large facial database generalize remarkably well to faces in free environments, even with a simple classifier.

In feature-based 3D face modelling approaches discussed above, personalized 3D face shapes were reconstructed from a set of facial features specified on facial images. The use of prior knowledge helps the systems to reduce the number of gallery views required. However, the prior knowledge of human face shapes will be unreliable if the input face shape is very different from the average shape, which causes the shape deformation to fail in converging to a plausible reconstruction result.

5.3. Image-Based Reconstructions

Image-based 3D face reconstructions carefully study the connection between image pixel intensities and its corresponding shape/texture properties. Unlike featurebased 3D face reconstructions' use of a few features on the face images is restricted, image-based 3D face reconstructions make use of almost every point on the face images and it is thought to closely resemble the reality of reflections.

Blanz and Vetter [33] proposed a successful face recognition system using 3D Morphable Model (3DMM) 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 where pixelwised inter-personal correspondence had been established using 3D version of optical flow on 3D surfaces. 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.

Georghiades et al. [34] proposed Illumination Cone Models (ICM) that successfully performed face recognition under pose and illumination variations using the techniques of photometric stereo. Their method is based on the fact that the set of images of an object with Lambertian surfaces in fixed pose but under all possible illumination conditions is a convex cone in the space of images. From a set of frontal face images under different near frontal illumination conditions, personalized face shape and surface reflectance information was reconstructed by minimizing the difference of the input gallery face images and the corresponding rendered images associated with surface gradients and reflecting properties. The procedure sequentially estimates lighting conditions, surface gradients, and diffuse reflectance coefficients and gradually converges to an optimal solution in a least square sense using singular value decomposition. Virtual views in novel illumination and viewing conditions were then synthesized and used in face recognition to match with the probe image with the closest virtual images in sampled poses and illuminations.

Castillo and Jacobs [35] discussed about a method to use the cost of stereo matching of gallery face image and probe face image to recognize faces. The stereo matching algorithm defines four planes which were left and right occluded planes and left and right matched planes. It involved fourteen transitions such as state preserving transitions and between state transitions. The cost of the stereo matching is defined as the sum of all the matching rows of the first image to the second image. Exhaustively performing stereo matching using every view in the gallery to the probe image, the match was selected when the cost of stereo matching was the smallest.

Sincy John, Ajit Danti et al. [38] discusses about passive methods in 3D face reconstruction techniques. The challenges addressed in recent studies are mainly focused on the faster reconstruction of 3D Images, improved accuracy of reconstructed images, human pose identification, image reproduction with higher resolution. Their paper attempts to consolidate the research works that have happened in the history of 3D face reconstruction. Also, they try to classify the existing methods based on the input for the process.

6. OTHER RELATED WORKS

Rui Huang et al. [36] proposed a Two-Pathway Generative Adversarial Network (TP-GAN) for photorealistic frontal view synthesis by simultaneously perceiving global structures and local details. Four landmark placed patch networks are proposed to attend to local textures in addition to the commonly used global encoder-decoder network. Except for the novel architecture, they make the ill-posed problem well constrained by introducing a combination of adversarial loss, symmetry loss and identity preserving loss. The combined loss function leverages both the frontal face distribution and pre-trained discriminative deep face models to guide an identity preserving inference of frontal views from profile images. Different from previous deep learning methods that mainly rely on intermediate features for recognition, their method directly leverages the synthesized identity preserving image for downstream tasks like face recognition and attribution estimation.

Luan Tran and Xi Yin et al. [37] address the problem that large pose discrepancy between two face images is one of the key challenges in face recognition. Conventional approaches for pose invariant face recognition either perform face frontalization on, or learn a pose-invariant illustration from, a non-frontal face image. They argue that it is more desirable to perform both tasks jointly to allow them to leverage each other. Their paper proposes Disentangled Representation Learning-Generative Adversarial Network (DR-GAN) with three distinct novelties. First, the encoder-decoder structure of the generator permits DR-GAN to find out a generative and discriminative representation, additionally to image synthesis. Second, this representation is explicitly disentangled from other face changes such as pose, through the pose code provided to the decoder and pose estimation in the discriminator. Third, DR-GAN can take one or multiple images as the input, and generate one unified illustration along with an arbitrary number of generated images.

7. CONCLUSIONS

As the prominent problem in face recognition, pose variation received more attention in the research community of computer vision and pattern recognition. A number of promising techniques have been proposed to tolerate image variations brought by pose changes. However, achieving pose invariance in face recognition still remains an unsolved challenge, which requires continuing attentions and efforts.

This survey started on discussions of the problem of pose varying face recognition, with current evaluation technologies, and performances of different techniques. Face recognition techniques relevant to handling pose variations were then classified into three broad categories, general algorithms, 2D techniques and 3D techniques.

Based on this review, several findings are summarised as follows. Prior knowledge of human faces plays an important role in handling pose variations in face recognition, especially with limited gallery images. The inclusion of this prior knowledge often requires extensive trainings and the performance is dependent on training data. The techniques without prior knowledge of human faces usually don't need any training process, which rely only on the available gallery images. Due to the insufficient data provided by the 2D gallery images, however, these techniques usually require more than one gallery image to successfully compensate pose variations. 3D face recognition approaches can generally handle larger pose variations than 2D techniques. Because pose variations are 3D transformations rather than 2D image transformations, 3D approaches are more promising to achieve better performance in pose varying face recognition. The existing 3D face reconstruction methods made suboptimal surface assumptions on human faces, which affects the reconstruction results. The most common assumption is Lambertian assumption, which only considers diffuse reflection of faces. Studies in human skins show specular and diffuse reflectivity of human faces are both histological characteristics different from person to person which can be and should be used as discriminating parameters in face recognition.

8. FUTURE WORK

A limited number of face images at different poses is used to train the model, where a number of separate generator models learn to map a single face image at any arbitrary pose to specific poses and the discriminator performs the task of face recognition along with discriminating fake images from the real ones.

REFERENCES

[1] M. Kirby and L. Sirovich, "Application of the KarhunenLoève procedure for the characterization of human face," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 1, pp. 103-108, 1990
[2] M. Turk and A. Pentland, "Eigenfaces for recognition," J Cogn Neurosci, vol. 3, no. 1, pp. 71-86, 1991.
[3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in Proc. IEEE Conf. CVPR, pp. 586-591, 1991.
[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711-720, 1997.
[5] S. Chen, X. Tan, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," Pattern Rec., vol. 39, no. 9, pp. 1725-1745, 2006.
[6] T. Stonham, "Practical face recognition and verification with WISARD," Aspects of Face Processing, pp. 426–441, 1984.
[7] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 6, pp. 764-779, 2002.
[8] R. Brunelli and T. Poggio, "Face recognition: features versus templates," IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 10, pp. 1042-1052, 1993.
[9] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in Proc. IEEE Conf. CVPR, pp. 84-91, 1994.
[10] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037-2041, 2006.
[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971987, 2002.
[12] D. J. Beymer, "Face recognition under varying pose," in Proc. IEEE Conf. CVPR, pp. 756-761, 1994.
[13] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 37, no. 5, pp. 12121225, 2007.
[14] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham, "Active shape models - their training and application," Comput. Vis. Image Understand., vol. 61, no. 1, pp. 38-59, 1995.
[15] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 681-685, 2001.