

Contents

8	A Review on Face and Gait Recognition: System, Data and Algorithms	1
	<i>Haiping Lu, Jie Wang and Konstantinos N. Plataniotis</i>	
8.1	Introduction	1
8.1.1	Face recognition	2
8.1.2	Gait recognition	3
8.1.3	Organization	3
8.2	Face and Gait Recognition System Overview	4
8.2.1	Model-based face recognition approach	5
8.2.2	Model-based gait recognition approach	6
8.2.3	Appearance-based face recognition approach	7
8.2.4	Appearance-based gait recognition approach	7
8.2.5	Fusion of face and gait for recognition	7
8.3	Face and Gait Data Sets	8
8.3.1	The PIE database	8
8.3.2	The FERET database	9
8.3.3	The USF gait database	10
8.4	Face and Gait Recognition Algorithms	11
8.4.1	Linear subspace learning algorithms	11
8.4.1.1	Principal component analysis	12
8.4.1.2	Bayesian method	12
8.4.1.3	Linear discriminant analysis	13
8.4.2	Nonlinear kernel-based subspace learning algorithms	15
8.4.2.1	Kernel principal component analysis	18
8.4.2.2	Kernel-based discriminant analysis (KDA)	19

ii CONTENTS

8.4.3	Multilinear subspace learning algorithms	20
8.4.3.1	Multilinear principal component analysis	22
8.4.3.2	Multilinear discriminant analysis	23
8.5	Concluding Remarks	24
8.5.1	Summary	24
8.5.2	Current state and future directions of face recognition	24
8.5.3	Current state and future directions of gait recognition	25
	References	27

8 A Review on Face and Gait Recognition: System, Data and Algorithms

Haiping Lu^{1*}, Jie Wang² and K. N. Plataniotis³

^{1,3}The Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto, M5S 3G4, Canada

²Epson Edge, 3771 Victoria Park Avenue, Toronto, M1W 3Z5, Canada

*haiping@comm.toronto.edu

ABSTRACT

This chapter reviews two important biometric recognition technologies that have received significant attention recently: face and gait recognition, where individuals are recognized by their faces and the way they walk, respectively. These two technologies are mainly motivated from security-related applications. We first describe typical face and gait recognition systems and three common recognition scenarios. Then, we present the two general approaches for solving the problem, the model-based approach and the appearance-based approach, as well as the fusion of face and gait, followed by discussions on commonly used face and gait databases. Next, a survey of important recognition algorithms are provided, including linear, nonlinear and multilinear subspace learning algorithms. Finally, the current state and future directions are summarized.

8.1 INTRODUCTION

For thousands of years, humans have used visually perceived body characteristics such as face and gait to recognize each other. This remarkable ability of human visual system has inspired researchers to build automated systems to recognize individuals from digitally captured facial images and gait sequences [1]. Face and gait recognition belong to the field of biometrics, a very active area of research in the computer vision and pattern recognition society, and face and gait are two typical physiological and behavioral biometrics, respectively. Compared with other biometric traits, face and gait have the unique property that they facilitate human recognition at a distance, which is extremely important in surveillance applications. Moreover, their unintrusive nature leads to high collectability and acceptability, making them very promising technologies for wide deployments. The collectability refers to the

ease of acquisition for measurement and the acceptability indicates the extent to which people are willing to accept the use of a particular biometric identifier in their daily lives [1].

8.1.1 Face recognition

Face recognition has a wide range of applications, such as biometric authentication, surveillance, human-computer interaction and multimedia management. Table 8.1 lists some of these applications in detail. On the other hand, the rapid developments in technologies such as digital cameras, the Internet and mobile devices also increase the popularity of face recognition [2]. A number of commercial face recognition systems have been deployed, such as Cognitec [3], Eyematic [4], and L-1 Identity Solutions [5]. Face recognition is one of the three identification methods used in e-passports and it has an important advantage over other popular biometric technologies: it is non-intrusive and easy to use [2]. Among the six biometric attributes considered in [6], facial features scored the highest compatibility in a machine readable travel documents (MRTD) system based on a number of evaluation factors, such as enrollment, renewal, machine requirements, and public perception [6]. In addition, in advanced human-computer interaction applications, it is very important for robots to be able to identify faces, expressions and emotions while interacting with humans.

Table 8.1 Applications of face recognition (adapted and expanded from [7] and [8]).

Areas	Examples of Applications
Biometrics, Mugshot Identification	Drivers Licenses, Entitlement Programs, Smart Cards Immigration, National ID, Passports, Voter Registration Welfare Fraud, Airline Industry, Bank Industry
Information Security	Desktop Logon, Secure Trading Terminals, Application Security, Database Security, File Encryption, Intranet Security, Internet Access, Medical Records
Law Enforcement and Surveillance	Automated Video Surveillance (e.g., airport security checkpoints), Portal Control, Postal-Event Analysis, Face Reconstruction from Remain Shoplifting and Suspect Tracking and Investigation
Access Control	Access to Private Buildings, Facility, Personal Computers, PDA and Cell Phone
Others	Human-Computer Interaction, Content-based Image Database Management, Information Retrieval, Multimedia Communication (e.g., generation of synthetic faces)

8.1.2 Gait recognition

Gait recognition [9, 10], the identification of individuals in video sequences by the way they walk, is strongly motivated by the need for automated person identification system at a distance in visual surveillance and monitoring applications in security-sensitive environments, e.g., banks, parking lots, museums, malls, and transportation hubs such as airports and train stations [11], where other biometrics such as fingerprint, face or iris information are not available at high enough resolution for recognition [12, 13]. Furthermore, night vision capability (an important component in surveillance) is usually not possible with other biometrics due to the limited biometric details in an IR image at large distance [12, 13].

Gait is a complex spatio-temporal biometric [10, 12] that can address the problems above. In 1975 [14], Johansson used point light displays to demonstrate the ability of humans to rapidly distinguish human locomotion from other motion patterns. Similar experiments later showed the capability of identifying friends or the gender of a person [15, 16], and Stevenage *et al.* show that humans can identify individuals based on their gait signature in the presence of lighting variations and under brief exposures [17].

Gait is a behavioral (habitual) biometric, in contrast with those physiological biometrics such as face and iris, and it is viewed as the only true remote biometric [18]. Capturing of gait is unobtrusive, which means that it can be captured without requiring the prior consent of the observed subject, and gait can be recognized at a distance (in low resolution video) [19]. In contrast, other biometrics either require physical contact (e.g., fingerprint) or sufficient proximity (e.g., iris). Furthermore, gait is harder to disguise than static appearance features such as face.

8.1.3 Organization

The problem of face and gait recognition is very challenging. The main challenge in vision-based face or gait recognition is the presence of a high degree of variability in human face images or gait sequences. For facial images [20], the intra-subject variations include pose (imaging angle), illumination, facial expression, occlusion, makeup, glasses, facial hair, time (aging) and imaging parameters such as aperture and exposure time. For gait, the intra-subject variations include pose (viewing angle), shoes, walking surface, carrying condition, clothing, time and also imaging device. Therefore, extracting the intrinsic information of a person's face or gait from their respective images or gait sequences, respectively, is a demanding task. Despite the difficulties, researchers have made significant advancement in this area. The rest of this chapter provides a survey on face and gait recognition. The next section is an overview of the face and gait recognition systems, where the key components are described and the two common approaches are introduced. The fusion of face and gait for recognition is also discussed in the next section. In Section 8.3, several commonly used face and gait databases are reviewed. Section 8.4 presents various feature extraction algorithms for face and gait recognition, ranging from linear, nonlinear to multilinear subspace learning algorithms. Finally, concluding remarks are drawn in Section 8.5.

8.2 FACE AND GAIT RECOGNITION SYSTEM OVERVIEW

Figure 8.1 depicts a typical face or gait recognition system. By observing a subject in the view, a digital camera captures a digital raw facial image or a digital raw gait video. This image or video is then pre-processed (e.g., filtered to remove noise) to extract a facial image or a gait sequence for feature extraction. In the feature extraction, face or gait features are extracted from the input image or image sequence, and these features are passed to the recognition module, where classifiers are employed to match them with the stored features in the face or gait database and a person is recognized with his/her ID as the output.

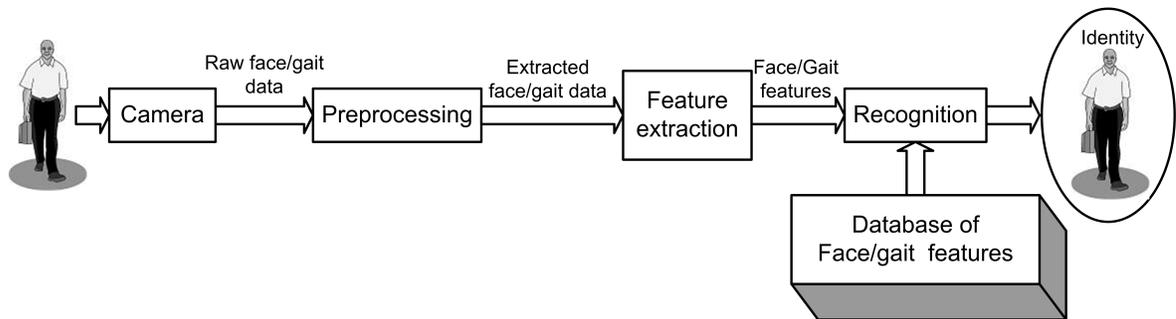


Fig. 8.1 A typical face or gait recognition system.

Face localization and normalization (face detection and alignment) are preprocessing steps before face recognition (facial feature extraction and matching) is performed. Face detection segments the face area from the background and provides (coarse) information about the location and scale of each detected face. Face alignment aims to achieve more accurate localization and normalizes faces. Some specific facial components, such as eyes, nose, mouth and facial outline, are further located, and then the input face image is aligned and normalized in geometry (such as size and pose) and photometry (such as illumination and gray scale).

For gait recognition, binary gait silhouettes are usually extracted through background subtraction, where a background model is estimated from the input raw gait sequences and then it is subtracted to get the silhouettes [21]. The extracted silhouettes are then cropped and resized to a standard size.

After the preprocessing, features useful for discriminating different persons are extracted from the normalized face or gait sample, and in the recognition stage, the extracted features are matched against those of enrolled ones in the database. Finally, the system outputs the identity of the input when a match is found with sufficient confidence or indicates an unknown identity otherwise.

There are three main recognition tasks: verification, identification and watch list [10]. Verification involves a one-to-one match that compares a query sample against a template of the claimed identity in the database. The claim is either accepted or rejected. The verification performance is usually measured by the receiver operating characteristic (ROC), which plots the false accept rates (FAR) versus the false rejection rate (FRR). Identification

involves one-to-many matches that compares a query sample against all the templates in the database to output the identity or the possible identity list of the input query. In this scenario, it is often assumed that the query sample belongs to the persons who are in the database. The identification performance is usually measured by the cumulative match characteristic (CMC) [22,23], which plots the identification rate against the rank. The watch list scenario involves one-to-few matches that compares a query sample against a list of suspects. In this task, the size of database is usually very small compared to the possible queries and the identity of the probe may not be in the database. Therefore, the recognition system should first detect whether the query is on the list or not and if yes, correctly identify it. The performance of watch list tasks is usually measured by the detection rate, the identification rate and the false alarm rate.

Over the past two decades, a large number of face or gait recognition algorithms have been developed. These algorithms can be categorized into two approaches, the model-based approach and the appearance-based approach [12, 24]. Although the model-based approach takes advantage of our prior knowledge on human face or gait, reliable recovery of model parameters from raw facial images or gait sequences is a very hard problem, especially in unconstrained conditions. Therefore, the appearance-based approach has been the most successful in reported literature. In the following, we briefly describe these two approaches, with more detailed discussions on the popular and more successful appearance-based approach in Section 8.4.

8.2.1 Model-based face recognition approach

The model-based face recognition approach aims to construct a model of the human face in order to capture the facial variations, based on prior knowledge of human faces. Then, the model is fitted to a given face image and parameters of the fitted model form the feature vector for recognition through similarity calculation between the query face and stored faces in the database [24].

In the so-called feature-based approach popular in the early days, local facial features such as eyes, nose, mouth, and chins are first accurately located. Properties of and relations between (such as areas, distances and angles) the features are used as descriptors for recognition. Typical examples include the Hidden Markov Model methods [25] and the elastic bunch graph matching algorithm [26]. Another model-based method is the active appearance models (AAM) [27] where face variations are learned through integrating shape and texture information with a 2-D morphable face model.

The main advantage of the model-based approach is that the model is based on intrinsic physical relationship with real faces and it is less sensitive to variations in illumination and viewpoint [24]. However, relatively high resolution and good quality face images are needed in this approach. Besides, the recognition performance relies heavily on the exact localization of facial features, but facial feature extraction techniques developed to date are not reliable enough for accurate recognition. For instance, most eye localization techniques assume some geometric

and textural models of the eyes and they fail if the eyes are closed. Moreover, such geometric properties alone are inadequate for face recognition because rich information contained in the facial texture or appearance is discarded.

8.2.2 Model-based gait recognition approach

The model-based gait recognition approach considers a human subject as an articulated object represented by various body poses. Model-based gait recognition algorithms are usually based on 2-D fronto-parallel body models [19, 28–30] and target to model human body structure explicitly, with support from the anthropometry and the biomechanics of human gait [31, 32]. Body model parameters, such as joint angles, are searched in the solution space through matching edges and region-based information (e.g., silhouettes). The searching methods are either exhaustive [28, 29] or in a Bayesian hypothesis-and-test fashion [19], where proper dealing with local extrema is an important problem. The estimated parameters are either used directly as features or fed into a feature extractor (e.g., frequency analyzer) to obtain gait features.

There are also works on coarser human body models. For instance, the work in [33] fits several ellipses to different parts (blobs) of the binary silhouettes and the parameters of these ellipses (e.g., location. and orientation) are used as gait features.

Recently, a full-body layered deformable model (LDM) is proposed in [34], inspired by the manually labeled body-part-level silhouettes [35]. The LDM has a layered structure to model self-occlusion between body parts and it is deformable so simple limb deformation is taken into consideration. In addition, it also models shoulder swing. The LDM parameters can be recovered from automatically extracted silhouettes and then used for recognition [36]. Figure 8.2 shows three examples of the LDM body poses (on the right) recovered from raw image frames (on the left) for illustration.

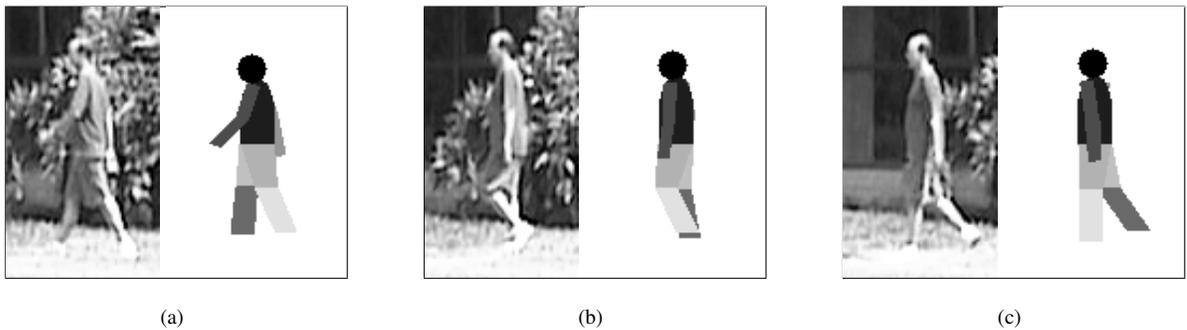


Fig. 8.2 Three examples of the human body poses recovered through the LDM. Left: the raw image frame. Right: the automatically reconstructed silhouette through the LDM recovery.

As in the case of face recognition, model-based gait recognition algorithms utilize our knowledge on gait and they are expected to be less sensitive to noise due to background clutters, cloths, shadows, etc. However, reliable estimation of model parameters is a very difficult task.

8.2.3 Appearance-based face recognition approach

In contrast with the model-based approach, the appearance-based face recognition approach operates directly on 2-D facial image and processes them as 2-D holistic patterns to avoid difficulties associated with 3-D modeling, and shape or landmark detection. Consequently, this class of methods tends to be easier to implement, more practical and reliable [37]. Furthermore, evidences have been shown in the studies in visual neuroscience that facial features are processed holistically and facial recognition is dependent on holistic processes involving an interdependency between featural and configural information [38].

In this approach, the whole face region is the raw input to a recognition system and each face image is commonly represented by a high-dimensional vector consisting of the pixel intensity values in the image, i.e., a point in a high-dimensional vector space. Thus, face recognition is transformed to a multivariate statistical pattern recognition problem. Although the embedding is high-dimensional, the natural constraints of the face data indicate that the face vectors lie in a lower-dimensional subspace (manifold). The popular subspace learning is such a method to identify, represent, and parameterize this subspace with some optimality criteria [24].

8.2.4 Appearance-based gait recognition approach

The appearance-based gait recognition approach considers gait as a holistic pattern and uses a full-body representation of a human subject as silhouettes or contours. Most of the gait recognition algorithms proposed are appearance-based [11, 33, 35, 39–49]. Some works use the silhouettes directly as the gait representation [23, 50] and some others use the average silhouettes as the gait representation [43, 51]. These gait recognition algorithms extract structural or shape statistics as features from silhouettes, e.g., width [41], contours [11, 40], projections [44], and motion patterns [39, 52–54]. There are also methods based on dense optical flow [55], which identify individuals by periodic variations (phase features) in the shape of their motion. In addition, as in face recognition, a gait subspace can be learnt for gait recognition as well [56].

8.2.5 Fusion of face and gait for recognition

Besides using face or gait for recognition individually, these two biometric characteristics can also be fused to achieve multi-modal recognition with enhanced results. The fusion can be done either at the feature (or sensor) level [57] or at the decision (or matching score) level [58, 59]. In [60], face and gait features are integrated at the feature level, where the face and gait features are extracted separately and then they are normalized and

concatenated to form a single feature vector for recognition. Improvement due to fusion over recognition with a single modality is observed in this work. In [61], face and gait cues are derived from multiple simultaneous views and transformed to the canonical pose, frontal face and profile gait silhouettes. Then view-normalized face and gait sequences are used individually for recognition and the recognition results are combined using a number of conventional rules including the MAX, MIN, MEAN, PRODUCT and MAJORITY rules. The results presented in [61] indicate that the MEAN and PRODUCT rules have better results. In [62], two ways of decision-level face and gait fusion are explored. The first uses the gait recognition algorithm as a filter to pass on a smaller set of candidates for face recognition. The second combines similarity scores obtained separately from the face and gait recognition algorithms, with the SUM, MIN and PRODUCT rules tested. Their results indicate that the second way is more effective.

8.3 FACE AND GAIT DATA SETS

The development of face or gait recognition algorithms largely depends on the availability of large and representative public databases of face images or gait sequences so that algorithms can be compared and advancements can be measured. In this section, we review three widely used face and gait databases: the Pose, Illumination, and Expression (PIE) database from the Carnegie Mellon University (CMU) [63], the Facial Recognition Technology (FERET) database [22] and the HumanID Gait Challenge data sets from the University of South Florida (USF) [23].

In typical pattern recognition problems of face and gait recognition, there are usually two types of data sets: the gallery and the probe [22, 23]. The gallery set contains the set of data samples with known identities and it is used for training. The probe set is the testing set where data samples of unknown identity are to be identified and classified via matching with corresponding entries in the gallery set.

8.3.1 The PIE database

Visually perceived human faces are significantly affected by three factors: the pose, which is the angle at they are viewed from, the illumination/lighting conditions, and the facial expression such as happy, sad and anger. The collection of the PIE database is motivated by a need for a database with a fairly large number of subjects imaged a large number of times to cover these three significant factors, i.e., from a variety of different poses, under a wide range of illumination variation, and with several expressions [63].

This database was collected between October 2000 and December 2000 using the CMU 3D Room and it contains 41,368 face images from 68 individuals. Face images with 13 different poses are captured using 13 synchronized cameras. For the illumination variation, the 3D Room is augmented with a flash system having 21 flashes. Images are captured with and without background lighting, resulting in $21 \times 2 + 1 = 43$ different illumination conditions.

In addition, the subjects were asked to pose with four different expressions. Figure 8.3(a) shows 144 sample face images for one subject in this database.

The PIE database can be used for a variety of purposes, including evaluating the robustness of face recognition systems against the three variations and three-dimensional modeling. In particular, this database has a very large number (around 600 on average) of facial images available for each subject, allowing us to study the effects of the number of training samples (per subject) on the recognition performance. In practice, a subset is usually selected with a specific range of pose, illumination and expression for experiments so that data sets with various difficulties can be obtained where a wider range of the three variations leads to a more difficult recognition task.



(a)



(b)

Fig. 8.3 Examples of one subject from (a) the PIE face database and (b) the FERET database.

8.3.2 The FERET database

The FERET database is a widely used database for face recognition performance evaluation. It was constructed through the FERET program, which aims to develop automatic face recognition systems to assist security, intelligence, and law enforcement personnel in the performance of their duties [22]. The face images in this database

cover a wide range of variations in pose (viewpoint), illumination, facial expression, acquisition time, ethnicity and age.

The FERET database was collected in 15 sessions between August 1993 and July 1996 and it contains a total of 14,126 images from 1,199 individuals with views ranging from frontal to left and right profiles. The face images were collected under relatively unconstrained conditions. The same physical setup and location was used in each session to maintain a degree of consistency throughout the database. However, since the equipment was reassembled for each session, images collected on different dates have some minor variation. Sometimes, a second set of images of an individual was captured on a later date, resulting in variations in scale, pose, expression, and illumination of the face. Furthermore, for some people, over two years elapsed between their first and last capturing in order to study changes in a subject's facial appearance over a year. Figure 8.3(b) shows eight sample face images for one subject in this database. This database has a large number of subjects and it becomes the de facto standard for evaluating face recognition technologies [64], especially in the small sample size scenario, where a smaller number of training samples per subject and a larger number of total subjects lead to a more difficult recognition task [65].

8.3.3 The USF gait database

In the area of gait recognition, the Gait Challenge data sets from the USF captures the variations of a large number of covariates¹ for a large group of people and it has emerged as a standard testbed for new gait recognition algorithms. Other databases are limited in size, variations, capturing conditions, or of high resolutions [9, 23].

The Gait Challenge data sets consists of 1870 sequences from 122 subjects and a set of twelve outdoor experiments are constructed to examine the effects of five covariates on performance. In the collection, the subjects are asked to walk in elliptical paths in front of the camera. The image frame is of size 720×480 and a subject's size in the back portion of the ellipse is on average 100 pixels in height. Two sample frames are shown in Fig. 8.4.

The covariates either effect gait or effect the extraction of gait features from images. They are: change in viewing angle (left or right), change in shoe type (A or B), change in walking surface (Concrete or Grass), carrying or not carrying a briefcase and temporal (time) differences, where the time covariate implicitly includes other changes naturally occur between video acquisition sessions such as change of shoes and cloths, change in the outdoor lighting conditions, and inherent variation in gait over time. These covariates are selected (based on logistical issues and collection feasibility) from a larger list that was arrived at based on discussions with HumanID researchers at CMU, Maryland, MIT, Southampton and Georgia Tech about the potentially important covariates for gait analysis. It is shown in [23] that the shoe type has the least impact on the performance, next is the viewpoint, the third is briefcase, then surface type (flat concrete surface and typical grass lawn surface), and time (six months)

¹Covariates are random variables you treat as concomitants or as other influential variables that also affect the response.



Fig. 8.4 Two sample frames from the USF Gait Challenge data sets.

difference has the greatest impact. The latter two are the most “difficult” covariates to deal with. In particular, it was found that the surface covariate impacts the gait period more than other covariates. Since its release, this database has made significant contributions to the advancement of the gait recognition technology.

8.4 FACE AND GAIT RECOGNITION ALGORITHMS

In this section, we survey popular appearance-based face and gait recognition algorithms, and the focus is on subspace learning algorithms for feature extraction. We start from the classical linear subspace learning algorithms in Section 8.4.1. Then, we move on to the nonlinear kernel-based subspace learning algorithms that map the input to a high-dimensional space for (hopefully) better separation in Section 8.4.2. Next, Section 8.4.3 reviews recent multilinear subspace learning algorithms that operate directly on tensorial representations.

8.4.1 Linear subspace learning algorithms

The linear algorithms reviewed here include the principal component analysis (PCA), the Bayesian method and the linear discriminant analysis (LDA). Linear subspace learning algorithms solve for a linear projection with some optimality criteria, given a set of training samples. The problem can be formulated mathematically as follows.

A set of M vectorial samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ is available for training, where each sample \mathbf{x}_m is an $I \times 1$ vector in a vector space \mathbb{R}^I . The linear subspace learning objective is to find a linear transformation (projection) $\mathbf{U} \in \mathbb{R}^{I \times P}$ such that the projected samples (the extracted features) $\{\mathbf{y}_m = \mathbf{U}^T \mathbf{x}_m\}$ satisfy an optimality criterion, where $\mathbf{y}_m \in \mathbb{R}^{P \times 1}$ and $P < I$. In classification, these features are fed into a classifier, e.g., the nearest neighbor classifier, and the similarity is usually calculated based on some distance measure.

8.4.1.1 Principal component analysis The PCA is one of the most influential linear subspace learning methods. The well-known eigenface method [66] for face recognition, which is built on PCA, started the era of the appearance-based approach to face recognition, and more generally to visual object recognition. The central idea behind PCA is to reduce the dimensionality of a data set consisting of a larger number of interrelated variables, while retaining as much as possible the variation present in the original data set [67]. This is achieved by transforming to a new set of variables, the so-called principal components (PCs), which are uncorrelated, and ordered so that the first few retain most of the original data variation. Thus, the PCA aims to derive the most descriptive features.

In practice, the variation to be maximized is measured by the total scatter through the total scatter matrix \mathbf{S}_T defined as follows,

$$\mathbf{S}_T = \sum_{m=1}^M (\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T, \quad (8.1)$$

where $\bar{\mathbf{x}} = \frac{1}{M} \sum_{m=1}^M \mathbf{x}_m$ is the mean of all the training samples. The PCA projection matrix \mathbf{U}_{PCA} is then composed of the eigenvectors corresponding to the largest P ($P < I$) eigenvalues of \mathbf{S}_T . The projection of a test sample \mathbf{x} in the PCA space is obtained as:

$$\mathbf{y} = \mathbf{U}_{PCA}^T (\mathbf{x} - \bar{\mathbf{x}}). \quad (8.2)$$

The PCA is an unsupervised learning technique that does not take underlying class structure information into account, even when such information is available for use. In such cases, both intra-subject and inter-subject variations are maximized in the PCA feature space. For classification purposes, however, large intra-subject (intra-class) variations have negative impact on classification performance. Figure 8.5(a) shows two classes of data samples, represented by “o” and “ Δ ”, and the obtained PCA feature basis. It is not difficult to see that by projecting the data samples onto the PCA basis, these two classes are mixed together and become difficult to be separated. Therefore, it is generally believed that, for classification purposes, the PCA can not perform as well as those supervised learning techniques such as the LDA [68].

8.4.1.2 Bayesian method The Bayesian method [69] treats a multi-class recognition problem as a binary pattern classification problem. First, a feature space of Δ vectors is defined as the differences between two samples: $\Delta = \mathbf{x}_{m_1} - \mathbf{x}_{m_2}$. We can then define two class of variations: the intra-subject variation Ω_I (corresponding to different appearances of the same individual) or an inter-subject variation class Ω_E (corresponding to variations between different subjects). Let $\{\Delta_I\}$ be the intra-subject difference set consisting of difference vectors between samples from the same subject and $\{\Delta_E\}$ be the inter-subject difference set consisting of difference vectors between samples from different individuals, two feature subspaces denoted as intra-subject space and inter-subject space are then extracted by performing the PCA on $\{\Delta_I\}$ and $\{\Delta_E\}$, respectively. In the PCA subspace, the class conditional distributions of the two classes $p(\Delta|\Omega_I)$ and $p(\Delta|\Omega_E)$ are estimated using a Gaussian model. Thus, to determine if two samples \mathbf{x}_i and \mathbf{x}_j belong to the same individual, their difference $\Delta_{ij} = \mathbf{x}_i - \mathbf{x}_j$ is first calculated.

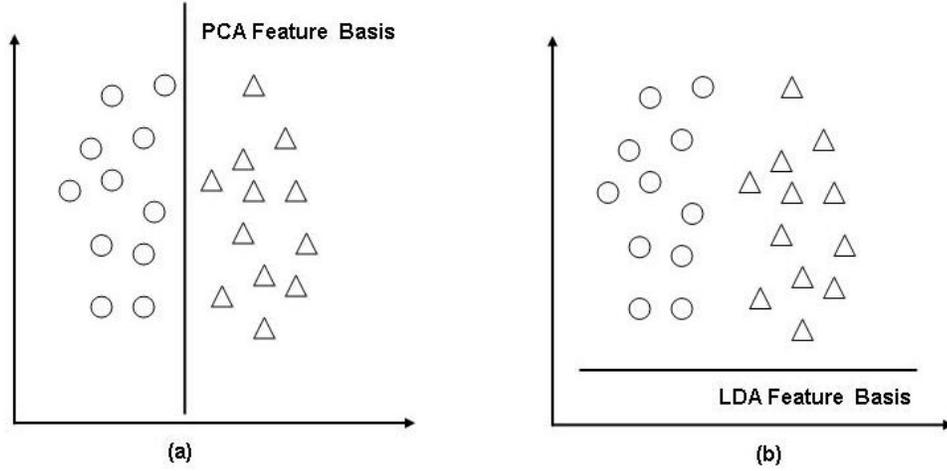


Fig. 8.5 2-D training samples from two classes \circ and \triangle and (a) the obtained PCA feature basis; (b) the obtained LDA feature basis.

The decision is then made according to the maximum-likelihood classification rule, i.e.,

$$\Delta_{ij} \in \Omega_I \quad \text{if} \quad p(\Delta_{ij}|\Omega_I) > p(\Delta_{ij}|\Omega_E) \quad (8.3)$$

or the *maximum-a-posteriori* rule, i.e.

$$\Delta_{ij} \in \Omega_I \quad \text{if} \quad P(\Omega_I)p(\Delta_{ij}|\Omega_I) > P(\Omega_E)p(\Delta_{ij}|\Omega_E), \quad (8.4)$$

where $P(\Omega_I)$ and $P(\Omega_E)$ are the priori probabilities.

8.4.1.3 Linear discriminant analysis The LDA is a classical supervised linear subspace learning method that has been very successful and applied widely in various applications. It aims to derive the most discriminative features and produces a class-specific feature space based on the maximization of the so-called Fisher's Discriminant Criterion, which is defined as the ratio of between-class scatter to within-class scatter:

$$\mathbf{U}_{LDA} = \arg \max_{\mathbf{U}} \frac{|\mathbf{U}^T \mathbf{S}_B \mathbf{U}|}{|\mathbf{U}^T \mathbf{S}_W \mathbf{U}|}, \quad (8.5)$$

where the \mathbf{S}_B and \mathbf{S}_W are the between-class and within-class scatter matrices, respectively, and they are defined as

$$\mathbf{S}_B = \sum_{c=1}^C M_c (\bar{\mathbf{x}}_c - \bar{\mathbf{x}})(\bar{\mathbf{x}}_c - \bar{\mathbf{x}})^T, \quad (8.6)$$

and

$$\mathbf{S}_W = \sum_{m=1}^M (\mathbf{x}_m - \bar{\mathbf{x}}_{c_m})(\mathbf{x}_m - \bar{\mathbf{x}}_{c_m})^T. \quad (8.7)$$

In the definitions above, C is the number of classes, c is the class index, and c_m is the class label for the m th training sample. M_c is the number of training samples in class c , and the mean for class c is

$$\bar{\mathbf{x}}_c = \frac{1}{M_c} \sum_{m, c_m=c} \mathbf{x}_m. \quad (8.8)$$

The maximization of (8.5) leads to the following generalized eigenvalue problem:

$$\mathbf{S}_B \mathbf{u}_p = \lambda_p \mathbf{S}_W \mathbf{u}_p. \quad (8.9)$$

Thus, \mathbf{U}_{LDA} consists of the generalized eigenvectors corresponding to the largest P generalized eigenvalues of (8.9). When \mathbf{S}_W is not singular, \mathbf{U}_{LDA} can be obtained as the eigenvectors corresponding to the largest P eigenvalues of $\mathbf{S}_W^{-1} \mathbf{S}_B$. The projection of a test sample \mathbf{x} in the LDA space is then obtained as:

$$\mathbf{y} = \mathbf{U}_{LDA}^T \mathbf{x}. \quad (8.10)$$

Figure 8.5(b) shows the LDA feature basis obtained from two classes of data samples (represented by “o” and “△” respectively). Obviously, compared to the PCA, the LDA provides a more ideal projection basis for classification purposes.

Although the LDA, as a class specific solution, is generally believed to be superior to the PCA for classification purposes, it is more susceptible to the small sample size problem, where the number of training samples per subject is much smaller than the dimensionality of the input sample space. In such cases, direct optimization of the ratio in (8.5) becomes impossible as \mathbf{S}_W is singular. In order to address this problem, the Fisherface method (FLDA) [68] proposes to apply the PCA as a pre-processing step to remove the null space of \mathbf{S}_W .

However, by removing the null space of \mathbf{S}_W , significant discriminatory information may be discarded since the maximum of (8.5) can be reached when

$$\mathbf{u}_p^T \mathbf{S}_W \mathbf{u}_p = 0 \text{ and } \mathbf{u}_p^T \mathbf{S}_B \mathbf{u}_p \neq 0. \quad (8.11)$$

In other words, the null space of \mathbf{S}_W spanned by \mathbf{u}_p such that $\mathbf{u}_p^T \mathbf{S}_W \mathbf{u}_p = 0$, may contain significant discriminatory information. To avoid the possible loss of discriminatory information residing in the discarded null space of \mathbf{S}_W , a direct linear discriminant analysis (DLDA) [70] was proposed to solve the small sample size problem by diagonalizing \mathbf{S}_B and \mathbf{S}_W directly. The premise behind the DLDA solution is that the discriminatory information resides in the intersection of the null space of \mathbf{S}_W , denoted as \mathbb{A} , and the complement space of the null space of \mathbf{S}_B , denoted as \mathbb{B}' , i.e., $\mathbb{A} \cap \mathbb{B}'$. At the same time, no significant information, in terms of maximization of (8.5), will be lost if the null space of \mathbf{S}_B (denoted as \mathbb{B}) is discarded. The DLDA can be performed by diagonalizing either \mathbf{S}_W first or \mathbf{S}_B first. Given the fact that under the small sample size scenario, the rank of \mathbf{S}_B , determined by $\text{rank}(\mathbf{S}_B) = \min(I, C - 1)$ is much smaller than I in most cases, the more attractive option is to diagonalize

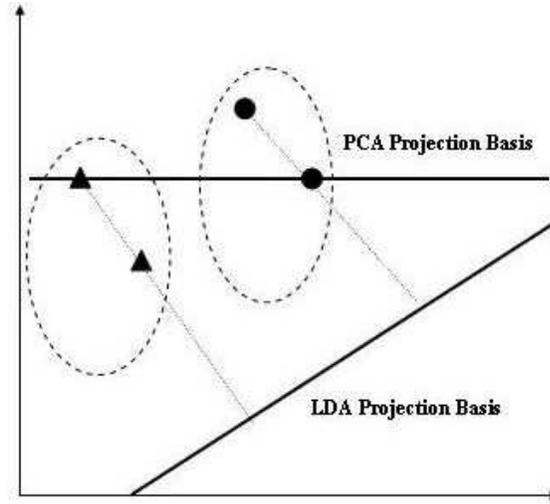


Fig. 8.6 An example to illustrate why, in some cases, the PCA produces better features than the LDA.

\mathbf{S}_B first. In such cases, \mathbb{B}' is first extracted which is spanned by the $C - 1$ eigenvectors of \mathbf{S}_B , denoted as $\mathbf{V}_b = [\mathbf{v}_{b1}, \dots, \mathbf{v}_{b(C-1)}]$, corresponding to the nonzero eigenvalues denoted as $[\lambda_{b1}, \dots, \lambda_{b(C-1)}]$. Following that, \mathbf{S}_W is projected to \mathbb{B}' giving $\mathbf{U}^T \mathbf{S}_W \mathbf{U}$, where $\mathbf{U} = \mathbf{V}_b \Lambda_b^{-1/2}$ is the transformation matrix from the \mathbb{R}^I to \mathbb{B}' , $\Lambda_b = \text{diag}([\lambda_{b1}, \dots, \lambda_{b(C-1)}])$ and $\text{diag}(\cdot)$ denotes the diagonalization operator. The null space of \mathbf{S}_W thus can be easily found by solving an eigenvalue problem of $\mathbf{U}^T \mathbf{S}_w \mathbf{U}$ in a low dimensional \mathbb{B}' space.

In general, the LDA-based solutions are believed to be superior to the PCA-based solutions in the context of pattern classification. However, it should be noted that when the number of training samples per subject is small or the training samples are not representative to those in the test, the superiority of LDA technique can not be guaranteed and PCA may outperform LDA, as shown in Figure 8.6 (adapted from [71]). There are two classes, each of which has a Gaussian-like class conditional distribution (represented by two ellipses in the figure). However, only two samples per class are provided for training. The PCA method calculates the projection basis such that the variance of the projected training samples is maximized, resulting in a horizontal projection basis as illustrated in the figure. The LDA method, however, extracts the projection basis such that the between-class variance of the projected samples is maximized while the within-class variance is minimized. It can be observed from the figure, in such cases, the projection basis obtained by the PCA solution is more desirable than that obtained by the LDA method for classification purposes.

8.4.2 Nonlinear kernel-based subspace learning algorithms

The LDA-based methods have been shown to be successful and cost-effective techniques widely used in pattern recognition applications. However, as the complexity of patterns increases, the performance of the LDA-based

Table 8.2 Commonly used kernel functions

Kernel name	Kernel function
Gaussian	$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{\sigma^2})$, $\sigma \in \mathbb{R}$
Polynomial	$k(\mathbf{x}_i, \mathbf{x}_j) = (a(\mathbf{x}_i \cdot \mathbf{x}_j) + b)^d$, $a, b, d \in \mathbb{R}$
Sigmoidal	$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a(\mathbf{x}_i \cdot \mathbf{x}_j) + b)$, $a, b \in \mathbb{R}$

solutions could deteriorate rapidly. This is due to the linear nature of the LDA-based methods assuming that the class conditional distribution of patterns is Gaussian with identical covariance structure. In such cases, the resulting class boundaries are restricted to linear hyper-planes. In appearance-based face and gait recognition, due to the large variations in the appearance, however, the distribution of patterns is far more complicated than Gaussian, usually multi-modal and non-convex [2]. To this end, nonlinear, kernel-based algorithms have been developed to handle complexly distributed data.

Originating from the well-known support vector machine [72], the so-called “kernel machine” technique is considered an important tool in the design of nonlinear feature extraction techniques. The premise behind the kernel machine technique is to find a nonlinear mapping from the original input space (\mathbb{R}^I) to a higher dimensional kernel feature space \mathbb{F}^F by using a nonlinear function $\phi(\cdot)$, i.e.,

$$\phi : \mathbf{x} \in \mathbb{R}^I \rightarrow \phi(\mathbf{x}) \in \mathbb{F}^F \quad I < F \leq \infty. \quad (8.12)$$

In the kernel feature space \mathbb{F}^F , the pattern distribution is expected to be simplified so that better classification performance can be achieved by applying traditional linear methodologies [73]. In general, the dimensionality of the kernel space is much larger than that of the original input space, sometimes even infinite. Therefore, an explicit determination of the nonlinear map ϕ is difficult or intractable. Fortunately, with the so-called “kernel trick”, the nonlinear mapping can be performed implicitly in the original input space \mathbb{R}^I by replacing dot products of the feature representations in \mathbb{F}^F with a kernel function defined in \mathbb{R}^I [73]. Thus, if $\mathbf{x}_i \in \mathbb{R}^I$ and $\mathbf{x}_j \in \mathbb{R}^I$ are two vectors in the original input space, the dot product of their feature representations $\phi(\mathbf{x}_i) \in \mathbb{F}^F$ and $\phi(\mathbf{x}_j) \in \mathbb{F}^F$ can be computed by a kernel function $k(\cdot)$ defined in \mathbb{R}^I , i.e.,

$$\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j). \quad (8.13)$$

The function selected as the kernel function should satisfy the Mercer’s condition [73]. Some commonly used kernel functions are summarized in Table 8.2. A toy example [20] is illustrated in Figure 8.7, where a second order

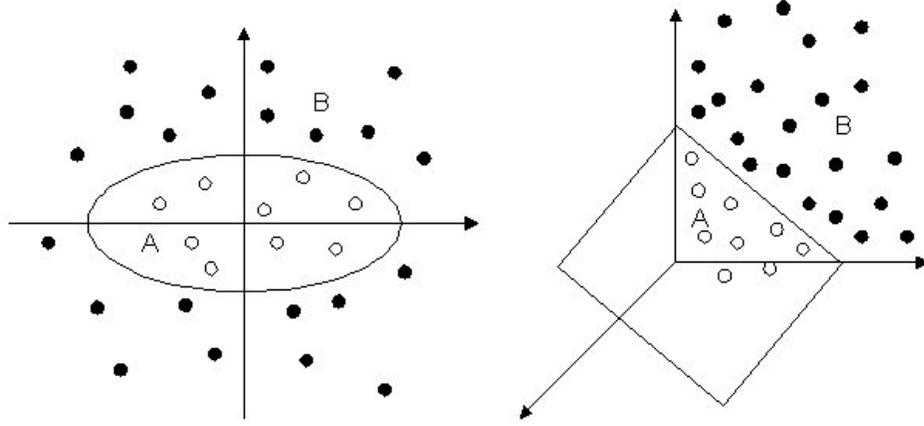


Fig. 8.7 A two-class classification example showing non-linear and linear decision boundaries. Left: samples in the 2-D input space; Right: samples in the 3-D kernel feature space

polynomial kernel function is used, i.e.,

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^2, \quad \phi(\mathbf{x}) = [\alpha_1^2, \sqrt{2}\alpha_1\alpha_2, \alpha_2^2]^T, \quad (8.14)$$

where $\mathbf{x} = [\alpha_1, \alpha_2]^T$. It can be easily observed that by mapping the samples into the 3-D kernel space, use of a linear hyperplane to separate the two classes becomes more efficient than using a nonlinear ellipsoidal decision boundary in the original space.

The kernel-based solutions have been shown to be effective techniques to handle complicated pattern classification problems. However, the performance of the algorithm is significantly affected by the selected kernel functions and the corresponding kernel parameters. A toy example is given in Figure 8.8 which depicts the data distribution in different kernel spaces. Figure 8.8(a) shows a 2-D data set of 400 samples drawn from two Gaussian distributions (denoted as “+” and “o” respectively), i.e., $\mathbf{x} \sim \mathcal{N}(\mu_i, \Sigma_i)$, $i = 1, 2$, where \mathbf{x} denotes the 2-D data, $\mu_1 = [-2, 0]^T$, $\mu_2 = [2, 0]^T$, $\Sigma_1 = \Sigma_2 = \text{diag}([2.25, 1])$. Each class contains 200 samples. Figure 8.8(b) shows the projection of the data in a second order polynomial kernel space, i.e., $k(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 \cdot \mathbf{x}_2)^2$. For better visual perception, the projection on the first two significant bases are depicted. Figures 8.8(c) and 8.8(d) illustrate the corresponding projection in a Gaussian kernel space when a Gaussian kernel function

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(\frac{-\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\sigma^2}\right) \quad (8.15)$$

is employed with $\sigma^2 = 1000$ for Fig. 8.8(c) and $\sigma^2 = 10$ for Fig. 8.8(d). It is not difficult to observe that different kernel functions and different kernel parameters significantly affect the geometric structure of the mapped data. If a polynomial kernel is used as shown in the Figure 8.8(b), the class separability in the kernel space is even worse than that in the original input space. Therefore, for kernel-based solutions, selecting an appropriate kernel function

is of vital importance to ensure good classification performance and methods have been proposed to tackle this problem [74]. The traditional approach is the cross validation framework and there are more systematic approaches for kernel parameter optimization [74].

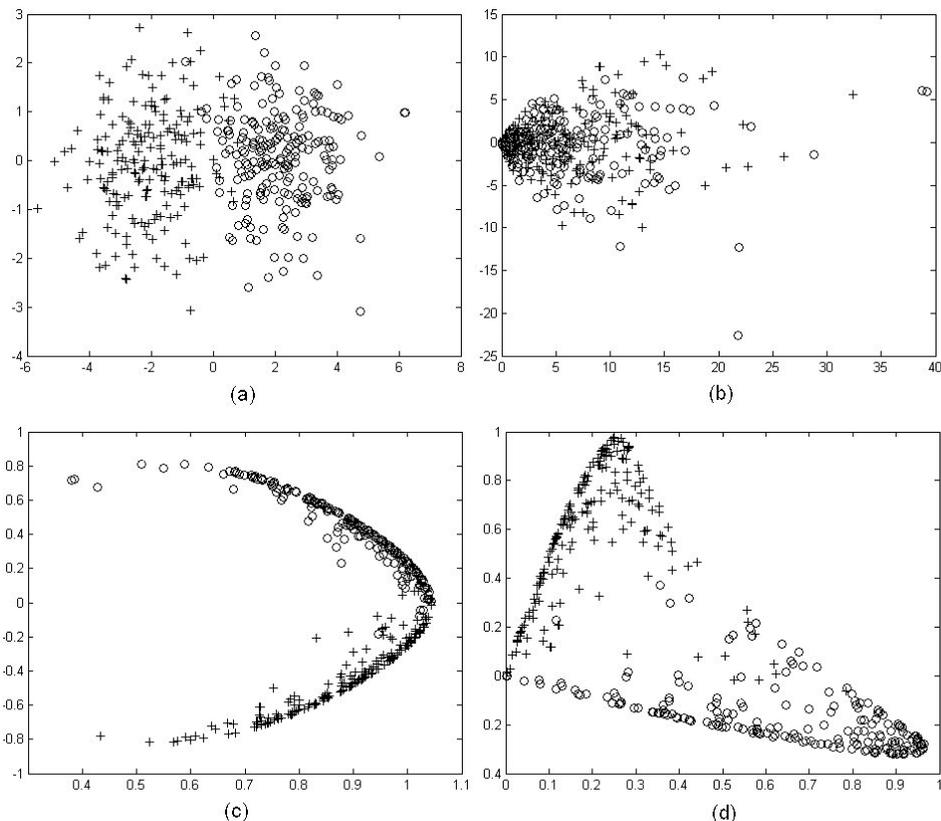


Fig. 8.8 An example showing projections of a set of 2-D data points in different kernel spaces. (a): 2-D samples drawn from two Gaussian distributions; (b): projections in the polynomial kernel space; (c): projections in the Gaussian kernel space with $\sigma^2 = 1000$; (d): projections in the Gaussian kernel space with $\sigma^2 = 10$.

In the context of the kernel trick, the key task of designing a kernelized feature extraction algorithm is to represent the linear feature extraction procedure using dot product forms. The corresponding kernel version of the algorithm can then be obtained by replacing the dot product with the kernel function. The kernel principal component analysis (KPCA) [75], the generalized discriminant analysis (GDA) [76] and the kernel direct discriminant analysis (KDDA) [77] are typical kernel-based learning algorithms commonly used for feature extraction and they are described below.

8.4.2.1 Kernel principal component analysis The KPCA [75] is actually an implementation of the traditional PCA algorithm in the kernel feature space. Let $\tilde{\mathbf{S}}_T$ be the total scatter matrix defined in \mathbb{F}^F , which could be

expressed as follows,

$$\tilde{\mathbf{S}}_T = \sum_{m=1}^M (\phi(\mathbf{x}_m) - \bar{\phi})(\phi(\mathbf{x}_m) - \bar{\phi})^T \quad (8.16)$$

where

$$\bar{\phi} = \frac{1}{M} \sum_{m=1}^M \phi(\mathbf{x}_m) \quad (8.17)$$

is the mean of gallery samples in \mathbb{F}^F . Thus, the KPCA subspace is spanned by the first P significant eigenvectors of $\tilde{\mathbf{S}}_T$, denoted as $\tilde{\mathbf{U}}_{KPCA}$, corresponding to the P largest eigenvalues. The KPCA feature representation of a given input \mathbf{x} is thus obtained by the dot product,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{U}}_{KPCA} \cdot (\phi(\mathbf{x}) - \bar{\phi}), \quad (8.18)$$

computed implicitly through the kernel function $k(\cdot)$ [75].

8.4.2.2 Kernel-based discriminant analysis (KDA) The GDA [76] and the KDDA [77] are two kernel-based discriminant analysis solutions which produce corresponding subspaces by maximizing the Fisher's criterion defined in \mathbb{F}^F ,

$$\tilde{\mathbf{U}}_{KDA} = \arg \max_{\tilde{\mathbf{U}}} \frac{|\tilde{\mathbf{U}}^T \tilde{\mathbf{S}}_B \tilde{\mathbf{U}}|}{|\tilde{\mathbf{U}}^T \tilde{\mathbf{S}}_W \tilde{\mathbf{U}}|} = [\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_P], \quad (8.19)$$

where $\tilde{\mathbf{u}}_p$ is the p th generalized eigenvector of

$$\tilde{\mathbf{S}}_B \tilde{\mathbf{u}}_p = \tilde{\lambda}_p \tilde{\mathbf{S}}_W \tilde{\mathbf{u}}_p, \quad (8.20)$$

and the corresponding between-class matrix $\tilde{\mathbf{S}}_B$ and within-class scatter matrix $\tilde{\mathbf{S}}_W$ are defined as

$$\tilde{\mathbf{S}}_B = \sum_{c=1}^C M_c (\bar{\phi}_c - \bar{\phi})(\bar{\phi}_c - \bar{\phi})^T \quad (8.21)$$

and

$$\tilde{\mathbf{S}}_W = \sum_{m=1}^M (\phi(\mathbf{x}_m) - \bar{\phi}_{c_m})(\phi(\mathbf{x}_m) - \bar{\phi}_{c_m})^T \quad (8.22)$$

where

$$\bar{\phi}_c = \frac{1}{M_c} \sum_{m, c_m=c} \phi(\mathbf{x}_m) \quad (8.23)$$

is the class mean in \mathbb{F}^F .

As discussed earlier, the LDA-based solutions often suffer from the small sample size problem. When applying the LDA techniques in the kernel feature space, the problem becomes even more severe due to the extremely high dimensionality of \mathbb{F}^F . Therefore, solving the small sample size problem is also a demanding task in the implementation of kernel-based discriminant analysis solutions. GDA attempts to solve the small sample size

problem by removing the null space of $\tilde{\mathbf{S}}_w$ with a PCA routine as is done in the FLDA method, while the KDDA implements the DLDA solution in the kernel space.

8.4.3 Multilinear subspace learning algorithms

The algorithms reviewed so far all take vectorial input. However, gray-level face images (*row* \times *column*) and binary gait silhouette sequences (*row* \times *column* \times *time*) are naturally multidimensional objects, which are formally called tensor objects. Therefore, the linear and nonlinear algorithms above need to reshape these tensors into vectors in a very high-dimensional space, which not only results in high computation and memory demand, but also breaks the natural structure and correlation in the original data. This motivated the recent development of the multilinear subspace learning algorithms [78–84], which extract features directly from the tensorial representation rather than the vectorized representation, and it is believed that more compact and useful features can be obtained this way.

Tensors are conventionally denoted by calligraphic letters [85], e.g., \mathcal{A} . The elements of a tensor are to be addressed by N indices, where N (the number of indices used in the description) defines the order of the tensor object and each index defines one mode [85]. Thus, vectors are first-order tensors (with $N = 1$) and matrices are second-order tensors (with $N = 2$). Tensors with $N > 2$ can be viewed as a generalization of vectors and matrices to higher order.

The elements of a tensor are denoted with indices in brackets. Indices are denoted by lowercase letters and span the range from 1 to the uppercase letter of the index, e.g., $n = 1, 2, \dots, N$. An N th-order tensor is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. It is addressed by N indices i_n , $n = 1, \dots, N$, and each i_n addresses the n -mode of \mathcal{A} . The n -mode product of a tensor \mathcal{A} by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is a tensor with entries:

$$(\mathcal{A} \times_n \mathbf{U})(i_1, \dots, i_{n-1}, j_n, i_{n+1}, \dots, i_N) = \sum_{i_n} \mathcal{A}(i_1, \dots, i_N) \cdot \mathbf{U}(j_n, i_n). \quad (8.24)$$

The scalar product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined as:

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} \mathcal{A}(i_1, i_2, \dots, i_N) \cdot \mathcal{B}(i_1, i_2, \dots, i_N) \quad (8.25)$$

and the Frobenius norm of \mathcal{A} is defined as

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}. \quad (8.26)$$

The “ n -mode vectors” of \mathcal{A} are defined as the I_n -dimensional vectors obtained from \mathcal{A} by varying the index i_n while keeping all the other indices fixed. A rank-1 tensor \mathcal{A} equals to the outer product of N vectors:

$$\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \dots \circ \mathbf{u}^{(N)}, \quad (8.27)$$

which means that

$$\mathcal{A}(i_1, i_2, \dots, i_N) = \mathbf{u}^{(1)}(i_1) \cdot \mathbf{u}^{(2)}(i_2) \cdot \dots \cdot \mathbf{u}^{(N)}(i_N) \quad (8.28)$$

for all values of indices. Unfolding \mathcal{A} along the n -mode is denoted as

$$\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N)}, \quad (8.29)$$

and the column vectors of $\mathbf{A}_{(n)}$ are the n -mode vectors of \mathcal{A} .

Based on the definitions above, a tensor can be projected to another tensor by N projection matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}$ as

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \dots \times_N \mathbf{U}^{(N)T}. \quad (8.30)$$

The projection of an n -mode vector of \mathcal{X} by an n -mode projection matrix $\mathbf{U}^{(n)T}$ is computed as the inner product between the n -mode vector and the rows of $\mathbf{U}^{(n)T}$. Figure 8.9 illustrate the 1-mode projection of a third-order tensor $\mathcal{X} \in \mathbb{R}^{10 \times 8 \times 6}$ in the 1-mode vector space by a projection matrix $\mathbf{U}^{(1)T} \in \mathbb{R}^{5 \times 10}$, resulting in the projected tensor $\mathcal{X} \times_1 \mathbf{U}^{(1)T} \in \mathbb{R}^{5 \times 8 \times 6}$. In the 1-mode projection, each 1-mode vector of \mathcal{X} of length 10 is projected by $\mathbf{U}^{(1)T}$ to obtain a vector of length 5, as indicated by the differently shaded vectors in Fig. 8.9.

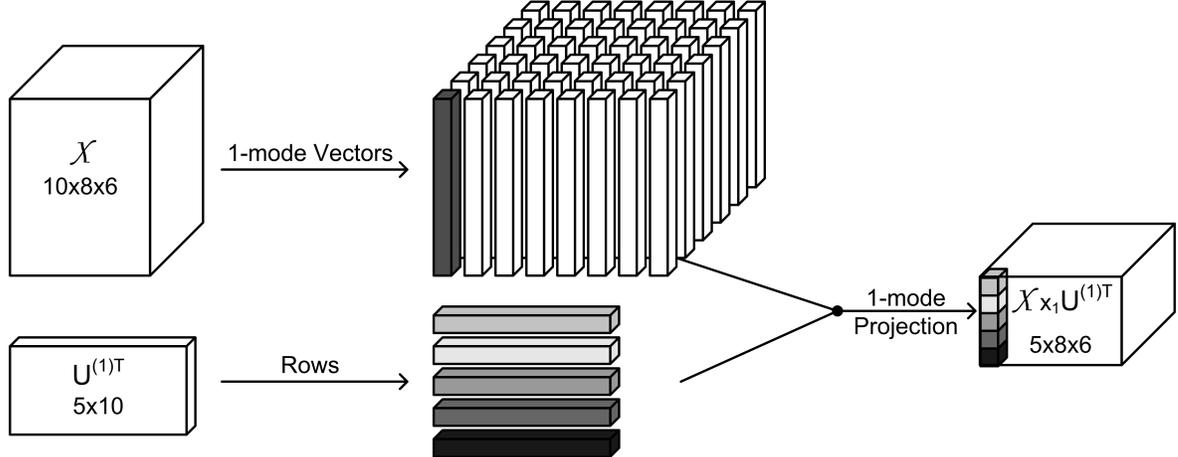


Fig. 8.9 An illustration of the 1-mode projection: $\mathcal{X} \times_1 \mathbf{U}^{(1)T}$.

The problem of multilinear subspace learning based on the tensor-to-tensor projection above can be mathematically defined as follows:

A set of M N th-order tensorial samples $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_M\}$ is available for training, where each sample \mathcal{X}_m is an $I_1 \times I_2 \times \dots \times I_N$ tensor in a tensor space $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. The objective of multilinear subspace learning through tensor-to-tensor projection is to find a tensor-to-tensor projection $\{\tilde{\mathbf{U}}^{(n)} \in \mathbb{R}^{I_n \times P_n}, n = 1, \dots, N\}$ mapping from the original tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$ into a tensor subspace $\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \dots \otimes \mathbb{R}^{P_N}$ (with $P_n < I_n$,

for $n = 1, \dots, N$):

$$\mathcal{Y}_m = \mathcal{X}_m \times_1 \tilde{\mathbf{U}}^{(1)T} \times_2 \tilde{\mathbf{U}}^{(2)T} \dots \times_N \tilde{\mathbf{U}}^{(N)T}, m = 1, \dots, M, \quad (8.31)$$

such that the projected samples (the extracted features) satisfy an optimality criterion, where the dimensionality of the projected space is much lower than the original tensor space. Here, \otimes denotes the Kronecker product. This problem is usually solved in an iterative alternating projection manner. In classification, these tensorial features can be fed directly into a classifier, e.g., the nearest neighbor classifier, and the similarity is calculated according to some tensorial distance measure, e.g. based on the Frobenius norm in (8.26), or they can be converted into vectors before the feeding so that more conventional distance measures and traditional classifiers can be used.

In the following, we describe two multilinear subspace learning algorithms. One is the multilinear principal component analysis (MPCA) and the other is a multilinear discriminant analysis algorithm.

8.4.3.1 Multilinear principal component analysis The recent MPCA algorithm is a multilinear extension of the PCA algorithm and it has been applied successfully to gait recognition [56]. The MPCA maximizes the following tensor-based scatter measure:

$$\Psi_{\mathcal{Y}} = \sum_{m=1}^M \|\mathcal{Y}_m - \bar{\mathcal{Y}}\|_F^2, \quad (8.32)$$

named as the total tensor scatter, where

$$\bar{\mathcal{Y}} = \frac{1}{M} \sum_{m=1}^M \mathcal{Y}_m \quad (8.33)$$

is the mean sample.

This problem cannot be solved deterministically and it is decomposed into N simpler linear optimization problems, resulting in an iterative alternating projection procedure. In each subproblem, the following eigenvalue problem is solved: Given all the other projection matrices $\tilde{\mathbf{U}}^{(1)}, \dots, \tilde{\mathbf{U}}^{(n-1)}, \tilde{\mathbf{U}}^{(n+1)}, \dots, \tilde{\mathbf{U}}^{(N)}$, the matrix $\tilde{\mathbf{U}}^{(n)}$ that maximizes $\Psi_{\mathcal{Y}}$ consists of the P_n eigenvectors corresponding to the largest P_n eigenvalues of the matrix

$$\Phi^{(n)} = \sum_{m=1}^M (\mathbf{X}_{m(n)} - \bar{\mathbf{X}}^{(n)}) \cdot \tilde{\mathbf{U}}_{\Phi^{(n)}} \cdot \tilde{\mathbf{U}}_{\Phi^{(n)}}^T \cdot (\mathbf{X}_{m(n)} - \bar{\mathbf{X}}^{(n)})^T, \quad (8.34)$$

where

$$\tilde{\mathbf{U}}_{\Phi^{(n)}} = \left(\tilde{\mathbf{U}}^{(n+1)} \otimes \tilde{\mathbf{U}}^{(n+2)} \otimes \dots \otimes \tilde{\mathbf{U}}^{(N)} \otimes \tilde{\mathbf{U}}^{(1)} \otimes \tilde{\mathbf{U}}^{(2)} \otimes \dots \otimes \tilde{\mathbf{U}}^{(n-1)} \right). \quad (8.35)$$

The obtained projection matrices $\{\tilde{\mathbf{U}}^{(n)}, n = 1, \dots, N\}$ can be viewed as $\prod_{n=1}^N P_n$ EigenTensors: $\tilde{\mathcal{U}}_{p_1 p_2 \dots p_N} = \tilde{\mathbf{u}}_{p_1}^{(1)} \circ \tilde{\mathbf{u}}_{p_2}^{(2)} \circ \dots \circ \tilde{\mathbf{u}}_{p_N}^{(N)}$, where $\tilde{\mathbf{u}}_{p_n}^{(n)}$ is the p_n th column of $\tilde{\mathbf{U}}^{(n)}$. However, not all of them are useful for recognition and they can be selected according to their class discriminability $\Gamma_{p_1 p_2 \dots p_N}$, where $\Gamma_{p_1 p_2 \dots p_N}$ for the eigentensor

$\tilde{\mathcal{U}}_{p_1 p_2 \dots p_N}$ is defined as

$$\Gamma_{p_1 p_2 \dots p_N} = \frac{\sum_{c=1}^C M_c \cdot [\bar{\mathcal{Y}}_c(p_1, p_2, \dots, p_N) - \bar{\mathcal{Y}}(p_1, p_2, \dots, p_N)]^2}{\sum_{m=1}^M [\mathcal{Y}_m(p_1, p_2, \dots, p_N) - \bar{\mathcal{Y}}_{c_m}(p_1, p_2, \dots, p_N)]^2}. \quad (8.36)$$

\mathcal{Y}_m is the feature tensor of \mathcal{X}_m in the projected tensor subspace, and the class mean feature tensor

$$\bar{\mathcal{Y}}_c = \frac{1}{M_c} \sum_{m, c_m=c} \mathcal{Y}_m. \quad (8.37)$$

The entries in \mathcal{Y}_m are then arranged into a feature vector \mathbf{y}_m ordered according to $\Gamma_{p_1 p_2 \dots p_N}$ in descending order and only the first P entries of \mathbf{y}_m are kept. This feature vector can then be used directly as the input to conventional classifier or it can be further combined with the LDA (or other vector-based learning algorithms) to produce an MPCA+LDA algorithm [86].

8.4.3.2 Multilinear discriminant analysis There are also multilinear extensions of the LDA algorithm [79–81]. The Discriminant Analysis with Tensor Representation (DATER)² was proposed to perform discriminant analysis directly on general tensor objects. Like in MPCA, the DATER algorithm solves for a tensor-to-tensor projection $\{\tilde{\mathbf{U}}^{(n)} \in \mathbb{R}^{I_n \times P_n}, P_n < I_n, n = 1, \dots, N\}$ that project a tensor $\mathcal{X}_m \in \mathbb{R}^{I_1 \times \dots \times I_N}$ to \mathcal{Y}_m . The DATER algorithm maximizes a tensor-based discrimination objective criterion: the tensor-based scatter ratio $\Psi_{B_{\mathcal{Y}}} / \Psi_{W_{\mathcal{Y}}}$, where the between-class scatter $\Psi_{B_{\mathcal{Y}}}$ is defined as:

$$\Psi_{B_{\mathcal{Y}}} = \sum_{c=1}^C M_c \|\bar{\mathcal{Y}}_c - \bar{\mathcal{Y}}\|_F^2, \quad (8.38)$$

and the within-class scatter $\Psi_{W_{\mathcal{Y}}}$ is defined as:

$$\Psi_{W_{\mathcal{Y}}} = \sum_{m=1}^M \|\mathcal{Y}_m - \bar{\mathcal{Y}}_{c_m}\|_F^2. \quad (8.39)$$

Similar to the case of the MPCA, a deterministic solution does not exist either and an iterative alternating projection procedure is followed. In addition, the method to obtain vectorial features $\{\mathbf{y}_m\}$ in MPCA above can also be applied to this algorithm. In its application to gait recognition [87], it is found that this algorithm does not converge and it appears to be very sensitive to parameter settings.

²Here, the name that was used when the algorithm was first proposed is adopted as it is more commonly refereed to in the literature.

8.5 CONCLUDING REMARKS

In this section, we give a summary of this chapter and then discuss the current state and future directions of face and gait recognition.

8.5.1 Summary

This chapter presents a comprehensive review on the face and gait recognition technologies. The motivations and applications of face and gait recognition have been described in detail. Face and gait are both biometric characteristics that can be easily collected and well accepted by the public. Therefore, they have great potential in various applications. Face and gait recognition systems typically acquire images/videos of a subject through a digital camera. The acquired images/videos are then preprocessed to either detect facial images or extract gait silhouettes, from which features are extracted and classified. The two general approaches of model-based and appearance-based approaches are then reviewed, with their advantages and limitations pointed out. In addition, existing methods for the fusion of face and gait traits are also discussed. After the system-level overview, we move on to describe three widely used databases for research in face and gait recognition. Next, the important subspace learning algorithms for appearance-based recognition are surveyed in detail, including classical linear subspace learning algorithms such as the PCA and the LDA, and their nonlinear (kernel-based) and multilinear extensions. The intent here is not to discuss thoroughly all the face and gait recognition algorithms. Instead, we concentrate on the most representative algorithms in our view.

8.5.2 Current state and future directions of face recognition

Over the past two decades, face recognition has received substantial attention from both research communities and the market, and the recognition performance has improved significantly. Although progress in face recognition is encouraging, this task remains to be very challenging for real applications, especially for less constrained situations (e.g., outdoor) or even unconstrained situations where there are considerable variations in the viewpoint, illumination, expression and time [2]. Current face recognition technologies may work well for cooperative frontal faces without exaggerated expressions and under illumination without much shadow, but recognition in an unconstrained daily life environment without user cooperation, such as identifying persons in crowds and in surveillance videos, is still a very challenging problem that offers great opportunities for researchers in all related areas [2]. The following are some new developments and trends in face recognition.

Three-dimensional face recognition [88], where the 3-D geometry of the human face is used, is believed to have the potential to achieve better accuracy than its 2-D counterpart, which is sensitive to changes in variations like pose, illumination, and facial expressions. The 3-D face shape is usually extracted by a range camera and the 3-D

model can also be used to improve accuracy of traditional image-based recognition by transforming the head into a preferred view. Furthermore, most range cameras acquire both a 3-D mesh and the corresponding texture, which allows the combination of the pure 3-D matcher output with traditional 2-D face recognition algorithms for better performance [89].

Some researchers believe that the use of video sequences, as opposed to a single image, will lead to much better recognition rates [12, 90]. Their argument is that integrating the recognition performance over a sequence would give a better result than considering just one single image from that sequence.

The Face Recognition Grand Challenge (FRGC) [91] is recently designed to advance face recognition by presenting a six-experiment challenge problem to researchers, along with data corpus of 50,000 images. The data consists of 3-D scans and high resolution still imagery taken under controlled and uncontrolled conditions. The experiments can be used to measure performance on controlled images, uncontrolled images, 3-D images, multi-still images and the matching between 3-D and still images. Having seen the contributions made by the public databases such as PIE and FERET, it is expected that this FRGC database will further advance the state-of-the-art for face recognition.

8.5.3 Current state and future directions of gait recognition

Gait is a new biometric characteristics receiving increasing attention. It has the unique advantage over the other biometrics in surveillance applications, where the recognition needs to be performed at a distance and only low-resolutions videos can be captured. On one hand, the state-of-the-art gait recognition algorithms have achieved high recognition rates on gait sequences captured indoors with controlled environment (such as controlled lighting, background clutter) or sequences captured outdoor under the same surface, with variation in viewing angle and shoe type. On the other hand, the recognition on sequences captured under uncontrolled environment, such as outdoor with different surfaces, different carrying conditions, and different time, is still very challenging, thus offering great opportunities to researchers. In particular, the recent development in multilinear subspace learning algorithms is encouraging since they can handle natural gait sequences directly.

The uniqueness of gait is lower than the other biometrics such as face and fingerprint. Thus, in practice, a gait recognition system may not be used alone. Instead, it can be deployed as part of a multi-modal biometric recognition system. Besides security surveillance applications, studies on gait, especially the model-based approach, can benefit the entertainment and communication industries, and also the medical field, such as clinical rehabilitation of patients of stroke or spinal cord injuries and diagnosis of disorders [12, 92].

References

1. A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
2. S. Z. Li and A. K. Jain, "Introduction," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds. Springer-Verlag, 2004, pp. 1–11.
3. Cognitec systems gmbh|the face recognition company. [Online]. Available: <http://www.cognitec-systems.de/>
4. Eyematic interfaces inc. [Online]. Available: <http://www.siggraph.org/s2002/exhibition/detail/225.html>
5. L-1 identity solutions. [Online]. Available: <http://www.l1id.com/>
6. R. Hietmeyer, "Biometric identification promises fast and secure processing of airline passengers," *The International Civil Aviation Organization Journal*, vol. 55, no. 9, pp. 10–11, 2000.
7. W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey," *ACM Computing Surveys*, pp. 399–458, 2003.
8. R.-L. Hsu, "Face detection and modeling for recognition," Ph.D. dissertation, Michigan State University, 2002. [Online]. Available: <http://biometrics.cse.msu.edu/Publications/Thesis/VincentHsu.FaceDetection.PhD02.pdf>
9. M. S. Nixon and J. N. Carter, "Automatic recognition by gait," *Proc. IEEE*, vol. 94, no. 11, pp. 2013–2024, Nov. 2006.
10. A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
11. L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.
12. R. Chellappa, A. Roy-Chowdhury, and S. Zhou, *Recognition of Humans and Their Activities Using Video*. Morgan & Claypool Publishers, 2005.

28 REFERENCES

13. A. Kale, "Algorithms for gait-based human identification from a monocular video sequences," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Maryland College Park, 2003. [Online]. Available: <http://www.cs.uky.edu/~amit/thesis.pdf>
14. G. Johansson, "Visual motion perception," *Science American*, vol. 232, no. 6, pp. 76–88, 1975.
15. J. Cutting and L. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, no. 5, pp. 353–356, 1977.
16. C. Barclay, J. Cutting, and L. Kozlowski, "Temporal and spatial factors in gait perception that influence gender recognition," *Perception and Psychophysics*, vol. 23, no. 2, pp. 145–152, 1978.
17. S. V. Stevenage, M. S. Nixon, and K. Vince, "Visual analysis of gait as a cue to identity," *Applied Cognitive Psychology*, vol. 13, no. 6, pp. 513–526, 1999.
18. A. K. Jain, R. Chellappa, S. C. Draper, N. Memon, P. J. Phillips, and A. Vetro, "Signal processing for biometric systems," *IEEE Signal Processing Mag.*, vol. 24, no. 6, pp. 146–152, Nov. 2007.
19. L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, Feb. 2004.
20. J. Lu, "Discriminant learning for face recognition," Ph.D. dissertation, University of Toronto, 2004. [Online]. Available: <http://www.dsp.utoronto.ca/juwei/Publication/JuweiThesisUT04.pdf>
21. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Coarse-to-fine pedestrian localization and silhouette extraction for the gait challenge data sets," in *Proc. IEEE Conf. on Multimedia and Expo*, July 2006, pp. 1009–1012.
22. P. J. Phillips, H. Moon, S. A. Rizvi, and P. Rauss, "The FERET evaluation method for face recognition algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
23. S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. W. Bowyer, "The human ID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
24. X. Lu, "Image analysis for face recognition," *personal notes*, May 2003, 36 pages. [Online]. Available: http://www.face-rec.org/interesting-papers/General/ImAna4FacRcg_lu.pdf
25. F. Samaria and S. Young, "HMM based architecture for face identification," *Image and Vision Computing*, vol. 12, pp. 537–583, 1994.

26. L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 775–779, July 1997.
27. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 681–685, June 2001.
28. C. Y. Yam, M. S. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1057–1072, May 2004.
29. D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes," *Computer Vision and Image Understanding*, vol. 90, no. 1, pp. 1–41, Jan. 2003.
30. D. K. Wagg and M. S. Nixon, "On automated model-based extraction and analysis of gait," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, May 2004, pp. 11–16.
31. D. A. Winter, *The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological*. University of Waterloo Press, 1991.
32. ———, *The Biomechanics and Motor Control of Human Movement*. John Wiley & Sons, 2005.
33. L. Lee, G. Dalley, and K. Tieu, "Learning pedestrian models for silhouette refinement," in *Proc. IEEE Conf. on Computer Vision*, Oct. 2003, pp. 663–670.
34. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A layered deformable model for gait analysis," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Apr. 2006, pp. 249 – 254.
35. Z. Liu and S. Sarkar, "Effect of silhouette quality on hard problems in gait recognition," *IEEE Trans. Syst., Man, Cybern. B*, vol. 35, no. 2, pp. 170–178, 2005.
36. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A full-body layered deformable model for automatic model-based gait recognition," *EURASIP Journal on Advances in Signal Processing: Special Issue on Advanced Signal Processing and Pattern Recognition Methods for Biometrics*, vol. 2008, 2008, article ID 261317, 13 pages, doi:10.1155/2008/261317.
37. R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.
38. P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: 19 results all computer vision researchers should know about," *Proc. IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov. 2006.
39. R. Cutler, C. Benabdelkader, and L. Davis, "Motion-based recognition of people in eigengait space," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, May 2002, pp. 254–259.

30 REFERENCES

40. L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Trans. Image Processing*, vol. 12, no. 9, pp. 1120–1131, Sept. 2003.
41. A. Kale, A. N. Rajagopalan, A. Sunderesan, N. Cuntoor, A. Roy-Chowdhury, V. Krueger, and R. Chellappa, "Identification of humans using gait," *IEEE Trans. Image Processing*, vol. 13, no. 9, pp. 1163–1173, Sept. 2004.
42. D. Tolliver and R. T. Collins, "Gait shape estimation for identification," in *Proc. Int. Conf. on Audio and Video-Based Biometric Person Authentication*, June 2003, pp. 734–742.
43. Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: averaged silhouette," in *Proc. Int. Conf. on Pattern Recognition*, vol. 4, Aug. 2004, pp. 211–214.
44. J. P. Foster, M. S. Nixon, and A. Prügel-Bennett, "Automatic gait recognition using area-based metrics," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2489–2497, Oct. 2003.
45. M. Soriano, A. Araullo, and C. Saloma, "Curve spreads - a biometric from front-view gait video," *Pattern Recognition Letters*, vol. 25, no. 14, pp. 1595–1602, 2004.
46. J. Han and B. Bhanu, "Performance prediction for individual recognition by gait," *Pattern Recognition Letters*, to appear.
47. J. E. Boyd, "Video phase-locked loops in gait recognition," in *Proc. IEEE Conf. on Computer Vision*, vol. 1, July 2001, pp. 696–703.
48. N. Cuntoor, A. Kale, and R. Chellappa, "Combining multiple evidences for gait recognition," in *Proc. IEEE Conf. on Multimedia and Expo*, vol. 3, July 2003, pp. 113–116.
49. M. G. Grant, J. D. Shutler, M. S. Nixon, and J. N. Carter, "Analysis of a human extraction system for deploying gait biometrics," in *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation*, Mar. 2004, pp. 46–50.
50. N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "Gait recognition using linear time normalization," *Pattern Recognition*, vol. 39, no. 5, pp. 969–979, 2006.
51. J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
52. R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 781–796, Aug. 2000.

53. Y. Liu, R. T. Collins, and Y. Tsin, "A computational model for periodic pattern perception based on frieze and wallpaper groups," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 3, pp. 354–371, Mar. 2004.
54. I. R. Vega and S. Sarkar, "Statistical motion model based on the change of feature relationships: human gait-based recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 10, pp. 1323–1328, Oct. 2003.
55. J. J. Little and J. E. Boyd, "Recognizing people by their gait: the shape of motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.
56. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Networks*, vol. 19, no. 1, pp. 18–39, Jan. 2008.
57. A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," in *Proc. SPIE Conf. on Biometric Technology for Human Identification II*, Mar. 2005, pp. 196–204.
58. A. K. J. A. Ross and J. Z. Qian, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.
59. J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
60. X. Zhou and B. Bhanu, "Feature fusion of face and gait for human recognition at a distance in video," in *Proc. Int. Conf. on Pattern Recognition*, vol. 4, Aug. 2006, pp. 529–532.
61. G. Shakhnarovich and T. Darrell, "On probabilistic combination of face and gait cues for identification," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, vol. 5, May 2002, pp. 169–174.
62. A. Kale, A. K. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, May 2004, pp. 901–904.
63. T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.
64. C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 5, pp. 725–737, May 2006.
65. J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Networks*, vol. 17, no. 1, pp. 166–178, Jan. 2006.
66. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
67. I. T. Jolliffe, *Principal Component Analysis, second edition*. Springer Series in Statistics, 2002.

32 REFERENCES

68. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711–720, July 1997.
69. B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, no. 11, pp. 1771–1782, 2000.
70. H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.
71. A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 2, pp. 228–233, 2001.
72. B. Schölkopf, *Support Vector Learning*. Munich: R. Oldenbourg Verlag, 1997. [Online]. Available: http://www.kernel-machines.org/papers/book_ref.ps.gz
73. K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, no. 2, pp. 181–201, Mar. 2001.
74. J. Wang, H. Lu, K. N. Plataniotis, and J. Lu, "Gaussian kernel optimization for pattern classification," *Pattern Recognition*, submitted in 2008.
75. B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
76. G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385–2404, 2000.
77. J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Networks*, vol. 14, no. 1, pp. 117–126, Jan. 2003.
78. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear principal component analysis through successive variance maximization," in *Proc. International Conference on Machine Learning*, July 2008, to appear.
79. S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 212–220, Jan. 2007.
80. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization for gait recognition," in *Proc. Biometrics Symposium 2006*, September 2007.
81. J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in Neural Information Processing Systems (NIPS)*, 2004, pp. 1569–1576.

82. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Multilinear principal component analysis of tensor objects for recognition," in *Proc. Int. Conf. on Pattern Recognition*, vol. 2, August 2006, pp. 776 – 779.
83. J. Ye, R. Janardan, and Q. Li, "GPCA: An efficient dimension reduction scheme for image compression and retrieval," in *The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 354–363.
84. H. Lu, , K. N. Plataniotis, and A. N. Venetsanopoulos, "Boosting LDA with regularization on MPCA features for gait recognition," in *Proc. Biometrics Symposium 2006*, September 2007.
85. L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors," *SIAM Journal of Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.
86. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Gait recognition through MPCA plus LDA," in *Proc. Biometrics Symposium 2006*, September 2006, pp. 1–6.
87. D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.-J. Zhang, "Human gait recognition with matrix representation," vol. 16, no. 7, pp. 896–903, July 2006.
88. A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision*, vol. 64, no. 1, pp. 5–30, 2005.
89. Three-dimensional face recognition. [Online]. Available: http://en.wikipedia.org/wiki/Three-dimensional_face_recognition
90. S. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Trans. Image Processing*, vol. 13, no. 11, pp. 1491–1506, Nov. 2004.
91. P. J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, , and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 947–954.
92. N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Gait recognition: a challenging signal processing technology for biometrics," *IEEE Signal Processing Mag.*, vol. 22, no. 6, Nov. 2005.