

Face Recognition across Pose: a Review

Xiaozheng Zhang and Yongsheng Gao

*Computer Vision and Image Processing Lab,
Institute for Integrated and Intelligent Systems, Griffith University, Australia
{x.zhang, yongsheng.gao}@griffith.edu.au*

Abstract

One of the major challenges encountered by current face recognition techniques lies in the difficulties of handling varying poses, i.e., recognition of faces in arbitrary in-depth rotations. The face image differences caused by rotations are often larger than the inter-person differences used in distinguishing identities. Face recognition across pose, on the other hand, has great potentials in many applications dealing with uncooperative subjects, in which the full power of face recognition being a passive biometric technique can be implemented and utilised. Extensive efforts have been put into the research toward pose-invariant face recognition in recent years and many prominent approaches have been proposed. However, several issues in face recognition across pose still remain open, such as lack of understanding about subspaces of pose variant images, problem intractability in 3D face modelling, complex face surface reflection mechanism, etc. This paper provides a critical survey of researches on image-based face recognition across pose. The existing techniques are comprehensively reviewed and discussed. They are classified into different categories according to their methodologies in handling pose variations. Their strategies, advantages/disadvantages and performances are elaborated. By generalising different tactics in handling pose variations and evaluating their performances, several promising directions for future research have been suggested.

Keywords: Face recognition, pose variation, survey, review

1 Introduction

As one of the most important biometric techniques, face recognition has clear advantages of being natural and passive over other biometric techniques requiring cooperative subjects such as fingerprint recognition and iris recognition. To benefit from the non-intrusive nature of face recognition, a system is supposed to be able to identify/recognise an uncooperative face in uncontrolled environment and an arbitrary situation without the notice of the subject. This generality of environment and situations, however, brought serious challenges to face recognition techniques, e.g., the appearances of a face due to viewing (or photo shooting) condition changes may vary too much to tolerate or handle. Though many face recognition approaches, for example [4, 7, 27, 35, 43, 53, 75], reported satisfactory performances, their successes are limited to the conditions of controlled environment, which are unrealistic in many real applications. In recent surveys of face recognition techniques [21, 89], 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. Consequently, a few promising methods have been proposed in tackling the problem of recognising faces in arbitrary poses, such as tied factor analysis [63], 3D morphable model [14], eigen light-field [33], illumination cone model [30], etc. However, none of them is free from limitations and is able to fully solve pose problem in face recognition. Continuing attentions and efforts are still necessary in the research activities towards ultimately reaching the goal of pose-invariant face recognition and achieving the full advantage of being passive for face recognition. Although several survey papers [1, 3, 20, 21, 56, 89] and books [52, 78, 90] on face recognition have been published which gave very good reviews on face recognition in general, there is no review specific on this challenging problem of face recognition across pose. This paper provides the first survey on face recognition across pose,

with comprehensive and up-to-date reviews on existing techniques and critical discussions of major challenges and possible directions in this research area.

In this review, techniques of face recognition across pose are broadly classified into three categories, i.e., general algorithms, 2D techniques, and 3D approaches. By “general algorithms”, we mean these algorithms didn’t contain specific tactics on handling pose variations. They were designed for general purpose of face recognition equally handling all image variations (e.g., illumination variations, expression variations, age variations, and pose variations, etc.). In each category, further classifications were also made and the details of categorisation is summarised in Table 1. Generally, there are two trends in developing face recognition techniques, i.e., 1) improving the capability and universality of general face recognition algorithms so that image variation can be tolerated, and 2) particularly designing mechanisms that can eliminate or at least compensate the difficulties brought by image variations (e.g., pose variations) according to its own characteristics, such as through 2D transformations or 3D reconstructions. The problem of face recognition across pose is elaborated in Section 2 with discussions of demands, challenges and evaluations. Section 3 presents a review on general face recognition algorithms with discussions on their pose sensitivities. In Sections 4 and 5, a comprehensive survey is provided on techniques that actively compensate pose variations in face recognition, dependent on whether they are 2D techniques (Section 4) or 3D approaches (Section 5). Finally, summarising discussions are given in Section 6.

Table 1. Categorisation of face recognition techniques across pose

Category	Approach
General algorithms	
Holistic approaches	Principal component analysis [43, 74, 75], Fisher discriminant analysis [7] Artificial neural network (Convolutional Networks [47]) Line edge maps [27], directional corner point [28]
Local approaches	Template matching [16], Modular PCA [61] Elastic bunch graph matching [79], local binary patterns [2]
2D techniques for face recognition across pose	
Real view-based matching	Beymer's method [12], panoramic view [71]
Pose transformation in image space	Parallel deformation [10], pose parameter manipulation [32] Active appearance models [25, 39], linear shape model [40] Eigen light-field [33]
Pose transformation in feature space	Kernel methods (kernel PCA [54, 80], kernel FDA [36, 82]) Expert fusion [42], correlation filters [50] Local linear regression [19], tied factor analysis [63]
Face recognition across pose with assistance of 3D models	
Generic shape-based methods	Cylindrical 3D pose recovery [26] Probabilistic geometry assisted face recognition [55] Automatic texture synthesis [85]
Feature-based 3D reconstruction	Composite deformable model [48], Jiang's method [38], multi-level quadratic variation minimisation [87]
Image-based 3D reconstruction	Morphable model [13, 14], illumination cone model [29, 30] Stereo matching [18]

2 Problem definition, challenges, evaluations and categorisations

Face recognition across pose refers to recognising face images in different poses by computers. It is of great interest in many face recognition applications, most notably those using indifferent or uncooperative subjects, such as surveillance systems. For example, face recognition is appealing in airport security to recognise terrorists and keep them from boarding plane. Ideally, the faces of terrorists are collected and stored in the database against which travellers' faces will be compared. The face of everyone going through a security checkpoint will be scanned. Once a match is found, cameras will be turned on to surveil people with a live video feed, and then the

authorities will verify the match and decide whether to stop the individual whose face matches one in the database. The most natural solution for this task might be to collect multiple gallery images in all possible poses to cover the pose variations in the captured images, which requires a fairly easy face recognition algorithm and will be discussed in detail in Subsection 4.1. In many real situations, however, it is tedious and/or difficult to collect these multiple gallery images in different poses and therefore the ability of face recognition algorithm to tolerate pose variations is desirable. For instance, if only a passport photo per person was stored in the database, a good face recognition algorithm should still be able to perform the above airport surveillance task. In such sense, face recognition across pose refers to recognising face images whose poses are different from the gallery (known) images. If a face recognition doesn't have a good pose tolerance, given a frontal passport photo, the system appears to require cooperative subjects who look directly at the camera [17] and face recognition is no longer passive and non-intrusive. Therefore, pose invariance or tolerance is a key ability for face recognition to achieve its advantages of being non-intrusive over other biometric techniques requiring cooperative subjects such as fingerprint recognition and iris recognition.

Due to the complex 3D structures and various surface reflectivities of human faces, however, pose variations bring serious challenges to current face recognition systems. The image variations of human faces under 3D transformations are larger than that conventional face recognition can tolerate. Specifically, innate characteristics of the faces, which distinguish one face from another, do not vary greatly from individual to individual, while magnitudes of image variations caused by pose variations are often larger than magnitudes of the variations of the innate characteristics. The challenging task faced by pose-invariant face recognition algorithms is to extract the innate characteristics free from pose variations. Generally, if more gallery images in different poses are

available, the performance of recognising a face image in an unseen pose will be better. Several experiments conducted in the literature have supported this observation. For instance in [47], Eigenfaces and SOM+CN approaches both performed better when 5 gallery images per person were available than when only 1 gallery image was available. The performance increase of Eigenfaces was from 61.4% to 89.5% and that of SOM+CN was from 70.0% to 96.2%. This kind of increases is due to the capability for face recognition algorithms of tolerating small pose variations. As the number of gallery images increases, the probability that the probe pose lies closely to one of the gallery poses increases. The recognition then degrades to a real view-based matching, although the probe pose could have small difference to the gallery pose. Multiple gallery images also help various pose compensation algorithms better compensate pose variations. For instance, MQVM [87] used two gallery images of frontal view and side view in feature-based reconstruction of 3D human faces for recognition. The inclusion of additional side view gallery image provides more depth information of human face structures, and consequently results in better reconstructed models than those using single gallery images. The inclusion of multiple gallery images puts restricted requirements for data collections, because many existing face database might only contain a limited number of (even single) gallery images such as a passport photo database (single gallery images) or police mug-shot database (one frontal image and one side view image per face). Therefore, the requirement of multiple gallery images (in different poses) limits the applicability of face recognition algorithms and the most generic scenario is to recognise a probe image in an arbitrary pose from only a single gallery image in another (arbitrary) pose, which is also more challenging than multiple gallery view scenario. For the recognition of faces from a single gallery per face, interested readers are redirected to a recent survey specifically on face recognition from a single gallery image per person [21], though it

didn't emphasise pose invariance. It is often beneficial if the pose angle of the input image can be estimated before recognition such as in Modular PCA [61] and Eigen Light Field [33]. Head pose can be estimated either simultaneously in the process of recognition (as done in 3D Morphable Model [14] and Cylindrical 3D Pose Recovery [26]) or separately in an independent process. The latter alternative has been recently reviewed in [59].

As many pose-invariant face recognition approaches have been proposed recently, the need of evaluating different algorithms on a fair basis increased. A number of face image databases have been established for the purpose to compare performances of different face recognition algorithms across pose. Currently, the most widely used database for face recognition across pose are FERET database [62] and CMU-PIE database [70]. FERET database contains about 200 faces with 9 pose variations within $\pm 40^\circ$ in yaw. Specifically, the poses are -37.9° (labelled as "bi"), -26.5° ("bh"), -16.3° ("bg"), -7.1° ("bf"), -1.1° ("ba"), 11.2° ("be"), 18.9° ("bd"), 27.4° ("bc"), 38.9° ("bb") in yaw, which were estimated using 3DMM [14]. CMU-PIE database contains 68 faces with 13 different poses. MQVM [87] has calculated the pose angles using the coordinate information provided with the database, which were -62° yaw & 1° tilt (labelled as "22"), -44° yaw & 11° tilt ("25"), -44° yaw ("02"), -31° yaw ("37"), -16° yaw ("05"), 0° yaw & -13° tilt ("07"), 0° yaw ("27"), 0° yaw & 13° tilt ("09"), 17° yaw ("29"), 32° yaw ("11"), 47° yaw ("14"), 47° yaw & 11° tilt ("31"), 66° yaw ("34"). Compared to FERET database, CMU-PIE database which was established more recently contains larger pose variations and vertical in-depth rotations, but fewer faces. The performances of face recognition algorithms reviewed in this paper on FERET database and CMU-PIE database are summarised in Table 2 and Table 3, respectively. On these two databases, different algorithms are able to be compared on a relatively fair basis and one can easily pick up the algorithms with the good performances. However, the

direct performance comparison of face recognition across pose is not to be solved by this survey, because no algorithm can satisfactorily handle pose variations in face recognition as Table 2 and Table 3 suggested. For instance, the highest recognition performance on CMU-PIE database covering all 13 poses is around 70-80%, which is still far below the requirement of practical use. This paper mainly focuses on the discussions of different methodologies for face recognition across pose, in hope of providing helpful technical insights and promising directions to interested researchers.

Table 2 Experiments and performances of face recognition algorithms across pose on FERET database. Pose angles are approximate (see text for the accurate angles). The citations indicate the papers reporting the results.

No. of faces	Pose variations	Gallery / probe	Approach	Accuracy
100	9 poses within $\pm 40^\circ$ in yaw	1 random / 8 remaining	Eignfaces [33]	39.4%
			ELF [33]	75%
200	7 poses: $0^\circ \pm 15^\circ \pm 25^\circ \pm 45^\circ$ in yaw	1 random / 6 remaining	KPDA [69]	44.32%
		4 random / 3 remaining	KPDA [69]	94.46%
100	7 poses: $0^\circ, \pm 22.5^\circ \pm 67.5^\circ \pm 90^\circ$ in yaw	1 (0°) / 2 ($\pm 22.5^\circ$) 2 ($\pm 67.5^\circ$) 2 ($\pm 90^\circ$)	TFA [63]	100% 99% 92%
194	10 poses within $\pm 40^\circ$ in yaw	1 (0°) / 9 remaining	3DMM [14]	95.8%

Table 3 Experiments and performances of face recognition algorithms across pose on CMU-PIE database. Pose angles are approximate (see text for the accurate angles). The citations indicate the papers reporting the results.

No. of faces	Pose variations	Gallery / probe	Approach	Accuracy
34	13 poses within $\pm 66^\circ$ in yaw and $\pm 15^\circ$ in tilt	1 random / 12 remaining	Eigenfaces [33]	16.6%
			ELF [33]	66.3%
	9 poses: $0^\circ \pm 15^\circ \pm 30^\circ \pm 45^\circ \pm 60^\circ$ in yaw	1 (0°) / 8 remaining	Eigenfaces [55]	20%
			Probabilistic geometry assisted FR [55]	86%
	5 poses $0^\circ, \pm 16^\circ \pm 62^\circ$ in yaw	1 (0°) / 2 ($\pm 16^\circ$) 2 ($\pm 62^\circ$)	TFA [63]	100% 91%
40	2 pose: 0° and 15° in yaw	1 (0°) / 1 (15°)	Eigenfaces [85]	37.5%
			Automatic texture synthesis [85]	97.5%
68	5 poses: $0^\circ, \pm 15^\circ, \pm 30^\circ$ in yaw	1 (0°) / 4 remaining	Eigenfaces [32]	51.5%
			ELF [32]	87.5%
			3D-MM [32]	95.75%
			PDM [32]	97.42%
	5 poses: $0^\circ, \pm 30^\circ, \pm 60^\circ$ in yaw	3 ($0^\circ, \pm 30^\circ$) / 5 ($0^\circ, \pm 30^\circ, \pm 60^\circ$)	Mosaicing [71]	96.88%
	13 poses within $\pm 66^\circ$ in yaw and $\pm 15^\circ$ in tilt	1 random / 12 remaining	Stereo matching [18]	73.5%
		2 ($0^\circ, 66^\circ$) / 11 remaining	Eigenfaces [87]	40.64%
			LBP [87]	74.27%
			MQVM [87]	93.45%
	13 poses within $\pm 66^\circ$ in yaw and $\pm 15^\circ$ in tilt \times 21 lighting	1 (0°) / 12 remaining \times 21 lighting	Eigenfaces [38]	26.3%
			Fisherfaces [38]	25.7%
			Jiang's method [38]	46.66%
	3 poses: $0^\circ 15^\circ 60^\circ$ in yaw \times 22 lighting	1 random / remaining	3DMM [14]	92.1%
	7 poses: $0^\circ, \pm 22.5^\circ, \pm 45^\circ$ in yaw, $\pm 15^\circ$ in tilt	1 (0°) / 6 remaining	Local linear regression [19]	94.6%

For real data collection, more pose variations require more cameras to be installed in various locations and more complicated calibration and timing. An alternate is to use synthetic images rendered from 3D face database such as USF-3D database, which enables the experiment to have thousands of different poses. USF-3D database contains 136 3D face scans with facial textures, which can be rotated to render as many poses as the experiment requires. Other face database also contain various face images in different poses, such as ORL database (available at <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>), MIT database (not publicly

available), Asian database (available at <http://nova.postech.ac.kr/special/imdb/imdb.html>), Bern University database (no longer available), MVU database (not publicly available), Yale B database (available at <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>), XM2VTS database (available at <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>, payment required), etc. Compared to FERET and CMU-PIE, these databases contain either fewer faces or smaller pose variations. ORL database contains 10 poses within $\pm 20^\circ$ per face of 40 faces, Bern University database contains 5 poses within $\pm 20^\circ$ per face of 30 faces, WVU database contains 7 poses within $\pm 60^\circ$ per face of 40 faces, Asian database contains 5 poses within $\pm 25^\circ$ per face of 46 faces, MIT database contains 10 poses within $\pm 40^\circ$ per face of 62 faces, etc. Reported experiments and performances of face algorithms using these above databases are summarised in Table 4. Although there are a number of face databases containing pose variations, it is always helpful to establish new face databases for face recognition across pose. Recent face recognition researches are starting to solve recognition of faces in an extremely large database, a database containing over thousands of faces is then desirable for the experiments for face recognition across pose targeting on that problem.

Table 4 Experiments and performances of face recognition algorithms across pose on USF-3D database, ORL database, Bern University database, XM2VTS database, MIT database, Asian database, MVU database, etc. Pose angles are approximate. The citations indicate the papers reporting the results.

Database	No. of faces	Pose variations	Gallery / probe	Approach	Result
ORL	40	10 random poses within $\pm 20^\circ$ in yaw and tilt	5 random / 5 remaining	Eigenfaces [47]	89.5%
				SOM+CN [47]	96.2%
				PDBNN [53]	96.0%
			1 random / 5 remaining random	Eigenfaces [47]	61.4%
				SOM+CN [47]	70.0%
Bern Univ.	30	5 poses: 0° , $\pm 20^\circ$ in yaw and tilt	1 (0°) / 4 remaining	Eigenfaces [27]	65.12%
				LEM [27]	72.09%
				DCP [28]	68.61%
				Cylindrical 3D pose recovery [26]	80%
XM2VTS	125	5 poses: $0^\circ \pm 30^\circ$ in yaw and tilt	1 (0°) / 4 remaining	Fisherfaces [42]	46%
				Expert fusion [42]	70%
	100	3 poses: 0° , $\pm 90^\circ$ in yaw	1 (0°) / 2 ($\pm 90^\circ$)	TFA [63]	91%
USF-3D (Synthetic images)	50	2025 poses within $\pm 40^\circ$ in yaw and $\pm 12^\circ$ in tilt.	1 (0°) / 2024 remaining	KPCA [50]	43.3%
				GDA [50]	36.0%
				Correlation filter [50]	79.7%
	100	2 poses: 0° , 24° in yaw	1 (0°) / 1 (24°)	Linear shape models [76]	100%
WVU	40	7 poses: 0° , $\pm 20^\circ$, $\pm 40^\circ$, $\pm 60^\circ$ in yaw	3 (0° , $\pm 40^\circ$) / 7 (0° , $\pm 20^\circ$, $\pm 40^\circ$, $\pm 60^\circ$)	Mosaicing [71]	97.84%
MIT	62	10 poses within $\pm 40^\circ$ in yaw and $\pm 20^\circ$ in tilt	1 (15°) / 9 remaining	Parallel deformation [10]	82.2%
Asian	46	5 poses: 0° , $\pm 15^\circ$, $\pm 25^\circ$ in yaw	1 (0°) / 4 remaining	Eigenfaces [68]	31.5%
				AAM [68]	68%

As mentioned in Section 1, all of the approaches for face recognition across pose reviewed were classified into three broad categories depending on their different treatments to pose variations. This categorisation, however, is not unique and alternative categorisations based on other criteria are also possible. These criteria include 1) single/multiple gallery image(s), 2) whether training is required, 3) computational complexity, 4) whether the algorithm is feature-based or appearance-based, etc. Although we intend to provide insights on the problem of face recognition across pose through categorisation based on pose variation treatment, these alternative categorisations might provide other useful information for interested readers. These categorisations have been summarised in Table 5.

Table 5 Categorisation of face recognition techniques across pose based on other categorisation criteria.

Category	Approach
Criteria 1: Single/multiple gallery image(s)	
Single	PCA [43, 74, 75], Template matching [16], SOM+CN [47], LEM [27], DCP [28], Modular PCA [61], EBGM [79], LBP [2], parallel deformation [10], pose parameter manipulation [32], cylindrical 3D pose recovery [26], probabilistic geometry assisted FR [55], automatic texture synthesis [85], Jiang's method [38], 3DMM [13, 14], ICM [29, 30], stereo matching [18], ELF [33], Linear shape model [76], DCCF [50], TFA [63], LLR [19], Expert fusion [42], Composite deformable model [48].
Multiple (≥ 2)	LDA [7], real-view matching [12], panoramic view [71], MQVM [87], View-based AAM [25, 68].
Criteria 2: Whether training is required	
No training	Template matching [16], LEM [27], DCP [28], LBP [2], cylindrical 3D pose recovery [26], probabilistic geometry assisted FR [55], automatic texture synthesis [85], ICM [29, 30], stereo matching [18], Composite deformable model [48], panoramic view [71], MQVM [87].
Training required	PCA [43, 74, 75], SOM+CN [47], Modular PCA [61], EBGM [79], parallel deformation [10], pose parameter manipulation [32], Jiang's method [38], 3DMM [13, 14], ELF [33], Linear shape model [76], DCCF [50], TFA [63], LLR [19], Expert fusion [42], LDA [7], real-view matching [12], View-based AAM [25, 68].
Criteria 3: Computational complexity	
Low	PCA [43, 74, 75], LDA [7], LEM [27], DCP [28], Modular PCA [61], LBP [2], cylindrical 3D pose recovery [26], probabilistic geometry assisted FR [55], automatic texture synthesis [85].
Intermediate	Template matching [16], real-view matching [12], SOM+CN [47], DCCF [50], parallel deformation [10], ELF [33], pose parameter manipulation [32], TFA [63], LLR [19], composite deformable model [48], view-based AAM [25, 68], linear shape model [76], expert fusion [42], Jiang's method [38], MQVM [87], panoramic view [71], stereo matching [18].
High	EBGM [79], 3DMM [13, 14], ICM [29, 30].
Criteria 4: Feature-based or appearance-based	
Feature-based	PCA [43, 74, 75], LDA [7], LEM [27], DCP [28], Modular PCA [61], LBP [2], SOM+CN [47], DCCF [50], parallel deformation [10], ELF [33], TFA [63], LLR [19], EBGM [79].
Higher-order feature-based	KPCA [66], GW-KPCA [54], GW-DKPCA [80], ESBMM-KFDA [36], CFDA [82].
Appearance-based	Automatic texture synthesis [85], template matching [16], real-view matching [12], pose parameter manipulation [32], composite deformable model [48], view-based AAM [25, 68], Jiang's method [38], ICM [29, 30].
Hybrid	Cylindrical 3D pose recovery [26], probabilistic geometry assisted FR [55], expert fusion [42], panoramic view [71], MQVM [87], 3DMM [13, 14].

3 General Face Recognition Techniques and Their Sensitivities to Pose Variations

A typical face recognition problem is to visually identify a person in an input image through examining his/her face. The first attempt to this task can trace back to more than 30 years ago [41]. After that, a number of face recognition methods have been proposed, among which Principal Component Analysis (PCA, also known as Eigenfaces) [43, 75], Fisher Discriminant Analysis (FDA, also known as Fisherfaces, Linear Discriminant Analysis, or LDA in short) [7], Self Organising Map and Convolutional Network (SOM+CN) [47], template matching [16], Modular PCA [61], Line Edge Maps (LEM) [27], Elastic Bunch Graph Matching (EBGM) [79], Directional Corner Point (DCP) [28] and Local Binary Patterns (LBP) [2] are some of the representative works. All of these methods attempt to extract classification patterns (or features) from 2D face images and to recognise input face images based on these patterns against the known face images in the database.

3.1 Holistic Approaches

Kirby and Sirovich [43] used principal component analysis (PCA) to efficiently represent face images by a small number of coefficients corresponding to the most significant eigen values. Turk and Pentland [74, 75] used Eigenfaces for face detection and identification. In particular, a set of eigen vectors and eigen values were first calculated through principal component analysis to form the eigen space of human faces (or “Eigenfaces”) from a training face image set. The gallery and probe images were projected to this eigen space and their eigen values are compared in the recognition stage. The Eigenfaces approach appears to be a fast, simple, and practical method, which has become the most widely used face recognition technique. However, it does not provide invariance over changes in poses and scales. Fisherfaces approach (or Fisher

discriminant analysis, FDA) [7] was applied to expressly provide the discrimination among classes, when multiple training data per class are available. Through the training process, the ratio of between-class difference to within-class difference is to be maximised to find a base of vectors that best discriminate the classes. The between-class difference is characterised using between-class scatter matrix S_B which calculates the summed differences between class mean μ_i and overall mean μ . The within-class difference is represented as a within-class scatter matrix S_W which calculates the summed differences between individual image x_k to class mean μ_i . The generalised eigen vectors and eigen values were then computed to maximise the ratio of S_B to S_W , expressed as $S_B w_i = \lambda_i S_W w_i$, $i = 1, \dots, m$, where w_i are the m largest generalised eigen vectors and λ_i are the corresponding generalised eigen values. Using this specific projection method, the training and recognition were performed similarly to those of Eigenfaces. To overcome the problem of within-class scatter matrix being singular, the face images were first projected using PCA to reduce the dimensionality to a lower level that FDA can handle. In this case, it requires multiple gallery images per class (person) or FDA will be identical to PCA. As holistic face recognition approaches, both FDA and PCA are very sensitive to pose variations [21], because in-depth rotations of 3D human faces almost always cause misalignment of image pixels which are the only classification clues for these holistic approaches.

The attractiveness of using artificial neural network (ANN) could be due to its nonlinearity in the network. One of the first artificial neural network techniques used for face recognition is the single layer network WISARD [72], which contains a separate network for each stored individual. Lin et al. [53] used a Probabilistic Decision Based Neural Network (PDBNN) which also used one network for one face and required multiple gallery images per person in training

the network. Lawrence et al. [47] proposed a hybrid neural network, which combined local image sampling, a self-organising map (SOM), and a convolutional network (CN). In this approach, the SOM was used for dimension reduction which maps a high dimensional sub-image space (e.g., $5 \times 5 = 25$) to a lower dimensional discrete space represented by nodes (e.g., 3 dimensional space with 5 nodes per dimension). Each 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 found as the closest match. The BMU and the nodes around it are adjusted towards the training data controlled by a neighbourhood function. During iteration, the neighbourhood function will reduce its size gradually to zero when the iteration time goes towards infinity. In the feature detection and classification stage, a convolutional network has been applied which contains iterative convolution and down-sampling layers. Each convolutional layer containing multiple planes is formed by convolving a fixed kernel with the previous layer. Then the layer is down-sampled by neighbour averaging. The planes of the final layer have only one element, which indicates the classification results. In general, however, neural network approaches encounter problems when the number of classes (i.e., individuals) increases. For pose-invariant face recognition, one individual may require several classes in different poses.

Edge information of faces can also be used for face recognition. A Line Edge Map (LEM) [27] approach was proposed, which gives 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 sequentially 1) extracting edges, 2) thinning, and 3) polygonal line fitting. To measure the similarity between two LEMs, a line segment Hausdorff distance was introduced, which computes two line segments' distance as root-sum-square of three distance components, i.e., parallel distance, orientation distance, and perpendicular distance. Then the typical

Hausdorff distance on point sets was extended to LEMs based on the defined individual line segment distance. In recognition, each face image was first converted to an LEM, followed by matching probe LEMs against gallery LEMs using the line segment Hausdorff distance. A face feature descriptor, namely Directional Corner Point (DCP) [28], was proposed which is extracted by detecting image corner points which are not necessarily facial components. A DCP is represented by its Cartesian coordinates and two directional attributes pointing to the point's anterior and posterior neighbouring corner points. The distance of two DCPs is measured by calculating the warping cost through translation, rotation and opening/closing operations and averaging the minimum warping costs as the dissimilarity score. Face image retrieval using DCPs is generally economical for storage and robust to illumination changes. Its robustness to illumination changes is inherited from edge maps, because a corner point can be considered as the "edge of edges". Both LEM and DCP are, however, 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 all of the above methods, the face recognition decisions are made considering the entire face images, which can be classified as holistic approaches. In contrast, local approaches only or 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 provides an early attempt to recognise faces by considering local regions represented in templates, which compares input images pixel-wisely against a template (usually from a gallery image) using a suitable metric such as the Euclidean distance. Bruneli and Poggio [16] 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 normalised cross correlation. The recognition decision was made using summed matching scores. One problem of template matching lies in the description of these templates. Since the recognition system has to be tolerant to certain discrepancies between gallery and probe images, this tolerance might average out the differences that make individual faces unique. Pentland et al. [61] 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 (e.g., eyes, nose, and mouth) to form “Eigenfeatures” (Figure 1). 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 sub-regions 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 effect of facial feature misalignment under pose variations, at the price of neglecting some useful image patterns such as freckles, birthmarks, and wrinkles which can be considered in holistic approaches. As MPCA relies on the predefined facial components (or facial features), the feature detection is crucial to this approach similar to other feature-based face recognition methods. In experiments, it didn’t provide any test on face recognition across pose due to the difficulty of automatically detecting facial components under rotated face images. Similarly, other holistic recognition methods can also become modular, such as modular FDA, with similar gains and losses. The local feature extraction approaches can only alleviate pose variations in certain extent, because in local regions,

image distortions brought by pose variations still exist. The benefit of localising the image matching is also at the cost of extra requirement of feature detections.

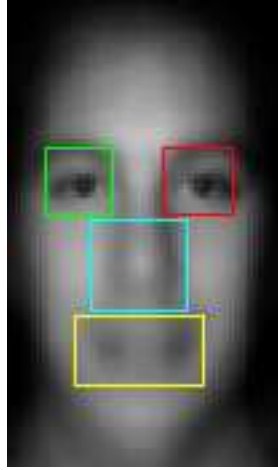


Figure 1 The Modular PCA builds multiple eigen spaces (eigenfeatures) in the regions of facial components (e.g., eyes, nose, and mouth) to achieve pose tolerances [61].

One successful local face recognition method is Elastic Bunch Graph Matching (EBGM) [79], in which human faces were described using Gabor wavelets in facial components (e.g., eyes, nose, and mouth) and an extended Dynamic Link Architecture (DLA) [44] for graph matching. In feature extraction, a Gabor jet on a point of a face image was introduced as a set of 40 Gabor wavelet coefficients obtained by convoluting 40 Gabor kernels with the local region around the point. The jet similarity measurement of two Gabor jets was defined by multiplying the magnitudes of the Gabor coefficients. These Gabor features were used for both facial component locating and recognition. In recognition, the Gabor features were extracted on facial components and the gallery and probe images was compared by calculating the similarity of the two sets of Gabor jets. Despite of the expensive computation, EBGM outperformed holistic approaches on the testing sets containing in-depth pose variations, which is largely due to Gabor features' robustness against image distortion and scaling [49]. In [69], elastic graph matching is extended and modified to apply a further Fourier transform on Gabor wavelet coefficients to be used as

features and to perform classifications using kernel-based projection discriminative analysis (KPDA) to achieve pose and expression tolerance.

Ahonen et al. [2] applied Local Binary Patterns (LBP) [60], a successful texture descriptor, to the task of face recognition. The local pattern is extracted by binarising the gradients of centre point to its 8 neighbouring points pixel-wisely and this binary pattern is used as image features for classification. Then the face image is divided into several sub-regions (or patches) and within 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 mainly 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 (or existence) of the pattern in a region. In our experiments, it is found that LBP can tolerate small pose variations and achieve perfect recognition rates when the rotations are less than 15 degrees. When the rotation becomes larger, however, the dividing face images into regions becomes problematic, because of the misalignment of image regions (e.g., a face region in a frontal image could become background in a 45° rotated image). A feature-based dividing could alleviate this effect, given an accurate feature detection result.

3.3 Discussions

In this section, some representative methods of face recognition have been reviewed with attentions on their performances under pose variations. The methodologies in each stage of face recognition are summarised in Table 6. More complete reviews on general face recognition

algorithms can be found in [1, 21, 89]. In methodology level, local approaches such as EBGM and LBP are more robust to pose variations than holistic approaches such as PCA and LDA. This is because local approaches are relatively less dependent on pixel-wised correspondence between gallery and probe images, which is adversely affected by pose variations. Their tolerance to pose variations is, however, limited to small in-depth rotations. Under intermediate or large pose variations, pose compensation or specific pose-invariant feature extraction are necessary and beneficial. The performances of local-region-based methods, e.g. template matching and Modular PCA, depend largely on the accuracy of facial component locating, which is also problematic on pose variant face images. These methods are not entirely robust to pose variations, because distortions exist in local image regions under pose variations.

Table 6 The methodologies of the general face recognition algorithms in each stage of face recognition.

FR Stage Approach	Region-based representation	Feature extraction	Dimension reduction	Classification
Eigenfaces [74, 75]	Holistic	Pixel intensity	Principal component analysis	Nearest neighbour
Fisherfaces [7]	Holistic	Pixel intensity	Linear discriminative analysis	Nearest neighbour
SOM+CN [47]	Evenly distributed image patches	Pixel intensity	Self organising map	Convolutional network
LEM [27]	Holistic	Line edge map	Line segment Hausdorff distance	Nearest neighbour
DCP [28]	Holistic	Local directional corner points	Minimum warping cost	Nearest neighbour
Template matching [16]	Patches around eyes, nose, and mouth	Pixel intensity	None	Normalised correlation
Modular PCA [61]	Regions around eyes, nose, and mouth	Pixel intensity	Principal component analysis	Nearest neighbour
EBGM [79]	Regions around 31 facial component points	Gabor wavelet	Normalised correlation	Averaging
LBP [2]	Evenly distributed image patches	Local gradient binary codes	Histogram	Weighted Chi square

For the evaluations of these general face recognition algorithms, the experiments mainly focused on recognition of frontal or near-frontal face images and few reports have conducted thorough experimentations on face recognition across pose. Most experiments of holistic approaches such as Eigenfaces, SOM+CN, LEM are limited to 20 degree rotations, where Eigenfaces yielded about 63% accuracy, SOM+CN's and LEM's performances are above 70%. These results show that small in-depth rotation affect the performances of the holistic face recognition algorithms adversely. Local algorithms were tested on datasets containing much larger pose variations, e.g., EBGM was tested on 68° and 90° rotated views in [79] and KPDA was tested on $\pm 45^\circ$ rotated views (mixed with other smaller rotated views) in [69]. However, their recognition rates are below 50%, which is far from the practical requirements. Table 7 summarises the advantages and disadvantages of these face recognition algorithms in terms of their pose tolerance. In the next two sections, the face recognition approaches explicitly handling pose variations are reviewed. Section 4 discusses 2D techniques that compensate pose variations while 3D methods are reviewed in Section 5.

Table 7 The advantages and disadvantages of the representative general face recognition algorithms to face recognition across pose.

Approach	Advantages	Disadvantages
Eigenfaces [74, 75]	Simple, fast	Sensitive to pixel misalignment, cannot separate image variances caused by identity and pose variation
Fisherfaces [7]	Maximising the separability of different identities	Sensitive to pixel misalignment, linear classes cannot adequately describe pose variations
SOM+CN [47]	Fast, tolerance to pixel misalignment due to quantisation	Linear mapping cannot adequately describe pose variations
LEM [27]	Simple, no training and facial component detection required	Sensitive to edge distortions caused by pose variation
DCP [28]	Fast, no training and facial component detection required	Sensitive to edge distortions caused by pose variation
Template matching [16]	Simple, Local regions around facial components provide some tolerance to pose variations	Sensitive to pixel misalignment in sub-image regions, dependent on facial component detection
Modular PCA [61]	Simple, fast, local regions around facial components provide some tolerance to pose variations	Sensitive to pixel misalignment in sub-image regions, dependent on facial component detection
EBGM [79]	Local regions around facial components and Gabor wavelet provide pose tolerance	Slow, distortions within local regions were not treated
LBP [2]	Simple, histogram in local regions tolerates pixel misalignment	Image dividing is problematic when pose variation is large

4 2D Techniques for Face Recognition across Pose

Due to the observation that most of the general face recognition approaches are sensitive to pose variations [21], a number of approaches have been proposed to explicitly handle pose variations. 2D techniques [10, 19, 25, 42, 71, 88] and 3D methods [11, 13, 62, 63] were used to handle or predict the appearance variations of human faces brought by changing poses. In this section, 2D techniques are classified into three groups, i.e., 1) pose-tolerant feature extraction [36, 50, 61], 2) real view-based matching [12, 71], and 3) 2D pose transformation [10, 25, 32, 39]. Approaches based on pose-tolerant feature extraction attempt to find face classifiers or pre-processing of linear/non-linear mapping in the image space that can tolerate pose variations. Real view-based matching captures and stores multiple (usually a large number of) real views to cover

exhaustively all possible poses for face recogniser. Because most of the face recognisers as reviewed in Section 3 are robust to small pose variations ($\sim 15^\circ$), a certain level of quantisation on the in-depth rotations is possible which can significantly reduce the number of real views. In case there are only a limited number of real views (or even only a single view) per person stored in the database in which real view-based matching is not possible, approaches using 2D pose transformation alter the appearances of the known face images to the unknown poses to synthesise virtual views to help the face recogniser to perform recognition across pose.

4.1 Real View-Based Matching

Despite of tolerating pose variations, one can actively compensate pose variations by providing gallery views in rotation to recognise rotated probe views. The natural way to realise a face recognition system against pose variations in this direction is to prepare multiple real view templates for every known individual. Because general face recognition algorithms as previously reviewed are able to tolerate small pose variations (e.g., 15 degree rotation), the number of required real gallery images can be significantly reduced by quantisation on the in-depth rotations. Beymer [12] designed a real view-based (RVB) face recognition system using a template matching of image-based single-view representation. Each input view was geometrically registered to the known person's templates by using locations of eyes and nose, which were automatically located by that system. The recogniser acquires 15 gallery face images to cover a range of pose variations with approximately $\pm 40^\circ$ in yaw and $\pm 20^\circ$ in tilt as shown in Figure 2. 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-centred probe face image with gallery face images in similar poses.

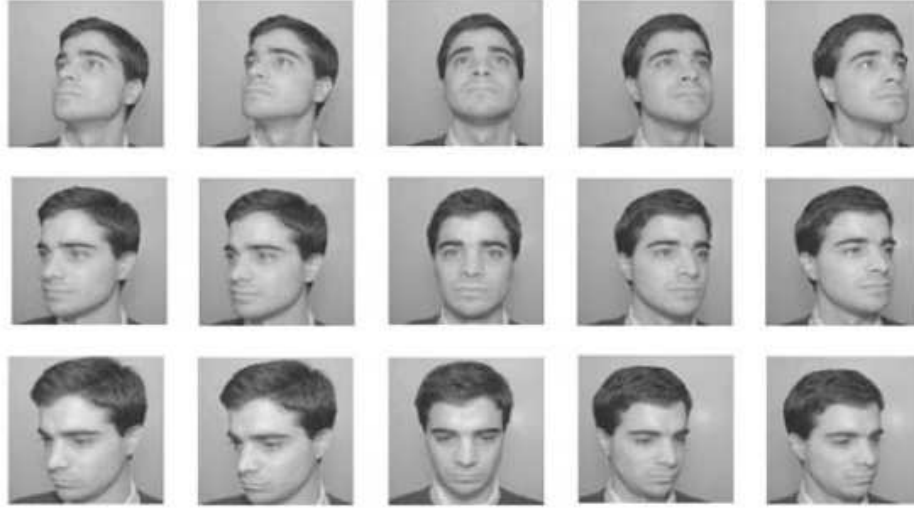


Figure 2 The view-based recogniser using real views stores a certain number of face images taken in different poses of the same person. [12]

Singh et al. [71] proposed a mosaicing scheme (MS) to form a panoramic view as shown in Figure 3 from multiple gallery images to cover the possible appearances under all horizontal in-depth rotations. The panoramic (namely composite) 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 different 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, which 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 the splined images were expanded and summed together to form the final composite face mosaic. In recognition, the synthesised 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 2v-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. The proposed face

mosaicing method, however, doesn't actively compensate pose variations and the recognition improvements are mainly contributed by 1) the use of multiple gallery images in different poses, and 2) the pose-invariance of the face recognition algorithm. In experiment, it was found that the optimal combination of gallery images is frontal image plus left and right views in 40 degree rotations. The main reason is that face recognition algorithms can normally tolerate small horizontal rotations and the input face images were matched against the part of the face mosaics for the nearest viewpoint.

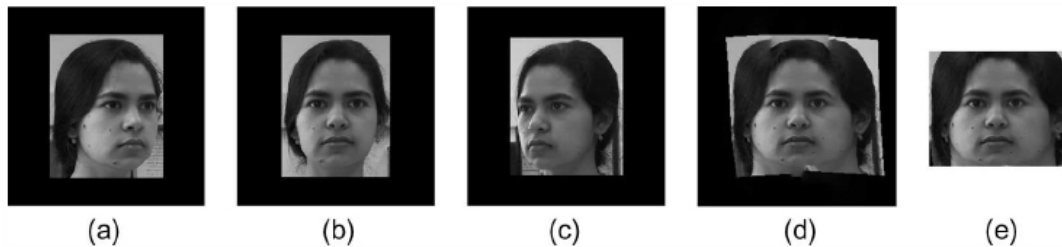


Figure 3 The process of face image mosaicing [71]. (a-c) three raw images in different poses, -20°, 0°, 20° in yaw, (d) panoramic view mosaiced from these three images, and (e) cropped panoramic image used in recognition experiment.

In general, face recognition methods of real-view based matching require multiple real views of each person as gallery. Either the raw gallery images or some transformations of them are considered in recognition to cover possible pose variations. These face recognition algorithms then rely on the capability of general (non-frontal) face recogniser in small pose tolerance to match the probe views in arbitrary poses exhaustively against all gallery images or transformed images, in hope that the closest appearance match belong to the same identity.

4.2 Pose Transformation in image space

As it is generally impractical or unfavourable to collect multiple images in different poses for real view-based matching, a feasible alternate is to synthesise virtual views to substitute the demand of real views from a limited number of known views (even from a single view). The

virtual view synthesis can be undertaken in 2D space as pose transformation or in 3D space as 3D face reconstruction and projection. The virtual view synthesis involving 3D models will be discussed in the next section, while various 2D pose transformation methods are discussed in this subsection, which include parallel deformation [10], pose parameter manipulation [32], and active appearance models [25, 39]. Besides, 2D pose transformation was performed on a model database containing image under different poses without virtual view synthesis in [58].

Beymer and Poggio [10] are probably the first researchers to specifically handling pose variations in face recognition. They proposed parallel deformation to generate virtual views covering a set of possible poses from a single example view using feature-based 2D warping [6]. A 2D non-rigid transformation on a prototype face from the real view in a standard pose to the real view in a target pose was recorded. To synthesise a virtual view of a gallery face (the face in the database to be matched against) in the same target pose, the real view in the standard pose was parallel deformed based on the recorded 2D transformation on the prototype face. Figure 4 shows a diagram of the process of parallel deformation, which synthesise a virtual image in the target pose from 3 real images, i.e., an image in the standard pose of the gallery face, images in the standard pose and the target pose of the prototype face. Figure 4(a) and 4(b) are the prototype face's two real views in different poses. A pixel-wise correspondence and a pose deformation path (Figure 4e) are recorded by applying a gradient-based optical flow on the two prototypical views. This pose deformation is then further deformed (referred as identity deformation) to the gallery face based on the differences of the face images in the standard pose between the gallery face and the prototype face, which are achieved by a manual feature-based 2D warping or an automatic face vectorisation. Applying the deformed deformation on the gallery image in the standard pose, a novel image in the target pose can be synthesised by directly taking the raw pixel

intensity from the gallery image as the textures of the novel image. In this process, the recorded non-rigid 2D transformation serves as the prior knowledge of the class of human faces, which provides reasonable predictions of the possible face appearances for rotated faces. 8 virtual views were synthesised per person from an example view in about 15 degree rotation away from the standard pose and 6 virtual views were synthesised by mirroring the corresponding face view with respect to the vertical axis using face symmetry information, covering -30 to 30 degree rotations in yaw and -15 to 15 degree rotation in tilt. Tested on a dataset containing 5 in-plane rotated views and 5 in-depth rotated views per person of 62 people, the proposed parallel deformation achieved an accuracy of 82.2% using manually labelled interpersonal correspondences. When using the automatic face vectorisation [11], the recognition rate was 75%.

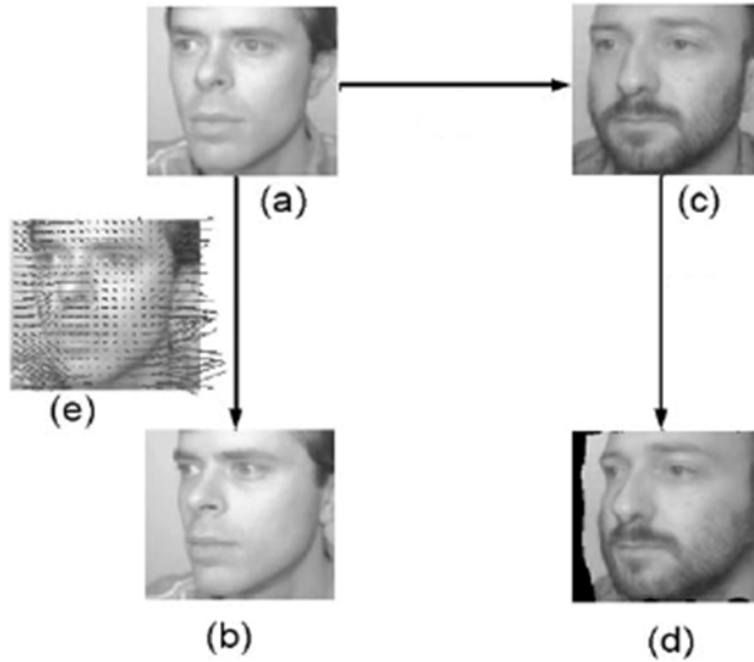


Figure 4 The process of parallel deformation. (a) the prototypical image in standard pose, (b) the prototypical image in target pose, (c) the gallery image at standard pose, (d) the synthesised novel image at target pose, and (e) the recorded deformation in prototype from the standard view to the virtual view $y_{p,r-v}^p$ [10]

Active Shape Model (ASM) originally proposed by Cootes et al. [23] is one of the most successful approaches in automatic face image representation [46], structure locating in medical images [22] and face recognition [25]. In ASM, Principal Component Analysis (PCA) was applied on the locations of facial components (e.g., facial contours, eyes, eyebrows, lips, etc.) presented as connected point distributions from a variety of manually labelled images (i.e., images with facial components marked on), 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, e.g., a face shape where the mouth location is between those of eyes and nose. To automatically adjust the point distribution to the new face image, a local searching strategy is applied on each point. First, a gradient-based local profile on the point is extracted along the local line segment perpendicular to the boundary of the point. Based on the training set, an average profile is calculated which captures the local texture variations around the point. This profile, in adjustment step, is used to find the location of that point in the new image whose local profile best fits this reference profile. To ensure the adjustment always follows the correct (or valid) path, the adjusted point distribution is then projected onto the previously trained eigen space. Those parameters whose values are larger than 3σ are set to 3σ which limits the deformation of the point distributions within the valid range of the assumed Gaussian distributions of the prior shape knowledge.

González-Jiménez and Alba-Castro [32] applied the concept of ASM with manual facial component locating to synthesise virtual views in different poses in their proposed Point Distribution Model (PDM). PCA was applied on the locations of facial components (e.g., facial contours, eyes, eyebrows, lips, etc.) presented as PDMs and they argued that the second

significant parameter is the “pose parameter” which controls the left-right rotations of faces. To build this eigen space, a variety of manually labelled images (i.e., images with facial components marked on) are used as the training set. Finding the top principal components using PCA doesn’t guarantee these parameters are specifically pose-related free from other variations, because these variations are mutually dependent in 2D face image space. Taking pose and expression as an example, some kind of image variations such as the movement of mouth corners can be explained by either pose variations (head tilting) or expression changes (smiling). To make the most principal components more specific to pose variations, the training set is intentionally chosen to include much more pose variations than other image variations. The 2D transformation was then achieved by only altering the pose parameter, leaving other personal information intact as shown in Figure 5. A probe image is labelled with a probe point distribution map (Figure 5a) and the gallery image is also labelled with a gallery point distribution map (Figure 5b). Both distribution maps are projected on to the point distribution model (previously trained eigen space) and the pose parameter of the gallery map is substituted by that from the probe map. Then the synthetic point distribution map is recovered based on these parameters (pose parameter is from probe while other parameters are all from the gallery), so that the synthesised mesh preserves all image information (e.g., identity) from the gallery face except pose, which is from the probe face. In this way, the pose variations are compensated and face recognition is performed using the typical EBGM recogniser (i.e., Gabor wavelet + normalised correlation) on two images in the same pose (e.g., probe). Tested on the CMU-PIE database with different 13 poses per face of 68 faces, their method achieved much higher recognition rates than [10]. For example, to recognise 15 degree rotated views and 30 degree rotated views from frontal gallery views, [32] achieved accuracies of 99.26% and 95.59% respectively. When the rotation angle increases, however, the

recognition rates drop to 67.5% (45 degree rotation) and about 20% (65 degree rotation). The performance of the proposed method on face views under tilt (vertical rotation) hasn't been reported, probably due to lack of training images in tilt from the CMU PIE database (the CMU-PIE database contains three different tilt rotations, i.e., ± 10 and 0 degrees).

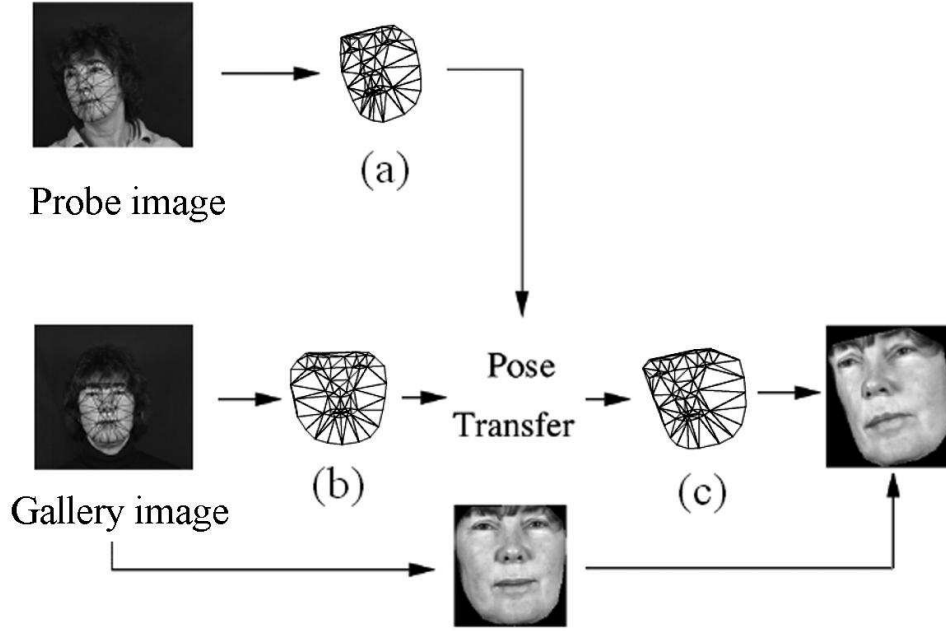


Figure 5 The process of generating virtual face views from training and input images by altering pose parameters and performing 2D image warping. The pose parameter was extracted from the mesh (shape) of the training image (b) and then replaced by the extracted pose parameter from input image (a) to form a mesh that has the same pose with input but the same identity information with training (c). [32]

As an extension of ASM, Active Appearance Models (AAM) [24] has been proposed to simultaneously model the variations of shape represented by point distributions and textures represented by pixel intensities. The shape variations were obtained in the same manner of ASM, using PCA on a training set of point distributions. For texture variations, each image in the training set was warped to a uniform shape (average point distribution) and the pixel intensities were then analysed 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 two eigen

spaces, a vector of similarity transformation parameters controlling shape transformations, and a vector of scaling and offset transformation parameters controlling texture transformations. In searching, all of the model, shape, and texture parameters were iteratively altered towards minimising the differences of the reconstructed face image and the real face image. Because the AAM is based on 2D image transformation, the in-depth rotation (pose) cannot be decoupled from the identity changes (face shape difference). To further explicitly model in-depth pose variations, a view-based AAM was proposed in [25] and was applied to face recognition across pose in [67]. In view-based AAM, the model parameter \vec{c} is approximated by a sum of triangular functions of rotation angle θ as $\vec{c} = \vec{c}_0 + \vec{c}_1 \cos \theta + \vec{c}_2 \sin \theta$, where $(\vec{c}_0, \vec{c}_1, \vec{c}_2)$ were learned using regression by estimating \vec{c} and giving θ in at least three different poses in the training set. Estimating pose from a new image is then performed by calculating the rotation angle θ from the estimated model parameter \vec{c} of the input image and $(\vec{c}_0, \vec{c}_1, \vec{c}_2)$ of the training set. In this process, the inter-person differences contained in \vec{c} are discarded and only the pose-related differences are modelled. To synthesise virtual views in a new pose (e.g., frontal view) from an input image in a certain pose (e.g., rotated view), the model parameter \vec{c} was first estimated from the input image and the closest matching image in the training data was found by minimising the difference of the two model parameters. Then the input model parameter was projected to the AAM of the closest match. The residual of the model parameters is retained to record the identity-related difference and the pose-related difference is altered by changing θ to a new value. This process is similar to that of the parallel deformation where the closest match is served as prior knowledge of the pose transformation, while the difference lies in that the choice of the reference face is unique in view-based AAM and arbitrary in parallel deformation. In [67], frontal virtual views were synthesised using this process from a single non-frontal face image

(ranging within $\pm 25^\circ$) based on a view-based AAM trained on three images per face (0° , $\pm 15^\circ$) of 40 faces. Then an adaptive PCA was applied on the synthetic frontal face images for recognition. On a face image set of 46 faces with 4 poses per face ($\pm 15^\circ$ and $\pm 25^\circ$), the recognition achieved 63% identification accuracy which is higher than the direct matching of rotated face with the frontal gallery image.

Vetter [76] further extended the concept of AAM by replacing the sparse point distributions with a pixel-wise correspondence between two images in different poses using optical flow. It differs from the typical AAM in two aspects. First, 2D shape information is represented by a dense point distributions and the dimension is comparable to that of a face image. Second, different linear shape models are learned distinctively in different poses where two models share the same set of model parameters thanks to the assistance of a set of 3D scans. Specifically, 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 synthesise new shape in that pose. The texture mapping is similar to AAM, except that the model parameters of texture model are independent to shape model. In experiment, the face recognition system using synthesised face images in target pose achieved 100% accuracy on 100 synthetic faces with 2 poses per face (24° in yaw as gallery and 0° as probe). Because the 3D face models were only

involved in model establishment and the main steps were in 2D image space free from 3D data, this approach was classified as a 2D technique and discussed in this section.

Gross et al. [33] proposed Eigen Light-Field (ELF) method to extend the capability of Vetter's method of handling multiple gallery images in different poses per face in face recognition across pose. They unified all possible appearances of faces in different poses within a framework of light field, which is in a 4D space (two viewing directions and two pixel positions). Assuming human faces as convex Lambertian objects, this light-field was highly redundant and consequently the light field coefficients were associated in different poses for the same identity. In training stage, a set of face images in different poses of different identities were first warped to a uniform shape based on a set of manually located feature points (e.g., eyes and mouth), where each pixel corresponded to a unique pixel location in the light field. The pose variant images were represented by a single concatenated vector for each identity and principal component analysis was performed on those concatenated vectors from different training identities. Because of the redundancy of the light-field, face images in different poses were represented using a single set of eigen vectors and eigen values to capture the variations due to identity changes. In recognition, input images (gallery image and/or probe image) were also warped and then projected onto the established eigen space by a least square method instead of direct dot product, because the dimensionality of input images is usually smaller than that of the light field (image dimension times number of poses). The recognition was then performed by comparing the projected eigen coefficients from gallery image(s) and probe image in Euclidean distance. This algorithm was tested on CMU-PIE [70] and the FERET [62] face databases. On FERET database with 9 poses within $\pm 40^\circ$ in yaw per face of 100 faces, ELF achieved 75% identification accuracy using any pose as gallery image and the remaining 8 poses as probe. On CMU-PIE database with

13 poses within $\pm 62^\circ$ in yaw and $\pm 20^\circ$ in tilt per face of 34 faces, ELF achieved 66.3% accuracy. This method also showed capability of improving recognition accuracy when more gallery images were available. However, since ELF requires a restricted alignment of 2D image to the light field space, it actually discarded face shape variations due to different identity which was a critical feature for face recognition. In this sense, this ELF method is parallel to those methods using generic face shape for pose recovery which will be discussed in Subsection 5.1.

Kahraman et al. [39] enhanced AAM's capability in pose tolerance by enlarging the training set with synthetic pose variant face images. In pose normalisation step, they recorded displacements of all landmarks of the AAM using a reference face and the landmark's coordinate ratios between rotated view and frontal view were assumed constant when transforming from the reference face to an input face. This assumption is also similar to parallel deformation which introduces the errors from different choices of the reference faces. Synthetic images in different poses were then generated from single frontal images by moving the landmarks along the recorded displacements. A single AAM was trained on the synthetic images covering 8 different poses rather than multiple AAMs trained in different poses as done in [25]. Frontal and non-frontal images within 45 degree rotations can then be transformed mutually by altering the parameters controlling pose variations after the AAM was aligned to each image. The experiments of face recognition were conducted on Yale B database but the focus is on illumination tolerance, not pose tolerance. The proposed modified AAM method advantages the original AAM in the single training and alignment process for all poses, while it suffers from incapability to handle larger pose variations because AAM cannot be reliably aligned to rotated views with occluded landmarks.

4.3 Pose Transformation in Feature Space

Pose tolerance can also be achieved in feature space instead of the explicit image space, where the feature-space transformed data cannot be visually displayed as face images as in the image space. These transformations in feature space are designed either to general image variations (e.g., kernel tricks) or specifically to pose variations (linear pose transformation). One possible feature space transformation for face recognition is kernel tricks which nonlinearly map face images into a higher dimensional non-linear feature space, so that the previously non-separable distributions caused by pose variations could be (better) linearly separable. This is supported by Cover's theorem [34], that nonlinearly separable patterns in an input space will become linearly separable with a high probability if the input space is transformed nonlinearly to a high-dimensional feature space. A number of kernel-based face recognisers were proposed to perform face recognition or other pattern recognition tasks, such as various Kernel PCAs [54, 66, 80] and Kernel FDA [36, 82]. In [66], 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 [54] pre-processed the facial images with Gabor wavelets and extended kernel polynomial functions to have fractional powers in Kernel PCA. Xie and Lam [80] proposed to train an eigenmask as an additional kernel function to adjust the contributions of different image pixels due to their importance (or discriminative power), e.g., pixels around eyes might be more important in face recognition than other pixels in cheeks so that they will be assigned higher weights.

Huang et al. [36] 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 Maximisation (ESBMM) algorithm. Experimental results on face

recognition across pose were reported on Yale Database B and CMU-PIE database and showed their method outperformed other algorithms such as PCA and KPCA. However, these experimental evaluations are quite limited. Yale Database B contains only small in-depth rotations within 24 degrees and 10 different faces which is not a convincing test bed for face recognition algorithms claiming to have pose handling abilities. In CMU-PIE database, the 13 face images of different poses are mixed with additional 43 images under different lighting conditions, which diluted the test's sharpness. Yang et al. [82] proposed to perform FDA on the KPCA transformed feature space and to differentiate regular and irregular features based on the singularity of the within-class scatter matrix. The regular features were performed under the standard FDA mechanism and the irregular features were treated under PCA. Then the two sets of coefficients were fused using summed normalised-distance for classification. 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 limited due to the fact that the actual nonlinear transformation forms caused by pose variations are unknown. The existing non-linear kernel functions only have random effects on face recognition across pose, i.e., it may equally possible to improve and to reduce the performance and this effect is unknown before experimental evaluations. The ideal pose transformations, if they are known, could be unlikely capable of being analytically formulated such that these transformations cannot be treated as kernels in KPCA or KFDA.

To explicitly model the pose variations, researchers proposed to train a linear transformation based on a set of images under pose variations to learn pose transformation free from identity-related image features. Kim and Kittler [42] proposed a hybrid approach, expert fusion, fusing four 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 generalised 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. The first two systems belong to this subcategory of linear pose-tolerant feature extraction; the third system belongs to non-linear pose-tolerant feature extraction (i.e., kernel-based method); and the last system is classified in 2D transformation using a 3D generic face model. Finally, these four systems are fused to form single classification decisions in Euclidean distance assuming they are mutually independent. After training on 170 people, the proposed fused experts achieved 70% accuracy on 30 degree rotated faces using single frontal views as gallery on 125 different people from XM2VTS database.

As the pose variation is a projection of 3D rigid transformation onto 2D image space, the global linear approximation is incapable of accurately describing image variations caused by pose changes, which results in unwanted distortions in certain face regions. To alleviate this problem, different localisations were proposed such as using evenly distributed patches [19] and using image regions around facial components [63]. These treatments are effective similar to local approaches of general face recognition algorithms using evenly distributed patches (e.g., SOM+CN [47], LBP [2], etc.) or using image regions around facial components (e.g., Template Matching [16], Modular PCA [61], EBGM [79], etc.). With the assistance of a generic cylindrical face model, Chai et al. [19] 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 minimise 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 a intensity averaging of overlapped pixels to form holistic frontal virtual views for recognition. On CMU PIE database with a rotation within 45 degrees, the proposed method showed superior performance over Eigen Light-field [33], achieving an average accuracy of 94.6%.

Prince et al. [63] proposed a linear statistical model, Tied Factor Analysis (TFA) model, to describe pose variations on face images and achieved state-of-the-art face recognition performances under large pose variations. The underlying assumption is that all face images of a single person in different poses can be generated from the same vector in identity space by performing identity-independent (but pose-dependent) linear transformations. From a set of training images in different known poses, the identity vectors and the parameters of the linear transformations were estimated iteratively using an EM algorithm. Figure 6(a) shows face image distributions in the observed feature space, which is either the raw space spanned by vectorised image pixels or a transformed space after simple pose-independent transformations (e.g., Gabor wavelet or radial basis functions). In this space, the locations of pose variant face images x and identity variant face images are mixed altogether, which makes the separation of identity from pose variations infeasible. To effectively separate identity from pose variations, an ideal identity feature space is proposed and shown in Figure 6(b). This identity space is spanned only by faces of different identities h , free from pose variations. The relationship between these two spaces is

under a linear transformation, which includes a multiplication F , an offset m and a Gaussian noise ε . Hence, a single point h in the identity space can be mapped to positions in observed feature space (e.g., x_1 , x_2 , and x_3) under different pose transformations (e.g., $x_1 = F_1 h + m_1 + \varepsilon_1$, etc.), which represent different face images of the same identity in different poses. Because of the inclusion of the offset and Gaussian noise, this linear transformation can better model the actual pose transformation projected to the 2D image space than the linear transformation in Expert Fusion [42]. However, the computation is more challenging due to the nonlinearity of the noise factor. The tied factor analysis approach also assumes a Gaussian distribution for the identity space and these pose-independent identity vectors were then used in face recognition through Maximum A Posteriori (MAP) mechanism by choosing the gallery image which corresponds to the maximum probability under this linear transformation scheme. This approach has advantages over applying fixed transformations before recognition, because the tied factor analysis explicitly searches transformations to achieve pose-independent feature extractions. Because the transformation was limited to linear due to computational feasibility, it could be insufficient to adequately describe pose variations which are non-linear transformations if mapped to 2D image space. In recognition experiments, the estimation of identity vectors were limited to two poses only (the gallery pose and a single probe pose) and on 100 faces from FERET database it achieved accuracies of 83% for 22.5 degrees, 59% for 67.5 degrees, and 41% for 90 degrees respectively against frontal gallery images. The performances were further improved to 100% for 22.5 degrees, 99% for 67.5 degrees, and 92% for 90 degrees, when the algorithm takes local Gabor data around manually labelled facial features instead of raw image data as input. This result is consistent to the previous reports [61, 79] that (1) Gabor features are robust to image

distortions [49], and (2) local features are more robust to pose variations than global images as discussed in the previous section.

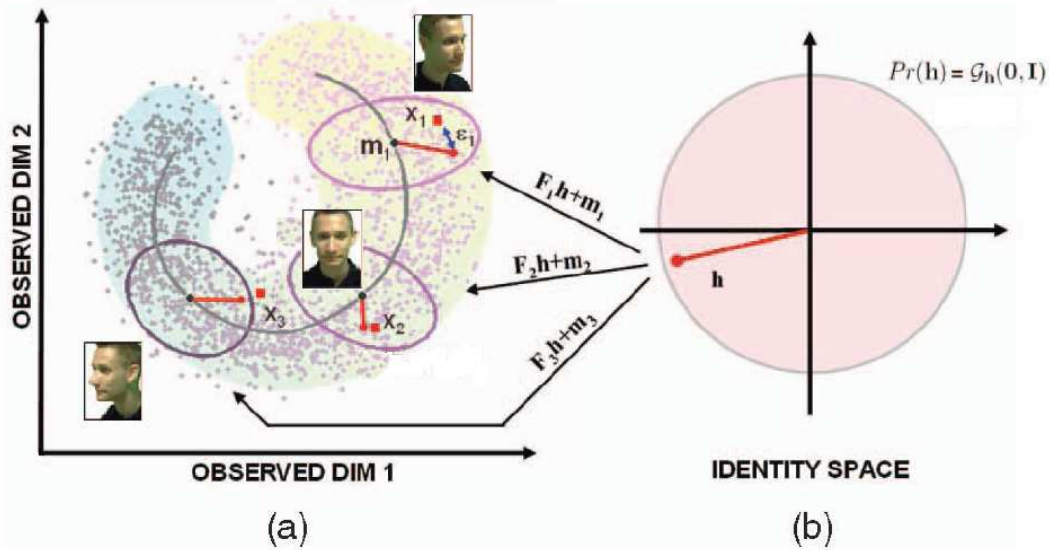


Figure 6 The explicit pose-dependent linear transformations and the pose-independent identity vector in the tied factor analysis in [63]. (a) In the observed space (e.g., image space), pose variant images of the same identity locate at different locations which results in low performance of face recognition across pose variations; (b) Under identity-independent linear 2D transformations, these images were traced back to the same vector in the identity space which represents solely the identity information free from pose variations.

Pose transformation can also be approximated in transformed space, other than in the original image space. In this sense, the TFA approach using wavelet coefficients as the input data also belongs to this subcategory. Besides, Levine and Yu [50] compared five correlation filters on face recognition in terms of their robustness to pose, illumination, and expression variations, which all perform image transformations including pose transformation approximation in Fourier transformed frequency space. The best performed correlation filter under pose variations, Distance-Classifier Correlation Filter (DCCF) [57], achieved 79% recognition rate using single gallery views on the USF 3D Human-ID database (USF-3D) containing synthetic images of 50 people in 2050 different poses within 40 degrees in yaw and 12 degrees in tilt. Similar to K-FDA,

DCCF searches an optimal correlation filter h to maximise a scatter cost function $J(h)$ in a Fourier transformed frequency space. Parallel to FDA in the image space, the cost function is defined as the ratio of a between-class scatter measure to a within-class scatter measure in the Fourier transformed space. Applied to face recognition across pose, this mechanism is equivalent to searching an optimal linear filter in frequency space which best describes the pose variations according to FDA classification criterion. This approach is nonlinear due to the involvement of Fourier transform. Generally, correlation filter-based methods are more sensitive to pose variations where nonlinear image distortion occurs than other image variations (e.g., illumination changes, expression, etc.) as shown in the experiments of [50]. Because correlation filters are linearly associated with a fixed nonlinear transform (Fourier), this observation on face recognition experiments could lead to a conclusion that pose variations cause severer nonlinearity than illumination and expression variations do, at least in Fourier transformed space.

In this subcategory, various methods have been proposed to transform the image space to a feature space where pose variation can be better tolerated, by 1) nonlinear mapping defined by various kernel functions, and 2) pose specific linear transformation in image space, Gabor coefficients, or frequency domain under Fourier transform. Kernel-based methods rely on predefined (fixed) nonlinear transform function to approximate pose variations which are not specifically adjusted to fit pose variations. Pose specific linear transformations tend to train parameters from pose variant face images, which can be used to specifically describe pose variations. However, pose variations, when projected to 2D image space, are not exactly linear transformations, because it contains a number of occlusions and warps. Consequently linear transformations are not capable to adequately describe pose variations in image spaces. One of the possible directions could be the design of a pose specific nonlinear transformation, which

will better approximate image transformations caused by pose variations, but will inevitably result in more tractability and computation challenges.

4.4 Summary and Discussions

The pose-invariant face recognition methods using 2D techniques have been classified into three groups, i.e., real view-based matching, pose transformation in image space, and pose transformation in feature space. The methodologies of these techniques have been summarised in Table 8 and the advantages/disadvantages of different methods have been summarised in Table 9. Real view-based methods are the most straight-forward techniques in handling face recognition across pose. They can make direct use of general face classifiers discussed in Section 3 to match input images with gallery images in the same rotated pose. The performances of real view-based matching methods are similar to those of frontal face recognition using general face recognition algorithms, because the only difference is its non-frontal matching. The limitation lies in that it requires a relatively large number of real images captured from all possible viewing directions, which restrains them from practical applications. Face recognition based on 2D pose transformation in image space is a successful extension of the real view-based face recognition. Instead of acquiring a large number of real images as gallery views, these techniques synthesise virtual views in possible poses from a limited number of real gallery views (often from a single gallery view) to substitute the real gallery views with help of reference face(s) as prior knowledge. The techniques used in virtual view synthesis are 2D image transformations based on pixel correspondence between the source images and the target images. Assuming image continuity in pose transformation, these techniques can effectively handle pose variations within small to median in-depth rotations usually limited to 45 degrees. However, large pose variations bring

image discontinuities in the 2D image space so that it cannot be reliably handled within 2D space. Under such circumstances, 3D approaches generally outperform 2D techniques and they are reviewed in next section (Section 5). Another issue in pose transformations is the suboptimal modelling of facial textures. Because pose variations are always associated with the changes of illumination, the same points on a face may appear differently in two face images taken from different viewpoints. Most 2D transformation methods, however, only considered shape transformation by finding the corresponding pixels between images and neglect that the pixel values may change as well. Among the pose transformation methods reviewed in this section, AAM and Vetter's linear shape model tend to actively model facial textures. The modelling of facial textures is however in a linear interpolation manner, which cannot adequately approximate the non-linear variations of reflected intensities from human face surfaces. An accurate yet computationally feasible approximation of face surface reflection may help to improve the performance of 2D transformation methods in recognising faces across pose.

Pose transformations in feature space tend 1) to implicitly improve linear separability of face images under pose variations by non-linear mapping prior to recognition, and/or 2) to explicitly model the pose transformation using linear approximations. The second strategy has the promise to find a non-linear mapping and space best suitable to pose variations, while the current research stage is primarily limited to fundamental mapping functions (e.g., radial basis functions). The question of whether there is a feature space where rotated faces are separable is still open to the research community. An answer to this question may lead to a clearer understanding of pose-invariant face recognition, similar to the findings of linear subspaces in illumination-invariant face recognition [5, 8, 64].

Table 8 The methodologies of the 2D techniques for face recognition across pose.

Approach	Pose tolerance and compensation	Face recognition algorithms
Real view matching [12]	Multiple real views	Template matching
Mosaicing [71]	Panoramic view from multiple real views in different poses	Log Gabor transform + modified C2 features + support vector machine
Parallel deformation [10]	Inter-person deformation + intra-person deformation across pose + double deforming the gallery view to generate virtual views	Template matching
PDM [32]	Facial components locating + PCA on point sets + altering the second principal component to compensate pose variations	EBGM matching
View-based AAM [25, 68]	PCA on point sets and warped image intensities + finding the closest matching model parameters to the input image + pose transformation in model parameters with a residual	Adaptive PCA
Linear shape model [76]	Dense point distributions + 2D projections of 3D linear shape model + pose transformation using 2D linear shape models in different poses	Similarity measure using 1) correlation, and 2) Euclidean distance
Eigen light-field [33]	Feature-based warping image to uniform shape + CPA on concatenated image vectors in different poses + image projection onto eigen space using least square method	Euclidean distance on projected eigen coefficients of gallery and probe images
Pose normalisation in AAM [39]	Synthesising virtual views in different poses with a reference face + enhanced AAM training + standard AAM searching	PCA, FDA
KPCA [66]	Kernel functions	Kernel principal component analysis (KPCA)
GW-KPCA [54]	Fractional power polynomial kernel functions	Gabor wavelet + KPCA
GW-DKPCA [80]	Double nonlinear mapping in kernel functions	Trained weigh mask + Gabor wavelet + KPCA
ESBMM-KFDA [36]	Adaptive kernel functions	Kernel Fisher's discriminant analysis
CFDA [82]	Kernel functions	KPCA + KFDA and PCA + fused sum-normalised distance
Expert fusion [42]	1) Linear pose transformation, 2) radial basis function, 3) 3D generic shape compensation	Fusion of 1) PCA + linear transformation + LDA 2) generalised discriminant analysis, and 3) linear discriminant analysis
DCCF [50]	Correlation filter in frequency domain	Fast Fourier transform + correlation filter + Inverse FFT + distance between peaks
LLR [19]	Evenly distributed patches as input data + linear approximation of pose variation on 2D image patches	FDA
TFA [63]	Gabor wavelets on local regions around facial components as input data + linear transformation with offset and noise factors	linear transformation + maximum a posteriori (MAP)

Table 9 The advantages and disadvantages of the face recognition algorithms using 2D techniques in face recognition across pose.

Approach	Advantages	Disadvantages
Real view matching [12]	Simple, straightforward, good performance	Need to collect a large number of gallery images per person covering all possible poses
Mosaicing [71]	Continuous pose coverage, single panoramic view required	Distortions exist, no vertical in-depth rotation (tilting)
Parallel deformation [10]	Simple, fast, sharp, single gallery image	Pose tolerance is small, the choice of reference face is arbitrary
PDM [32]	Simple, fast, single gallery image, good separation of pose and identity in statistics	Manual interference, performance largely dependent on PCA training
View-based AAM [25, 68]	Considering both shape and texture, single gallery image, intermediate pose coverage	Searching is not always reliable, the choice of reference image may introduce identity-related errors.
Linear shape model [76]	Detailed shape description, linking shape variations in different poses	Automatic correspondence is not reliable on non-feature points, many models are required to cover a range of poses
Eigen light-field [33]	Capable of handling multiple gallery images, single eigen space for different poses	Discarding shape variations by warping which could be critical features for recognition
Pose normalisation in AAM [39]	Single AAM for all poses	The choice of reference face shape is arbitrary, the pose normalisation assumption is coarse.
Kernel tricks [36, 54, 66, 80, 82]	Nonlinear transformation encapsulated in dimension reduction, simple, fast	The existing kernel functions are not specific to pose variations, the choice of kernel functions are limited.
DCCF [50]	Nonlinearity by Fourier transform, translation invariant	Correlation filter cannot adequately describe image variations caused by pose variations
Linear pose transformation in expert fusion [42]	Simple, characterising pose variations using explicit transformation	Linear transformation cannot adequately describe image variations caused by pose variations
LLR [19]	Localisation alleviates inaccuracy of linear approximation of pose transformation	Linear transformation cannot adequately approximate pose variations even in local regions, overlapping of patches may cause problem
TFA [63]	Consideration of noise factor and offset in linear pose transformation, localisation around facial components	Linear transformation cannot adequately describe image variations caused by pose variations

5 Face Recognition across Pose with Assistance of 3D Models

Recently, face recognition with assistance of 3D models becomes one of the successful approaches, especially when dealing with pose and illumination variations. 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. The 3D face models used in face recognition could be generic 3D shape models [26, 55, 85], personalised 3D scans [37, 40, 77, 83], and personalised 3D models reconstructed from 2D images [14, 18, 30, 38, 48, 86, 87]. Face recognition using personalised 3D scans belongs to 3D face recognition and is out of the scope of this review, whose focus is 2D image-based face recognition. We redirect the interested readers to the excellent reviews specifically on 3D face recognition [15, 65]. Face recognition techniques using generic shapes consider the uniform face shape as a tool for the transformation of image pixels. Personalised 3D face (shape) models can be reconstructed using feature-based (Section 5.2) or image-based techniques (Section 5.3). Feature-based 3D face reconstructions [38, 48, 87] utilise facial features (e.g., eyes, nose, mouth, and etc.) extracted from 2D images to predict the volumetric information of the input face. Image-based 3D face reconstructions [14, 18, 30] consider facial textures (e.g., pixel intensities) as critical clues and used them in reconstruction.

5.1 Generic Shape-Based Approaches

A simple and efficient pose recovery methodology (cylindrical 3-D pose recovery) 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 (Figure 7). Given a rotated input image, this method first detected the locations of two eyes, the vertical symmetric line, and face boundary. By calculating the relationships of the horizontal distances of eyes to symmetric line and face width, the rotation angle can be estimated by geometric transformations. Facial textures were then mapped by transforming the rotated view to a frontal pose on a cylinder. In the

implementation, this process is integrated into image normalisation and the processing time can be neglected compared to the rest of the processes in the recognition. Using LEM and Eigenfaces as face classifiers, this pose recovery demonstrated to be able to improve face recognition performances under pose variations.

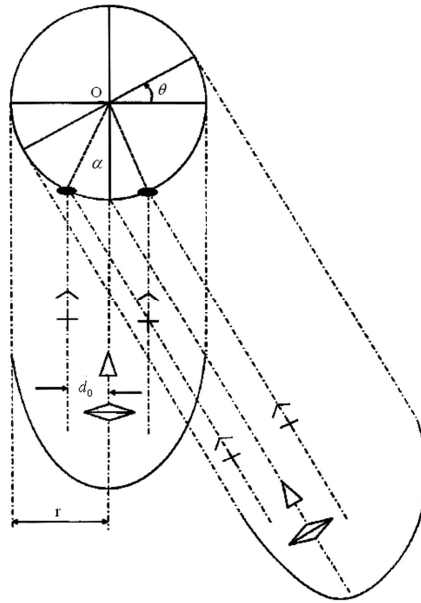


Figure 7 The pose recovery from a non-frontal view to a frontal view using a cylindrical face shape. [26]

Liu and Chen [55] 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. 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 different poses were different, so that a normalised pixel-wise Euclidean distance was used for recognition which only considers the overlapped region of two texture maps on the ellipsoid. A probabilistic model was trained to assign different weights to different pixels according to their discriminating powers in recognition, which can further improve the performance of face recognition across

pose. In experiments on CMU-PIE database with 9 poses and 34 faces, the probabilistic geometry assisted face recognition algorithm achieved an average of 86% identification accuracy. Yang and Krzyzak [81] have recently incorporated this geometrical mapping technique into a complete face detection and recognition system, where face detection is based on skin colour and pose estimation is based on facial features.

Besides using simple geometries (e.g., cylinder, ellipsoid, etc.) as generic 3D face models, Zhang et. al [85] 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. A gallery face image was aligned using two eyes' locations to the 3D generic shape and standard computer graphics procedure was applied to render virtual face views in different poses. By considering diffuse and specular reflectivity of the face surface, the texture mapping can generate simulated highlights on the rotated face views. In experiment on CMU-PIE database with one frontal gallery image and one 15 degree-rotated probe image per face of 40 faces, the ATS approach achieved 97.5% identification accuracy.

Generally, 3D approaches are computationally complex compared to their 2D counterparts. However, approaches using generic 3D shapes do not have this disadvantage. For instance, the generic shape-based pose recovery method [26] is very efficient and the pose recovery step can be neglected in the processing time. In this sense, 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 non-linear 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.

Under a relatively large pose variation, the differences between the generic shape and the individualised shape usually result in a decrease of recognition accuracy. Due to this aspect, other 3D approaches try to build personalised 3D face shapes even though it could be computationally demanding.

5.2 Feature-Based 3D Face Reconstructions

3D reconstruction is an active research area in computer vision, which inversely estimates 3D shape information from 2D images. Generalised 3D reconstruction considers all of the shape modelling, the surface reflectivity descriptions and the estimation of environmental parameters (e.g., lighting conditions). The clues for reconstructing 3D objects in 2D images are usually image features (e.g., edges and corners) and image intensities. In the context of face recognition through 3D reconstructions, these two groups are feature-based 3D face reconstructions and image-based 3D reconstructions, respectively. Feature-based 3D face reconstructions [38, 48, 51] reviewed in this subsection estimate personalised face shapes from the 2D locations of facial features (facial components such as eyes, nose, etc and image features such as edges or corners). Other 3D face reconstructions use image intensities and the reflectance models to extract shape and/or texture information from 2D images, in which more complicated processing is usually involved. These image-based reconstructions will be reviewed in the next subsection (Subsection 5.3).

Lee and Ranganath [48] 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 minimising a cost function consisting of edge fitting errors,

colour region displacements and deformation energy. The edge model defines the outlines of the face as well as various facial features such as the eyebrows, eyes, nose, mouth and ears. In the colour model, seven facial regions with colour information were considered. The regions were eyebrows, eye, nostrils and mouth. In the wire frame model, the face surface was divided into 100 triangles, which are defined by 59 vertices. To overcome false convergence on local minima, multiple evenly distributed models were assigned at the initial stage and the model with the lowest cost was chosen as the initial model. Then a typical gradient descent method was applied for minimisation of the cost function. 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 synthesised 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 synthesised image with the probe real image pixel-wisely in Euclidean distance. In experiments on a dataset of 15 faces with 11 different conditions per face (6 poses + 3 lighting + 2 scales), it achieved 56.2% recognition accuracy using a single gallery image per person and 92.3% accuracy using 10 gallery images per person. For the latter experiment setting using 10 gallery images per person, the pose-invariant face recognition algorithm was degraded to real view-based matching as the number of gallery images (10) is almost equal to the number of testing conditions (11).

Jiang et al. [38] used facial features to efficiently reconstruct personalised 3D face models from a single frontal face image for recognition. Their method is based on the automatic detection of facial features on the frontal views using Bayesian shape localisation. 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. Personalised 3D face shapes were reconstructed and the facial textures were directly mapped onto the face shape to synthesise virtual views in novel conditions as shown in Figure 8. Because the facial features all have semantic meanings, this method is also capable to synthesise virtual views with different expressions through changing locations of the facial features on the reconstructed 3D models. On CMU-PIE database, the method was shown to improve both PCA and LDA recognition algorithms, especially for LDA in half-profile views. This method, however, cannot effectively improve the recognition performance of near-profile views, due to the unreliable synthesis of the profile virtual views. This indicates that the facial features on the frontal views are not associated with the height information of face shapes. For instance, a narrow nose may or may not be higher than a broad nose. Therefore, a side view per person is desirable for more accurate estimation of the surface heights of the face. Compared to the composite deformable model [48], this model used the distributions of facial feature points on training face shapes as the space for new input shapes to project onto. Composite deformable model, on the other hand, didn't consider such prior knowledge and limited shape variations by introducing a general deformation cost defined by purely geometric changes of the 3D model. Such a deformation cost might be inappropriate of describing face shape variations due to identity changes. Jiang's method limited the identity-related shape variations within a training distribution by introducing a fairly complicated training mechanism.

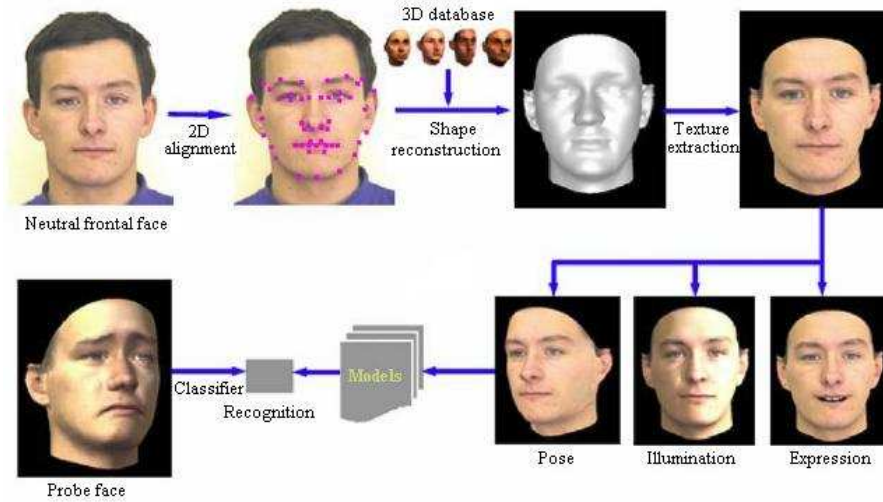


Figure 8 The process of 3D reconstruction and view synthesis in [38]. 100 3D faces with labelled images were used to relate 3D structure with 2D facial features. Neutral frontal images were automatically labelled and 3D structures were estimated using the prior knowledge from 100 training faces. Then the 3D structures with 2D textures were altered to generate novel virtual views in different poses, illumination conditions, and expressions and these views were used as models in recognition.

Using two orthogonal gallery images per face, Zhang et al. [84, 87] proposed to reconstruct the personalised 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 minimising 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 as shown in Figure 9. Specifically, the global cost function was defined as the second order smoothness of the surface, which was expressed of second partial derivatives on the x and y coordinates. Face shapes represented as vectors were varied in seeking the minimal of the cost function while maintaining the facial feature points on the right locations specified on the frontal and side-view gallery images. This process 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 analysed 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 synthesised and local binary patterns (LBP) were used for recognition in a view-based manner. In experiment on CMU-PIE database containing 13 poses per face of 68 faces with frontal lighting, this method achieved 93.45% recognition accuracy on 11 testing poses using two poses (frontal and side-view) as gallery. Compared to [38], MQVM used two views in different poses for 3D face shape reconstruction which is beneficial to face recognition across pose, because an additional view in a different viewpoint will provide more shape information otherwise unavailable from a single viewpoint. This extension, however, put an additional requirement for face database which might limit the applicability in general face recognition scenario.

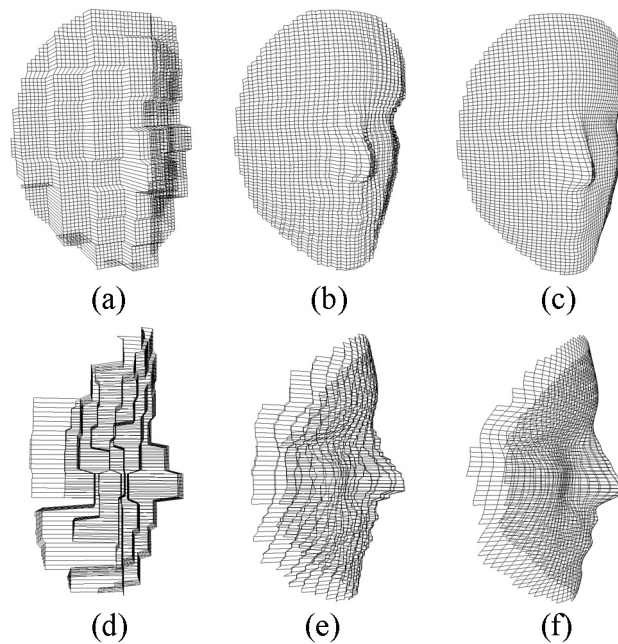


Figure 9 The hierarchical process of Multi-level Quadratic Variation Minimisation in [87]. (a, d) Reconstructed shape in coarse resolution, (b, e) in intermediate resolution level, and (c, f) in fine resolution.

In feature-based 3D face modelling approaches discussed above, personalised 3D face shapes were reconstructed from a set of facial features specified on facial images. The use of prior knowledge as in [38] helps the systems to reduce the number of gallery views required. However, the prior knowledge of human face shapes (usually obtained by analysing a set of existing 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. Because the facial features are usually sparse (about 100 points can be specified compared to 10,000 image pixels on 100x100 face images), they are unlikely capable of providing sufficient information for fine structure reconstruction such as eye balls and lips. A pixel-wise feature set should be used to achieve better reconstruction quality which will be discussed in the next subcategory of image-based reconstruction.

5.3 Image-Based Reconstruction

Image-based 3D face reconstructions carefully study the relationship between image pixel intensities and its corresponding shape/texture properties. From a set of pixel intensities, 3D face geometry and face surface properties can be estimated using appropriate reflectance models, which associate shape and texture information with reflected intensities. Unlike feature-based 3D face reconstructions' limited use of a few features on the face images, 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 [13, 14] 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 pixel-wised 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. The morphable model was then fitted into a single face image in an arbitrary condition by iteratively minimising pixel differences of image intensities and reconstructed virtual intensities using the set of parameters controlling the variations of shape, texture, illumination, pose, specularity, camera parameters, etc. Using stochastic Newton optimisation method, the process first makes use of several facial features defined on both image and 3D model to find a rough alignment and then relies more and more on the comparison of pixel intensities. The principal components of shape model and texture model were obtained in this process which was then used to reconstruct personalised 3D models and used for recognition using a modified angular (dot product) similarity measure based on linear discriminative analysis. The recognition was then performed using the extracted shape and texture parameters between gallery and probe as shown in Figure 10. In experiments on CMU-PIE database with 3 poses (0° , 15° , and 60° in yaw) by 22 illumination conditions per face of 68 faces, the recognition has achieved 92.1% using one of the 66 (3×22) images as gallery and the rest as probe for each person. On FERET database with 9 poses ranging within $\pm 40^\circ$ in yaw per face of 194 faces, the recognition algorithm achieved 95.8% using frontal view as gallery and the rest as probe.

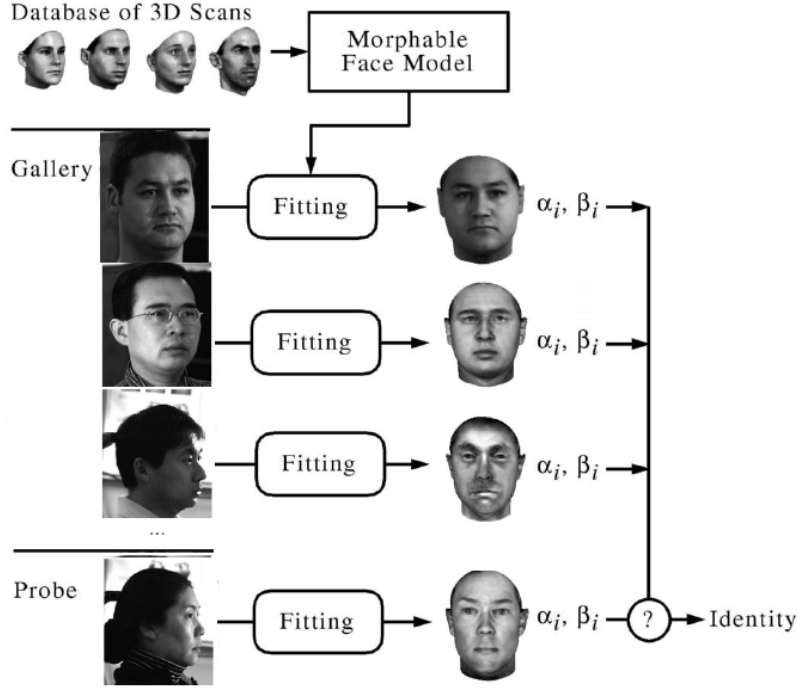


Figure 10 The process of face recognition based on 3D morphable model [14]. The shape and texture prior knowledge characterised by principal components was learned from a database of 3D face scans. 3D morphable model was then fitted to single input images for both gallery and probe. Personalised shape and texture coefficients (i.e., $\vec{\alpha}$ and $\vec{\beta}$ respectively) were extracted which are free from external pose and illumination conditions. These identity-related parameters were then used in recognition.

Georghiades et al. [30] proposed Illumination Cone Models (ICM) which 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, personalised face shape and surface reflectance information was reconstructed by minimising 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 (i.e. light directions and intensities), 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 synthesised and used in face recognition to match the probe image with the closest virtual images in sampled poses and illuminations. Their recognition approach was tested on the Yale Database B consisting 4,050 images of 10 faces under 45 illumination conditions \times 9 different poses ($\pm 24^\circ$ in-depth rotation). It achieved 96.25% recognition accuracy using the frontal image as gallery and other 8 poses as probe. This approach relies only on the pixel intensities from multiple images under different lighting conditions and a fixed pose and the reconstruction process doesn't require any form of prior knowledge of human faces. The assumption of Lambertian surfaces, however, causes the reconstruction results to have a bas-relief ambiguity [9], i.e., the reconstructed shape and estimated lighting are not unique. To resolve this ambiguity, certain forms of prior knowledge on human faces are used such as left-right symmetry, similar heights of forehead and chin, and relationship of the surface heights and width.

In [31], the illumination cone model was extended by incorporating Torrance-Sparrow model [73] into the process of 3D reconstruction of human faces to resolve the bas-relief ambiguity [9] associated with photometric stereo using Lambertian model [45]. Using the results of [30] as the initial estimate, the difference of the real face images and the rendered images using the estimated parameters based on a simplified Torrance-Sparrow model was minimised using the steepest descent method. This algorithm is able to inversely estimate a set of spatially varying diffuse reflectance coefficients with a uniform specular reflectance coefficient, while the estimation of a full set of spatially varying reflectance coefficients remains open. Tested on the same experiment setting as in [30], the face reconstruction method using Torrance-Sparrow

model achieved slightly higher recognition rates than the method based on Lambertian model. Other directions of the photometric stereo in face recognition include introducing a more general illumination model, i.e. spherical harmonics [5, 64] and representing different faces within a single class [91].

Besides photometric stereo which reconstruct face models from a set of 2D images in the same pose under different lighting conditions, stereo vision techniques can also be applied which reconstructs 3D face models from 2 face images in different poses. Castillo and Jacobs [18] proposed to use the cost of stereo matching of gallery face image and probe face image to recognise faces. The stereo matching algorithm used in this method defined 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 (say left) to the second (right) 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. Tested on PIE database with 13 poses per face of 68 faces, this method achieved 73.5% recognition accuracy using any one pose as gallery and the remaining 12 poses as probe.

5.4 Summary and Discussions

In this section, face recognition approaches with assistance of 3D face models have been reviewed. These approaches have been classified into three subcategories, i.e., 1) generic shape-based approaches, 2) feature-based 3D reconstructions, and 3) image-based 3D reconstructions. Compared to 2D approaches discussed in Section 4, 3D approaches try to approximate the image variations caused by pose variations in 3D space rather than limiting them within the image plane.

The different methodologies are summarised in Table 10. The simplest strategy is to apply a uniform (or generic) face shape to approximate various face shapes, which give these generic shape-based approaches the benefit of efficiency. However, each individual face shape may deviate from the generic face shape greatly due to interpersonal face shape differences, which cannot be overcome by improving the generic shape. Consequently, image distortions exist in these recovered face images in different poses, which affect the performance of face recognition across pose.

To better approximate face shape, personalised 3D models were reconstructed from a set of facial features (feature-based) or from pixel-wise image intensities (image-based). Generally, feature-based reconstructions require feature point locating, which is always based on image contents. For instance, edge information was used as features in [48] and was extracted by comparing the neighbouring pixels' intensities. Feature-based reconstructions limit the use of image intensities (textures) in 2D image plane to extract features for reconstruction, which is efficient and simple compared to considering image intensities in 3D space of image-based reconstructions. However, facial features are sparse compared to face image dimension. Consequently the feature-based reconstructions are at most accurate near facial features and could be inaccurate in other non-feature regions because they are usually interpolated from adjacent facial features. Image-based reconstructions rely on pixel-wise appearances of face images to reconstruct 3D face models, where the reflection mechanism of human face surfaces is crucial. These approaches are generally capable of generating more detailed face structures than feature-based reconstructions, because each pixel has been considered in reconstruction. The price is more complex procedures and sometimes unreliability. Table 11 summarises the advantages and disadvantages of the above discussed approaches.

Table 10 The methodologies of face recognition algorithms with assistance of 3D models.

Approach	Pose tolerance and compensation	Face recognition algorithms
Cylindrical 3D pose recovery [26]	Pose estimation using facial components and generic 3D shape + texture mapping by pose transformation on cylindrical shape	Eigenfaces, LEM
Probabilistic geometry assisted FR [55]	Pose estimation using universal mosaic modelling + texture mapping by pose transformation on ellipsoid shape	Weights assignments using probabilistic models + normalised Euclidean distance
Automatic texture synthesis [85]	Texture mapping by pose transformation on averaged face shape + reflection analysis and synthesis on Phong model	View-based PCA
Composite deformable model [48]	Fitting a deformable model to input image by minimising fitting error and deformation cost + texture coefficients by linearly projecting the input image on gallery image textures	Nearest neighbours on the estimated texture coefficients
Jiang's method [38]	Constructing an eigen space on training 3D face models + Projecting 2D input point distribution on the subspace of the eigen + Reconstruct 3D point distribution on eigen space of 3D points + direct texture mapping	Linear discriminant analysis
Multi-level Quadratic Variation Minimisation [87]	Reconstructing 3D shape by minimising surface roughness controlled by facial feature points + extracting texture coefficients by fitting input images to Phong model + synthesising virtual views by rotating 3D model and generating virtual textures	Local binary patterns
3D Morphable Model [14]	Training shape and texture eigen spaces on 3D dataset + aligning 3D morphable model to 2D input image by minimising a weighted sum of feature displacement, image dissimilarity and external parameters variations from their averages + extracting shape and texture model coefficients	Modified dot product on shape and texture model coefficients based on linear discriminant analysis
Illumination Cone Model [30]	Extracting surface normals and reflectance coefficients from front images under different lighting + integrability enforcement to reconstruct 3D surface from normal directions + virtual view synthesis in different poses and lighting	View-based exhaustive searching in all possible virtual images in Euclidean distance
Stereo Matching [18]	Align images according to epipolar geometry + stereo matching + extracting matching cost	Nearest neighbour of the matching cost

Table 11 The advantages and disadvantages of the face recognition algorithms with assistance of 3D face models in face recognition across pose.

Approach	Advantages	Disadvantages
Cylindrical 3D pose recovery [26]	Simple, efficient	Inaccurate face shape approximation
Probabilistic geometry assisted FR [55]	Simple, efficient	Inaccurate face shape approximation
Automatic texture synthesis [85]	Simple, facial textures were approximated	Rigid face shape approximation doesn't fit to all faces
Composite deformable model [48]	Personalised 3D face shape was reconstructed	Deformation is arbitrary
Jiang's method [38]	Efficient, deformation is based on face shape variations	Texture mapping doesn't consider appearance variations
Multi-level Quadratic Variation Minimisation [87]	No prior knowledge is required, two gallery views provide better 3D shape information	Requires manual locating of facial features
3D Morphable Model [14]	Only single image is required for reconstruction, both shape and texture modelling are based on prior knowledge of shape and texture variations, reconstruction is performed pixel-wisely comparing image intensities	Unstable, identity-related shape and texture coefficients were affected during cost function minimisation
Illumination Cone Model [30]	No prior knowledge is required so that identity-related parameters were preserved	Requires multiple images under certain restrictions, surface approximation discarded specular reflection
Stereo Matching [18]	Simple, single gallery is required	Image-based matching doesn't consider appearance changes due to pose variations

Though image-based reconstructions often involve complex processing in considering reflection of human faces, they made the most use of image information by exhaustive treatment of all image pixels. Compared to feature-based reconstructions which at best can guarantee accurate reconstruction around facial features, image-based reconstructions have the potential to achieve pixel-wisely accurate reconstruction results. Feature-based reconstructions also suffer from the inaccuracy of feature detections, while feature detection is no longer required in image-based reconstructions. Because image-based 3D reconstructions consider pixel-wise reflection mechanisms in estimating shape and texture information, they are generally more sensitive and consequently vulnerable to image variations, such as shadows and spatial misalignment. To

alleviate spatial misalignment, 3D morphable model used a feature-based approach as the initial stage of the 3D reconstruction, and illumination cone model requires rigorous alignment of multiple photometric stereo images under a fixed viewpoint.

The reflection mechanism of human faces is crucial to image-based 3D face reconstructions. In contrast, existing approaches tend to make simplistic approximations on face surfaces. Most of existing methods assume face surfaces as Lambertian surfaces, which only consider diffuse reflection and neglect specular reflection. In fact, human faces reflect both diffusely and specularly and reflectance models beyond Lambertian assumption should be taken into considerations to achieve better reconstruction performance by making more realistic surface approximations. In feature-based 3D face reconstruction, the texture estimation is also suboptimal by primarily using the Lambertian assumption to approximate human face skins. Similar to 2D pose transformation, image intensities of real views are usually directly mapped onto the reconstructed 3D shape without considering the intensity variations caused by pose changes.

6 Conclusions and Further Discussions

As the prominent problem in face recognition, pose variation received extensive attentions in the research community of computer vision and pattern recognition. A number of promising techniques have been proposed to tolerate and/or compensate 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 paper first reviewed these techniques, providing a comprehensive survey and critical discussions on major challenges and possible future research directions towards pose-invariant face recognition. This paper started on

discussions of the problem of face recognition across pose, with elaborations on the challenges, current evaluation methodologies, and performances of different approaches. Face recognition techniques relevant to handling pose variations were then classified into three broad categories, i.e., general algorithms, 2D techniques and 3D approaches. Representative general algorithms have been reviewed with an emphasis on their sensitivities to pose variations. 2D techniques and 3D approaches which actively compensate pose variations have been comprehensively reviewed in the last two sections with discussions on their advantages and limitations.

Based on this review, several insights 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 (e.g., one example gallery image per person). The image variations caused by pose changes can be learned from known face images or models in the 2D and 3D approaches, which are then applied to new input image(s) to simulate real pose transformations. 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. Consequently, these techniques could better preserve discriminative features of the gallery images, free from the influences of the training data. Due to the insufficient information 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 face recognition across pose. 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 reflectivities 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.

On the other hand, a comprehensive consideration of the complicated face surface reflection mechanism and external lighting parameters brings serious ill-conditions to 3D face modelling, because the number of unknown parameters is excessive and the problems are intractable. Several attempts have been made to extend Lambertian assumption to include specular reflection, at the price of resolving non-linear optimisation problems. These early attempts towards precise descriptions on face surface reflection are only limited to coarse approximations of specular reflection, while ignoring other factors such as inter-reflections and subsurface scattering. It is still an open question on how to incorporate these complicated image formation approximations into face recognition to improve its pose tolerance while keeping the problem tractable. The potential solutions towards this direction rely on both better image formation models specifically suitable to face modelling and task specific computational tools that reliably and efficiently solve non-linear optimisation problems.

The strategy of nonlinear mapping has the promise to find a feature space best suitable to pose variations, while the current research stage is preliminarily limited to fundamental mapping functions (e.g., radial basis functions). The question of whether there is a feature space where rotated faces are separable is still open. An answer to this question may lead to a clearer understanding of pose-invariant face recognition problem, similar to the findings of linear subspaces in illumination-invariant face recognition.

References

- [1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Rec. Lett.*, vol. 28, no. 14, pp. 1885-1906, 2007.
- [2] 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.
- [3] W. A. Barrett, "A survey of face recognition algorithms and testing results," in *Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 301-305, 1998.
- [4] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Network*, vol. 13, no. 6, pp. 1450-1464, 2002.
- [5] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218-233, 2003.
- [6] T. Beier and S. Neely, "Feature-based image metamorphosis," *SIGGRAPH 92 (Computer Graphics)*, vol. 26, pp. 35-42, 1992.
- [7] 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.
- [8] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 245-260, 1998.
- [9] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *Int. J. Comput. Vis.*, vol. 35, no. 1, pp. 33-44, 1999.
- [10] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proc. Int. Conf. Comput. Vis.*, pp. 500-507, 1995.
- [11] D. Beymer, "Feature correspondence by interleaving shape and texture computations," in *Proc. IEEE Conf. CVPR*, pp. 921-928, 1996.
- [12] D. J. Beymer, "Face recognition under varying pose," in *Proc. IEEE Conf. CVPR*, pp. 756-761, 1994.
- [13] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *SIGGRAPH*, pp. 187-194, 1999.
- [14] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063-1074, 2003.
- [15] K. W. Bowyer, C. Kyong, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1-15, 2006.
- [16] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042-1052, 1993.
- [17] S. Cass and M. Riesenman, "Improving security, preserving privacy," *IEEE Spectrum*, vol. 39, no. 1, pp. 44-49, 2002.
- [18] C. D. Castillo and D. W. Jacobs, "Using stereo matching for 2-D face recognition across pose," in *Proc. IEEE Conf. CVPR*, pp. 1-8, 2007.
- [19] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Proc.*, vol. 16, no. 7, pp. 1716-1725, 2007.

- [20] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc IEEE*, vol. 83, no. 5, pp. 705-741, 1995.
- [21] 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.
- [22] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The Use of Active Shape Models For Locating Structures in Medical Images," *Image Vis. Comput.*, vol. 12, no. 6, pp. 355-366, 1994.
- [23] 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.
- [24] 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.
- [25] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image Vis. Comput.*, vol. 20, pp. 657-64, 2002.
- [26] Y. Gao, M. K. H. Leung, W. Wang, and S. C. Hui, "Fast face identification under varying pose from a single 2-D model view," *IEE Proc. Vis. Image Signal Process.*, vol. 148, no. 4, pp. 248-253, 2001.
- [27] 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.
- [28] Y. Gao and Y. Qi, "Robust visual similarity retrieval in single model face databases," *Pattern Rec.*, vol. 38, pp. 1009-1020, 2005.
- [29] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: generative models for recognition under variable pose and illumination," in *Int Conf Auto Face Gesture Rec*, pp. 277-284, 2000.
- [30] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643-660, 2001.
- [31] A. S. Georgiades, "Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo," in *Proc. ICCV*, vol. 2, pp. 816-823, 2003.
- [32] D. González-Jiménez and J. L. Alba-Castro, "Toward pose-invariant 2-D face recognition through point distribution models and facial symmetry," *IEEE Trans. Info. Foren. Sec.*, vol. 2, no. 3-1, pp. 413-429, 2007.
- [33] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 449-465, 2004.
- [34] S. Haykin, *Neural Networks—A Comprehensive Foundation*, 2nd ed. New York: Prentice-Hall, 1999.
- [35] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328-340 2005.
- [36] J. Huang, P. C. Yuen, W. S. Chen, and J. H. Lai, "Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 847-862, 2007.
- [37] R. Ishiyama, M. Hamanaka, and S. Sakamoto, "An appearance model constructed on 3-D surface for robust face recognition against pose and illumination variations," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 326-334, 2005.

- [38] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition," *Pattern Rec.*, vol. 38, no. 6, pp. 787-98, 2005.
- [39] F. Kahraman, B. Kurt, and M. Gokmen, "Robust face alignment for illumination and pose invariant face recognition," in *Proc. IEEE Conf. CVPR*, pp. 1-7, 2007.
- [40] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640-649, 2007.
- [41] T. Kanade, "Picture processing by computer complex and recognition of human faces," Department of Information Science, Kyoto University, Japan, 1973.
- [42] T. K. Kim and J. Kittler, "Design and fusion of pose-invariant face-identification experts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 9, pp. 1096-1106, 2006.
- [43] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human face," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103-108, 1990.
- [44] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. van der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans Comput*, vol. 42, pp. 300-311, 1993.
- [45] J. Lambert, *Photometria Sive de Mensura et Gradibus Luminus, Colorum et Umbrae*. Eberhard Klett, 1760.
- [46] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 743-756, 1997.
- [47] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Trans Neural Network*, vol. 8, no. 1, pp. 98-113, 1997.
- [48] M. W. Lee and S. Ranganath, "Pose-invariant face recognition using a 3D deformable model," *Pattern Rec.*, vol. 36, no. 8, pp. 1835-1846, 2003.
- [49] T. S. Lee, "Image representation using 2-D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959-971, 1996.
- [50] M. D. Levine and Y. Yu, "Face recognition subject to variations in facial expression, illumination and pose using correlation filters," *Comput. Vis. Image Understand.*, vol. 104, no. 1, pp. 1-15, 2006.
- [51] C. Li, G. Su, Y. Shang, and Y. Li, "Frontal face synthesis based on multiple pose-variant images for face recognition," in *International Conference on Advances in Biometrics*, pp. 948-957, 2007.
- [52] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*. New York. Springer, 2005.
- [53] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans Neural Network*, vol. 8, no. 1, pp. 114-132, 1997.
- [54] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 572-581, 2004.
- [55] X. Liu and T. Chen, "Pose-robust face recognition using geometry assisted probabilistic modeling," in *Proc. IEEE Conf. CVPR*, vol. 1, pp. 502-509, 2005.
- [56] Y. Lu, J. Zhou, and S. Yu, "A survey of face detection, extraction and recognition," *Computing and Informatics*, vol. 22, no. 2, pp. 163-195, 2003.

- [57] A. Mahalanobis, B. V. K. V. Kumar, and S. R. F. Sims, "Distance-classifier correlation filters for multiclass target recognition," *Applied Optics*, vol. 35, no. 17, pp. 3127–3133, 1996.
- [58] M. K. Müller, A. Heinrichs, A. H. J. Tewes, A. Schäfer, and R. P. Würtz, "Similarity rank correlation for face recognition under unenrolled pose," in *International Conference on Biometrics*, pp. 67-76, 2007.
- [59] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 607-626, 2009.
- [60] T. Ojala, M. Pietikainen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, 2002.
- [61] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Conf. CVPR*, pp. 84-91, 1994.
- [62] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The Feret database and evaluation procedure for face recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295-306, 1998.
- [63] S. J. D. Prince, J. Warrell, J. H. Elder, and F. M. Felisberti, "Tied Factor Analysis for Face Recognition across Large Pose Differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 970-984, 2008.
- [64] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1322-1333, 2002.
- [65] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A Survey of 3D Face Recognition Methods," in *International Conference on Audio- and Video-based Biometric Person Authentication*, vol. 3546, pp. 891-899, 2005.
- [66] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computing*, vol. 10, no. 5, pp. 1299-1319, 1998.
- [67] D. Shan and R. Ward, "Face recognition under pose variations," *Journal of the Franklin Institute*, vol. 343, no. 6, pp. 596-613, 2006.
- [68] T. Shan, B. C. Lovell, and S. Chen, "Face Recognition Robust to Head Pose from One Sample Image," in *18th International Conference on Pattern Recognition*, vol. 1, pp. 515-518, 2006.
- [69] H. C. Shin, J. H. Park, and S. D. Kim, "Combination of warping robust elastic graph matching and kernel-based projection discriminant analysis for face recognition," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1125-1136, 2007.
- [70] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615-1618, 2003.
- [71] 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. 1212-1225, 2007.
- [72] T. Stonham, "Practical face recognition and verification with WISARD," *Aspects of Face Processing*, pp. 426–441, 1984.
- [73] K. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *Journal of the Optical Society of America*, vol. 56, no. 7, pp. 916-925, 1967.

- [74] M. Turk and A. Pentland, "Eigenfaces for recognition," *J Cogn Neurosci*, vol. 3, no. 1, pp. 71-86, 1991.
- [75] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. CVPR*, pp. 586-591, 1991.
- [76] T. Vetter, "Synthesis of novel views from a single face image," *Int. J. Comput. Vis.*, vol. 28, no. 2, pp. 103-116, 1998.
- [77] Y. Wang and C.-S. Chua, "Robust face recognition from 2D and 3D images using structural Hausdorff distance," *Image Vis. Comput.*, vol. 24, no. 2, pp. 176-185, 2006.
- [78] H. Wechsler, *Reliable Face Recognition Methods: System Design, Implementation and Evaluation*. Springer, 2006.
- [79] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775-779, 1997.
- [80] X. Xie and K.-M. Lam, "Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image," *IEEE Trans. Image Proc.*, vol. 15, no. 9, pp. 2481-2492, 2006.
- [81] F. Yang and A. Krzyzak, "Face recognition under significant pose variation," in *2007 Canadian Conference on Electrical and Computer Engineering*, pp. 1313-1316, 2007.
- [82] J. Yang, A. F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 230-244, 2005.
- [83] L. Zhang, A. Razdan, G. Farin, J. Femiani, M. Bae, and C. Lockwood, "3D face authentication and recognition based on bilateral symmetry analysis," *Visual Computer*, vol. 22, no. 1, pp. 43-55, 2006.
- [84] X. Zhang, Y. Gao, and M. K. H. Leung, "Multilevel quadratic variation minimization for 3D face modeling and virtual view synthesis," in *Proc. Int. Multimedia Model. Conf.*, pp. 132-138, 2005.
- [85] X. Zhang, Y. Gao, and M. K. H. Leung, "Automatic texture synthesis for face recognition from single views," in *Proc. ICPR*, vol. 3, pp. 1151-1154, 2006.
- [86] X. Zhang, Y. Gao, and B.-L. Zhang, "Recognising rotated faces from two orthogonal views in mugshot databases," in *Proc. ICPR*, vol. 1, pp. 195-198, 2006.
- [87] X. Zhang, Y. Gao, and M. K. H. Leung, "Recognizing rotated faces from frontal and side views: an approach towards effective use of mugshot databases," *IEEE Trans. Info. Foren. Sec.*, vol. 3, no. 4, pp. 684-697, 2008.
- [88] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915-928, 2007.
- [89] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399-459, 2003.
- [90] W. Zhao and R. Chellappa, *Face Processing: Advanced Modeling and Methods*. Academic Press, 2005.
- [91] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs, "Appearance characterization of linear Lambertian objects, generalized photometric stereo, and illumination-invariant face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 230-245, 2007.

