ELSEVIER

Face recognition from a single image per person: A survey

Xiaoyang Tan[a, b], Songcan Chen[a, c, *], Zhi-Hua Zhou[b], Fuyan Zhang[b]

[a]*Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China*
[b]*National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China*
[c]*State Key Laboratory of Pattern Recognition, Institution of Automation, Chinese Academy of Sciences, Beijing 100080, China*

Abstract

One of the main challenges faced by the current face recognition techniques lies in the difficulties of collecting samples. Fewer samples per person mean less laborious effort for collecting them, lower cost for storing and processing them. Unfortunately, many reported face recognition techniques rely heavily on the size and representative of training set, and most of them will suffer serious performance drop or even fail to work if only one training sample per person is available to the systems. This situation is called "one sample per person" problem: given a stored database of faces, the goal is to identify a person from the database later in time in any different and unpredictable poses, lighting, etc. from just one image. Such a task is very challenging for most current algorithms due to the extremely limited representative of training sample. Numerous techniques have been developed to attack this problem, and the purpose of this paper is to categorize and evaluate these algorithms. The prominent algorithms are described and critically analyzed. Relevant issues such as data collection, the influence of the small sample size, and system evaluation are discussed, and several promising directions for future research are also proposed in this paper.
© 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Face recognition; Single training image per person

1. Introduction

As one of the few biometric methods that possess the merits of both high accuracy and low intrusiveness, face recognition technology (FRT) has a variety of potential applications in information security, law enforcement and surveillance, smart cards, access control, among others [1–3]. For this reason, FRT has received significantly increased attention from both the academic and industrial communities during the past 20 years. Several authors have recently surveyed and evaluated the current FRTs from different aspects. For example, Samal et al. [4] and Valentin et al. [5] surveyed the feature-based and the neural-network-based techniques, respectively, Yang et al. reviewed face detection techniques

[6], Pantic and Rothkrantz [7] surveyed the automatic facial expression analysis, Daugman [3] pointed out several critical issues involved in an effective face recognition system, while the most recent and comprehensive survey is possibly from that of Zhao et al. [1], where many of the latest techniques are reviewed.

The aim of face recognition is to identify or verify one or more persons from still images or video images of a scene using a stored database of faces. Many research efforts [1] have been focused on how to improve the accuracy of a recognition system, however, it seems that most of them ignore the potential problem that may stem from the face database at hand, where there may be only one sample image per person stored, possibly due to the difficulties of collecting samples or the limitations of storage capability of the systems, etc. Under this condition, most of the traditional methods such as eigenface [8,9] and fisherface [10–13] will suffer serious performance drop or even fail to work (see more details in Section 2). This problem, called

* Corresponding author. Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China. Tel.: +86 25 8489 2805; fax: +86 25 8489 3777.

   E-mail address: s.chen@nuaa.edu.cn (S. Chen).

the *one sample per person problem* (or, *one sample problem* for short), is defined as follows. Given a stored database of faces with only one image per person, the goal is to identify a person from the database later in time in any different and unpredictable poses, lighting, etc. from the individual image. Due to its challenge and significance for real-world applications, this problem has rapidly emerged as an active research sub-area of FRT in recent years, and many ad hoc techniques have been developed to attack this problem, such as synthesizing virtual samples [14–16], localizing the single training image [17,18], probabilistic matching [19–21] and neural network methods [22,23].

The significant contribution of this paper is to give a comprehensive and critical survey of those ad hoc methods that recognize face from one individual image. We believe this work would be a useful complement to Refs. [1–7], where most of the techniques surveyed do not consider the one sample problem. Indeed, through a more focused and detailed study of the techniques addressing this problem, we hope this survey can provide more insights into the underlying principles, interrelations, advantages, limitations, and design tradeoffs of these techniques. Relevant issues such as data collection, the influence of the small sample size, and system evaluation are also discussed.

In the following section we first try to establish common ground regarding what the one sample per person problem is and why and when this problem should be considered. Specifically, we also discuss what the problem is not. Then in Section 3 we continue to review state of the art techniques addressing this problem. From this, we hope some useful lessons can be learned to help address this problem more effectively. In Section 4, we discuss a few issues concerning performance evaluation. Finally, we conclude this paper with a discussion of several promising directions for the one sample problem in Section 5.

2. The one sample per person problem

In this section, we discuss what the problem of one sample per person indeed is. At first, we give some background relating directly to the generation of the one sample problem. Then we describe how the problem influences the existing FR algorithms and some challenges it arises to the face recognition algorithm designs. At last, we discuss why and when we should consider this problem.

2.1. Background

The origin of the one sample problem can be traced back to the early period when the geometric-based methods were popular, where various configural features such as the distance between two eyes are manually extracted from the single face image and stored as templates for later recognition [24]. One image per person is not a problem at all for these methods.

However, in some application scenarios where a large amount of face images are available (e.g. in law enforcement), one may need more intelligent and less laborious way to process faces. This directly leads to the birth of the so-called appearance-based techniques. Armed with modern intelligent tools from diverse disciplines such as computer vision, pattern recognition, machine learning and neural network, appearance-based techniques circumvent the laborious procedure of geometrical feature extraction with a vectorlized representation of face image, and greatly improve the effectiveness and efficiency of face recognition systems. Consequently, these methods have become one of the dominant techniques in the field of face recognition since the 1990s.

However, one of the key components of appearance-based methods is their learning mechanism, whose performance is heavily affected by the number of training samples for each face [25]. Most of current FR techniques assume that several (at least two) samples of the same person are always available for training. Unfortunately, in many real-world applications, the number of training samples we *actually* have is by far smaller than that we *supposedly* have. More specifically, in many application scenarios, especially in large-scale identification applications, such as law enforcement, driver license or passport card identification, there is usually only one training sample per person in the database. In addition, we seldom have opportunity to add more samples of the same person to the underlying database, because collecting samples is costly and even we can do so, there is the question of how many samples to add and in what way. Those situations have been less studied in the field so far.

Therefore, it makes sense to distinguish the face recognition techniques using only one training sample per person between those using multiple (at least two) training samples of the same person. In this paper, these two categories of face recognition problems are named the *one sample problem* and the *multiple samples problem*, respectively. At the first sight, the difference between them seems to lie in how many number of training images they possess for each person. In this sense, the one sample problem appears to be a special case of the multiple samples problem. *Is that true?* *Can algorithms handling multiple samples problem be simply used to deal with the one sample problem as well?* We will discuss these questions in the following section.

2.2. The challenges of one sample problem

In this section, we discuss the influence and challenges brought by the one sample per person problem.

Broadly speaking, one sample problem is directly related to the small sample size problem in statistics and pattern recognition (see Refs. [25–27] for a general discussion on this topic). As mentioned before, the basis of appearance-based methods is their learning mechanisms, while the classic families of learning mechanisms (or classifiers) basically

Fig. 1. Some samples from one subject in ORL dataset.

need sufficiently large training set for a good generation performance [28], partly due to the high-dimensional representation of face images (recall that in the appearance-based domain, face images are vectorlized directly with the gray value of each image pixel). For example Ref. [19], for a 100×100 face image being vectorlized into a $10\,000$ dimensions feature space, theoretically the number of training images for each person should be at least ten times that of the dimensionality [25], that is, $100\,000$ images in total per person. Intuitively, it is hardly conceivable that we human being need so many photos of a person in order to develop a good model of his appearance.

To address this problem, dimensionality reduction techniques can be employed. One of the most successful techniques used in face recognition is principal-component analysis (PCA). The method based on PCA technique is named eigenface in literatures [8,9]. Formally, each n dimensional face image x can be represented as a linearly weighted sum of a set of orthogonal basis u_i ($i = 1, \ldots, n$): $x = \sum_{i=1}^{n} \alpha_i u_i \approx \sum_{i=1}^{m} \alpha_i u_i$ (typically $m \ll n$), by solving the eigenproblem $CU = U\Lambda$, where C is the covariance matrix for the N training samples and can be rewritten as follows [29]:

$$
\begin{aligned}
C &= \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)(X_i - \mu)^{\mathrm{T}} \\
&= \frac{1}{2N} \sum_{l(X_i)=l(X_j)} (X_i - X_j)(X_i - X_j)^{\mathrm{T}} \\
&\quad + \frac{1}{2N} \sum_{l(X_i)\neq l(X_j)} (X_i - X_j)(X_i - X_j)^{\mathrm{T}} \\
&\triangleq C_{\mathrm{I}} + C_{\mathrm{E}}.
\end{aligned}
\tag{1}
$$

That is, the total scatter matrix C equals the sum of intra-person scatter matrix C_{I} and inter-person scatter matrix C_{E}. Under the situation of only one training sample per person, $C_{\mathrm{I}} = 0$, therefore, Eq. (1) reduces to C_{E}. The eigenspace estimated using only C_{E} is not reliable, however, because it cannot effectively capture the major identification difference among other transformation errors and random noise [29].

To illustrate how the performance of eigenface is influenced by different number of training samples per person, we take ORL dataset [30] as test bed. The ORL dataset contains images from 40 individuals, each providing 10 different images. See Fig. 1 for 10 sample images of one person. In the experiment, we fix the testing face but vary the number of training faces for each person. More specifically, we use the
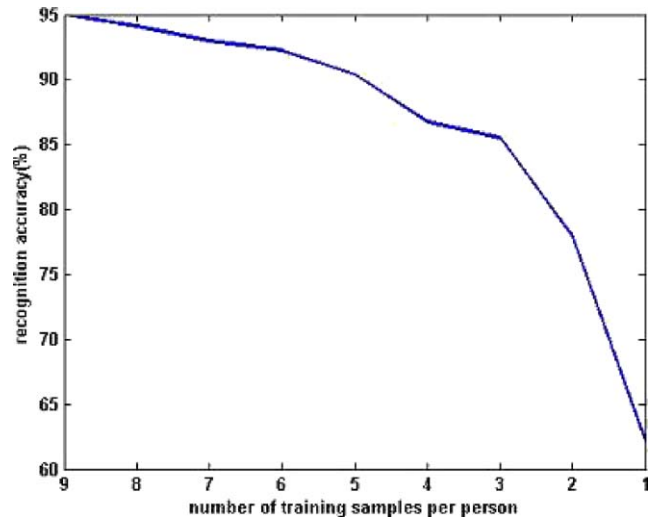

Fig. 2. The average top 1 recognition performance as a function of the number of training samples per person.

last face image of each person (Fig. 1) for testing, and randomly choose the first n images of each person ($n < 9$) for training. The above procedure is repeated 20 times. Fig. 2 shows the average top 1 recognition rate as a function of the number of training samples per person. We can see from Fig. 2 that the performance of eigenface drops with the decreasing number of training samples for each person. In the extreme case, if only one training sample per person is used, the average recognition rate of eigenface falls to below 65%, a 30% drop from 95% when nine training samples per person are given.

Based on the standard eigenface technique, researchers have developed various extended algorithms during the last decades, including probabilistic-based eigenface [31], linear discriminative analysis (LDA) based subspace algorithms [10–13], support vector machine (SVM) based method [32], feature line method [33], Evolution pursuit [34], and Laplacianfaces [35], et al. All of these approaches claimed to be superior to eigenface. However, it may not be the case if only one training sample per person is available, due to the fact that most of them will either reduce to the basic eigenface approach or simply fail to work in that case. Detailed explanations are given below:

(1) The goal of most LDA-based subspace algorithms is to find the most discriminative projection directions

in eigenspace, by maximizing the ratio of inter-person variations to intra-person variations. However, LDA-based method works well only when a lot of representative training samples per person are given, otherwise, its performance may be even worse than eigenface [36]. When there is only one training sample per person stored in the database, LDA-based subspace methods fail to work because the so-needed intra-person variation cannot be obtained. As a remedial measure, Zhao et al. [13] proposed to replace the intra-person scatter matrix using a constant matrix and by doing so, the LDA-based method actually reduces to eigenface method.

(2) Probabilistic-based [31] turns face recognition problem into a two-class problem by estimating the probability of the difference between a test image and a prototype image belongs to intra-person variation or inter-person-variation. Based on the same reason given above, the distribution of intra-person cannot be estimated in the situation of one sample per person and thus the method also reduces to eigenface.

(3) Both evolution pursuit [34] and more recent Laplacianfaces [35] depend on large number of training samples per person to reliably estimate the so-needed low-dimensional manifold. In the situation of one sample per person, both methods also reduce to their starting points, i.e., eigenface.

(4) SVM-based method [32] and feature line method [33] are actually the classification method in eigenspace. If only one sample per person is given, neither of them works.

In summary, most of state of the art face recognition methods will suffer a lot from the one sample per person problem, and some of them even fail to work. In other words, this problem has indeed become a blind area for most current face recognition techniques. However, only may ask the question: *is investigating this problem really worthwhile*?

2.3. The significance of the one sample problem

We have shown that the performance of most face recognition algorithms can be seriously influenced by the limited number of training sample per person. However, one might still question whether it deserves further investigation. In this section, we will discuss these problems from two aspects.

On one hand, as mentioned above, the extreme case of one sample per person really commonly happens in real scenarios, and therefore, this problem needs be carefully addressed so as to make face recognition techniques applicable in those situations.

On the other hand, despite likely bad news for most FR techniques, storing only one sample per person in the database has several advantages, which are desired by most real world applications:

(1) *Easy to collect samples*, *either directly or indirectly*: One common component of face recognition systems is the face database, where the "template" face images are stored. Construction of such a face database is a very laborious and time-consuming work. This problem can now be effectively alleviated if only one image per person is needed to be sampled. For the same reason, the deployment of face recognition system would also become much easier. Furthermore, in those application scenarios where direct image sampling is very difficult (if not impossible), one sample per person has its distinctive prevalence. Consider an application in surveillance of public place such as airports and train stations, where a large number of people need to be identified. In this case, we can construct the needed face database efficiently by scanning photographs attached on most certificates such as passports, identification cards, student ID, driver license ID and so on, rather than really taking photos for each people.

(2) *Save storage cost*: The storage costs of face recognition system will be reduced when only one image per person is needed to be stored in the database.

(3) *Save computational cost*: The computational expense for large-scale applications could be significantly reduced, because the number of training samples per person has direct effect on the costs of operations involved in face recognition, such as preprocessing, feature extraction and recognition.

In summary, the above observations reveal that one sample problem is unavoidable in real world scenarios and it has equally impressive advantages. In addition, developing a clear insight into this particular problem will have broader-ranging implications not only for face recognition but for solving the more general small sample problems as well. Therefore, recognition from one sample per person is an important problem for both practice and research. It provides new challenges and new opportunity to the face recognition community. By addressing this problem, the application areas of FRT could be much extended and the underlying techniques could also be enhanced.

Meanwhile, it should be noted the essence of one sample problem is not a problem concerning how many number of training samples each person has, but that concerning how to improve robustness performance against different variations under this extremely small sample size condition. Fortunately, this problem has drawn attention from more and more researchers, and numerous techniques have been developed to deal with it. We will categorize and review these techniques in the next section.

3. Recognizing from one sample per person

In this section, we review existing methods dealing with robust face recognition from a single intensity image.

Table 1
Categorization of methods for face recognition from a single image

Approach	Representative works
Holistic methods	
Extensions of principal-component analysis (PCA)	
$(PC)^2A$	Enrich face image with its projections [37]
2DPCA	Two-dimensional PCA [38]
Noise model	Use noise model to synthesize new face [39]
Discriminant eigenface	Select discriminant eigenfaces for face recognition [40]
Enlarge the size of training set	
Construct new representations	ROCA [41], Imprecisely location method [19], $E(PC)^2A$ [42]
Generate novel views	View synthesis using prior class-specific information [14–16]
Local methods	
Local feature-based	
DCP	Use directional corner points (DCP) features for recognition [43]
Feature graph	Graph matching methods [44–47]
Local appearance-based	
Local probabilistic subspace	Local probabilistic subspace method [19]
Neural network method	SOM learning based recognition [22,23]
Hidden Markov Model	HMM method [48]
Subpattern-based FLDA	Modified LDA method [17,18]
Analytic-to-holistic approach	Hybrid local features [49]
Local binary model	Face recognition with local binary patterns [50]
Fractalfeatures	Fractal-based face recognition [51]
Hybrid methods	
Virtual samples+local features	Local probabilistic subspace method [19]

We have broadly classified these methods into three categories, according to the type of features used by various methods; some methods clearly overlap category boundaries and are discussed at the end of this section.

(1) *Holistic methods*: These methods identify a face using the whole face image as input. The main challenge faced by these methods is how to address the extremely small sample problem.
(2) *Local methods*: These methods use the local facial features for recognition. Care should be taken when deciding how to incorporate global configurational information into local face model.
(3) *Hybrid methods*: These methods use both the local and holistic features to recognize a face. These methods have the potential to offer better performance than individual holistic or local methods, since more comprehensive information could be utilized.

Table 1 summarizes algorithms and representative works for face recognition from a single image. Below, we discuss the motivation and general approach of each category first, and then, we give the review of each method, discussing its advantages and disadvantages.

3.1. Holistic methods

In these methods, each face image is represented as a single high-dimensional vector by concatenating the gray values of all pixels in the face. The advantages of this representation are two folds. First, it implicitly preserves all the detailed texture and shape information that are useful for distinguishing faces. Second, it can capture more global aspects of faces than local feature-based descriptions [52]. On the other hand, there are two unfavorable consequences of this representation under the condition of one sample per person: (1) it makes the dilemma between high dimensions of image data and small samples more serious; and (2) since only one vector exists for each class, the within-class variation needed by many PR techniques cannot be estimated directly any more.

Accordingly, these problems can be addressed by roughly two ways under the holistic representation framework. The first way is trying to squeeze as much information as possible from the single face image, either in the high-dimensional face space or more commonly, in the dimensionality-reduced eigenspace, as an extension of the standard PCA technique. As the second way, one can incorporate prior information by constructing novel views or different representations for each image, so that the actual training set can be effectively enlarged.

3.1.1. Extensions of principal-component analysis (PCA)

As illustrated above, one generally cannot expect to get good generalization performance from the standard eigenface technique if only one sample per person is given. However, it is possible to use this method for any sample size based on the viewpoint of computation. Therefore, extending this method for higher robustness performance becomes a natural choice.
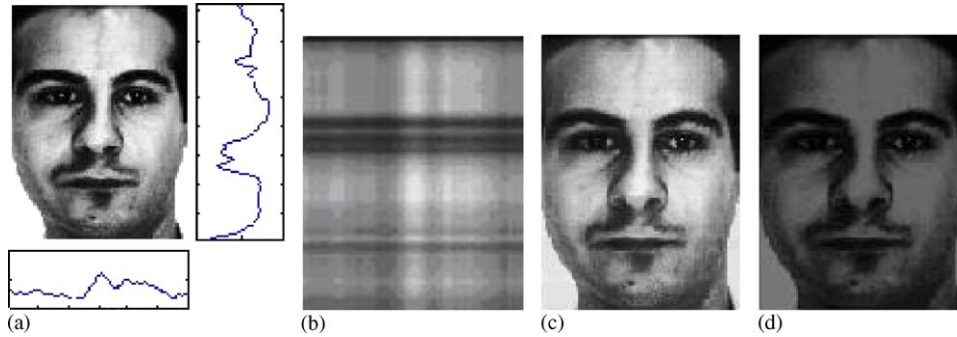
Fig. 3. Some sample images in $(PC)^2A$ method: (a) original face image and its horizontal and vertical profiles; (b) first-ordered projection map; (c) first-ordered projection-combined image; and (d) second-ordered combined image.

Wu and Zhou presented a method named $(PC)^2A$ to enrich the information of face space [37]. Their work was motivated in part by the projection method from the face detection field [53]. Let $I(x, y)$ be the intensity value of an $m \times n$ image at position (x, y), the horizontal and vertical projections of the image are defined as $HI(x) = \sum_{y=1}^{n} I(x, y)$ and $VI(y) = \sum_{y=1}^{n} I(x, y)$, respectively. As can be seen from Fig. 3a, these two projections reflect the distribution of the salient facial features that are useful for face recognition. Then, the obtained projections are used to synthesize a new image (Fig. 3b), defined as $M_p(x, y) = VI(x)HI(y)/\bar{J}$, where $\bar{J}$ is the average intensity of the image. This new projection image is then combined with the original face image to complete the information-enriching procedure (see Fig. 3c). As a result, unimportant features for face recognition are faded out and the important features become more salient after the preprocessing. After that, the traditional eigenface technique is used for face recognition. Their method had been tested on a subset of FERET database [54] with one training image for each person, and they reported that this method achieves 3–5% higher accuracy than the standard eigenface technique through using 10–15% fewer eigenfaces. The $(PC)^2A$ method was extended by Chen et al. later with a method named $E(PC)^2A$ (Enhanced $(PC)^2A$ [42]). Their idea is to generalize and further enhance $(PC)^2A$ based on n-order images, defined as $I(x, y)^n$. A second-ordered projected image is shown in Fig. 3d. It is reported that the enhanced versions of $(PC)^2A$ is more effective and more efficient than its counterpart.

The essence of both $(PC)^2A$ and $E(PC)^2A$ is trying to enrich the information of eigenspace by perturbing spatially the single training sample. Another similar method is the single vector decomposition (SVD) perturbation introduced by Zhang et al. [55]. According to SVD theorem [56,57], an intensity face image I can be written as $I = USV^T$, where $U^T U = V^T V$ (U and V are both orthogonal matrixes, and E is identity matrix), S is a diagonal matrix consisting of singular values of I. Then the perturbed image can be defined as $I_\alpha = US^\alpha V^T$ or $I_\beta = U(S + \beta E)V^T$, where α, β are perturbing factors. After that, the obtained derived

image can be combined with the original image for later recognition.

In traditional eigenface technique, how to reliably estimate the covariance matrix under the small sample size condition remains unsolved. A potential solution known as two-dimensional PCA (2DPCA) has recently been proposed by Yang et al. [38]. This method uses straightforward 2D image matrices rather than 1D vectors for covariance matrix estimation, thus claimed to be more computationally cheap and more suitable for small sample size problem. More specifically, let each face image A_j ($j = 1, \ldots, N$, N is the number of training samples) be a $m \times n$ random matrix, and the average image of all training samples be $\bar{A}$, then the image covariance (scatter) matrix G_t can be evaluated according to $G_t = 1/N \sum_{j=1}^{N}(A_j - \bar{A})^T(A_j - \bar{A})$. By maximizing the criterion function: $J(X) = X^T G_t X$, one can obtain a set of optimal projection axis $\{X_1, X_2, \ldots, X_d\}$, which are then used for feature extraction. The effectiveness and robustness of 2DPCA has been demonstrated on several well-known face image databases.

The methods reviewed above actually handle the one sample problem in an indirect way, that is, the variations of expression, illumination or pose are not explicitly addressed. Therefore, their robustness performance is somehow predictable. While in real world scenarios, various certificate photographs are usually corrupted by various scratches, blur or discoloration. Jung et al. [39] developed an authenticating system to handle these problems with one training image per person. The basic idea is trying to synthesize multiple new face images which imitate the corrupted images for recognition. The imitation is done by a noise model with three noise parameters controlling the degree of contrast, brightness and Gaussian blur, respectively. The synthesizing procedure is shown in Fig. 4, where, by changing the values of noise parameters, several corrupted images corresponding to one sample are imitated, which essentially improves the representative of the given sample. The authors scanned 137 face images from identification cards with 300dpi to test their method. Experimental results show that the error rate is only 1.32%, indicating that the method can
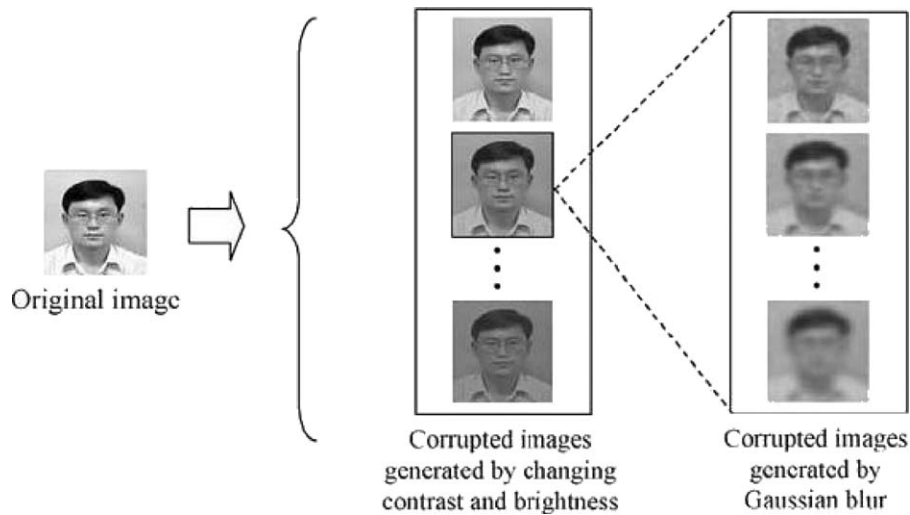
Fig. 4. Synthesizing new samples with the noise model [58].

significantly improve the similarity between the corrupted images and the training images.

The fisherface approach [10] can be regarded as an extension of eigenface as well, in that it tries to find the most discriminative subspace in the eigenspace. However, this method fails when each person just has only one training face sample because of nonexistence of its within-class scatter. Wang et al. [40] presented a method to solve this problem by incorporating prior information of the within-class scatter from other people, based on the assumption that human being exhibits similar intraclass variation. For that purpose, a generic training set with multiple samples per person are collected and used to provide the so-needed intra-person variation. Once the intra-person variation has been obtained, a method similar to fisher criterion is employed to select the most discriminative eigenvectors for recognition. This method was tested on a subset of FERET database with 256 subjects and a 10% performance margin over traditional eigenface method was obtained. The authors also investigated the influence of the generic training set on recognition performance and suggested that a larger sample size would be preferable. However, generic set can provide a useful or harmful bias during the learning of a new task, thus the problem of how to collect suitable generic set so as to obtain the optimal performance should be carefully considered. It is worthy mentioning that the idea of incorporating prior information is widely used to enlarge the actual training set when needed, which will be discussed in the following section.

3.1.2. Enlarge the training set

Another main stream to solve the one sample problem using holistic features is to enlarge the size of training set. This can be done by synthesizing virtual samples. Here virtual samples include both new representations for the single training sample and new visual samples that are not previously existed in the database. It follows then that for virtual samples generation, one can choose to either construct new representations or create novel visual images, based on the single training image. As we will see later, the fundamental difference between the two methods lies in whether prior information is used: the representation construction method focuses on mining more information from the face image at hand, while visual sample generation method concentrates on learning extra information from the domain, besides the given training set. Note that, however, the boundary between the two methods is not so crisp, since novel visual image can also be thought of as a new representation of the given face. Both the two types of methods are discussed in more detail in the subsection below.

3.1.2.1. Construct new representations. Representation is crucial for the success of a recognition system. Different representations are robust against different types of noises and appearance changes. However, it is generally difficult to choose the optimal representation that maximizes recognition performance. Instead, one can combine various representations of the same image so as to exploit their specific merit. Motivated by this, Torre et al. [41] presented a method named representational oriented component analysis (ROCA), basing on one sample per person. In particular, each gallery image is firstly preprocessed to mitigate the effects of light direction changes, and then, several linear and non-linear filters are applied on each image to produce its 150 representations in total. Next, an OCA classifier is built on each representation. Finally, all the OCA classifiers are combined with a weighted linear sum method to give the final decision. Several numerical methods have been introduced to improve the OCA classifier's generalization [41]. Experimental results on 760 images from the FRGC v1.0 dataset (http://www.beebiometrics.org) show that an over 20% performance improvement over the best individual classifier can be achieved.

Image perturbation is another convenient way to construct new representations (samples) for each training image. Martinez introduced a perturbation-based approach for generating new samples [19,20], which is actually a byproduct when handling the imprecisely location problem in face image preprocessing. More specifically, given a face image x, an error range can be set for horizontal and vertical localization error, respectively. Each time, by changing the horizontal or vertical coordinate value (i.e., perturb spatially the given image), a new sample is generated, accounting for the localization error at that position. As a result, a large number of new samples are created for each person. Subsequently, the standard eigenspace technique is used for feature extraction and recognition. Experimental results on a test set of 200 images with 50 persons show that this method is superior to the classical PCA approach.

Other algorithms belonging to this class include E(PC)^{2}A and SVD perturbation method described above. Theoretically, one can generate any number of imitated face images for each face by perturbing its n-ordered images. Consequently, the problem of face recognition with one training image per person becomes a common face recognition problem. In addition, these methods can potentially be used to counter the curse of dimensionality, due to the large amount of additional virtual samples. However, as pointed out by Martinez [19], one drawback of these methods that cannot be ignored is that these generated virtual images may be highly correlated and therefore the new samples should not be considered as independent training images.

3.1.2.2. Generate novel views. Ideally, the really-needed virtual samples should be diverse and representative enough, i.e., they occupy different locations in the face space and represent specific variations of face images. One possible way to achieve the goal is to exploit prior information that can be learned from prototypical examples in the domain. As a simple way to create new samples, one can introduce some geometrical transformations on the original face images, such as rotation, scale transformation, and bilaterally symmetric transformation. Yang et al. [59] described an approach known as symmetrical PCA: in the beginning, two virtual image sets, i.e., the even and odd symmetrical image sets are constructed, respectively, which are then input to a standard PCA algorithm for feature extraction. The different energy ratio of the obtained even and odd symmetrical principal components is employed as criterion for feature selection, due on their different sensitivities to pattern variations. On a large database of 1005 persons with single sample for each person, they reported a 90.0% recognition accuracy using only 20 eigenfaces. Another work in this direction is reported by Gutta et al. [60], who used mirror image of one face to handle the half face occlusion problem.

More complicated ways to create virtual samples with prior knowledge mainly aim to address the face recognition problem in uncontrolled conditions (e.g. in an outdoor environment), where pose and illumination may change

significantly. We can reasonably expected that, given a training face image, synthesizing faithfully its multiple new images under different poses and different illumination conditions would greatly improve the generalization of a face recognizer. However, this is a very challenging task if only one training sample per person is given. One way to accomplish that task is by exploiting prior information about how face images transform, which can be learned through extensive experience with other faces.

Poggio and Vetter considered a computational model known as linear classes in order to synthesize virtual views that can be used as additional samples [61]. Their idea is to learn class-specific image-plane transformations from examples of objects of the same class (which can be collected beforehand as generic training set), and then apply them to the real image of the new object for the purpose of virtual sample generation. In particular, A 3D face image can be represented by a vector $X = (x_1, y_1, z_1, x_2, \ldots, y_n, z_n)^T$, where n is number of feature points and x_i, y_i, z_i is the corresponding x, y, z-coordinates of each point. Further, assume $X \in R^{3n}$ is the linear combination of q 3D face images of other persons of the same dimensionality, such that $X = \sum_{i=1}^{q} \alpha_i X_i$, where α_i is the linear coefficient. The linear coefficients are then learned for each face, and with the same coefficients, a rotated view of X can be generated by regarding it as a linear combination of the rotated views of the other objects.

The above idea has been successfully extended to 2D image by Beymer et al. [14]. They presented a method called parallel deformation to generate novel views of a single face image under different poses. The transformation operator is still needed to be learned from other prototypical face images. Let the difference image of an original face X to the reference face be ΔX, and the difference images of other prototypical face X_i to the same reference face be ΔX_i. By applying the linear class assumption, we have: $\Delta X = \sum_{i=1}^{q} \alpha_i \Delta X_i$. The transformation coefficient α_i can be obtained by minimized the criterion function $J(\alpha) = \|\Delta X - \sum_{i=1}^{q} \alpha_i (\Delta X_i)\|$. After the transformation has been learned, one can use it to generate novel images for each single face. In the face recognition system implemented by Beymer et al., 14 virtual images are generated for each real training image, as shown in Fig. 5. With both the real and virtual samples, their experiments achieved a 85% recognition rate using a test set of 620 images with 60 persons, compared to 32.1% with only real samples. These results clearly demonstrate the effectiveness of this method. Indeed, Niyogi et al. [16] have shown that incorporating prior knowledge is mathematically equivalent to introducing a regularizer in function learning, thus implicitly improving the generalization of the recognition system. Nevertheless, there are still several disadvantages of this method that needs to be mentioned. First, this method does not distinguish different transformation coefficients according to different transformations, thus one cannot generate virtual image of given pose at will. Second, since the number of virtual images needed to be generated
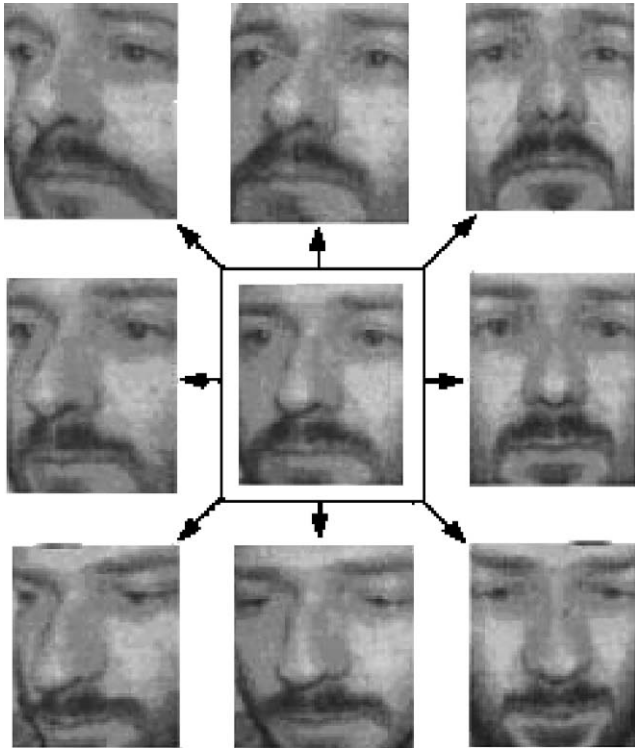
Fig. 5. A real view (center) surrounded by virtual views generated using the parallel deformation method.

depends on the number of modes of variation to be modeled, this approach has difficulty as the number of modes of variation grows large. Third, to enable texture rendering for novel image, one needs to know the correspondence between the real image and the referenced image. Although such correspondence can be learned using the so-called optical flow technique [62], this technique may fail to work under the condition of partial occlusions or toolarge variation. A possible solution to this problem that reconstructs partially damaged face images from other faces was presented in Ref. [63].

Besides novel images compensating for pose variations, illumination changed images can also be synthesizing for robust face recognition using such methods as active shape model (ASM [64]) and illumination cones method [65–67]. However, these methods generally need more than one samples of the same person for training, thus are beyond the scope of this paper. Readers are referred to the excellent survey of Zhao et al. [1] for detailed description about this topic. In addition, the generated virtual samples can be used to construct class-specific face subspaces, as done in Refs. [68,69].

3.1.3. Discussion

Although much success has been achieved by holistic methods, only one single feature vector is used to represent each face image. Such a representation is known to be sensitive to large appearance changes due to expression,

illumination, pose and partial occlusion, especially when Euclidean structure on the face space is assumed, as in most cases. One possible way to handle this problem is to adopt a more flexible non-metric similarity measure for image matching, such as l_p distance (or p-distance), with $0 < p < 1$, as suggested by Donahue et al. [70]. It has been found that non-metric distance measures are less affected by extreme differences than Euclidean distance, thus being more robust to outliers [71]. Another way to attack this problem is to use local facial representations, due to the observation that local features are generally not as sensitive as global features to appearance changes. In the following section, we turn to local methods.

3.2. Local methods

Local methods which use local facial features for face recognition are a relatively mature approach in the field with a long history [6,22,24,44–47,72–75]. Compared with holistic methods, local methods may be more suitable for handling the one sample problem due to the following observations: firstly, in local methods, the original face is represented by a set of *low* dimensional local feature vectors, rather than one single full *high*-dimensional vector, thus the "curse of dimensionality" can be alleviated from the beginning. Secondly, local methods provide additional flexibility to recognize a face based on its parts, thus the common and class-specific features can be easily identified [76]. Thirdly, different facial features can increase the diversity [77] of the classifiers, which is helpful for face identification.

Despite those advantages, incorporating global configurational information in faces is extremely critical for the performance of a local method. Generally, there are two ways for that purpose. First, the global information can be explicitly embedded into the algorithm using such data structure as graph, where each node represents a local feature, while the edge connecting two nodes accounts for the spatial relationship between them. Face recognition are then formulated as a problem of graph matching. Alternative approach to incorporating global information is by classifier combination technique [78]. In particular, separate classifier is employed on each local feature to calculate a similarity score, and then all the similarity scores of different features can be integrated to obtain a global score for final decision.

Similar to the taxonomy that successfully used in Ref. [1], we classify local methods into two categories, i.e., the local feature-based method and the local appearance-based method. The former detects local features first, and then extracts features on the located feature points. The latter simply partitions the face image into sub-regions, based on which local features are directly extracted.

3.2.1. Local feature-based methods

Most of earlier face recognition methods [72–75] belong to this category. In these methods, usually only a single
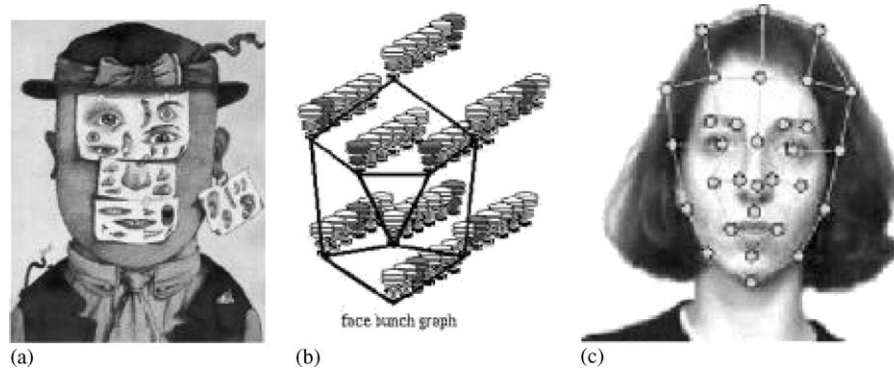
Fig. 6. Illustration of brunch graph from: (a) artistic point of view; (b) scientific point of view; and (c) a brunch graph matched to a face [46].

image per person is used to extract geometrical measures such as the width of the head, the distances between the eyes, and so on. The extracted features are then stored in the database as templates for later matching usage. In early 1990s, Brunelli and Poggio described a face recognition system, which can automatically extract 35 geometrical features to form a 35-dimensional vector for face representation, and the similarity matching is performed with a Bayes classifier [24]. A good recognition rate of 90% on a database of 47 subjects was reported. The storage cost of such systems is very low compared to appearance-based method. However, these methods are usually criticized for two reasons: (1) geometrical-features are hard to be extracted in some complicated cases; and (2) geometrical features alone is not enough to fully represent a face, while other useful information such as the gray level values of the image is totally lost.

The above two problems of geometrical methods actually suggest two research directions. The first direction focuses on how to detect facial features more robustly and more accurately. This direction is subject to a lot of studies [6,24,79–82]. Brunelli and Poggio [24] presented a method that uses a set of templates to detect the eye position in a new image, by looking for the maximum absolute values of the normalized correlation coefficient of these templates at each point in the test image; Rowley et al. [79] endeavored to train several specific feature detectors corresponding to each facial part (e.g. eyes, nose, mouth and profile); Wu et al. [80] introduced a method to automatically locate the region of eyeglasses, using an offline trained eye region detector; Lanitis et al. [81] and Jain et al. proposed to construct statistical model for face shape. Despite all the efforts, there is still a long way to go before the method becomes really mature.

In the second direction, more powerful local feature representation methods rather than the purely geometrical ones are pursued. Manjunath et al. [44] proposed a method for facial feature detection and representation, based on Gabor wavelet decomposition [83] of the face. For each detected feature points, two kinds of information are stored, i.e., location information S and feature information q. Feature information contained in each feature point is defined as a vector, $q_i = [Q_i(x, y, \theta_1), \ldots, Q_i(x, y, \theta_N)]$, where N is the number of q's predefined nearest neighbors, $Q_i(x, y, \theta_j)$ representing the spatial and angular distance from the ith feature points to its jth nearest neighbor. To model the relationship among the feature points, a topological graph is constructed according to the following rule: two feature points in some spatial range with minimal distance will be connected with an edge. After the topological graph has been constructed, face recognition is formulated as a graph matching problem. In particular, the total similarity between two graphs is estimated by decomposing it into two parts: one focuses on the similarity of local features, the other on the global topology similarity. The effectiveness of this method was validated on a face dataset of 86 subjects, containing variations of expression and pose, and 86% and 94% recognition accuracies were reported in terms of the top one and top three candidate matches, respectively, showing good robust performance. However, one drawback of this method is that once the topology graph is constructed, no more modification is allowed. In fact, face images are easy to change due to different variations, thus a fixed topology graph scheme is not adequate.

Based on this observation, Lades et al. [45] proposed a deformable topology graph matching method, now known as Elastic Bunch Graph Matching (EBGM [46]). As in Ref. [44], a topology graph is constructed for each face first, with each node attached one or several Gabor jets. Each component of a jet is a filter response of a specific Gabor wavelet extracted at a predefined critical feature point. These locally estimated Gabor features are known robust against illumination change, distortion and scaling [83]. This is the first key factor in EBGM method. Another key point of this method lies in the graph matching, whose fist step is similar to that in Ref. [44], i.e., both the local and global similarities are considered. The novelty lies in the second step, where a deformable matching mechanism is employed, that is, each node of the template graph is allowed to vary its scale and position (thus named as brunch graph, see Fig. 6) according to the appearance variations on a specific face. For the

above two reasons, elastic matching method exhibits high robustness against appearance changes and has become one of the most successful algorithms in FERET competition [54] of 1996. A good comparative study on eigenface, neural work-based method and elastic matching method has recently been conducted by Zhang et al. [84]. However, there are two obvious disadvantages of this method. First, it may require more computational effort than other method such as eigenface, thus being more difficult to implement in practice. Second, only information at key positions of the image (e.g. eyes, nose, moth, et al.) is used for recognition. Although this is a crucial factor that contributes to the robustness of the method, it is not clear how it can effectively handle the situation when the key positions are occluded.

Both of the above problems of EBGM have close connection with the graph-based representation of human face. Hopefully, relaxing the rigid graph constrains would help to alleviate these problems. Regarding each face as a bag of local features extracted from different feature points instead of a list of graph nodes may be an acceptable compromise between robustness and efficiency. Kepenekci et al. [47] presented an implementation of this idea, based on Gabor features. Instead of predefining fixed number of feature points in the given face image as EBGM, they used a set of Gabor filter matrix to scan local facial regions, and those feature points with higher response of Gabor filter are automatically chosen to be candidates for face representation. Since the resulting feature points are different face to face, the possibility of finding class-specific features is increased. For each feature point, besides the Gabor response values, its location is also recorded, thus implicitly considering the spatial structure of the face. This data structure makes the matching procedure very simple: just by summing the similarity scores of each pair of corresponding local features first, and then assigning the label of the training image with largest total similarity score to the given test image. Experimental results on the ORL dataset show a 95.25% top 1 recognition accuracy, with only one image per person used for training. In the standard FERET test [54], the performance of this method is reported to be comparable to that of EBGM, while its computational complexity is much smaller than the latter. However, there exist some risks in its too flexible way for detecting feature points, due to possible nonexistence of corresponding features in a local area.

Unlike the methods mentioned above, where local feature points are isolated and extra mechanism (e.g. topology graph) needs to be used to describe the neighboring relationship, Gao et al. [43] presented a method that directly integrates such a mechanism into the representation of local feature points, and a novel geometrical feature descriptor, named directional corner point (DCP), is obtained. More specifically, a DCP is a feature point companying with two parameters which provide the isolated point with additional structural information about the connectivity to their neighbors. As shown in Fig. 7, where P is the DCP at hand, and two additional attributes i.e., δ_1, δ_2, are used to model its
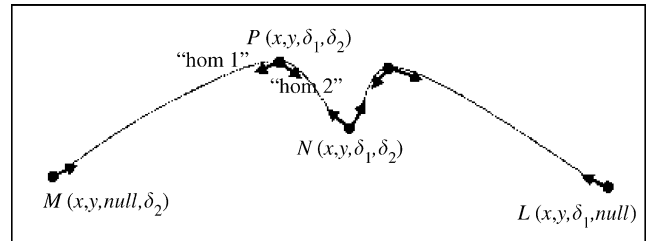

Fig. 7. An illustration of DCPs [43].

local spatial relationship to its anterior neighboring corner point M and posterior neighboring corner point N, respectively. The DCP descriptor is expected to be both economical for storage and less sensitive to illumination changes. This claim was validated by the experimental results on the AR databases [85] with a satisfactory top 5 recognition of 96.43% for both the images with left light on and right light on, under the condition of only one single image per person was used as templates.

In summary, local feature points based methods have been proven to be an effective method dealing with the one sample problem. However, their performance depends critically upon the accurate localization of feature points. Unfortunately, this is not a trivial task in practice, especially in some situation where the shape or appearance of a face image can be changed a lot. For example, the over-illumination would cause strong specular reflection on the face skin and thus making the shape information on faces suppressed or lost. Another example is the large expression changes such as "screaming", which has effects on both the upper and the lower face appearance. These situations may cause much trouble to the local feature-based methods. The performance of the DCP method, for example, drops to 61.61% and 27.68% under the above two conditions, respectively. One possible method to circumvent this problem is to use local appearance-based methods, where accurate localization of feature points is generally not needed.

3.2.2. Local appearance-based methods

In this section, before reviewing local appearance-based methods, we give a high-level block diagram of these methods, for readers' convenience. As shown in Fig. 8, four steps are generally involved in these methods, i.e., local region partition, feature extraction, feature selection and classification. Details of these steps are given below.

Step 1: Local regions partition. At the beginning, local regions should be defined. This involved two factors, i.e., the shape and the size of the local regions. The simplest and most widely-used region shape is rectangular window [17,22,23,50], as shown in Fig. 9a. The windows can be either overlapped with each other [22,48,50] or not [17,22,23]. Besides rectangle, other shapes such as ellipse (Fig. 9b) [19] and strip (Fig. 9c) [86] can also be used. The size of local areas has direct influence on the number of local features and the robustness of the underlying method [23].
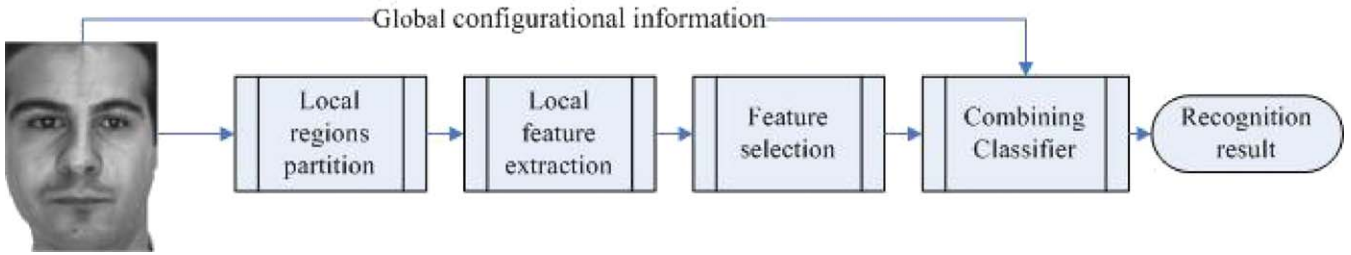
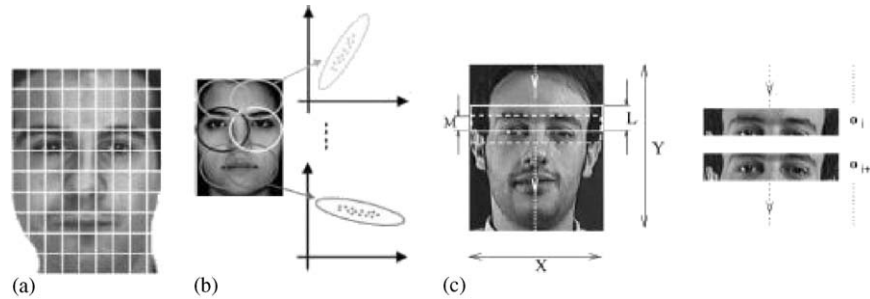Fig. 8. Overall framework of local appearance-based methods.

Fig. 9. Typical local shapes used by the local appearance-based methods: (a) rectangular shape; (b) ellipse shape; and (c) strip shape.

Step 2: Local feature extraction. Once the local regions are defined, one has to decide how to represent the information of them. This is very critical for the performance of a recognition system. The commonly used features include gray-value features [17–19,22,23] and a variety of derived features, such as Gabor wavelet [44–47], Harr wavelet [48], fractal features, and so on. It's hard to give a "winner" feature descriptor that is suitable for all the applications. In general, gray-value feature is the simplest feature without loss of texture information, while Gabor features and other derived features are more robust against illumination change and some geometrical translations [44,45].

Step 3: Feature selection. If lots of features are generated in the previous step, additional feature selection stage is usually needed for effectiveness and efficiency consideration. PCA [8] is a commonly used feature selection method guaranteeing minimal loss of information; LDA [17,18] can be used for selecting the features with most discriminative power; some local statistics such as the degree of texture variation [87,88] are also used for feature selection.

Step 4: Classification. The final step is face identification. Combining Classifiers is the most common way for that purpose. In particular, each component classifier is applied on one local feature, and then, the final decision is made by majority voting or linearly weighted summing.

Note that the above four steps are not a must for each method. Some steps, such as the feature selection, may be cancelled or combined with other steps according to some specific situation.

Now, we will give a detailed review of the local appearance-based methods. Martinez presented a local

probabilistic approach to recognize partially occluded and expression variant face from a single sample per class [19]. As mentioned in previous section, lots of virtual samples accounting for localization errors are first generated using an image-perturbation method, and then, each face (including the generated face images) is divided into six ellipse-shaped local areas. Next, all the local patches at the same position of each face are grouped into a face subspace separately (thus six subspaces in total). For a more compact and efficient representation, each face subspace is further transformed into an eigenspace, where the distribution is estimated by means of a mixture model of Gaussians using the EM algorithm. In the identification stage, the test images are also divided into six local areas and are projected onto the above computed eigenspaces, respectively. A probabilistic rather than a voting approach is used to measure the similarity of a given match. Experiments on a set of 2600 images show that the local probabilistic approach does not reduce accuracy even when $\frac{1}{6}$ of the face is occluded (on the average). However, the mixture of Gaussians used in this method is parametric in nature, which heavily depends on the assumption that the underlying distribution can be faithfully represented with the given samples. Although lots of samples are synthetically generated as the way described above, the computational and storage costs along with the procedure of generating virtual samples may be very high (e.g. 6615 samples per individual in Ref. [19]) when the face database is very large.

Based on the above observation, Tan et al. [23] extended the local probabilistic approach by proposing an alternative way of representing the face subspace with Self-Organizing
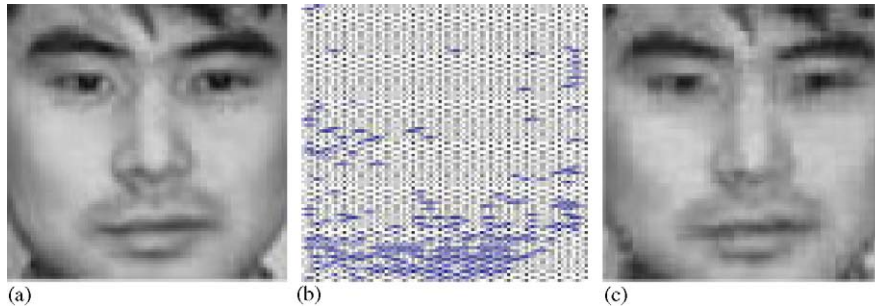
Fig. 10. Example of: (a) an original face image; (b) its projection (SOM-face); and (c) the reconstructed image.

Maps (SOM [89]). More specifically, each face image I is partitioned into M different local sub-blocks $R_i|_{i=1}^{M}$ first, then a SOM network is trained using all the obtained sub-blocks from all the available training images irrespective of classes. After the SOM map has been trained, each sub-block R_i of the same face image I are mapped to its corresponding best matching units (BMUs) by a nearest neighbor strategy, whose location in the 2D SOM topological space is denoted as a location vector $l_i = \{x_i, y_i\}$. All the location vectors from the same face can be grouped as a set, i.e., $I = \{l_i\}_{i=1}^{M} = \{x_i, y_i\}_{i=1}^{M}$, which is called the face's "SOM-face" representation (Fig. 10b). There are several advantages of this representation. Firstly, since possible faults like noise in the original face image can be eliminated in the process of SOM training, this representation is robust to noise. Secondly, it is a compact way to represent face. Finally but most importantly, unlike other methods such as PCA, this representation is intuitively comprehensible in that each element of a SOM-face does have its real physical meaning through the weight vector preserved in the corresponding node of the SOM map, which can be further interpreted as a local facial patch in the input space. A similar previous SOM-based method has been proposed by Lawrence et al. [22], however, their main objective is not to provide a general representation for one sample per person problem, instead, they focused on improve the robustness of recognition system using a five-layered convolutional neural network (CNN).

As explained before, LDA-based subspace methods may fail to work under the condition of only one sample per class. Chen et al. [17] proposed a method to make LDA applicable to this extreme small sample size condition. By first partitioning each face image into a set of corresponding sub-images with the same dimensionality (Fig. 9a), multiple training samples (composed of all the partitioned sub-images) for each class are produced and thus FLDA can be applicable to the set of newly-produced samples. This method has been tested on a subset of FERET dataset containing 200 persons with one training image per person, and an 86.5% recognition accuracy is achieved. Huang et al. [18] also proposed a LDA-based method to handle the one sample problem. Their method is similar to the spatial perturbation method developed by Martinez [19] except that only

some facial components such as mouth and eyes instead of the whole face are perturbed. In this way, the number of training patterns per class is enlarged and LDA thus can be performed.

Pose change is one of the most important and difficult issues for the use of automatic face recognition. To deal with this problem under the condition of one sample per person, Kanade et al. proposed a multi-subregion based probabilistic approach [90]. Similar to the Bayesian method proposed by Moghaddam et al. [31], the face recognition problem at hand is formulated as a two-class pattern classification problem, i.e., whether the two faces are from the same identity or not. In particular, they divided each face into a set of subregions and then tried to construct a probabilistic model of how each local subregion of a face changes its appearance as the pose changes. Such a model is difficult to estimate reliably if adequate training information is absent. Therefore, an accessorial face dataset with many viewing angles are collected for training. After the so-needed probabilistic model is obtained under a Gaussian assumption, the utility of each subregion for the task of face recognition can be calculated and further combined for the final decision. Experiments on the CMU PIE database [91] show that the recognition performance retains within a less than 10% difference until the probe pose begins to differ more than 45° from the frontal face.

Lam et al. also presented a method for pose-invariant recognition from one sample per person [49]. Their idea is to combine analytic-based and appearance-based features in a two-stage strategy (see also Ref. [92] for more discussion). In the first stage, 15 feature points were manually located on a face, which were then used to estimate a head model to characterize the rotation of the face. In the second stage, the correlation of local appearance features of the eyes, nose, and mouth were computed for face identification. Note that before the correlation computation, the previously-estimated head model was employed to provide the local components a suitable compensation for possible geometrical distortion caused by head rotation. Experiments on the ORL face dataset [30] show that under different perspective variations, the overall recognition rates of their method are over 84% and 96% for the first and the first three likely matched faces, respectively.

The methods mentioned above don't explicitly consider the relationship between local features. It is conceivable that using that information would be beneficial for a recognition system. A possible way for that purpose is to construct a flexible geometrical model over the local features as done in EBGM [44–46]. Motivated by this, Huang et al. [93] and Heisele et al. [94] proposed a component-based detection/recognition method with gray-scale local features. Unfortunately, their methods need a lot number of training samples per person, taken from different poses and lighting directions, and thus are not suitable for the problem considered in this paper here.

Another interesting way to incorporate global information is the hidden Markov model (HMM)-based method. Rather than treating face image as a static topology graph with local features as nodes, HMM-based method characterizes face pattern as a dynamic random process with a set of parameters. Samaria et al. [86] illustrated the usefulness of HMM techniques in face recognition. In their method, a face pattern is divided into five overlapped regions, including the forehead, eyes, nose, mouth and chin as shown in Fig. 9c. Then the HMM technique is introduced by regarding each region as one hidden state of a HMM model. A face pattern is then regarded as an observation sequence consisted of five states, each of which can be modeled by a multivariate Gaussian distribution, and the probabilistic transitions between states can be learned from the boundaries between regions. After the HMM has been trained, a face can be recognized by calculating the output probability of its observation sequence. One drawback of this method lies in its greedy requirement in training samples to ensure the reliability of parameter estimation. Le et al. [48] presented a method to make HMM technique applicable in one sample circumstances. Two factors contributed to the feasibility and effectiveness of their method. First, they generated a large collection of observation vectors from each image, in both vertical and horizontal directions, thus enlarging the training set. Second, the Haar wavelet transform was applied to the image to lessen the dimension of the observation vectors and improve the robustness performance. Their experiment results tested on the frontal view AR face database show that the proposed method outperforms the PCA, LDA, and local feature analysis (LFA [95]) approaches.

Ahonen et al. [50] described a local-appearance-based approach, with a single, spatially enhanced feature histogram for global information representation. In their method, three different levels of locality are designed, i.e., pixel-level, regional level and holistic level. The first two levels of locality are realized by dividing the face image into small regions (Fig. 9a), from which the local binary pattern (LBP) features [96] are extracted for efficient texture information representation. The holistic level of locality, i.e., the global description of the face, can be obtained by concatenating the regional LBP features extracted. The recognition is performed using a nearest neighbor classifier in the computed feature space with Chi square as a dissimilarity measure. Their experiments on the FERET dataset show good robust performance using one sample per person for training. Besides LBP, other features that widely used in computer vision field can also be used in face recognition, such as fractal features. For example, Komleh et al. [51] presented a method based on fractal features for expression-invariant face recognition. Their method is tested on the MIT face database with 100 subjects. One image per subject was used for training while 10 images per subject with different expressions for testing. Experimental results show that the fractal features are robust against expression variation.

Relevant studies in psychophysics and neuroscience have revealed that different facial features have different degree of significance to face recognition [1]. For example, it has been found that hair, face outline, eyes and mouth are more important for perceiving and remembering faces, while the nose plays a relatively unimportant role [97]. Inspired by this finding, Brunelli and Poggio used four masks respectively to get the regions of eyes, nose, mouth and the whole face for recognition [24]. Subsequently, Pentland et al. [98] extended Brunelli and Poggio's work by projecting each local feature onto its corresponding eigenspace and using the obtained *eigenfeatures* for recognition. Experimental results of both studies indicate that these facial features are indeed important for face identification, which are conformed to the outcome of Shepherd et al. [97] in the field of neuroscience. However, defining in advance the same regions for all the classes seems to be inconsistent with our intuition that each class should have its own class-specific features, which are really meaningful to the recognition. In this sense, automatic feature selection, as a specific case of feature weighting techniques [99], may be more appropriate.

Automatic feature selection, however, is generally very difficult for high-dimensional, unstructured face image. Tan et al. [100] proposed a possible solution to this problem. Their idea is to transform the high-dimensional face image data to lower dimensional space first, and then select features in the latter space. The task of feature selection may be much easier due to the simplification of the feature space. This method is actually a direct extension to their SOM-face algorithm [23]. More specifically, after representing each face image as a set of nodes in low-dimensional SOM topological space, some simple statistics based on the class distribution of face image data (e.g. the number of faces absorbed by each node) is computed and used to identify important local features of face images. Their experimental results on AR database reveal that up to 80% sub-blocks of a face can be safely removed from the probe set without loss of the classification accuracy. This could be particularly useful when a compact representation of face is needed, such as in the application of smart card, where the storage capability is very limited. Geng and Zhou [101] also proposed a method named SEME to automatically select the appropriate regions for face recognition, based on the idea from ensemble learning field that the facial regions which are both accurate and diverse should be chosen as candidates.

Martinez described a local area weighting based method for expression-invariant face recognition [19–21]. His work was motivated in part by the fact that different face expressions influence different parts of the face more than others. This information can be incorporated into the classifier as prior knowledge. To do that, the author built a learning mechanism to search for those areas that are less affected by a given emotion within a group of people, and an accessorial dataset were employed for this purpose. Once the so-needed weights have been obtained, they can be used for recognition within the same group of individuals. Their experiments on the AR database with 50 persons show a significant increase on the recognition rate than the un-weighted version, even when the expression is largely changed on the probe image.

3.2.3. Discussion

We have reviewed and classified local methods dealing with the one sample problem into two major categories, i.e., feature-based and appearance-based method. However, in some specific situation, these two categories are not so distinct, due to the fact that local regions are actually consisted of a series of pixels, among which interesting feature points could be detected. Furthermore, other categorical standards may be applicable, such as the types of local features, the recognition method, or the way to model global information. Nevertheless, we believe that since the manner to extract local feature is the start point of any local method and has global influence on the followed processes, using it as the categorical guideline is both sufficient and appropriate.

Before ending this section, it should be noted that although local methods are proven to be an effective way to handle the one sample problem, some common problems are still unsolved in these method. For example, it is not clear which kind of local features and in which way to incorporate global information is more appropriate for a given application scenario. In addition, most methods mentioned above are only robust against some variations while not against others. For instance, EGBM is known robust against changes of expression, illumination and pose, but not against occlusion; Local probabilistic method and SOM-face are robust against large expression variation and occlusion, but not against pose changes. A possible way to further improve the robustness performance of a recognition system, as pointed out by Zhao et al. [1], may lie in the combination of different methods, called hybrid methods here.

3.3. Hybrid methods

Hybrid methods are those approaches using both holistic and local features. The key factors that influence the performance of hybrid methods include how to determine which features should be combined and how to combine, so as to preserve their advantages and avert their disadvantages

Table 2
Comparison of the local features and global features' sensitiveness to variations

Variation factors	Local features	Holistic features
Small variations	*Not sensitive*	*Sensitive*
Large variations	*Sensitive*	*Very sensitive*
Illuminations [103]	*Very sensitive*	*Sensitive*
Expressions [19,23]	*Not sensitive*	*Sensitive*
Pose [94]	*Sensitive*	*Very sensitive*
Noise [104]	*Very sensitive*	*Sensitive*
Occlusion [19,23]	*Not sensitive*	*Very sensitive*

at the same time. These problems have close relationship with the multiple classifier system (MCS [78]) and ensemble learning [102] in the field of machine learning. Unfortunately, even in these fields, these problems remain unsolved. In spite of this, numerous efforts made in these fields indeed provide us some insights into solving these problems, and these lessons can be used as guidelines in designing a hybrid face recognition system. For example, components of a hybrid system, either feature or classifier, should be both accurate and diverse, such that a complementary advantages can be feasible.

In fact, local features and global features have quite different properties and can hopefully offer complementary information about the classification task. Table 2 summarizes qualitatively the difference between the two types of features. We can see from the table that local features and global ones are separately sensitive to different variation factors. For instance, illumination changes may have more influence on local features, while expression changes have more impact on holistic features. For these observations, hybrid methods that use both holistic and local information for recognition may be an effective way to reduce the complexity of classifiers and improve their generalization capability.

Despite the potential advantages, the work in this the category is still relatively few, possibly due to the difficulties mentioned above, while typical hybrid methods in traditional sense (i.e., multiple samples per person), such as flexible appearance models [105], hybrid LFA [95], are generally not suitable for handling the one sample problem. We hope more researching efforts could be engaged in this method, and in doing so, we believe that the potential power of hybrid method would put forth sooner or later.

Before ending the discussion of this topic, we note that, from some point of view, local methods can be regarded as a specific kind of hybrid method, since global information is usually incorporated in some way into the algorithm. In local probabilistic method [19], for example, novel training samples are first generated for each person with the holistic method, and then local method is utilized for recognition.

Table 3
Summary of reported experimental results of various algorithms

Method	Face dataset	No. of testing persons	No. of testing images	Accessorial dataset used?	Top one match score(%)	Bench-mark score(%)	Primary variations in the testing dataset
Parallel deformation [14]	N/A	62	620	Y	85.0	32.1	Pose
Local probabilistic subspace [19]	AR	100	600	Y	82.3	70.2	Expression, Time
–	AR	100	400	N	71.0	33.0	Occlusion
SOMface [23]	AR	100	600	N	93.7	70.2	Expression, Time
–	AR	100	400	N	76.0	33.0	Occlusion
2DPCA [38]	AR	100	600	N	74.8	70.2	Expression, Time
1D-DHMM [48]	AR	120	1440	N	89.8	67.2	Expression, Illumination, Time
$(PC)^2A$ [37]	FERET	200	200	N	83.5	83.0	N/A
$E(PC)^2A$ [42]	FERET	200	200	N	85.5	83.0	N/A
SVD Perturbation [55]	FERET	200	200	N	85.0	83.0	N/A
Modular FLDA [17]	FERET	200	200	N	86.5	83.0	N/A
Component LDA [18]	FERET	70	350	N	78.6	32.0	Expression, Illumination
EBGM [46]	FERET	1196	1196	N	95.0	79.7	Expression
LBP [50]	FERET	1196	1196	N	97.0	79.7	Expression
Discriminant PCA [40]	FERET	256	914	Y	72.0	74.0	Expression, Time
Analytic-to-holistic approach [49]	ORL	40	160	N	84.0	74.7	Pose
Face-specific subspace [68]	Yale	15	150	N	95.3	74.7	Expression, Illumination

4. Performance evaluation

To quantitatively assess and fairly compare the methods that aim at addressing the one sample per person problem, algorithms should be tested on the same benchmark dataset according to a standard testing procedure. Unfortunately, such a requirement is seldom satisfied in practice. Although numerous algorithms have been developed, most of them have been tested on different datasets in a different manner. In this section, we review the recognition performance reported by these algorithms and a few issues that need to be carefully considered when evaluating algorithms under the condition of only one sample per person.

Table 3 summarizes the reported performance among several reviewed methods dealing with the one sample problem. According to the suggestion from Ref. [58], in this table, several statistics directly relating to the experimental results are selected to describe the performance of the given algorithms. These include the name of the dataset used, primary variation contained in the probe set, the total number of testing images and persons, whether accessorial dataset is used to help for training, and the performance percentages for the top one rank score. To help understand the difficulty of the given problem, the performance of standard eigenface algorithm is also given here as benchmark. Note that information about training set except accessorial dataset is not mentioned here, since for most algorithms considered here, the number of persons in the gallery set is the same as that in the probe set, and the number of training image per person is only one. It is also worthy mentioning that from the perspective of performance evaluation, however, the experimental statistics listed here may not fully characterize the behavior of these methods, and furthermore, it

is really difficult to declare a "winner" algorithm from this table.

Firstly, according to FERET, basically two types of face recognition methods, i.e., gallery insensitive and gallery sensitive, can be classified. The gallery set consists of a set of known individuals, while the probe set is input images to be labeled. Note that gallery set is not necessarily the training set. If an algorithm's training procedure is completed prior to the start of the test, such an algorithm is called gallery insensitive, otherwise gallery sensitive. FERET design principles force each algorithm to be gallery insensitive to ensure that each algorithm has a general representation for faces, not a representation tuned to a specific gallery. This requirement is desired for commercial systems, since users are usually not care about how the system at hand is trained. However, if one cannot adjust system parameters according to the gallery set, some problems may be encountered. From the perspective of pattern recognition, it is always assumed that some close relationship exists between the training set and the probe set. For example, intuitively, a recognition system trained on a dataset of American people may not be directly used for the face identification of Chinese people. In other words, only a large amount of representative training samples were pre-collected such that the gallery is fully covered, the gallery-insensitiveness could be really possible. This seems to deviate the goal of the algorithm evaluation to a pursuit of high quality training set, and algorithms having good training set are obvious in the ascendant of those not. Therefore, we urge members of the community to give a detailed description of training set (including accessorial set), if they want their algorithms to be evaluated fairly as a learning mechanism rather than a commercial system. Besides, the turning parameters involved should not be ignored

as well, since they may have a direct effect on the classification performance.

Secondly, most of current algorithms don't report how their performance is influenced by the number of training images per person. This is particular important for one sample problem, since in this case, the efficiency of learning from only a single face image is critical. Indeed, we should not ignore the following question: *at least how many training images are required to achieve certain performance in a particular task*? Fortunately, this question has begun to drawn attention from researchers most recently [106].

Thirdly, for some applications, especially real-time application, time taken is a very important factor. In Ref. [107], Ting Shan builds a real-time automatic person identification system using a gallery of still images containing only one frontal view image per person.

Finally, the evaluation criteria should also take the specific application scenarios into account. Different applications present different constraints and need different processing requirements. For example, when recognizing National ID card, credit card, passport, driver license, et al., the shooting condition is usually controllable and the quality of photographs is relatively good; while in the situation of video surveillance, the obtained image may be both small in size and blurred, and the background of pictures can be very complicated as well. Therefore, the probe set should be statistically as similar as possible to the images in real world.

In summary, for a fair and effective performance evaluation of the methods of one sample per person, great attention should be paid on the protocols, the training and testing data sets. If existed evaluation protocol such as FERET is applied, special care should also be taken on whether it is suitable for the one sample problem. Needless to say, some suggestions in FERET are appropriate for the problem considered here. For example, in its close-universe model, not only '*is the top match correct*' but also '*is the correct answer in top n matches*?' are considered, which indeed agrees with the general application scenarios of one sample per person.

5. Discussion and conclusion

Face recognition from one sample per person is an important but a challenging problem both in theory and for real-world applications. In this paper, we attempt to provide a comprehensive survey of current researches on this problem. Firstly, we discussed what the one sample problem is, how it affects the current face recognition algorithms, and revealed its important implications for both face recognition and artificial intelligence research. Secondly, we described the current state of the art in dealing with this problem with a three-categories-based taxonomy, i.e., holistic, local and hybrid methods. This, we believe, could help us to develop a clear understanding of this particular problem. Thirdly, we reviewed the relative recognition performance of current

algorithms and discussed several issues that need to be carefully considered when deciding how to evaluate a proposed method. Specifically, we urge the community to report the effect of training-set size on the performance of the underlying algorithm. Finally, it is worthy mentioning that some closely related problems are deliberately ignored in this paper, such as face detection and normalization of the images. For these topics, we can refer to Refs. [1,6] for a detailed discussion.

Obviously, the one sample per person problem and the multiple samples per person problem are two-aspects of the same problem. Separately considering the two problems is motivated in part by the observation that the major part of current research still focuses on the latter problem, while the former does not gain enough deserved attention from the face recognition community. We have revealed in this paper that face recognition algorithms under the assumption of multiple training samples per person may not be applied directly to solve the one training sample per person problem. The relationship between such two problems is beyond the number of training samples per person, and it lies in such aspects as application scenarios, algorithm designing methodology and performance evaluation. Table 4 briefly summarizes these two problems. We can see from this table that there are both commonness and diversity between them. Diversity indicates the necessity of studying the one sample problem, while the commonness suggests the potential value of multiple samples problem in dealing with the one sample problem and vice versa. We believe that bridging these two aspects of face recognition problem could help us designing more efficient algorithms. For example, we can generate additional samples as the way in Ref. [14] from a single sample per person, and use them for facilitating the learning of manifold structure of face images in low-dimensional space, where the requirement of training samples is usually demanding.

Although the one sample problem is by far not solved, some promising research directions can be suggested. Firstly, the solution of this problem may benefit from other related fields, especially from the field of biometrics, where other biometric features, such as iris, fingerprints and speech can be used along with the single training sample to construct more robust recognition system [108]. Secondly, one can seek alternative methods that have the capability to provide more information to the recognizer so as to compensate for the limited representative of a single image. Among them, we believe that at least two methods likely influence the future developments of the one sample problem, i.e., the 3D recognition techniques and techniques that can exploit information from other faces. 3D enhanced approaches can help to overcome sensitivity to geometric and lighting changes of faces. In Ref. [109], a 3D spherical harmonic basis morphable model (SHBMM) is proposed, with which, both face synthesis and recognition can be feasible even if only one single image under unknown lighting is available. The idea of using prior knowledge, also

Table 4
Comparison of recognition from one sample per person and from multiple samples per person

	One sample per person problem	Multiple samples per person problem
Major applications	Smart cards (national ID, drivers' licenses, passports, credit card, et al.), law enforcement, surveillance	Information security, entertainment, human-machine interaction, et al.
Advantages	Lower costs of collecting, storing, and processing samples	More training samples available, higher robustness, plenty of statistical tools for use
Disadvantages	Small size of training samples, lower robustness performance, fewer available methods and tools.	Higher collecting, training and storing costs

referred to as *learning to learn* [110] in the machine learning field, has been widely used and studied in the last decade. Recent results [111] show interesting cases in handwritten digit recognition, where, by exploiting prior knowledge, a recognition system trained on only a single example for each class can achieve better performance than that of the one using thousands of training examples per character.

Thirdly, one of the major contributions from Turk M et al. [9] is that they successfully bring the face recognition into the field of pattern recognition in an efficient and convenient manner, such that a large amount of mature methods in PR, including Bayesian method, PCA, LDA, SVM, et al., can be immediately used for face recognition. However, if only a single sample per person is presented to the system, most of pattern recognition methods are helpless. Therefore, we think that the final solution of this problem depends heavily on the advances of related areas, such as pattern recognition, machine learning and machine vision.

Finally, although there do exist some results from statistical learning theory [112], more study needs to be carried out on how the generalization capacity of an algorithm is influenced by the training set size (especially with extremely small size). It is also meaningful to theoretically investigate how much information can be gained when additional training samples could be presented to the system.

Nevertheless, these difficulties do not mean that addressing this problem cannot be achieved under the current technique framework. As shown in Table 3, considerable efforts in this field are very encouraging. Actually, recent studies [113] have shown that a single training image, especially the front-view images, contains plentiful information that could be used for recognition. Moses et al. [114] also experimentally revealed that human has the ability to generalize the recognition of faces to novel images, from just a single view.

Acknowledgments

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. December Issue (2003) 399–458.

[2] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (5) (1995) 705–740.

[3] J. Daugman, Face and gesture recognition: overview, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 675–676.

[4] A. Samal, P.A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey, Pattern Recognition 25 (1) (1992) 65–77.

[5] D. Valentin, H. Abdi, et al., Connectionist models of face processing: a survey, Pattern Recognition 27 (9) (1994) 1209–1230.

[6] M.H. Yang, D. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 34–58.

[7] M. Pantic, L.J.M. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1424–1445.

[8] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. Soc. Am. A 4 (3) (1987) 519–524.

[9] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–86.

[10] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

[11] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 18 (8) (1996) 831–836.

[12] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, IEEE Trans. Neural Networks 14 (1) (2003) 117–126.

[13] W. Zhao, R. Chellappa, P.J. Phillips, Subspace linear discriminant analysis for face recognition, Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.

[14] D. Beymer, T. Poggio, Face recognition from one example view, Science 272 (5250) (1996).

[15] T. Vetter, Synthesis of novel views from a single face image, Int. J. Comput. Vision 28 (2) (1998) 102–116.

[16] P. Niyogi, F. Girosi, T. Poggio, Incorporating prior information in machine learning by creating virtual examples, Proc. IEEE 86 (11) (1998) 2196–2209.

[17] S.C. Chen, J. Liu, Z.-H. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognition 37 (7) (2004) 1553–1555.

[18] J. Huang, P.C. Yuen, W.S. Chen, J.H. Lai, Component-based LDA method for face recognition with one training sample, AMFG (2003) 120–126.

[19] A.M. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intell. 25 (6) (2002) 748–763.

[20] A.M. Martinez, Recognizing expression variant faces from a single sample image per class, Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2003, pp. 353–358.

[21] A.M. Martinez, Matching expression variant faces, Vision Res. 43 (9) (2003) 1047–1060.

[22] S. Lawrence, C.L. Giles, A. Tsoi, A. Back, Face recognition: a convolutional neural-network approach, IEEE Trans. Neural Networks 8 (1) (1997) 98–113.

[23] X. Tan, S.C. Chen, Z.-H. Zhou, F. Zhang, Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble, IEEE Trans. Neural Networks 16 (4) (2005) 875–886.

[24] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Trans. Pattern Anal. Mach. Intell. 15 (10) (1993) 1042–1062.

[25] A.K. Jain, B. Chandrasekaran, Dimensionality and sample size considerations in pattern recognition practice, in: P.R. Krishnaiah, L.N. Kanal (Eds.), Handbook of Statistics, vol. 2, 1982, pp. 835–855.

[26] A.K. Jain, B. Chandrasekaran, Dimensionality and sample size considerations in pattern recognition practice, in: P.R. Krishnaiah, L.N. Kanal (Eds.), Handbook of Statistics, vol. 2, North-Holland, Amsterdam, 1987, pp. 835–855.

[27] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 252–264.

[28] R.P.W. Duin, Small sample size generalization, in: G. Borgefors (Eds.), SCIA'95, Proceeding of the Ninth Scandinavian Conference on Image Analysis, vol. 2, Uppsala, Sweden, 6–9 June 1995, pp. 957–964.

[29] X. Wang, X. Tang, Unified subspace analysis for face recognition, Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003, pp. 679–686.

[30] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: F.L. Sarasota (Ed.), Proceedings of Second IEEE Workshop on Applications of Computer Vision, December 1994, pp. 138–142.

[31] B. Moghaddam, A. Pentland, Probabilistic visual learning for object representation, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 696–710.

[32] P.J. Phillips, Support vector machines applied to face recognition, Adv. Neural Inform. Process. Syst. 11 (03) (1998) 809.

[33] S.Z. Li, J. Lu, Face recognition using the nearest feature line method, IEEE Trans. Neural Networks 10 (2) (1999) 439–443.

[34] C. Liu, H. Wechsler, Evolutionary pursuit and its application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 22 (6) (2000) 570–582.

[35] X. He, X. Yan, Y. Hu, et al., Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.

[36] A. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 228–233.

[37] J. Wu, Z.-H. Zhou, Face recognition with one training image per person, Pattern Recognition Lett. 23 (14) (2002) 1711–1719.

[38] J. Yang, D. Zhang, A.F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.

[39] H.C. Jung, B.W. Hwang, S.W. Lee, Authenticating corrupted face image based on noise model, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, 272pp.

[40] J. Wang, K.N. Plataniotis, A.N. Venetsanopoulos, Selecting discriminant eigenfaces for face recognition, Pattern Recognition Lett. 26 (10) (2005) 1470–1482.

[41] F. Frade, De la Torre, R. Gross, S. Baker, V. Kumar, Representational oriented component analysis (ROCA) for face recognition with one sample image per training class, in: Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, June 2005, pp. 266–273.

[42] S.C. Chen, D.Q. Zhang, Z.-H. Zhou, Enhanced (PC)2A for face recognition with one training image per person, Pattern Recognition Lett. 25 (10) (2004) 1173–1181.

[43] Y. Gao, Y. Qi, Robust visual similarity retrieval in single model face databases, Pattern Recognition 38 (7) (2005) 1009–1020.

[44] B.S. Manjunath, R. Chellappa, C.V.D. Malsburg, A feature based approach to face recognition, in: Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 1992, pp. 373–378.

[45] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, Malsburg von der, R. Wurtz, Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Comput. 42 (3) (1993) 300–311.

[46] L. Wiskott, R. Fellous, N. Kruger, C. von Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (July 1997) 775–779.

[47] B. Kepenekci, F.B. Tek, G. Bozdagi Akar, Occluded face recognition based on Gabor wavelets, ICIP 2002, September 2002, Rochester, NY, MP-P3.10.

[48] H.-S. Le, H. Li, Recognizing frontal face images using hidden Markov models with one training image per person, Proceedings of the 17th International Conference on Pattern Recognition (ICPR04), vol. 1, 2004, pp. 318–321.

[49] K.-M. Lam, H. Yan, An analytic-to-holistic approach for face recognition based on a single frontal view, IEEE Trans. Pattern Anal. Mach. Intell. 20 (7) (1998) 673–686.

[50] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, Computer Vision, ECCV 2004 Proceedings, Lecture Notes in Computer Science, vol. 3021, Springer, Berlin, 2004, pp. 469–481.

[51] H.E. Komleh, V. Chandran, S. Sridharan, Robustness to expression variations in fractal-based face recognition, Proceedings of ISSPA-01, vol. 1, Kuala Lumpur, Malaysia, 13–16 August 2001, pp. 359–362.

[52] A.J. O'Toole, H. Abdi, Low-dimensional representation of faces in higher dimensions of the face space, Opt. Soc. Am. 10 (3) (1993) 405–411.

[53] C. Kotropoulos, I. Pitas, Rule-based face detection in frontal views, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. 4, 1997, pp. 2537–2540.

[54] P.J. Phillips, H. Moon, S. Rizvi, P. Rauss, The FERET evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1103.

[55] D.Q. Zhang, S.C. Chen, Z.-H. Zhou, A new face recognition method based on SVD perturbation for single example image per person, Appl. Math. Comp. 163 (2) (2005) 895–907.

[56] G.H. Golub, C.F. Loan, Van Matrix Computations, Johns Hopkins University Press, Baltimore, Maryland, 1983.

[57] Z. Hong, Algebraic feature extraction of image for recognition, Pattern Recognition 24 (1991) 211–219.

[58] J.P. Phillips, E.M. Newton, Meta-analysis of face recognition algorithms, in: Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, May 2002, pp. 235–241.

[59] Q. Yang, X. Ding, Symmetrical PAC in face recognition, ICIP02 V (2) (2002) 97–100.

[60] S. Gutta, H. Wechsler, Face recognition using asymmetric faces, in: D. Zhang, A.K. Jain (Eds.), ICBA2004, Lecture Notes in Computer Science, vol. 3072, 2004, pp. 162–168.

[61] T. Poggio, T. Vetter, Recognition and structure from one 2D model view: observations on prototypes, object classes and symmetries, Artificial Intelligence Laboratory, MIT, Cambridge, MA, A.I. Memo No. 1347, 1992.

[62] D. Beymer, T. Poggio, Image representations for visual learning, Science 272 (5270) (1996) 1905–1909.

[63] B.W. Hwang, S.W. Lee, Reconstruction of partially damaged face images based on a morphable face model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (3) (2003) 365–372.

[64] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685.

[65] A.S. Geoghiades, P.N. Belhumeur, D.J. Kriegman, Illumination-based image synthesis: creating novel images of human faces under differing pose and lighting, in: Proceedings, Workshop on Multi-View Modeling and Analysis of Visual Scenes, 1999, pp. 47–54.

[66] A.S. Geoghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 643–660.

[67] A.S. Geoghiades, D.J. Kriegman, P.N. Belhumeur, Illumination cones for recognition under variable lighting: faces, in: Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 1998, pp. 52–58.

[68] S.G. Shan, W. Gao, D. Zhao, Face identification based on face-specific subspace, Int. J. Image Syst. Technol. 13 (1) (2003) 23–32 (Special issue on face processing, analysis and synthesis).

[69] G. Wen, S. Shan, et al., Virtual face image generation for illumination and pose insensitive face recognition, Proceedings of ICASSP2003, vol. IV, Hong Kong, 2003, pp. 776–779.

[70] M. Donahue, D. Geiger, R. Hummel, T. Liu, Sparse representations for image decompositions with occlusions, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 1996, pp. 7–12.

[71] D.W. Jacobs, D. Weinshall, Y. Gdalyahu, Classification with non-metric distances: image retrieval and class representation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (6) (2000) 583–600.

[72] M.D. Kelly, Visual identification of people by computer, Technical Report AI-130, Stanford AI Project, Stanford, CA, 1970.

[73] T. Kanade, Picture processing by computer complex and recognition of human faces, Technical Report, Department of Information Science, Kyoto University, 1973.

[74] A.J. Goldstein, L.D. Harmon, A.B. Lesk, Identification of human faces, Proc. IEEE 59 (5) (1971) 748–760.

[75] Y. Kaya, K. Kobayashi, A basic study on human face recognition, in: S. Watanabe (Ed.), Frontiers of Pattern Recognition, 1972, pp. 265–289.

[76] P.R. Villeala, J. uan Humberto Sossa Azuela, Improving pattern recognition using several feature vectors, Lecture Notes in Computer Science, Springer, Heidelberg, 2002.

[77] L.I. Kuncheva, C.J. Whitaker, Feature subsets for classifier combination: an enumerative experiment, in: J. Kittler, F. Roli (Eds.), Lecture Notes in Computer Science, vol. 2096, Springer, Berlin, 2001, pp. 228–237.

[78] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[79] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998) 23–38.

[80] C. Wu, C. Liu, et al., Automatic eyeglasses removal from face image, IEEE Trans. Pattern Anal. Mach. Intell. 26 (3) (2004) 322–336.

[81] A. Lanitis, C.J. Taylor, T.F. Cootes, Automatic face identification system using flexible appearance models, Image Vis. Comput. 13 (5) (1995) 393–401.

[82] A.K. Jain, Y. Zhong, S. Lakshmanan, Object matching using deformable templates, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.

[83] T.S. Lee, Image representation using 2-d Gabor wavelets, IEEE Trans. Pattern Anal. Mach. Intell. 18 (10) (1996) 959–971.

[84] J. Zhang, Y. Yan, M. Lades, Face recognition: eigenface, elastic matching and neural nets, Proc. IEEE 85 (9) (1997) 1423–1435.

[85] A.M. Martínez, R. Benavente, The AR face database, CVC Technical Report No. 24, June 1998.

[86] F. Samaria, Face segmentation for identification using hidden Markov models, in: British Machine Vision Conference, BMVA Press, 1993, pp. 399–408.

[87] T. Kölsch, D. Keysers, H. Ney, R. Paredes, Enhancements for local feature based image classification, in: International Conference on Pattern Recognition (ICPR 2004), vol. I, Cambridge, UK, August 2004, pp. 248–251.

[88] Y.K. Kim, J.B. Ra, Adaptive learning method in self organizing map for edge preserving vector quantization, IEEE Trans. Neural Networks 6 (1) (1995) 278–280.

[89] T. Kohonen, Self-Organizing Map, second ed., Springer, Berlin, 1997.

[90] T. Kanade, A. Yamada, Multi-subregion based probabilistic approach toward pose-invariant face recognition, in: Proceeding of IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), July 16–20, Kobe, Japan, 2003, pp. 954–959.

[91] T. Sim, S. Baker, M. Bsat, The CMU pose illumination and expression (PIE) database, in: Proceedings of the Fifth International Conference on Automatic Face and Gesture Recognition, 2002.

[92] I. Craw, N. Costen, et al., How should Faces for automatic recognition?, IEEE Trans. Pattern Anal. Mach. Intell. 21 (8) (1999) 725–736.

[93] J. Huang, B. Heisele, V. Blanz, Component-based face recognition with 3D morphable models, in: Proceedings, International Conference on Audio- and Video-Based Person Authentication.

[94] B. Heisele, T. Serre, M. Pontil, T. Poggio, Component-based face detection, in: Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 657–662.

[95] P. Penev, J. Atick, Local feature analysis: a general statistical theory for object representation, Netw.: Comput. Neural Syst. 7 (1996) 477–500.

[96] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.

[97] J.W. Shepherd, G.M. Davidies, H.D. Ellis, Studies of cue saliency, in: G.M. Davies, H.D. Ellis, J.W. Shepherd (Eds.), Perceiving and Remembering aces, Academic Press, London, UK, 1981.

[98] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, Seattle, WA, 1994, pp. 84–91.

[99] D. Wettschereck, D.W. Aha, T. Mohri, A review and comparative evaluation of feature weighting methods for lazy learning algorithms, Technical Report AIC-95-012, Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence, Washington, DC, 1995.

[100] X. Tan, S.C. Chen, Z.-H. Zhou, F. Zhang, Feature selection for high dimensional face image using self-organizing maps, in: Proceedings of the Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'05), Hanoi, Vietnam, Lecture Notes in Artificial Intelligence, vol. 3518, 2005, pp. 500–504.

[101] X. Geng, Z.-H. Zhou, Face recognition based on selective ensemble of multiple eigenspaces, in: Proceedings of the International Symposium on Computer Vision, Object Tracking and Recognition, Beijing, China, August 2004 (in Chinese).

[102] Z.-H. Zhou, J. Wu, W. Tang, Ensembling neural networks: many could be better than all, Artif. Intell. 137 (1–2) (2000) 239–263.

[103] P.W. Hallinan, et al., Two-and Three-dimensional Patterns of the Face, A K Peters, Ltd., Natick, MA, 1999.

[104] N.P. Costen, T.F. Cootes, C.J. Taylor, Compensating for ensemble-specific effects when building facial models, Image Vision Comput. 20 (2002) 673–682.

[105] A. Lanitis, C.J. Taylor, T.F. Cootes, Automatic face identification system using flexible appearance models, Image Vision Comput. 13 (1995) 393–401.

[106] P. Jonathon, P.J. Flynn, T. Scruggs, K.W. Bowyer, Overview of the face recognition grand challenge, IEEE Conference on Computer Vision and Pattern Recognition, vol. I, San Diego, June 2005, pp. 947–954.

[107] Ting Shan, Automatic multiple-view person identification system under unconstrained environment using one sample per person, Ph.D. Confirmation Seminar in the School of Information Technology &

Electrical Engineering (ITEE), University of Queensland, August 2005.

[108] T. Choudhury, B. Clarkson, T. Jebara, A. Pentland, Multimodal person recognition using unconstrained audio and video, International Conference an Audio and Video-based Biometric Authentication, Washington, DC, 1999, pp. 176–181.

[109] L. Zhang, D. Samaras, Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 351–363.

[110] S. Thrun, L. Pratt, Learning to learn: introduction and overview, in: S. Thrun, L. Pratt (Eds.), Learning to Learn, Kluwer Academic Publishers, Dordrecht, 1998.

[111] E. Miller, N. Matsakis, P. Viola, Learning from one example through shared densities on transforms, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2000, pp. 464–471.

[112] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[113] R. Gross, J. Shi, J. Cohn, Quo Vadis face recognition? Third Workshop on Empirical Evaluation Methods in Computer Vision, December 2001, pp. 119–132.

[114] Y. Moses, S. Ullman, S. Edelman, Generalization to novel images in upright and inverted faces, Perception 25 (1996) 443–462.