ORIGINAL ARTICLE

Low-resolution face recognition: a review

**Zhifei Wang · Zhenjiang Miao · Q.M. Jonathan Wu ·
Yanli Wan · Zhen Tang**

Abstract Low-resolution face recognition (LR FR) aims to recognize faces from small size or poor quality images with varying pose, illumination, expression, etc. It has received much attention with increasing demands for long distance surveillance applications, and extensive efforts have been made on LR FR research in recent years. However, many issues in LR FR are still unsolved, such as super-resolution (SR) for face recognition, resolution-robust features, unified feature spaces, and face detection at a distance, although many methods have been developed for that. This paper provides a comprehensive survey on these methods and discusses many related issues. First, it gives an overview on LR FR, including concept description, system architecture, and method categorization. Second, many representative methods are broadly reviewed and discussed. They are classified into two different categories, super-resolution for LR FR and resolution-robust feature representation for LR FR. Their strategies and advantages/disadvantages are elaborated. Some relevant issues such as databases and evaluations for LR FR are also presented. By generalizing their performances and limitations, promising trends and crucial issues for future research are finally discussed.

Z. Wang (✉) · Z. Miao · Z. Tang
Institute of Information Science, Beijing Jiaotong University, Beijing, China
e-mail: yichen0201@163.com

Z. Wang · Z. Miao · Z. Tang
Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China

Q.M. Jonathan Wu
Department of Electrical and Computer Engineering, University of Windsor, Windsor, Canada

Y. Wan
Institute of System Engineering and Control, Beijing Jiaotong University, Beijing, China

1 Introduction

Face recognition (FR) has been widely studied for decades due to its great potential applications. Many technologies focus on dealing with complex conditions such as aging, occlusion, disguise, and variations in pose, illumination, and expression [1]. Although the recognition accuracy of face recognition in controlled environments with cooperative subjects is satisfactory, the performance in real applications such as surveillance is still an unsolved problem partially due to low-resolution (LR) image quality [2]. With the growing installation of surveillance cameras in many places, there are increasing demands for face recognition in surveillance applications from small-scale stand-alone cameras in banks and supermarkets, to large-scale multiple networked close-circuit televisions in public streets [3]. In such cases, subjects are far from cameras, and face regions tend to be small. This issue is called low-resolution face recognition (LR FR).

In this paper, LR FR aims to recognize faces from small size or poor quality images with varying pose, illumination, expression, etc. Traditional methods [4–9] based on high-resolution (HR) face images could not perform well when face images have relatively LR. Compared with other non-invasive biometric authentication techniques such as gait recognition at a distance [10], LR FR is a more difficult task due to the variability of human facial features. The challenges associated with LR FR can be attributed to the following factors:

Misalignment Inaccurate alignment will severely affect the performance of HR as well as LR face recognition systems [11] and it is difficult to perform automatic alignment on LR images.

Noise affection As a generalized issue, LR problem causes a lot of chain effects for face recognition. The degradation in resolution together with pose, illumination, and expression variations adds complexity to the recognition [12]. In other words, these variations produce much more noises for LR FR.

Lack of effective features LR leads to the loss of large amounts of information. Most effective features used in HR FR such as Gabor [13] and local binary pattern (LBP) [14] may fail in LR case, especially very LR case, e.g., 6×6. Novel features insensitive to resolution are essential for LR FR.

Dimensional mismatch Different resolutions between gallery images and probe ones in LR FR systems cause dimensional mismatch in traditional subspace learning methods [15].

Due to the challenges and significances for real applications, LR FR has gradually become an active research subarea of face recognition in recent years, and about 150 publications have reported its related contributions. Many promising methods have been proposed, such as multi-modal tensor super-resolution (M^2TSR) [12], simultaneous super-resolution and recognition (S^2R^2) [16], discriminative super-resolution (DSR) [3], RQCr color features for degraded images [17], local frequency descriptor (LFD) [18], coupled locality preserving mappings (CLPMs) [19], multi-dimensional scaling (MDS) [20], and coupled kernel embedding (CKE) [21]. These representative LR FR methods are listed in Table 1. Some comparisons between HR FR and LR FR, including advantages and disadvantages are summarized.

Generally, the straightforward way to solve LR FR problem is super-resolution (SR), which first reconstructs HR faces from several LR faces and then performs recognition with the super-resolved HR images. This kind of methods is steadily developed within the last decade. Since 2005, many researchers have studied simultaneous SR and recognition for including facial features into an SR method as prior information. In recent years, resolution-robust feature representation methods have been gradually considered. However, all of them are limited to different constraints and do not completely solve LR FR problem. Future researches are still necessary for some related problems, such as accurate alignment and unified feature spaces, toward ultimately reaching the goal of resolution-robust face recognition.

Several survey papers [23–28] and books [29, 30] with good reviews on general face recognition have been published. S.Z. Li et al. [31, 32], and Wheeler et al. [33] definitely proposed issues, challenges, and prospects of biometric sensing and recognition at a distance for practical system deployments. However, none of them made specific reviews on LR FR. Thus, the contribution of this paper is to make a comprehensive survey on LR FR with detailed reviews on the existing methods and provide discussions on some open issues within this area. As the topic has attracted researchers since the year 2000, this review generalizes the main contributions during the last decade. No review of this nature can possibly cite every paper that has been published; therefore, we include only what we believe to be representative samples of important works and broad trends from recent years.

The rest of this paper is organized as follows. Section 2 provides a brief overview on LR FR with introductions of concept description, system architecture, method categorization, etc. In Sects. 3 and 4, a detailed review on LR FR methods is presented from two aspects: Super-Resolution for LR FR (Sect. 3) and Resolution-Robust Feature Representation for LR FR (Sect. 4). Section 5 describes the performance evaluations on LR FR methods. Section 6 gives a discussion on existing problems and future trends. Section 7 concludes the review.

2 Overview on LR FR

In this section, an overview on LR FR is given. First, the concept of LR FR is discussed after introducing two concepts: *the best resolution* and *the minimal resolution*. Then system architecture including the main strategies considered for LR FR is briefly described. Finally, LR FR methods are categorized, and some representative works are illustrated.

2.1 Concept descriptions

In an ideal scene, an HR face image with an abundance of pixels and details is important for recognition. That is to say, image resolution determines the capacity to discriminate fine features of image to a great extent. Before investigating LR FR, two concepts are introduced for describing the effects of image resolution on face recognition.

The best resolution, on which the optimal performance can be obtained with the perfect trade-off between recognition accuracy and performing speed. The problem of finding "what is the best spatial resolution for face recognition?" was originally raised by Kurita et al. [34]. They observed that faces characterized by global features could often be more easily recognized in lower resolutions. Pyramidal data structure was used to represent image sets in different resolutions from LR to HR with magnification factors from 1 to 64, with the order of 2, and they found that their classifier did not perform the best at the highest resolution. This is a phenomenon worth special attention.

The minimal resolution, also called *the threshold resolution*, above which the performance remains steady with

Table 1 Comparison between HR FR and LR FR

	HR FR	LR FR
Methods	PCA [4], LDA [5], ICA [6], LPP [7], NPE [22], EBGM [13], Bayesian FR [8]	M^2TSR [12], S^2R^2 [16], DSR [3], RQCr [17], LFD [18], CLPMs [19], MDS [20], CKE [21]
Advantages	More information, less noises, higher robustness, more features for use	Lower storing and computing costs
Disadvantages	Higher storing and computing costs	Less information, more noises, fewer available methods and tools
Applications	Entrance guard security, entertainment, etc.	Long distance surveillance, law enforcement, etc.

the resolution decreasing from the best resolution, but below which the performance deteriorates rapidly. Wang et al. [35] demonstrated that face image information could be divided into discriminative information (the individual information compared with other faces) and structure information (the common information of all face images under the same resolution). In addition, similar to the structure information, a new concept called face structural similarity was proposed in [36]. In fact, all of them are from an interesting phenomenon in the real world. When observing a person moving toward us, what we see first is a person moving nearer, then the identity of the person when the distance exceeds a fixed value. Here, the fixed value is just the minimal resolution for the human visual system. In other words, we first obtain the structure information, and then the discriminative information as the resolution exceeds the minimal resolution.

Investigation of the minimal resolution can be found in [35, 37–39]. Lemieux et al. [38] studied principal component analysis (PCA) system with AR database, and observed that the minimal resolution of face size is about 21×16 pixels. Wang et al. [35] explored PCA and linear discriminant analysis (LDA) system with AR database, and indicated the minimal resolution of the face size as 64×48 pixels. Boom et al. [37] investigated PCA and LDA system with FRGC database, and found the minimal resolution of face size to be 32×32 pixels. Fookes et al. [39] examined PCA and elastic bunch graph matching (EBGM) system with XM2VTS database, and obtained the minimal resolution of face size of about 42×32 pixels. From the above investigations, we can draw a conclusion that the minimal resolution depends on different methods and databases.

LR FR we address is automatically recognizing people by their face images in LR. An LR face image means that a face size is smaller than 32×24 pixels (with an eye-to-eye distance about 10 pixels), typically taken by surveillance cameras (maximum resolution 320×240, QVGA) without subject's cooperation and normally contains noises and motion blurs. In an LR image, exact delineation of facial features is not so trivial both for humans and machines. Generally, LR images discussed here roughly fall into three
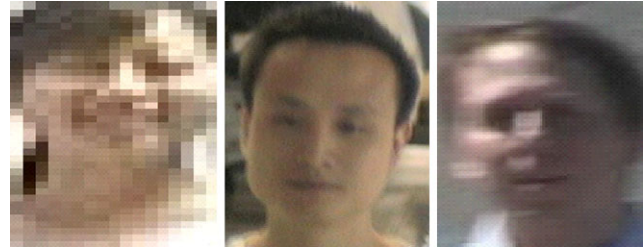
Fig. 1 Some examples of LR face images

cases [18] as follows. Such examples of LR are demonstrated in Fig. 1.

(1) *Small size*, for which the probe images have the insufficient number of pixels. According to the reports of FRVT2000 [40] on resolution experiment, the metric used to quantify resolution is eye-to-eye distance in pixels [41]. Nevertheless, the distance is not usually adopted in most LR FR systems but replaced by face size. In the conventional methodology, small size is sufficient for face recognition. However, when the size of face captured from surveillance camera is smaller than 32×24 pixels, or the size of face down-sampled from HR static image is smaller than 16×16 and even 6×6 pixels, most conventional methods will be of no effect [3, 16, 42].

(2) *Poor quality*, which is from the fact that probe images are provided in the resized form with blur (e.g., out of focus, interlace, and motion blur), variation of illumination and loss of details. Therefore, the underlying resolutions of the images are comparatively low [43], which means that even if the size of face is 200×200 pixels, there is no guarantee that the face is in HR. From the reports of MBGC2009 [44], the levels of "focus" and "illumination" are taken as image quality measures. However, they just consider LR problem from vision perspective rather than recognition purpose.

(3) *Small size & Poor quality*, which is from the combination of them naturally. In addition, Han et al. [45] introduced a new form of LR, called low gray-scale resolution.
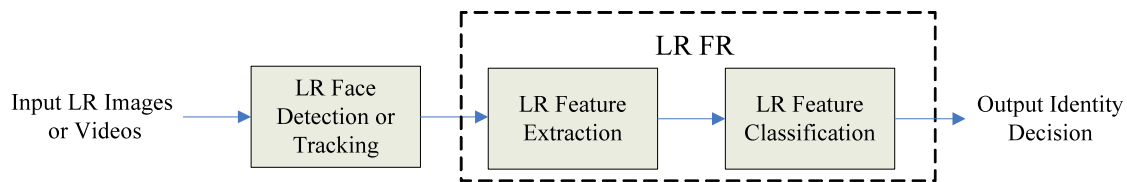
Fig. 2 The system architecture of LR FR

2.2 System architecture

Similar to conventional HR FR system, LR FR system also includes three main parts. That is LR face detection or tracking, LR feature extraction and LR feature classification, as illustrated in Fig. 2. In general, the latter two are collectively referred to as LR FR, which is the focus in this review.

LR face detection/tracking means that pre-processing, detection, tracking and segmentation of the faces are automatically performed on LR images or videos. State-of-the-art face detection methods such as AdaBoost [46], which are usually able to detect face images larger than 20×20, could fail in LR case. Therefore, building an efficient and effective face detection system is necessary for LR FR especially in surveillance applications. At present, two strategies are considered for this problem as follows.

The first is multicamera active vision systems, which are used for applications with large coverage areas and high detection efficiency [47–50]. Generally, they include two kinds of cameras. One is wide field of view (WFOV) cameras covering a large area to allow the detection and localization of subjects by a combination of motion detection, background modeling and skin detection. The other is narrow field of view (NFOV) cameras actively controlled to capture HR face images using pan, tilt, and zoom (PTZ) commands.

The second is to improve the existing HR face detection methods for solving LR problem. For example, Hayashi et al. [42] proposed the first method to detect very LR faces. They adopted upper-body images and frequency-band limitations of features based on AdaBoost-based detector. The detection rate was improved from 39 % to 73 % on 6×6 with MIT+CMU frontal database. Recently, Zheng et al. [51] developed a modified census transform technique by using boosting classifiers for detecting LR faces in color images.

Besides these two strategies, selection of the most suitable faces is also used for LR problem.

LR FR initially extracts resolution-robust features, and performs classification by matching the features to obtain the identity decision. The steps are similar to HR FR system from general framework. However, as opposed to HR FR, LR FR needs to consider the particular problem of dimensional mismatch. In practical face recognition applications, it is reasonable to assume that all gallery images are in HR. From classification perspective, LR will obviously
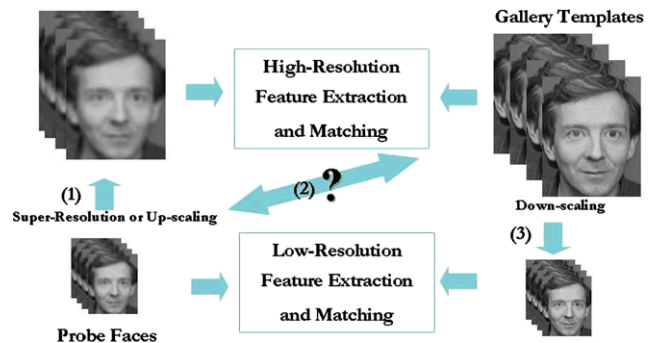

Fig. 3 Three general ways used for LR FR

cause the mismatch problem between gallery/probe pairs, as illustrated in Fig. 3. To deal with the problem, three general ways can be considered as follows:

(1) *Up-scaling* (*or interpolation*), such as cubic interpolation, is conventionally adopted in most of the subspace-based face recognition methods. For LR images, it does not introduce any new information but potentially brings noises or artifacts. Therefore, the process of up-scaling can be feasible under high-resolution or middle-resolution, but may drop in performance confronting with much lower resolution. Thus, it is generally not a good way for solving LR problem. For further refined solution, *super-resolution or hallucination* [52] can be employed to estimate HR faces from LR ones. However, it usually requires a lot of images, which belongs to the same scene with precise alignment, and it also needs large computation cost.

(2) *Unified feature space*, also called *inter-resolution* (*IR*) *space* [19], is used to project HR gallery images and LR probe ones into a common space. This idea seems to be direct and reasonable for solving LR problem. However, it is difficult to find the optimal inter-resolution space. And the two bidirectional transformations from both HR to IR and LR to IR may bring much more noises.

(3) *Down-scaling* seems to be a feasible solution for the mismatch problem. Unfortunately, it reduces the amount of available information, especially the high-frequency information mainly for recognition. However, down-scaling on both training/gallery and test/probe may improve the performance under very LR case such as 7×6 [3].

Here, we briefly summarize the three ways.

In the first way, most super-resolution methods are taken as indirect way. They firstly obtain super-resolved HR images from LR images and then for recognition.

The second way is to build unified feature space with optimal mapping techniques, or resolution-robust feature extraction techniques. They can be taken as direct way that performs on the original LR images.

Since the third way, i.e., down-scaling techniques are poor in performance for solving LR problem, and few researches focused on them, which are not our main considerations in this review.

2.3 Categorization and representatives

Based on the above analyses, LR FR methods can generally be classified into two categories, (1) *Indirect method*: super-resolution for LR FR, (2) *Direct method*: resolution-robust feature representation for LR FR. Of course, some methods may overlap category boundaries.

(1) Indirect method: super-resolution (SR) is initially used to synthesize the higher-resolution images from the LR ones, and then traditional HR FR methods could be applied for recognition. The landmark works in this category are face hallucination (Baker et al. [52]) and simultaneous SR and recognition (S^2R^2) (Hennings-Yeomans et al. [16]). Two criteria are considered for SR applications: visual quality and recognition discriminability. However, most of these methods [36, 53–55] aimed to improve face appearance but failed to optimize face images from recognition perspective. Recently, a few attempts were made to achieve these two criteria under very LR case [3].

(2) Direct method: resolution-robust feature representation is the process of directly extracting the discriminative information from LR images. The landmark works are color feature (Choi et al. [17]) and coupled locality preserving mappings (CLPMs) (Li et al. [19]). These methods can be separated into two groups further. One is feature-based method in which the resolution-robust features, such as texture [18], and subspace [56] information, are used to represent faces. However, some features used in traditional HR FR methods are sensitive to resolution. The other is structure-based method, e.g., multidimensional scaling (MDS) [20] in which the relationships between LR and HR are explored in resolution mismatch problem.

This categorization might provide useful information for potential readers. However, it is not unique. Alternative categorizations based on other criteria are also possible, such as relatively LR case and very LR case, single-modality based (against LR only) and multimodality based (against LR and other variations such as pose). Table 2 summarizes representative works of LR FR within the two categories. In Sects. 3 and 4, the general principle and typical methods of each category will be discussed. They are followed by a review of specific methods, including discussions of their pros and cons and suggestions for future researches.

3 Super-resolution for LR FR

Many researchers want to build face recognition systems with LR images obtained by web cameras or close-circuit television. However, the overall performance of LR FR needs great improvement. Compared with the development of resolution-robust face recognition methods, super-resolution (SR), or hallucination methods have gained much more attentions, due to many problems that degrade the quality of face images in LR case. In this section, some typical SR methods specifically satisfying the requirement for face recognition will be reviewed. In the last decade, most of the conventional SR methods called *vision-oriented SR* were taken as the indirect way, which is reconstruction followed by recognition. Recently, some researchers focused on simultaneous SR and recognition, and SR mainly for recognition, called as *recognition-oriented SR* obtaining promising results for LR classification.

3.1 Vision-oriented super-resolution for LR FR

The simplest way to increase resolution is direct interpolation of input images with methods such as nearest neighbor, bilinear, and bicubic. However, its performance is usually poor since no new information is added into the process [36]. In contrast to the interpolation, SR increases resolutions of images or video frames using the relationships among several images. Generally, SR can be divided into two classes [73]: *reconstruction-based method* (from input images alone) [74–77], and *learning-based method* (from other images) [36, 52–55, 57–59, 73, 78–87].

The reconstruction-based method reconstructs HR images based on sampling theory by simulating the image formation process. However, the method has some fatal shortcomings. Baker et al. [57] pointed out that the method inherits limitations when the magnification factor increases. Lin et al. [88] proposed the problem "Do fundamental limits exist for the reconstruction-based SR?". They further gave explicit bounds of the magnification factor based on perturbation theory analysis. Recently, Nasrollahi et al. [89] tried to solve the problem and improve the magnification factor from about two to almost four by using multilayer perceptron.

Most of the reconstruction-based methods are more suitable for synthesizing local texture, and they do not incorporate any specific prior information (e.g., face domain) about

Table 2 Categorization of LR FR methods

Categorization		Representatives
Indirect method: Super-resolution for LR FR	Vision-oriented SR	Face Hallucination [52, 57];
		Two-step Statistical Approach [53, 58];
		Eigentransformation [36];
		Extended Morphable Face Model [59].
	Recognition-oriented SR	Simultaneous SR and Recognition (S^2R^2) [16, 60–62];
		Multi-Modal Tensor SR (M^2TSR) [12];
		Discriminative SR (DSR) [3, 63];
		Support Vector Data Description (SVDD) [64, 65].
Direct method: Resolution-robust feature representation for LR FR	Feature-based method	Color-based Feature [17, 66–68];
		Texture-based Feature: Local Frequency Descriptor (LFD) [18, 69];
		Kernel Class-dependence Feature Analysis (KCFA) [56].
	Structure-based method	Eigenspace Estimation (EE) [15];
		Coupled Locality Preserving Mappings (CLPMs) [19, 70];
		Multi-Dimensional Scaling (MDS) [20, 71, 72];
		Coupled Kernel Embedding (CKE) [21].

the super-resolved images. Therefore, they are usually applicable to generic object or scenes rather than face images. Few researchers focused on enhancing face images, though Yu et al. [74] proposed a new method for enhancing LR face videos. Moreover, the performance of the reconstruction-based methods is severely influenced by the following factors [88]: the level of noise existing in the LR images, the accuracy of point spread function (PSF) estimation and the accuracy of alignment. In other words, higher level of noise or poorer PSF estimation and alignment will result in less improvement in resolution.

The learning-based method, also known as *face hallucination* [52], which is the focus in this review, is always used to enhance resolutions of face images compared with the reconstruction-based method. In essence, the learning-based SR is used to learn the relationships between LR and HR corresponding to different face images in a training set, and then use these learnt relationships to predict fine details for LR probe images (stored by image pixels, image patches, or coefficients of alternative representations). Establishing a good learning model to obtain the prior knowledge is the key to the learning-based method. At present, the commonly used learning models include the PCA model [36], image pyramid model [57], Markov model [82], etc.

How can we evaluate the quality of hallucinated HR images for face recognition? Three goals should be reached step by step. The first goal is to obtain HR images from visual perspective only. This is also the basic target for SR. Then the HR images are expected to be more like face images. Finally, we hope the HR face images are more like someone's face from recognition perspective.

For realizing the three goals, Liu et al. [58] introduced two different data constraints: soft constraint and hard constraint. The former was to beautify faces and make the results more like the mean face, corresponding to the first two goals. And the latter was to faithfully reproduce facial details to be exactly the same as the input face, similar to the third goal. Enlightened by Liu's work, Zou et al. [3] designed two new constraints. They included a data constraint and a discriminative constraint. The data constraint was $\|I_H - \mathbf{R}I_L\|^2$, to estimate the reconstruction error in the HR image space to make use of the information from HR training images. It was opposed to $\|\mathbf{D}I_H - I_L\|^2$ used in the LR image space in the conventional methods. The discriminative constraint was to use class label information to boost recognition performance. For the detailed discussions about the two constraints, please refer to Sect. 3.2.2.

Zou et al. [3] further categorized learning-based SR methods into two classes, namely *maximum a posteriori* (*MAP*)*-based method* [52–54, 57, 58, 73, 79–84] and *example-based method* [36, 55, 59, 78, 85–87]. It is noticeable that some methods overlap category boundaries. Other categories can also be discussed, such as single-frame-based and multiframe-based, intensity-based and frequency-based, global-based and local-based as well as global&local. In this paper, MAP-based and example-based are taken as the main categorization and other categories as a supplement.

3.1.1 MAP-based method

In this method, the goal is to find an optimal solution maximizing the posterior probability $p(I_H|I_L)$ to obtain super-

resolved HR images, i.e.,

$$\overline{I_H} = \arg\max_{I_H} p(I_H|I_L)$$

$$= \arg\max_{I_H} \frac{p(I_L|I_H)p(I_H)}{p(I_L)}. \tag{1}$$

In this formula, since $p(I_L)$ is a constant for LR image I_L is known already, the model can be simplified as follows:

$$\overline{I_H} = \arg\max_{I_H} p(I_L|I_H)p(I_H). \tag{2}$$

$p(I_L|I_H)$ is the probability of obtaining I_L when HR image I_H is given, and depends on the distribution of the noise deriving from the process of down-sampling. $p(I_H)$ denotes the prior of I_H. Therefore, the key of MAP-based method is to estimate $p(I_H)$.

For $p(I_H)$, different algorithms have different solutions. Baker et al. [57] first proposed the idea of face hallucination and led the precedent of learning-based method. They estimated $p(I_H)$ by using an image Gaussian pyramid under Bayesian formulation. The method obtained high-frequency components from a parent structure based on training face images; however, it intrinsically relied on a complicated statistical model. Similar to Baker's work, Capel et al. [83] also used MAP estimators, with the difference that Capel divided a face image into six unrelated parts, and applied PCA on them separately. Dedeoglu et al. [81] extended Baker's work to hallucinate face video by exploiting spatiotemporal constraints, and they reported a very high ($\times16$) magnification factor for LR case.

Based on Baker's work [57] and Freeman's work [84], Liu et al. [53, 58] proposed a two-step statistical method. It integrated a parametric model called global face image I_H^g recording common facial properties, with a nonparametric model called local feature I_H^l carrying individualities, to generate the HR image. Then the model (2) could be naturally transformed into

$$\overline{I_H} = \arg\max_{I_H^g, I_H^l} p\big(I_L|I_H^g\big) p\big(I_H^g\big) p\big(I_H^l|I_H^g\big). \tag{3}$$

They applied PCA linear inferences to maximize $p(I_L|I_H^g) \times p(I_H^g)$ and get an optimal global face image I_H^g, and a Markov random field prior was used to maximize $p(I_H^l|I_H^g)$ for obtaining a local feature I_H^l. However, this method depends on an explicit down-sampling function, which is sometimes unavailable in practice.

Enlightened by Liu's work, many methods treating face hallucination as a two-step problem have been proposed [73, 80]. They all perform as the following two-step process. First, a global face image containing low-frequency information is obtained, which looks smooth and lacks some detailed features. Second, a residue face image keeping high-frequency information is synthesized. And then the residue image is piled onto the global image to get the final super-resolved face images. For example, Li et al. [80] used a MAP criterion for reconstructing both the global image and the residual image. Jia et al. [73] proposed a unified tensor space representation for hallucinating low-frequency and middle-frequency information, and then recovered high-frequency part by patch learning.

In addition, some methods are performed in transformed feature space rather than pixel density domain. Zhang et al. [82] performed SR in frequency domain with inferring discrete cosine transform (DCT) coefficients instead of estimating pixel intensities in spatial domain. Alternating component (AC) coefficients in DCT were inferred by the Markov network of low-level vision. Subspace methods are also applied to restrict the reconstructed HR image locating within face subspace, such as PCA [54] and kernel PCA subspace [79]. However, most of these SR methods only focus on frontal faces, and fail to deal with unconstrained variations in pose, illumination, and expression.

3.1.2 Example-based method

In this method, the HR image $\overline{h}$ is reconstructed as a linear or nonlinear combination of the HR training images h_i by finding an optimal solution from the given LR image l, i.e., it can be mathematically written as

$$\overline{h} = \sum_i \alpha_i h_i. \tag{4}$$

The key of example-based method is to determine the weight coefficients α_i. They minimize the error caused by the linear or nonlinear approximation of the LR training images l_i (the pairs of h_i) for l as follows:

$$\alpha = \arg\min_{\alpha_i} \left\| l - \sum_i \alpha_i l_i \right\|^2. \tag{5}$$

For obtaining α_i, different algorithms have different models.

Wang et al. [36] proposed a representative example-based method and treated the hallucination problem as a transformation between LR and HR. They used PCA to fit an input LR face image as a linear combination of LR training images. The HR image was then synthesized by replacing the LR training images with their HR counterparts while retaining the same combination coefficients α_i. However, the linear PCA model could not capture distinct structures of the input face efficiently and only focused on global estimation without paying attention to local details. Thus, the results seemed unclear, lacked detailed features, and caused some distortions. Moreover, they designed a mask to avoid artifacts on hair and background, and performed hallucination in the interior region of face. In fact, local modeling and appropriate smoothing can be adopted to handle these artifacts

properly. That is to say, the idea of two-step in [53] can be used to compensate high-frequency features for the work.

Compared with Wang's work [36] operated in eigenface space, Liu et al. [55] performed as the two-step way in patch-tensor space. LR image was first partitioned into overlapped patches. HR patches were inferred respectively based on TensorPatch model and then fused together using a local distribution structure to form the hallucinated result. To further enhance the quality of the HR image, the coupled PCA method was developed for residue compensation. While the method added more details to the face, it also introduced more artifacts. Therefore, whether to adopt residue compensation techniques and when to do them is critical for super-resolution (SR).

Besides eigenface and tensor space, manifold learning techniques are used for example-based SR. Manifold learning theory suggests that the subspace of face images has an embedded manifold structure. The high-dimensional structure formed by HR face images is homeomorphic with a geometric structure in LR space. It means that the features of LR and HR face images share a common topological structure, and thus, they are coherent through the structure. Therefore, some ideas of manifold learning such as local linear embedding (LLE) [78] and locality preserving projection (LPP) [85]-[86] are introduced into SR and are discussed as follows.

Chang et al. [78] introduced the idea of LLE into SR with neighbor embedding. They assumed that training LR and HR images form manifolds with similar local geometry in two distinct feature spaces and used the training image pairs to estimate the weight coefficients for reconstruction. However, they treated SR as a patch-based single-step technique without compensation for local image details.

Zhuang et al. [85] developed locality preserving hallucination method based on LPP [7]. It combined LPP and radial basis function (RBF) together to hallucinate a global HR face. Compared with Wang's work [36], the hallucinated global HR face contained more detailed features. However, there were more noises in the local features such as contour, nostril, and eyebrow, because LPP resulted in the loss of nonfeature information. To improve the details of the synthesized HR face, they developed a residue compensation method based on patch by neighbor embedding [78].

Inspired by Zhuang's work, Ma et al. [86] employed additional constraints on neighborhood reconstruction for face hallucination. The input LR image determined the position at which the neighbors of a patch were chosen in training step, and then the hallucinated patches were reconstructed using optimal weights of the training image position-patches. The method did not incorporate any residue compensation step into hallucination. They gave the reason and discussed why the conventional two-step methods usually adopted the residue compensation step. According-

ing to them, some detailed facial information was lost in the first global reconstruction step.

In addition, some methods performed example-based SR on single-frame LR face image [59, 87]. For example, Park et al. [59] performed SR within PCA feature space with an extended morphable face model. They defined the model by the pixel correspondence between a reference face and other faces. By using the model, all face images were separated into extended 3D-shape and texture. Then the PCA-based SR method was implemented on both shapes and textures of LR input to reconstruct the corresponding HR shapes and textures respectively, and they were further synthesized into the result.

Recently, Hu et al. [87] also developed a single-frame SR method, like Liu's work [53]. They used both global and local constraints for hallucination; the difference was that their global model was derived from the nonrigid warping of reference face examples and the learning of the pixel structure. The warping could capture a moderate range of face variations. And the effects of warping errors were reduced by the adaptive weighting in the local prior model. Thus, the method could infer more faithful individual structures of the target HR face.

3.1.3 Discussion

Most of the vision-oriented SR methods have attempted to minimize mean-squared error (MSE) or maximize signal-to-noise ratio (SNR) between the original HR and the reconstructed SR images. As we all know, the performances of face recognition systems mostly rely on the ability to identify key facial features, which are typically captured by high-frequency components. However, high-fidelity reconstruction of low-frequency content in SR may dominate the image [90]. Therefore, obtaining a lower MSE or a higher SNR does not necessarily contribute to a better performance. That is to say, the primary goal of vision-oriented SR methods is to obtain a good visual reconstruction, but not usually designed from recognition perspective. As resolution decreases, SR becomes more vulnerable to unconstrained variations. It also introduces noises and distortions that affect recognition, especially when the probe identities could not be included in the process of training selection. Thus, one problem of what relationships exist between SR and recognition is generated.

Some researchers discussed the problem and made attempts to explore the potential of SR in recognition. Baker et al. [52] stated that no new information had been added during resolution enhancement. Also, face recognition methods could be developed to theoretically work as well on the LR images as they did on the hallucination results. Gunturk et al. [54] tried to reconstruct the necessary information required by face recognition system, especially with
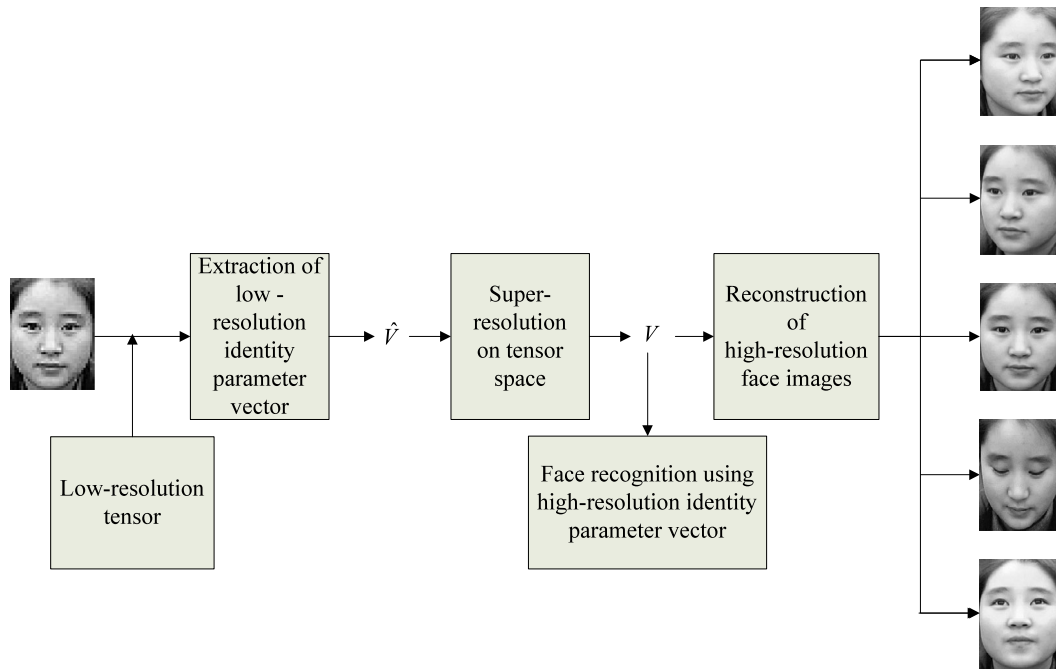
Fig. 4 The illustration of multimodal SR and recognition process in tensor space

the consideration of the statistics of noises and motion estimation errors. However, the method is unsuitable for pose variations, and is also heavily time-consuming. Wang et al. [36] explored whether hallucination could contribute to recognition. They found that the hallucinated images performed much better than the LR images but performances of both dropped when the face size decreased from 32×24 to 16×12 pixels with XM2VTS database. However, the improvement in recognition seemed not as significant as that in face appearance. They gave a relatively reasonable explanation that human visual system could better interpret the added high-frequency details in the reconstruction process.

In summary, most of the existing vision-oriented SR methods are not completely suitable for recognition. A promising way to further improve the robustness performance of SR for recognition is to embed SR into recognition. In the following subsection, we turn to recognition-oriented SR.

3.2 Recognition-oriented super-resolution for LR FR

Recognition-oriented SR is not to use SR before recognition in the conventional way. It embeds the elements of SR methods into face recognition. Specifically, it fuses the models of the image formation process and the prior information, together with feature extraction and classification to design methods for recognition [3, 12, 16, 64, 91–93]. Compared with vision-oriented SR, recognition-oriented SR maybe more suitable for LR FR due to the following two

observations. Firstly, it simultaneously performs *SR and feature extraction* with the direct goal of recognition. Secondly, it performs *feature SR* with the aim of reconstructing not only the low-frequency content (structure information) but also the high-frequency content (discriminative information) for recognition.

3.2.1 Simultaneous super-resolution and feature extraction

The method has drawn much attention. Here, we discuss two representative methods in detail as follows: multimodal tensor SR (M^2TSR) [12] and simultaneous SR and recognition (S^2R^2) [16]. Especially for S^2R^2, it is the first framework for realizing SR and recognition simultaneously.

Jia et al. [12] made some pioneering explorations in this field and proposed M^2TSR method, though it was still sequential and did not achieve complete simultaneity. It initially computed a maximum likelihood vector in the HR tensor space. Although it did not simultaneously perform against pose and illumination variation as illustrated in Fig. 4, face hallucination and recognition were unified in this way. The consideration of multimodality could contribute to LR FR. However, its disadvantage was that the tensor manipulations for reconstruction demanded high computation expenses.

Hennings-Yeomans et al. [16, 60, 61] showed that the performances of conventional SR methods were degraded under very LR case. Thus, they proposed S^2R^2 method to combine identification with reconstruction for dealing with LR problem by introducing the constraints between LR and

Fig. 5 The framework of S²R²
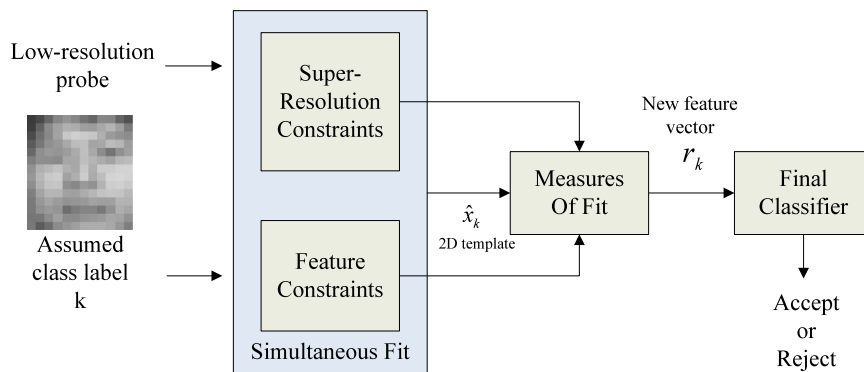
HR images in a regularization form, as illustrated in Fig. 5. Formula (6) denotes the base model of S²R². y_p, $f_g^{(k)}$, and x denote the input LR probe image, the gallery image in the kth class, and the output HR image, respectively; B, L, and F represent operators for down-sampling, smoothness and feature extraction, respectively; besides, α and β are the regularization parameters. The goal of the S²R² model is to obtain a suboptimal output HR image x for satisfying the need of vision and recognition simultaneously.

$$\|Bx - y_p\|^2 + \alpha^2\|Lx\|^2 + \beta^2\|Fx - f_g^{(k)}\|^2. \qquad (6)$$

In addition, the base S²R² model was improved by involving the cases of multiframes or multicameras version $B^{(i)}$. Furthermore, the base SR prior model $l^{(k)}$ and feature extraction (F_L) were modified based on multiresolutions version (l or L). The modified model is shown in (7), where α, β, γ are the regularization parameters, and B denotes the image formation process. The base S²R² model and the improved version tested on CMU Multi-PIE database on 6×6 obtained the accuracies of 62.8 % and 73 %, in comparison with the PCA baseline method at 47.1 %.

$$\|B^{(i)}x - y_p^{(i)}\|^2 + \alpha^2\|Lx - l^{(k)}\|^2$$
$$+ \beta^2\|Fx - f_g^{(k)}\|^2 + \gamma^2\|F_LBx - f_L^{(k)}\|^2. \qquad (7)$$

Compared with the general indirect SR methods, S²R² improves identification accuracy and gets promising results on 6×6. However, the parametric optimization needs to be repeated for each gallery image in the database, especially for large databases; thus, their formulation is quite time-consuming. Also, this method assumes that gallery and probe images are in the same pose, frontal or localized perfectly, directly resulting in its inefficiency under many general scenarios. Therefore, how to obtain the appropriate regularization parameters and reduce the computational complexity are two important issues in this model.

3.2.2 Feature super-resolution

The method is also called *feature hallucination*, which was innovatively proposed by Li et al. [93] to reconstruct HR features instead of HR images for face recognition. The kernel version of support vector data description (SVDD) [94] was used to synthesize HR discriminative features both for vision and recognition perspective [64, 65]. SVDD approximated the support of objects belonging to the normal class. Its main idea was to find a ball that could achieve two conflicting goals simultaneously. One was that it should be as small as possible and the other was that it should contain as much training data as possible with equal importance. However, the method is only for frontal faces and its generalization ability remains doubtful.

Besides, some methods are used for video applications. For example, Arandjelovic et al. [91] proposed an extended generic model called shape-illumination manifold (gSIM) framework by separating illumination and down-sampling effects for feature SR. Their experiments on both the Cambridge database [95] and the Toshiba database reported promising results for face recognition. However, the method requiring video sequences at enrollment makes it impractical for surveillance scenarios.

In addition, we introduce two representative recognition-oriented SR methods: nonlinear mappings on coherent features (NMCF) [92] and discriminative SR (DSR) [3]. Both of them introduce classification discriminability into SR process.

Huang et al. [92] proposed NMCF method with canonical correlation analysis (CCA) to establish coherent features between LR and HR images represented by PCA. Motivated by Zhuang's work [85], they also applied the radial basis function (RBF) mapping to build the regression model by adopting the advantages of RBF, such as fast learning and generalization ability. NMCF was evaluated on 12×12 with FERET database and obtained the accuracy of 84.4 % compared with 36.9 % of the PCA baseline method.

Recently, Zou et al. [3, 63] proposed the DSR method with two constraints (new data constraint and discriminative
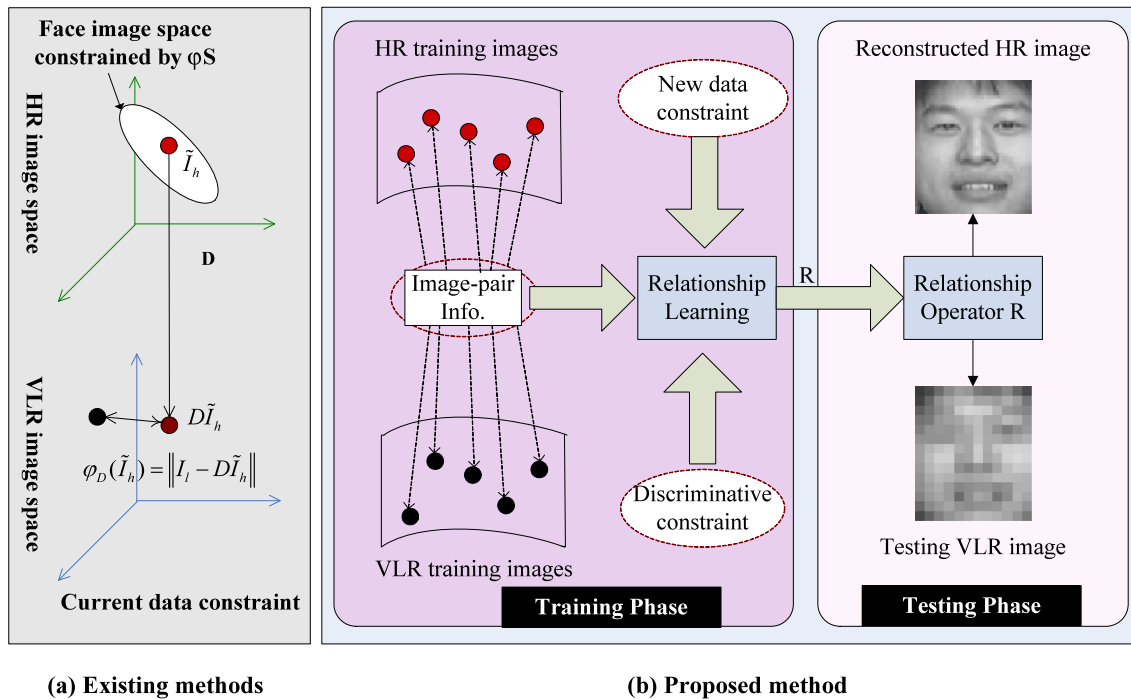
(a) Existing methods   **(b) Proposed method**

Fig. 6 The framework of discriminative SR for recognition with two new constraints

constraint). It modeled the SR problem as a regression problem in the kernel space under very LR case such as 16×12 and 7×6, as illustrated in Fig. 6. Formula (8) shows the mapping relationships under the two constraints. The former is the new data constraint, while the latter is the discriminative constraint. The conventional SR method employs the data constraint $\|\mathbf{D}I_h^i - I_l^i\|^2$ ($\mathbf{D}$ is the down-sampling operator) to make full use of the information in LR space for SR. However, this data constraint may not work well for very LR case because of the limited information carried by LR space. So, the data constraint is changed into $\|I_h^i - \mathbf{R}I_l^i\|^2$, where $\mathbf{DR} = \mathbf{I}$. The discriminative constraint is to use the class label information of the training data for improving the discriminability. DSR shows its superiority from both visual quality and recognition performance. For example, super-resolution results on 16×12 with Extended Yale B database are shown in Fig. 7. Also, DSR obtained the recognition accuracy of 73.5 % on 7×6 with CMU PIE database in comparison with 40.5 % in the PCA baseline method.

$$\overline{\mathbf{R}} = \arg\min_{\mathbf{R}} \frac{1}{N} \sum_{i=1}^{N} \|I_h^i - \mathbf{R}I_l^i\|^2 + \lambda \|\mathbf{DR} - \mathbf{I}\|_F^2$$
$$+ \gamma \big(mean\{\|I_h^i - \mathbf{R}I_l^j\|^2 \mid class(I_h^i) = class(I_l^j)\}$$
$$- mean\{\|I_h^i - \mathbf{R}I_l^j\|^2 \mid class(I_h^i) \neq class(I_l^j)\}\big). \quad (8)$$

A brief discussion on the similarities and differences of DSR and NMCF is as follows: Both DSR and NMCF require a training set containing LR and HR image pairs to learn the nonlinear mappings from LR to HR feature space, followed by the reconstruction of SR images or features. Compared with NMCF, DSR performs more efficiently when LR images are used for training/gallery sets. Conversely, when HR training/gallery sets are used, the performance of NMCF is better than DSR.

3.2.3 Discussion

Some successes have been achieved by recognition-oriented SR methods such as S²R² [16] and feature SR [93]. However, they just provide the framework for recognition-oriented SR, and their recognition performances largely depend on different reconstruction regularization models and feature extraction techniques. Some common problems are still unsolved in these methods. For example, it is unclear what kind of reconstruction regularization method is more appropriate for recognition. In addition, feature extraction is known to be sensitive to large appearance changes due to pose, illumination, expression, etc. To combine super-resolution and feature extraction perfectly is also a big issue for the future work. A possible way to handle these problems is to adopt more robust feature extraction techniques, which is the focus in the following section.

3.3 Summary of super-resolution for LR FR

A summary of super-resolution (SR) for LR FR is as follows:

Fig. 7 Comparison of different methods in Zou et al. [3]. (**a**) Input 16×12, (**b**) Bi-cubic, (**c**) Baker et al. [57], (**d**) Wang et al. [36], (**e**) Chakrabarti et al. [79], (**f**) Ma et al. [86], (**g**) Zou et al. [3], (**h**) Original 64×48

(a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)

(1) Most of the vision-oriented SR methods focus on obtaining a good visual reconstruction rather than a higher recognition rate; however, the essence of the recognition-oriented SR methods is to satisfy the need of recognition with LR images.

(2) Both vision-oriented SR and recognition-oriented SR are sensitive to different variations such as pose, and require lots of training samples of the same scene.

(3) In general, although SR methods require large computation costs, they have the potential advantages in very LR cases like 6×6.

4 Resolution-robust feature representation for LR FR

In contrast to super-resolution for LR FR, researches on resolution-robust feature representation in LR problem started around year 2008. The difficulties of finding the effective features in LR case render face recognition more complicated. Some typical features in HR case such as texture, shape, and color may fail in the LR case. Therefore, the only feasible option is to explore the potential of these features for LR FR. In addition, LR results in a dimensional mismatch problem under the subspace framework. Building interresolution space may provide a promising direction for solving this problem. Resolution-robust feature representation method is classified into two groups: *feature-based method* [17, 18, 56, 69, 96] and *structure-based method* [15, 19–21, 43, 97–100].

4.1 Feature-based method

This method identifies an LR face directly using the features extracted from probe images in resized forms. However, all the existing resolution-robust features are improved from the features used in HR FR, such as the improved color space [17] and the improved local binary pattern descriptor [18]. Similar to the categorization that successfully used in [24], we further classify feature-based method into two categories, that is, the *global feature-based method* and *local feature-based method*.

4.1.1 Global feature based method

In this method, the whole LR probe image, represented by a single high-dimensional vector containing the global low-frequency information, is taken as input. The advantage of this method is to implicitly preserve all the detailed texture and shape information, which is useful for recognizing LR faces. On the other hand, this method is also easily affected by variations such as pose and illumination like they do as HR FR. Here, two kinds of global methods are introduced: *color features* and *improved dimensionality reduction techniques*.

Color features are the representative global features. Choi et al. [17] first demonstrated that color-based features could significantly improve LR FR recognition performance compared with gray-based features. The idea was based on the boosting effects of color features on low-level vision [101]. A new metric called variation ratio gain (VRG) as shown in (9) was further defined to prove the significance of color effect on LR face images within the subspace face recognition framework. In VRG, $J^{\text{lum+chrom}}(\gamma)$ and $J^{\text{lum}}(\gamma)$ represent *variation ratio* parameterized by face resolution (γ) for color-augmentation-based feature subspace and intensity-based ones respectively. Here, $J(\gamma)$ is the ratio between the variations of extra-personal covariance matrices and those of intrapersonal covariance matrices for classification tasks. As illustrated in Fig. 8, as γ decreases, VRG (γ) gets larger, indicating that color components can compensate a decreased extra-personal variation by intensity component with LR. Based on the phenomenon, RQCr color space was selected for LR FR. Experiments on the hybrid database collected from CMU PIE, Color FEERT, and XM2VTS with probe resolution 15×15 tested on "RQCr" and "R" space achieved accuracies of 68 % and 54 %, respectively.

$$\text{VRG}(\gamma) = \frac{J^{\text{lum+chrom}}(\gamma) - J^{\text{lum}}(\gamma)}{J^{\text{lum}}(\gamma)} \times 100. \qquad (9)$$

VRG(γ) demonstrates the role of color in LR classification. However, no theories can prove that RQCr is more efficient for LR case in comparison with other color spaces. Therefore, to efficiently use color-based features for boosting intensity-based features is still an open issue. It is known that reducing the correlation of different color components is certainly helpful to HR FR, and even LR FR. Yang et al. [102] investigated the potential efficiency of color spaces, and proposed various normalized spaces such as the improved YRB space to enhance face recognition.

Choi et al. [67, 68] improved their work and proposed a color feature selection method by boosting-learning framework. Thirty-six different color components were used to form a color-component pool, and a weighted fusion scheme was used to fuse the selected color features at the feature
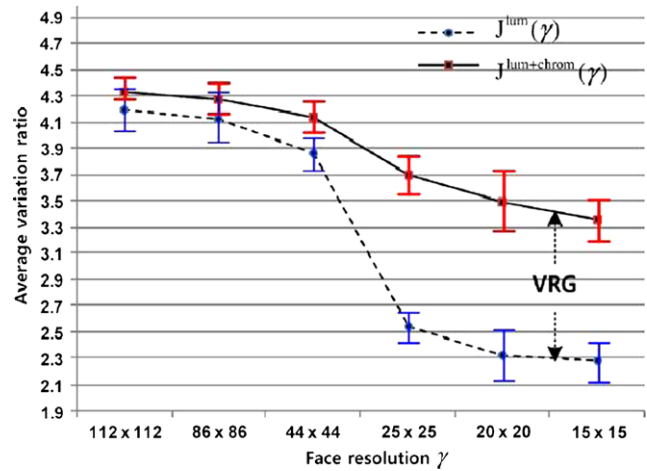

Fig. 8 Average variation ratios with respect to six different resolutions

level. The method was successfully evaluated on very LR images with SCface database. It improved the accuracy with RQCr space from 49.61 % to 62.78 % with the new color pool. The experiment indicated that the framework of color fusion was perhaps beneficial to LR FR. Furthermore, they adopted LBP features in color space for LR FR [66]. However, the role of color features for LR images is degraded by serious illumination variations despite of their successes in face recognition.

In addition, *improved dimensionality reduction techniques* are also proposed for the LR problem. Abiantun et al. [56] adopted the kernel class-dependence feature analysis (KCFA) method [103] for dealing with very LR case on the FRGC database Experiment 4. KCFA used a set of minimum average correlation energy filters to exploit higher-order correlations between training samples in the kernel space, and obtained the accuracy of 27.1 % on 8×8 compared with the PCA baseline method of 12 % on HR images. Wang et al. [96] proposed a new graph embedding method called FisherNPE for resolution-robust feature extraction, based on LDA and neighborhood preserving embedding (NPE) preserving both global and local structures on the data. Also, Bayesian probabilistic similarity analysis [8] of intensity differences between LR and HR images was used for classification. Photon-counting LDA [104] was proposed for coping with the LR FR, modeling the image pixels with Poisson distribution by the semiclassical theory of photon detection.

4.1.2 Local feature based method

In this method, the LR probe image is represented by a set of low-dimensional vectors containing the local high-frequency information. Compared with global method, local feature based method provides additional flexibility to recognize a face based on its parts, and is more robust to variations.

For example, Hadid et al. [105] proposed a novel discriminative feature space for detecting and recognizing LR faces from video, employing LBP representation and SVM classifier. Ahonen et al. [69] adopted local phase quantization (LPQ) method based on the assumption of point spread function. The method used the phase information of Fourier transformed images for LR FR, revealing that LPQ information in the high-frequency domain was almost invariant to blur. Afterwards, Lei et al. [18] made an improvement on LPQ and proposed local frequency descriptor (LFD) using not only phase information, but also magnitude information. Furthermore, the relative relationships between phase information were adopted without the assumption of point spread function instead of the absolute value. Also, a uniform pattern mechanism [14] was introduced to improve the performance.

4.1.3 Discussion

Finding resolution-robust features is a conventional issue for face recognition in the LR or HR case. Although many researchers concerned resolution-robust feature representation, the performance is far from perfect due to different complicated variations. Most methods mentioned above are only against one variation and not against multiple. For example, compared with local features, global features are more sensitive to illumination variation. With regards to pose and expression variation, local features are more susceptive than global features. A possible way to further improve the robustness may lie in the combination of local-based and global-based features. However, what features should be combined and how to combine them for concentrating their advantages are the future issues for LR FR.

4.2 Structure-based method

Compared with the feature-based method concerning resolution-robust features, the structure-based method focuses on constructing the relationships between LR and HR feature space for facilitating direct comparison of LR probe images with HR gallery ones from a classification perspective. The method aims to build the holistic framework for LR matching by especially solving the particular problem in LR FR, namely dimensional mismatch. Here, we introduce three kinds of structure-based methods. *Coupled mappings* [19] aim to find the structure relationships. *Resolution estimation* [43] determines the kinds of structures chosen for building LR FR system. Finally, the *sparse representation based method* [98] is adopted for representing LR probe images using HR training images from a structure perspective.

4.2.1 Coupled mappings

Choi et al. [15] first pointed out the dimensional mismatch problem, and proposed eigenspace estimation (EE) techniques for obtaining a common LR feature space for matching between LR and HR. Then Li et al. [19, 70] proposed a more general framework called unified feature space based on coupled mappings (CMs) (10). In CMs model, l_i and h_i represent an LR face image and an HR one, respectively, and $\mathbf{A}_L$ and $\mathbf{A}_H$ are two coupled mapping matrices. For LR FR, the mapping between each LR image and the corresponding HR image is expected to be as close as possible in the new unified feature space. Obviously, EE is one special case of CMs with $\mathbf{A}_H$ of down-sampling and $\mathbf{A}_L$ of identity matrix. Although CMs provides a promising framework for learning the relationships between LR and HR, it has an obvious shortcoming in poor discriminability for classification. Therefore, Li et al. introduced the locality preserving objective [7] into nonparametric CMs model, and proposed coupled locality preserving mappings (CLPMs) method as shown in (11). It significantly improved the performance by involving the weight relationships (W_{ij}) among data points. Evaluation on FERET database obtained the accuracy of 90.1 % on 12×12 in comparison with PCA baseline method with 61.8 %. However, CLPMs still exhibits sensitivity to the parameters and pose variations.

$$J(\mathbf{A}_L, \mathbf{A}_H) = \sum_{i=1}^{N_t} \left\| \mathbf{A}_L^T l_i - \mathbf{A}_H^T h_i \right\|^2, \tag{10}$$

$$J(\mathbf{A}_L, \mathbf{A}_H) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \left\| \mathbf{A}_L^T l_i - \mathbf{A}_H^T h_j \right\|^2 W_{ij}. \tag{11}$$

Other works with the aim of improving CMs have been developed. Zhou et al. [100] improved the classification discriminability of CMs by introducing linear discriminant relationships [5] between intraclass scattering and interclass scattering into CMs. Ren et al. [106] adopted canonical correlation analysis with local discrimination criterion [107] to compute the two coupled mapping matrices. With the process of regularization and piecewiseness on feature space, the method showed its superiority compared with CMs/CLPMs in both recognition accuracy and time complexity. Furthermore, Ben et al. [108] used the ideas of CMs to couple gait feature with LR face images and map them onto a common space for LR FR.

Recently, Ren at al. [21] further introduced the kernel tricks into CLPMs and proposed coupled kernel embedding (CKE) method for dealing with LR FR. In the CKE model (12), Ψ and Φ represent two different nonlinear mappings such as the Gaussian-quadratic kernel function. Experiments on the CMU Multi-PIE database obtained the accuracy of 84 % on 6×6. Although the Rank-1 accuracy on
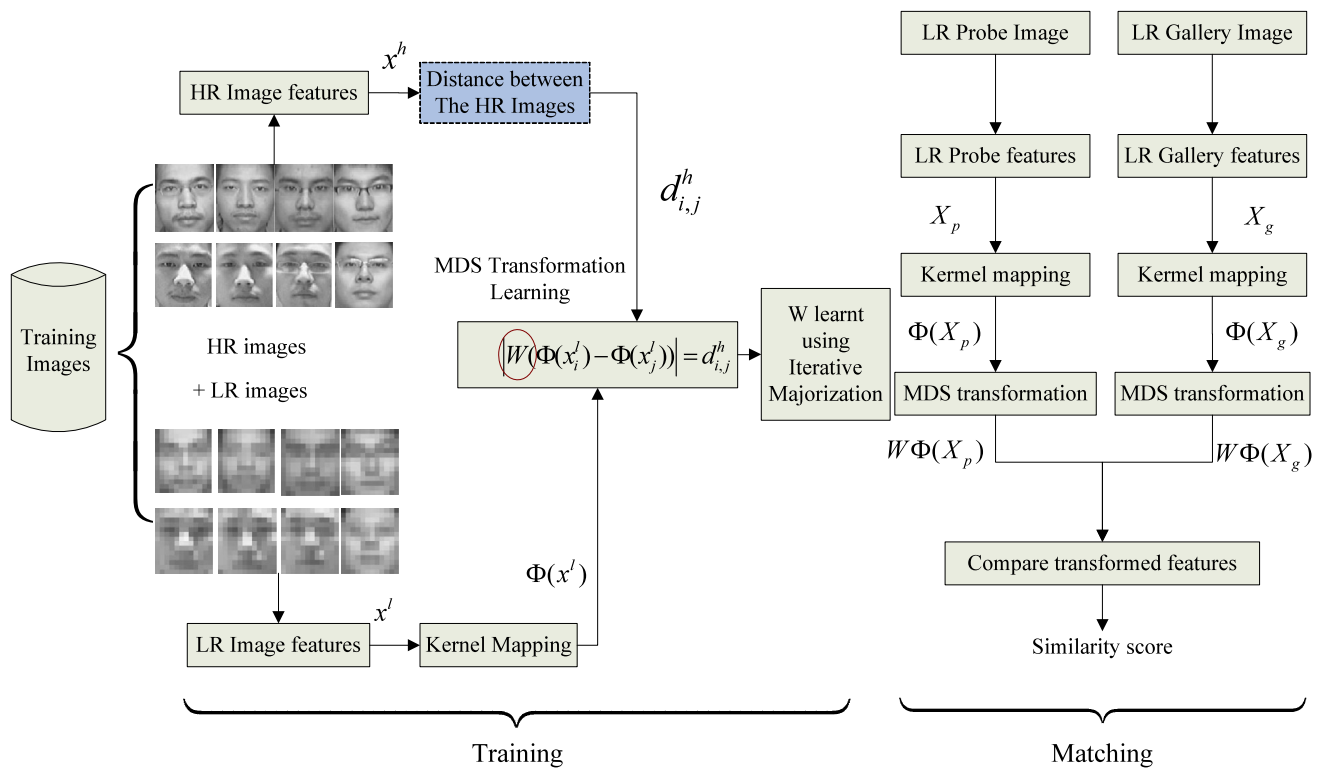
Fig. 9 The framework of MDS transformation learning method

SCface database is only 11 %, it still outperformed the LR baseline method with 4 %. By the kernel tricks, on one hand, CKE improved the classification performance; on the other hand, it increased the time complexity.

$$J(\mathbf{A}_L, \mathbf{A}_H) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \left\| \mathbf{A}_L^T \Psi(l_i) - \mathbf{A}_H^T \Phi(h_j) \right\|^2 W_{ij}. \quad (12)$$

In resolution mismatch problem, there exist three relationships (LR vs. LR, LR vs. HR, and HR vs. HR) involved in data. CMs/CLPMs only considered LR vs. HR. Deng et al. [109] further considered the other two relationships and adopted regularized coupled mappings with two new color spaces to get more information. However, the efficiency of the method empirically depends on the regularized parameters.

Similar to the work in [19, 109], Biswas et al. [20, 71, 72] skillfully utilized the three relationships between LR and HR, as illustrated in Fig. 9. During training, they embedded LR images into a new Euclidean space in order to achieve the best distances between their HR counterparts using multidimensional scaling (MDS) [110], and a transformation matrix W was obtained. During the test, LR gallery and probe images were transformed independently using the learned transformation matrix. Then the matching process was performed. It should be emphasized that

they highlighted the pose problem involved in LR recognition. This is an important contribution for researches on LR FR. They evaluated MDS on CMU Multi-PIE (8×6) and SCface database (12×10), and obtained the accuracies of 52 % and 71 %, respectively. In their experiments for LR FR, MDS performed better than sparse representation based super-resolution [111].

For further extension, here we provide a more general CMs model (13), including super-resolution (SR) and resolution-robust feature extraction. $\mathbf{F}$ represents feature extraction or subspace dimensionality reduction techniques. Then different stable features can be integrated into the framework. When $\mathbf{A}_L$ is replaced by SR constraints with the settings of $\mathbf{A}_H = \mathbf{I}_M$, the new model will be turned into SR. That is to say, SR is one special case of the model. Although S^2R^2 [16] is essentially different from the general indirect SR methods, it can also be represented by the new model when $\mathbf{A}$ and $\mathbf{F}$ are simultaneously used from both reconstruction and recognition perspectives. When $\mathbf{F}$ is provided as the form of kernel processing, the model will be turned into CKE [21] and MDS [20]. Finally, if $\mathbf{F}$ is ignored and just $\mathbf{A}$ is preserved, it returns to the base CMs model. In short, the new model is more general. However, to efficiently compute the two coupled mapping matrices $\mathbf{A}_H$ and $\mathbf{A}_L$ is also the key to the new model like the CMs

model.

$$J(\mathbf{A}_L, \mathbf{A}_H) = \sum_{i=1}^{N_t} \left\| \mathbf{A}_L^T \mathbf{F}_{l_i} - \mathbf{A}_H^T \mathbf{F}_{h_i} \right\|^2. \qquad (13)$$

In fact, ideas similar to CMs are also applied to other problems of face recognition. Lin et al. [112] proposed the idea of common discriminant feature extraction to solve the heterogeneous face recognition problems such as matching between visual (VIS) image and near infrared (NIR) image, and photo-sketch recognition. Recently, Lei et al. [113] proposed coupled spectral regression (CSR) to address VIS-NIR recognition. Furthermore, they improved CSR method in [114], which was also evaluated on LR images.

4.2.2 Resolution estimation

The method is another way proposed for dealing with the dimensional mismatch problem. It determines the kinds of structures chosen for building the LR FR system. Wong et al. [43] proposed two innovations for the LR problem. One was the concept of an underlying resolution, which did not rely on the size of face image. The other was that the local features sensitive to resolution were exploited for LR classification. Based on the innovations, they proposed a resolution detection and compensation framework for dynamically choosing the appropriate face recognition system. A similar method was proposed by Pedro et al. [115]. They developed the concept of estimating the acquisition distance in three different scenarios (close, medium, and far distance). And the distance was taken as the weight to fuse two systems (PCA-SVM system and DCT-GMM system) at the score-level. They demonstrated that training with medium distance images was a good way to control the performance degradation due to the varying distance.

In a way, the compensation frameworks show the potential of multiple face recognition systems for addressing LR FR. For example, color feature selection framework [67] is just a typical one. A combination of different classifiers [116, 117] also provided multimodal fusion at the classifier-level for the LR problem. In addition, a combination of several common methods was also proposed for dealing with the LR problem in [97]. They first adopted sparse representation [98] to describe patches represented by LBP features with different sizes. Then AdaBoost was used to select the most discriminative patches for classification. However, compared with feature selection [118], the method tested on Extended Yale B database showed poor performance. Thus, it remains doubtful whether such combination is efficient for LR FR, even just for HR FR.

4.2.3 Sparse representation based method

Sparse representation is first proposed by Wright et al. [98] for coping with robust classification problem. It is recently

warmed and has become one of the standard methods of face recognition within the literature followed by many researchers. They cast the recognition problem as one of classifying among multiple linear regression models. If the number of features was sufficiently large, and the sparse representation was correctly computed, they demonstrated that the choice of features was no longer critical. It was right even in the down-sampled images, though their work was not specialized for the LR case. However, like most of the other methods, sparse representation also requires the training/gallery samples covering different variations such as pose, illumination, and expression. It will be a big obstacle for real applications.

Furthermore, Yang and Wright et al. [111] adopted sparse representation for super-resolution (SRSR) on face images. More recently, inspired by their work, Bilgazyev et al. [90] performed SRSR on high-frequency components learned by wavelet decomposition-based rules. They reported that the recognition performance outperformed SRSR [111] and S^2R^2 [16] for the CMU PIE database. Moreover, Shekhar et al. [99] proposed an LR FR method with especially handling illumination variations based on sparse representation. In short, sparse representation may provide a new theoretical framework for dealing with LR FR problem in the future.

4.2.4 Discussion

The key of the coupled mappings is to obtain an interresolution space or unified feature space for solving the mismatch problem. The interresolution may be intuitively less than HR but greater than LR. However, EE [15] takes a new LR space as the interresolution space. Also, CLPMs [19] achieves the best performance in the interresolution less than the resolution of probe images. Thus, a question will be naturally generated as which the interresolution is suitable for LR FR. It is difficult to answer due to many factors such as different feature representations, face databases, and applications. Other strategies in the structure-based method, such as resolution estimation, system compensation, and sparse representation may provide the promising directions for addressing LR FR.

4.3 Summary of resolution-robust feature representation for LR FR

A summary of resolution-robust feature representation for LR FR is as follows:

(1) In general, resolution-robust feature representation methods are mainly applicable in relatively LR cases rather than very LR cases like 6×6.
(2) The feature-based methods can be used for multiple resolutions from HR to LR, but they need online training. However, the structure-based methods are more suitable

Table 3 Comparison between super-resolution and resolution-robust feature representation

Comparison issues	Super-resolution	Resolution-robust feature representation
Main purpose	For visual quality	For recognition discriminability
Resolution tolerance	Very LR	Relatively LR
Performing mode	Indirect way	Direct way
Computational complexity	High	Low
Unconstrained variations	Sensitive	Very sensitive
Requirements of training samples	Many	Few

for offline training, but they are mainly used for a single resolution application with the balance between efficiency and speed.

(3) Obviously, the combination of feature-based and structure-based methods will contribute to very LR cases. More characteristics about resolution-robust feature representation methods are shown in Table 3, which also provides a comparison between super-resolution and resolution-robust feature representation.

5 Evaluations on LR FR methods

In order to have a clear idea on various LR FR methods, it is important to evaluate them based on certain evaluation criteria with some standard LR face databases. Unfortunately, such a requirement is seldom satisfied in practice due to the lack of general criteria and databases originally developed for LR FR. At present, LR FR methods are just evaluated based on HR FR criteria and databases.

As for the evaluation criteria for face recognition, generally speaking, face recognition can be described in terms of the following two tasks. One is face verification where the input is a face image and an identity label, and the output is a binary decision, yes or no, to confirm the identity label. Thus, face verification is a 1-to-1 problem. The other is face identification where the input is a face image, and the output is to assign the identity label assigned to the face by pointing out the subject. Evidently, face identification is a 1-to-N problem and also popularly called face recognition. Moreover, some new tasks have been proposed, such as screening and watch list, which are the transformed versions of verification or identification [62].

In this review, face identification (recognition) is our focus. In identification, the *cumulative match characteristic* (CMC) curve is usually adopted to plot the percentage of identification accuracy (IDA) vs. Rank. (IDA is the number of correctly assigned labels to the total number of input faces.) All these metrics are generally derived from HR FR and can also be used in LR FR with different resolutions. Other metrics, such as the minimal resolution and execution time, are occasionally applied to evaluate LR FR methods. However, the performances of different methods depend on

different databases to some extent. Therefore, a few standard LR face databases are necessarily built for fair comparisons, which is the future work. In this section, we evaluate some representative methods on image-based standard databases and video-based real environment, respectively, to find the problems existing in LR FR. The performances and the limitations of these methods are finally summarized.

5.1 Evaluation on image-based standard databases

While HR FR is investigated in depth, many image-based standard databases have been established to compare performances of these methods, such as AR [119], Yale [120], Extended Yale B [121], and CAS-PEAL [122]. However, there is currently no database for LR FR, so that for evaluations on most of the existing LR FR methods, face images with frontal view, neutral expression, and illumination variations are selected and preprocessed such as downsampling and blurring instead of the actual LR images taken by surveillance cameras. Currently, the widely used databases for LR FR are FERET [123]/Color FERET [124], CMU PIE [125]/CMU Multi-PIE [126], FRGC [127], and SCface [128].

FERET [123] consists of one gallery set and four probe sets (fafb, fafc, dup1, dup2). There are 1,196 images of 1,196 subjects in the gallery set and the four probe sets contain 1,195, 194, 722, and 234 images, respectively. Fafb probe images are obtained as frontal view and expression variation, and usually taken as the experiment data in most LR FR methods such as CLPMs [19], LFD [18], S^2R^2 [16], and NMCF [92]. As for other methods such as M^2TSR [12], the experimental data are collected from the four probe sets. Moreover, Color FERET [124] is also used for evaluating LR FR methods such as RQCr [17]. It is worth noting that S^2R^2 is just evaluated on 6×6, and LFD and RQCr are tested on blurred images.

CMU PIE [125] includes 41,368 images of 68 subjects (21 samples/subject). Among them, 3,805 images have coordinate information of facial feature points. Some methods such as EE [15], RQCr [17] and DSR [3], evaluated on this database, select the frontal view images with illumination variation for experiments from the 3,805 images. CMU Multi-PIE [126] is a recent extension of CMU PIE database.

Table 4 Experiments and performances of LR FR on FERET database

No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
1196	12×12	1196/1195	Frontal, Expression	CLPMs [19]	90.1 %	61.8 % (PCA)
1196	33×30	1196/1195	Frontal, Blur	LFD [18]	86 %	48 % (LBP)
865	6×6	N/A	Frontal, Expression	S^2R^2 [16]	62 %	60 % (LDA)
295	14×9	1/4	Pose, Illumination	M^2TSR [12]	74.6 %	51.4 % (Tensor)
1196	12×12	1196/1195	Frontal, Expression	NMCF [92]	84.4 %	36.9 % (PCA)
140	15×15	1/4	Frontal, Blur, Color	RQCr [17]	68 %	54 % (R)

Note: "Frontal View" is abbreviated as "Frontal" in Tables 4–9 with the consideration of the table size. "1196/1195" means that 1,196 individuals for the gallery set and 1,195 individuals for the probe set fafb. Other "gallery/probe" shows the number of gallery/probe for one individual

Table 5 Experiments and performances of LR FR on CMU PIE database

No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
68	11×11	3/39	Frontal, Illumination, Neutral Expression	EE [15]	58.4 %	48.3 % (PCA)
68	15×15	1/20	Frontal, Illumination	RQCr [17]	68 %	54 % (R)
68	7×6	8/13	Frontal, Illumination, Expression	DSR [3]	73.5 %	40.5 % (PCA)

It has a total of 337 subjects (compared with 68 subjects in CMU PIE) who participated in one to four different recording sessions, separated by at least a month (unlike CMU PIE, where all images of each subject are captured on the same day in a single session). As in CMU PIE, facial pose, expression, and illumination variations due to flashes from different angles are recorded. The frontal view images with neutral expression and illumination variations are chosen for LR FR methods, e.g., S^2R^2 [16], MDS [20], CKE [21], and evaluated on CMU Multi-PIE, which is also similar to CMU PIE.

FRGC [127] consists of 50,000 images divided into training and validation sets. The training set is designed for training methods and the validation set is used for assessing performance of methods in a laboratory setting. The validation set includes 16,028 images of 466 subjects. The FRGC database consists of six experiments. Among the six experiments, in experiment 1, the gallery consists of a single controlled still image of a person and each probe consists of a single controlled still image. Experiment 2 studies the effect of using multiple still images of a person on performance. In experiment 2, each biometric sample consists of the four controlled images of a person taken in a subject session. Experiment 4 measures recognition performance from uncontrolled images. In experiment 4, the gallery consists of a single controlled still image, and the probe set consists of a single uncontrolled still image. S^2R^2 [16], KCFA [56], and DSR [3] methods are evaluated on FRGC database experiments 1, 2, and 4, respectively.

SCface [128] is a database of static images of human faces. Images were taken in an uncontrolled indoor environment using five video surveillance cameras of various qualities. The database contains 4,160 static images (in visible and infrared spectrum) of 130 subjects. Images from different quality cameras mimic the real-world conditions and enable robust testing, emphasizing law enforcement and surveillance scenarios. MDS [20], DSR [3], and CKE [21] methods are evaluated on this database.

The performances of LR FR methods tested on FERET, CMU PIE, CMU Multi-PIE, FRGC, and SCface are summarized in Tables 4, 5, 6, 7, and 8, respectively. Moreover, other databases such as XM2VTS [129], UMIST [130], ORL [131], and KFDB [132] are also used to evaluate LR FR methods such as RQCr [17], NMCF [92], and SVDD [64], which are shown in Table 9. Most methods listed in Tables 4–9 aim at face identification (recognition) with CMC (Identification accuracy vs. Rank-1) as the evaluation criterion except for KCFA in Table 7 aiming at face verification with receiver operating characteristic (ROC) curve (Verification accuracy vs. False acceptance rate).

On these databases, different methods are able to be compared on a relatively fair basis. And one can easily pick up the methods with good performances. The direct performance comparison of LR FR methods is not provided in this review mainly due to the lack of standard LR databases and evaluation protocols; however, the performances of baseline methods shown in the above tables can be taken as references for rough comparisons. Here, we discuss the results and attempt to give some relatively reasonable comparisons. Compared with other recent databases, FERET database mainly covers expression variation, which is relatively old. Tested on FERET, resolution-robust feature representation methods such as CLPMs [19] slightly

Table 6 Experiments and performances of LR FR on CMU multi-PIE database

No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
337	6×6	7/13	Frontal, Illumination, Neutral Expression	CKE [21]	84 %	53 % (LPP)
337	8×6	N/A	Frontal, Illumination, Neutral Expression	MDS [20]	52 %	40 % (PCA)
224	6×6	1/13	Frontal, Illumination, Neutral Expression	S^2R^2 [16]	62.8 %	47.1 % (LDA)

Table 7 Experiments and performances of LR FR on FRGC database

No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
300	6×6	1/20	N/A	S^2R^2 [16]	55 %	44 % (CFA)
311	7×6	8/2	Near Frontal, Illumination, Neutral Expression	DSR [3]	56.5 %	38.5 % (PCA)
466	8×8	8014/8014	Near Frontal, Expression, Severe Illumination	KCFA [56]	27.1 %	12 % (PCA)

Table 8 Experiments and performances of LR FR on SCface database

No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
130	16×16	N/A	Pose, Blur, Expression	CKE [21]	11 %	4 % (LPP)
130	12×10	1/4	Near Frontal, Expression	MDS [20]	71 %	19 % (PCA)
130	16×14	5/5	Pose, Blur, Expression	DSR [3]	22.5 %	14.5 % (PCA)

Table 9 Experiments and performances of LR FR on other databases

Database	No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
XM2VTS	133	15×15	1/7	Frontal, Illumination	RQCr [17]	68 %	54 % (R)
UMIST	20	14×11	10/5	Pose	NMCF [92]	93 %	90 % (PCA)
ORL	40	8×8	5/5	Pose, Illumination, Expression	NMCF [92]	95 %	84.5 % (PCA)
KFDB	N/A	16×16	N/A	Frontal, Neutral Expression	SVDD [64]	93 %	84 % (N/A)

outperform recognition-oriented super-resolution methods such as S^2R^2 [16] and NMCF [92]. The result further supports that the former is more suitable for simple conditions, e.g., single expression variation. However, for relatively complicated databases such as FRGC and SCface, the situation will be reversed. That is to say, recognition-oriented super-resolution methods obtain much better performance, although the resolution-robust feature representation method MDS [20] tested on SCface is greatly superior to DSR [3]. This result is attributed to the use of frontal view probe images. In addition, the evaluations on CMU PIE or CMU Multi-PIE demonstrate that recognition-oriented super-resolution methods such as S^2R^2 [16] and DSR [3] are more easily against unconstrained variations, e.g., illumination and expression. However, the resolution-robust feature representation method CKE [21] also obtains promising results on CMU Multi-PIE mainly with the help of the kernel trick.

From the above analyses, we can know that it is very difficult to rank all the methods based on the existing and widely used image-based standard databases. Also, we find that no method can satisfactorily handle the LR problem in face recognition under all complicated variations. For example, M^2TSR [12] is the only method specially designed to deal with pose and illumination problem in LR classification. However, its performance on the FERET database with 14×9 is only 74.6 %, which is still far below the requirement of practical use. Therefore, this review mainly focuses on the discussions of different methodologies for LR FR, in hope of providing helpful technical insights and promising directions for interested researchers.

5.2 Evaluation on video-based real environment

After evaluations on image-based standard databases, let us conduct some further experiments on video-based real environment to evaluate the performances of various LR
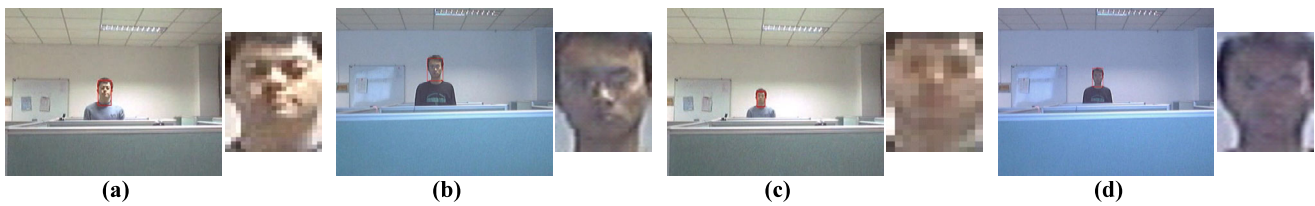
| **(a)** | **(b)** | **(c)** | **(d)** |

Fig. 10 The illustration of the surveillance camera system for different distances and illuminations. In each scenario, the *left* is a frame acquired by the surveillance camera, while the *right* is an up-sampled version of the frame. (**a**) 2.0 meters distance with good illumination, (**b**) 2.0 meters distance with poor illumination, (**c**) 3.0 meters distance with good illumination, (**d**) 3.0 meters distance with poor illumination

FR methods. Some typical video databases such as CMU MoBo [133], Honda/UCSD [134], and CLEAR2006 [135] are widely used for video-based face recognition even for LR FR. Recently, some researchers attempted to build LR databases to mimic real environment. Huang et al. [136] presented a database named "labeled faces in the wild" (LFW), containing images that were collected from the web. Although it has natural variations in pose, illumination, expression, etc., there is no guarantee that such a database can accurately capture all variations found in the real world [137]. Besides, most objects in LFW only have one or two images, which might not be enough to conduct different face recognition experiments. Yao et al. [138] created a face video database, UTK-LRHM, obtained from long distances and with high magnifications, both indoors and outdoors under uncontrolled surveillance conditions. Also, they developed a wavelet transform based multiscale processing algorithm, which was used to deal with image degradations related to long-distance acquisition and was successful in improving recognition rate. Ni et al. [139] manually put together a remote face database including the face images with variations due to occlusion, blur, pose, and illumination, which were taken from long distances and under unconstrained outdoor environments. The remote database was just evaluated on two state-of-the-art face recognition methods including baseline methods such as PCA, LDA, SVM, and the recently developed methods, e.g., sparse representation. But it did not form into a complete database for LR FR. In a word, there is still no LR benchmark database for public comparisons at present.

To evaluate the performances of the existing LR FR methods for real applications, we construct a video-based face database with uncooperative subjects in an uncontrolled indoor environment using a video camera (QVGA, 320×240). Images from the low-fidelity quality camera mimic the real-world LR FR conditions factually. The testing environment mainly includes two conditions, good/poor illumination, and 2.0/3.0 meters distance, which are illustrated in Fig. 10. The database contains 800 training/gallery images (20 images per subject) and 160 testing/probe videos (4 videos per subject) from 40 subjects. Here, we do not discuss the process of detection for capturing and tracking face

frames as that is out of the scope of this paper, though it is very important for video-based recognition.

In our experiments, S^2R^2 [16] and CLPMs [19] are used for the real environment evaluation, as they are the representatives in recognition-oriented super-resolution method and resolution-robust feature representation method, respectively. For S^2R^2, all the HR training images and the corresponding down-sampled LR images are used for computing Fisherface features for representation, and learning regularization parameters and the discriminant coefficient. For the details of the system implementation, please refer to [16]. In the testing procedure, since the current S^2R^2 method needs large computation and fails to reach the requirement of real time, frame images are taken as the probe samples at a fixed distance such as 2.0 and 3.0 meters instead of performing on video directly. For CLPMs, all the HR training images and the corresponding down-sampled LR images are used for computing the two coupled mapping matrices. For the details of the training process, please refer to [19]. Compared with S^2R^2, CLPMs is more suitable for real-time application. In experiments, only one aligned HR image (64×48) with frontal view and good illumination is taken as gallery for each subject. For each subject, five probe images are randomly selected at 2.0 and 3.0 meters distance, and their face sizes are resized into 32×24 and 21×15, respectively.

We design three group experiments for evaluating the effects of distance (resolution), illumination, and misalignment on the abilities of the two methods for dealing with the LR FR problem. For testing the effect of distance (resolution), the other two factors are taken as good illumination/manual alignment. In the same way, 2.0 meters distance/manual alignment and 2.0 meters distance/good illumination are set for testing the effects of illumination and alignment, respectively.

In 2.0 meters distance, S^2R^2 and CLPMs obtain the promising identification accuracies (IDA vs. Rank-1) of 80 % and 87.5 %, which are shown in Fig. 11. Compared with 2.0 meters case, the performances of S^2R^2 and CLPMs decreased rapidly in 3.0 meters distance, and the decrease amplitude of CLPMs is larger than that of S^2R^2. It is worth noting that both of their decrease amplitudes are even close to that of the LR case. In the experiments for testing the
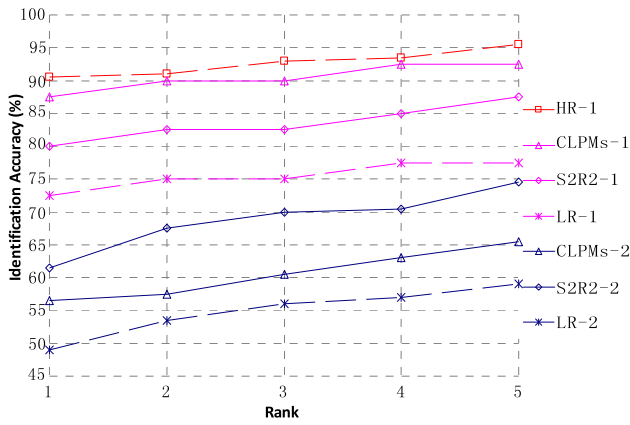
Fig. 11 The illustration of cumulative matching characteristic (CMC) performance (IDA vs. Rank1-5) due to the effect of *distance* (*resolution*) under good illumination/manual alignment. "1" denotes 2.0 meters distance and "2" denotes 3.0 meters distance. HR denotes matching at base high-resolution and LR denotes matching at probe low-resolution, and they all use the Fisherface method
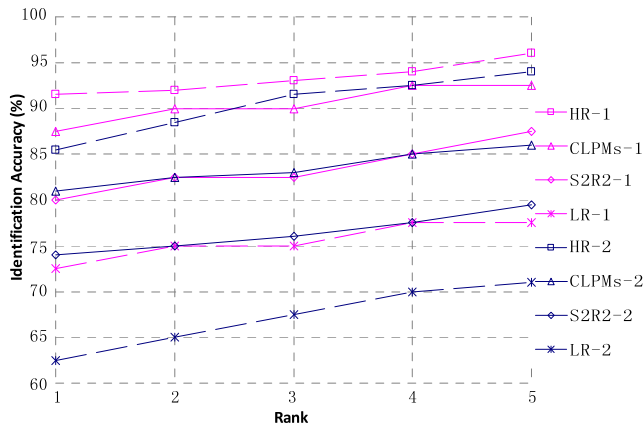

Fig. 12 The illustration of CMC performance due to the effect of *illumination* under 2.0 meters distance/manual alignment. "1" denotes good illumination and "2" denotes poor illumination
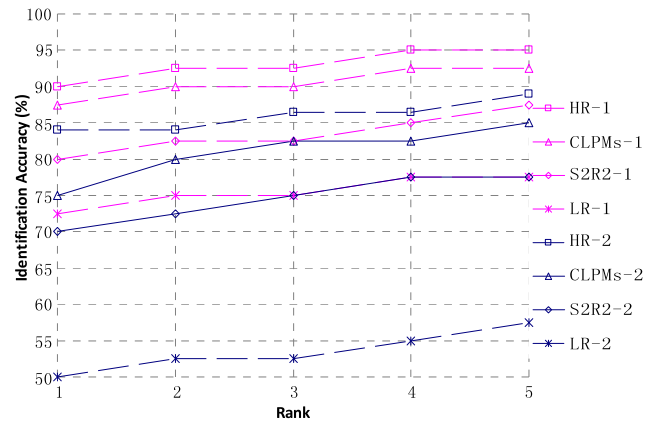

Fig. 13 The illustration of CMC performance due to the effect of *alignment* under 2.0 meters distance/good illumination. "1" denotes manual alignment and "2" denotes misalignment
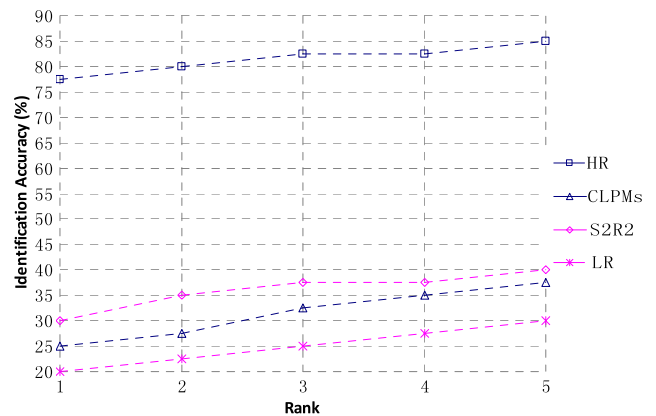

Fig. 14 The illustration of CMC performance due to the effect of *distance* (*resolution*), *illumination*, and mis*alignment* together. That is, the methods are tested on 3.0 meters distance, poor illumination, and misalignment

effect of illumination, the performances of all methods are decreased with the amplitude of about 6 % except for the LR case about 10 % when the illumination is changed from good into poor case, which are illustrated in Fig. 12. For testing the effect of misalignment, the probe images have to be manually aligned with the positions of two eyes and mouth. The performances of all methods are improved by alignment, which is shown in Fig. 13. And the result of CLPMs with alignment is even close to the HR case. However, S^2R^2 without alignment is even inferior to the LR case with alignment. It is enough to demonstrate that alignment is very important for face recognition, especially for LR FR. Finally, we gather the three factors (distance, illumination, and misalignment) for testing the methods, that is, 3.0 meters distance, poor illumination, and misalignment. From Fig. 14, the results of S^2R^2 and CLPMs are very poor, and they only

obtain the accuracies of 30 % and 25 %, which are a little better than that of the LR case.

From the results of our tests on the real-world environment, we can see that the existing LR FR methods have not performed well under real-world scenarios. The representatives of LR FR methods such as S^2R^2 and CLPMs are severely affected by complicated conditions, e.g., distance, illumination, misalignment. In such cases, compared with S^2R^2, the performance of CLPMs is much poorer, which is probably attributed to no stable features involved in CLPMs. However, S^2R^2 is also unsuitable for real-time application mainly due to the complication of the model parameter learning. In a word, the existing LR FR methods should be greatly improved so as to be used for real applications.

5.3 Discussion

Based on the evaluations on image-based standard databases and video-based real environment, we can have an idea

Table 10 Comparison between evaluation on image-based and video-based database

Database	S^2R^2	CLPMs	Comparison issues
Image-based (e.g. FERET)	62 % (6×6)	90.1 % (12×12)	Manual Alignment/Frontal View/Expression/Image/Illumination/Blur
Video-based (real environment)	30 % (21×15)	25 % (21×15)	Misalignment/Pose/Expression/Video/Illumination/Motion Blur/Noise

on the performances of the existing LR FR methods. As for image-based standard databases, the down-sampled images contain some variations in pose, illumination, expression, which can mimic real testing environment to some extent. However, the probe images with frontal view and neutral expression are generally selected and aligned for experiments in current testing procedures, which is unrealistic for surveillance applications without subjects' cooperation especially for moving subjects. As for video-based real environment, LR images captured from surveillance cameras typically contain the misalignment problem, which will severely affect the performances of face recognition systems. Besides, other particular factors in video-based testing, e.g., noises, motion blurs, and detection of LR face images will further make the problem much more complicated. And the requirement of real-time applicability is also an obstacle for most LR FR methods. In short, environmental conditions (e.g., noise, motion blur, distance, illumination), individualities of different subjects (e.g., pose, expression), and other factors (e.g., misalignment) will bring a lot of effects on LR FR. Table 10 gives a concise comparison of the representative methods S^2R^2 and CLPMs evaluated on image-based and video-based database, respectively. From the table, we can see that the performance evaluated on the video-based database is obviously inferior to image-based database. And the promising results on image-based database are largely attributed to the preprocessing steps such as alignment and the selection of frontal view images. Therefore, we can conclude that the existing LR FR methods are not fully satisfied for LR application, especially for video-based real environment. Future directions addressing LR FR problem are generalized in detail in the following section.

6 Existing problems and future trends

Compared with other subareas such as pose-invariant face recognition and illumination-invariant face recognition, LR FR is a new sub-area in face recognition. Thus, its first priority is to guarantee efficiency and accuracy rather than real time and low computational complexity. Although researchers have exerted efforts on improving LR FR methods, some specific problems still exist in real applications,

especially in surveillance scenarios. Based on the four main parts of LR FR systems, including preprocessing, facial representation, feature extraction, and feature classification, we point out four challenges and corresponding four future directions, namely *automatic alignment*, *insensitivity to multiple variations*, *resolution-robust feature extraction, and discriminative nonlinear coupled mapping*.

6.1 Automatic alignment

Alignment is one of the most important preprocessing issues in face recognition, especially in LR FR. In most LR FR methods such as S^2R^2 [16], RQCr [17] and CLPMs [19], face images are manually aligned and cropped with the positions of eyes and mouth. However, the way of manual alignment is very difficult in practical LR FR applications such as video surveillance. Furthermore, state-of-the-art face detectors such as AdaBoost [46] remain poor in detecting LR face images. To cope with this problem, some researchers attempted to propose automatic alignment techniques to register LR faces in raw images. Jia et al. [73] developed a pixel-wise alignment method by iteratively warping the probe face to its projection in the eigenface space. Park et al. [59] proposed an extended shape model with two constraints (face detection errors and shape estimation errors) to solve the misalignment problem. In a word, the development of automatic face alignment methods will facilitate the application of LR FR in real world.

6.2 Insensitivity to multiple variations

The performance of face recognition highly depends on unconstrained variation, e.g., pose, illumination, expression. Many methods have been proposed for reducing the effects of the changes, such as the Gabor filter function and the DCT on an edge map. However, most of the successful techniques could not be efficiently applied to LR data. For the LR case, the variations are the largest noises to some extent. Most of the existing LR FR methods assumed the constrained cases such as frontal pose, good illumination, and neutral expression, while few methods focused on facial representation against different variations.

For example, Shekhar et al. [99] proposed a synthesis-based method for handling the illumination-invariant LR

FR problem. Chang et al. [140] just examined the effect of resolution reduction with illumination variations. Arandjelovic et al. [91] adopted an extension of the generic shape-illumination method, which was invariant to changes in pose, illumination, and subject motion pattern. For pose variation, only two methods namely MDS [71, 72] and M^2TSR [12] definitely pointed out the pose problem, which shows that pose variation is one of the biggest obstacles in LR FR. 3D face model reconstruction [141] maybe an effective way for dealing with the problem. It should be emphasized that the M^2TSR method addressed not only the pose-invariant problem, but also illumination-robust recognition. However, a majority of the methods only considered single variation by preprocessing or selecting the most suitable samples before recognition. In fact, multiple variations also cause many effects on HR FR, which are still not fully resolved.

6.3 Resolution-robust feature extraction

Most of the effective features used in HR FR such as texture and color may fail in LR case. Thus, it is difficult to find resolution-robust features for LR FR, especially under facial and environmental variations such as pose, illumination, and expression. Choi et al. [17] made an attempt to adopt color features such as RQCr space and fusion of different color spaces for LR FR. It is known that color features are sensitive to illumination variation, which will be a big obstacle for LR FR application. Lei et al. [18] proposed a novel texture descriptor named LFD based on LBP and Fourier transformation. Abiantun et al. [56] used a new dimensionality reduction method called KCFA for dealing with LR feature extraction. Moreover, combination of global and local features, fusion of robust features, and effective subspace learning methods are future directions.

6.4 Discriminative nonlinear coupled mapping

Different resolutions between HR gallery images and LR probe images cause a dimensional mismatch problem in the traditional classification framework. In fact, this problem is the most essential problem in LR FR compared with HR FR. Other problems such as misalignment, noise affection, and lack of effective features commonly exist in face recognition system no matter HR or LR case. Li et al. [19] proposed the CMs/CLPMs model to build a unified feature space. Furthermore, this review gives the general CMs model including super-resolution and robust features for increasing the discriminability and generalization. Besides the relationships between LR and HR, other two relationships such as LR vs. LR and HR vs. HR were considered in [109]. To sum up, three problems in CMs need to be solved. The first is to improve its ability in discriminability. The second is how

to efficiently solve the eigenvalue decomposition problem and obtain the two optimal mapping matrices. The third is to generalize CMs from single LR to multiple LR applications even for across all resolutions. In addition, kernel tricks have great potentials in the application of LR FR such as CKE [21], MDS [20], KCFA [56], SVDD [64], and DSR [3]. Thus, the kernel can probably be used for describing nonlinear mappings.

Except for the four main directions, the way of *Multi-Frames* should be particularly considered for improving LR FR systems used in video surveillance applications. Most LR FR methods focus on single-frame (image-based) recognition, i.e., from only one LR input. Few methods deal with the multiframes (video-based) case. Hadid et al. [142] found that hidden Markov model (HMM) based methods with long sequences performed better than with short ones in both LR and HR, which means abundant information for recognition is included in video. Dedeoglu et al. [81] proposed the concept of video hallucination by exploiting spatiotemporal regularities. Wheeler et al. [143] adopted a sequence of video frames represented by the active appearance model (AAM) for LR FR. In a way, video-based feature extraction or video-to-video matching will provide a promising way for addressing the LR problem. Furthermore, multimodal biometric recognition systems, including LR FR may be used for recognition at a distance in the future. For example, multimodal face recognition is performed with LR, pose, and sketch images [144]. Fusion of LR face and finger veins is used for multimodal biometric recognition [145]. Integration of face and gait was proposed for human recognition at a distance in video [146].

7 Conclusions

This paper devotes to providing a comprehensive survey on LR FR. After giving an overview of the LR FR concept and introducing some structural categories of the existing methods, many representative methods have been reviewed and evaluated in detail. Discussions on major challenges as well as future research directions toward complete LR FR are provided.

Although significant progress has been made in the last decade, we believe that a robust LR FR system should be effective under the following variations:

- multiple resolutions, at a distance,
- noise, motion blur,
- orientation, pose, partial occlusion,
- illumination, expression.

From the discussions above, the conclusion can be drawn that LR is a challenging and interesting subarea in face recognition. The existing methods in LR FR fail to form a unified theoretical framework, and there are no standard databases and criteria to evaluate the performances of the LR FR methods. What are the effective features representing LR faces for recognition? How can we improve super-resolution processing to satisfy both vision and recognition purpose? Is there a unified feature space where LR faces are separable? All of them will spur researchers to create more effective methods. Answers to these questions may lead to a clearer understanding of LR FR even general LR object recognition, which are similar to the findings of nonlinear mappings in pose-invariant face recognition and linear subspaces in illumination-invariant face recognition.

References

1. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–459 (2003)
2. Pnevmatikakis, A., Polymenakos, L.: Far-field, multi-camera, video-to-video face recognition. In: Delac, K., Grgic, M. (eds.) Face Recognition, pp. 467–486 (2007)
3. Zou, W.W.W., Yuen, P.C.: Very low resolution face recognition problem. IEEE Trans. Image Process. **21**(1), 327–340 (2012)
4. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, June 1991, pp. 586–591 (1991)
5. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
6. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. IEEE Trans. Neural Netw. **13**(6), 1450–1464 (2002)
7. He, X.F., Yan, S.C., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using Laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 328–340 (2005)
8. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognit. **33**(11), 1771–1782 (2000)
9. Ruiz-del-Solar, J., Navarrete, P.: Eigenspace-based face recognition: a comparative study of different approaches. IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. **35**(3), 315–325 (2005)
10. Zhang, J.P., Pu, J., Chen, C.Y., Fleischer, R.: Low-resolution gait recognition. IEEE Trans. Syst. Man Cybern., Part B, Cybern. **40**(4), 986–996 (2010)
11. Wagner, A., Wright, J., Ganesh, A., Zhou, Z.H., Mobahi, H., Ma, Y.: Toward a practical face recognition system: robust alignment and illumination by sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 372–386 (2012)
12. Jia, K., Gong, S.G.: Multi-modal tensor face for simultaneous super-resolution and recognition. In: Proc. IEEE 10th Int. Conf. on Computer Vision (ICCV), Beijing, China, Oct. 2005, vol. 2, pp. 1683–1690 (2005)
13. Wiskott, L., Fellous, J.-M., Kruger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 775–779 (1997)
14. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
15. Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Feature subspace determination in video-based mismatched face recognition. In: Proc. IEEE 8th Int. Conf. on Automatic Face and Gesture Recognition (FG), Amsterdam, The Netherlands, Sep. 2008, pp. 1–6 (2008)
16. Hennings-Yeomans, P.H., Baker, S., Vijaya Kumar, B.V.K.: Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Alaska, USA, June 2008, pp. 1–8 (2008)
17. Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Color face recognition for degraded face images. IEEE Trans. Syst. Man Cybern., Part B, Cybern. **39**(5), 1217–1230 (2009)
18. Lei, Z., Ahonen, T., Pietikainen, M., Li, S.Z.: Local frequency descriptor for low-resolution face recognition. In: Proc. IEEE 9th Int. Conf. on Automatic Face and Gesture Recognition (FG), Santa Barbara, CA, USA, Mar. 2011, pp. 161–166 (2011)
19. Li, B., Chang, H., Shan, S.G., Chen, X.L.: Low-resolution face recognition via coupled locality preserving mappings. IEEE Signal Process. Lett. **17**(1), 20–23 (2010)
20. Biswas, S., Bowyer, K.W., Flynn, P.J.: Multidimensional scaling for matching low-resolution face images. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 2019–2030 (2012)
21. Ren, C.X., Dai, D.Q., Yan, H.: Coupled kernel embedding for low-resolution face recognition. IEEE Trans. Image Process. **21**(8), 3770–3783 (2012)
22. He, X.F., Cai, D., Yan, S.C., Zhang, H.-J.: Neighborhood preserving embedding. In: Proc. IEEE 10th Int. Conf. on Computer Vision (ICCV), Beijing, China, Oct. 2005, vol. 2, pp. 1208–1213 (2005)
23. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)
24. Chen, S., Tan, X., Zhou, Z.H., Zhang, F.: Face recognition from a single image per person: a survey. Pattern Recognit. **39**(9), 1725–1745 (2006)
25. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D face recognition: a survey. Pattern Recognit. Lett. **28**(14), 1885–1906 (2007)
26. Zhang, X.Z., Gao, Y.S.: Face recognition across pose: a review. Pattern Recognit. **42**(11), 2876–2896 (2009)
27. Zou, X., Kittler, J., Messer, K.: Illumination invariant face recognition: a survey. In: Proc. IEEE 1st Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS), Crystal City, VA, USA, Sep. 2007, pp. 1–8 (2007)
28. Ersotelos, N., Dong, F.: Building highly realistic facial modeling and animation: a survey. Vis. Comput. **24**(1), 13–30 (2008)
29. Li, S.Z., Jain, A.K.: Handbook of Face Recognition, 2nd edn. Springer, New York (2011)
30. Zhao, W., Chellappa, R.: Face Processing: Advanced Modeling and Methods. Academic Press, New York (2005)
31. Ao, M., Yi, D., Lei, Z., Li, S.Z.: Face recognition at a distance: system issues. In: Tistarelli, M., Li, S.Z., Chellappa, R. (eds.) Handbook of Remote Biometrics, Aug. 2009, pp. 155–168. Springer, Berlin (2009)
32. Li, S.Z., Schouten, B., Tistarelli, M.: Biometrics at a distance: issues, challenges, and prospects. In: Tistarelli, M., Li, S.Z., Chel-

lappa, R. (eds.) Handbook of Remote Biometrics, Aug. 2009, pp. 3–22. Springer, Berlin (2009)

33. Wheeler, F.W., Liu, X.M., Tu, P.H.: Face recognition at a distance. In: Li, S.Z., Jain, A.K. (eds.) Handbook of Face Recognition, 2nd edn., pp. 353–381. Springer, Berlin (2011)

34. Kurita, T., Otsu, N., Sato, T.: A face recognition method using higher order local autocorrelation and multivariate analysis. In: Proc. IEEE 11th Int. Conf. on Pattern Recognition (ICPR), The Hague, The Netherlands, vol. B, pp. 213–216 (1992)

35. Wang, J.D., Zhang, C.S., Shum, H.Y.: Face image resolution versus face recognition performance based on two global methods. In: Proc. IEEE 6th Asian Conf. on Computer Vision (ACCV), Jeju Island, Korea, Jan. 2004 (2004)

36. Wang, X.G., Tang, X.O.: Hallucinating face by eigentransformation. IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. **35**(3), 425–434 (2005)

37. Boom, B.J., Beumer, G.M., Spreeuwers, L.J., Veldhuis, R.N.J.: The effect of image resolution on the performance of a face recognition system. In: Proc. IEEE 9th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV), Grand Hyatt, Singapore, Dec. 2006, pp. 1–6 (2006)

38. Lemieux, A., Parizeau, M.: Experiments on eigenfaces robustness. In: Proc. IAPR/IEEE 16th Int. Conf. on Pattern Recognition (ICPR), Quebec, Canada, Aug. 2002, vol. 1, pp. 421–424 (2002)

39. Fookes, C., Lin, F., Chandran, V., Sridharan, S.: Evaluation of image resolution and super-resolution on face recognition performance. J. Vis. Commun. Image Represent. **23**(1), 75–93 (2012)

40. Blackburn, D.M., Bone, M., Phillips, P.J.: Face recognition vendor test 2000, FRVT2000. Evaluation Report. www.nist.gov/itl/iad/ig/frvt-2000.cfm (Feb. 2001)

41. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 large-scale experimental results. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 831–846 (2010)

42. Hayashi, S., Hasegawa, O.: A detection technique for degraded face images. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), New York, USA, June 2006, vol. 2, pp. 1506–1512 (2006)

43. Wong, Y.K., Sanderson, C., Mau, S., Lovell, B.C.: Dynamic amelioration of resolution mismatches for local feature based identity inference. In: Proc. IAPR/IEEE 20th Int. Conf. on Pattern Recognition (ICPR), Istanbul, Turkey, Aug. 2010, pp. 1200–1203 (2010)

44. MBGC2009. http://www.nist.gov/itl/iad/ig/mbgc.cfm

45. Han, H., Shan, S.G., Chen, X.L., Gao, W.: Gray-scale super-resolution for face recognition from low gray-scale resolution face images. In: Proc. IEEE 17th Int. Conf. on Image Processing (ICIP), Hong Kong, China, Sep. 2010, pp. 2825–2828 (2010)

46. Paul, V., Michael, J.: Rapid object detection using a boosted cascade of simple features. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Kauai, Hawaii, USA, vol. 1, pp. 511–518 (2001)

47. Prince, S.J.D., Elder, J., Hou, Y., Sizinstev, M., Olevsky, Y.: Towards face recognition at a distance. In: Proc. Int. Conf. on Crime and Security, the Institute of Engineering and Technology (IET), Savoy Place, London, England, June 2006, pp. 570–575 (2006)

48. Wheeler, F.W., Weiss, R.L., Tu, P.H.: Face recognition at a distance system for surveillance applications. In: Proc. IEEE 4th Int. Conf. on Biometrics: Theory, Systems, and Applications (BTAS), Washington, DC, USA, Sep. 2010, pp. 1–8 (2010)

49. Marchesotti, L., Marcenaro, L., Regazzoni, C.: Dual camera system for face detection in unconstrained environments. In: Proc. IEEE 10th Int. Conf. on Image Processing (ICIP), Barcelona, Catalonia, Spain, Sep. 2003, vol. 1, pp. 681–684 (2003)

50. Choi, H.-C., Park, U., Jain, A.K.: PTZ camera assisted face acquisition, tracking & recognition. In: Proc. IEEE 4th Int. Conf. on Biometrics: Theory, Systems, and Applications (BTAS), Washington, DC, USA, Sep. 2010, pp. 1–6 (2010)

51. Zheng, J., Ramirez, G.A., Fuentes, O.: Face detection in low-resolution color images. In: Proc. 7th Int. Conf. on Image Analysis and Recognition (ICIAR), Povoa de Varzim, Portugal, June 2010, pp. 454–463 (2010)

52. Baker, S., Kanade, T.: Hallucinating faces. In: Proc. IEEE 4th Int. Conf. on Automatic Face and Gesture Recognition (FG), Grenoble, France, Mar. 2000, pp. 83–88 (2000)

53. Liu, C., Shum, H.Y., Zhang, C.S.: A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In: Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, pp. 192–198 (2001)

54. Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., Mersereau, R.M.: Eigenface-domain super-resolution for face recognition. IEEE Trans. Image Process. **12**(5), 597–606 (2003)

55. Liu, W., Lin, D.H., Tang, X.O.: Hallucinating faces: tensorpatch super-resolution and coupled residue compensation. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, June 2005, vol. 2, pp. 478–484 (2005)

56. Abiantun, R., Savvides, M., Vijaya Kumar, B.V.K.: How low can you go? Low resolution face recognition study using kernel correlation feature analysis on the FRGCv2 database. In: Biometrics Symposium, Maryland, USA, Sep. 2006 (2006)

57. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. IEEE Trans. Pattern Anal. Mach. Intell. **24**(9), 1167–1183 (2002)

58. Liu, C., Shum, H.Y., Freeman, W.T.: Face hallucination: theory and practice. Int. J. Comput. Vis. **75**(1), 115–134 (2007)

59. Park, J.-S., Lee, S.-W.: An example-based face hallucination method for single-frame, low-resolution facial images. IEEE Trans. Image Process. **17**(10), 1806–1816 (2008)

60. Hennings-Yeomans, P.H., Baker, S., Vijaya Kumar, B.V.K.: Recognition of low-resolution faces using multiple still images and multiple cameras. In: Proc. IEEE 2nd Int. Conf. on Biometrics: Theory, Systems, and Applications (BTAS), Virginia, USA, Sep. 2008, pp. 1–6 (2008)

61. Hennings-Yeomans, P.H., Vijaya Kumar, B.V.K., Baker, S.: Robust low-resolution face identification and verification using high-resolution features. In: Proc. IEEE Int. Conf. on Image Processing (ICIP), Cairo, Egypt, Nov. 2009, pp. 33–36 (2009)

62. Hennings-Yeomans, P.H.: Simultaneous super-resolution and recognition. Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (Oct. 2008)

63. Zou, W.W.W., Yuen, P.C.: Learning the relationship between high and low resolution images in kernel space for face super resolution. In: Proc. IAPR/IEEE 20th Int. Conf. on Pattern Recognition (ICPR), Istanbul, Turkey, Aug. 2010, pp. 1152–1155 (2010)

64. Lee, S.-W., Park, J., Lee, S.-W.: Low resolution face recognition based on support vector data description. Pattern Recognit. **39**, 1809–1812 (2006)

65. Lee, S.-W., Park, J., Lee, S.-W.: Face reconstruction with low resolution facial images feature vector projection in kernel space. In: Proc. IAPR/IEEE 18th Int. Conf. on Pattern Recognition (ICPR), Hong Kong, China, Aug. 2006, vol. 3, pp. 1179–1182 (2006)

66. Choi, J.Y., Plataniotis, K.N., Ro, Y.M.: Using colour local binary pattern features for face recognition. In: Proc. IEEE 17th Int. Conf. on Image Processing (ICIP), Hong Kong, China, Sep. 2010, pp. 4541–4544 (2010)

67. Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Boosting color feature selection for color face recognition. IEEE Trans. Image Process. **20**(5), 1425–1434 (2011)

68. Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: A comparative study of preprocessing mismatch effects in color image based face recognition. Pattern Recognit. **44**(2), 412–430 (2011)

69. Ahonen, T., Rahtu, E., Ojansivu, V., Heikkila, J.: Recognition of blurred faces using local phase quantization. In: Proc. IAPR/IEEE 19th Int. Conf. on Pattern Recognition (ICPR), Tampa, Florida, USA, Dec. 2008, pp. 1–4 (2008)

70. Li, B., Chang, H., Shan, S.G., Chen, X.L.: Coupled metric learning for face recognition with degraded images. In: Advances in Machine Learning, pp. 220–233. Springer, Berlin (2009)

71. Biswas, S., Aggarwal, G., Flynn, P.J.: Pose-robust recognition of low-resolution face images. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, USA, June 2011, pp. 601–608 (2011)

72. Biswas, S., Aggarwal, G., Flynn, P.J., Bowyer, K.W.: Pose-robust recognition of low-resolution face images. IEEE Trans. on Pattern Analysis and Machine Intelligence, accepted for future publication (2013)

73. Jia, K., Gong, S.G.: Generalized face super-resolution. IEEE Trans. Image Process. 17(6), 873–886 (2008)

74. Yu, J.G., Bhanu, B.: Super-resolution restoration of facial images in video. In: Proc. IAPR/IEEE 18th Int. Conf. on Pattern Recognition (ICPR), Hong Kong, China, Aug. 2006, vol. 4, pp. 342–345 (2006)

75. Schulz, R.R., Stevenson, R.L.: Extraction of high-resolution frames from video sequences. IEEE Trans. Image Process. 5(6), 996–1011 (1996)

76. Hardie, R.C., Barnard, K.J., Armstrong, E.E.: Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. IEEE Trans. Image Process. 6(12), 1621–1633 (1997)

77. Elad, M., Feuer, A.: Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images. IEEE Trans. Image Process. 6(12), 1646–1658 (1997)

78. Chang, H., Yeung, D., Xiong, Y.: Super-resolution through neighbor embedding. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, June 2004, vol. 1, pp. 275–282 (2004)

79. Chakrabarti, A., Rajagopalan, A.N., Chellappa, R.: Super-resolution of face images using kernel PCA-based prior. IEEE Trans. Multimed. 9(4), 888–892 (2007)

80. Li, Y., Lin, X.Y.: An improved two-step approach to hallucinating faces. In: Proc. IEEE 3rd Int. Conf. on Image and Graphics (ICIG), Hong Kong, China, Dec. 2004, pp. 298–301 (2004)

81. Dedeoglu, G., Kanade, T., August, J.: High-zoom video hallucination by exploiting spatio-temporal regularities. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, June 2004, vol. 2, pp. 151–158 (2004)

82. Zhang, W., Cham, W.K.: Learning-based face hallucination in DCT domain. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Alaska, USA, June 2008, pp. 1–8 (2008)

83. Capel, D.P., Zisserman, A.: Super-resolution from multiple views using learnt image models. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, vol. 2, pp. 627–634 (2001)

84. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. Int. J. Comput. Vis. 40(1), 25–47 (2000)

85. Zhuang, Y.T., Zhang, J., Wu, F.: Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation. Pattern Recognit. 40, 3178–3194 (2007)

86. Ma, X., Zhang, J.P., Qi, C.: Hallucinating face by position-patch. Pattern Recognit. 43(6), 2224–2236 (2010)

87. Hu, Y., Lam, K.-M., Qiu, G.P., Shen, T.Z.: From local pixel structure to global image super-resolution: a new face hallucination framework. IEEE Trans. Image Process. 20(2), 433–445 (2011)

88. Lin, Z.C., Shum, H.Y.: Fundamental limits of reconstruction-based super-resolution algorithms under local translation. IEEE Trans. Pattern Anal. Mach. Intell. 26(1), 83–97 (2004)

89. Nasrollahi, K., Moeslund, T.B.: Extracting a good quality frontal face image from a low-resolution video sequence. IEEE Trans. Circuits Syst. Video Technol. 21(10), 1353–1362 (2011)

90. Bilgazyev, E., Efraty, B., Shah, S.K., Kakadiaris, I.A.: Sparse representation-based super-resolution for face recognition at a distance. In: Proc. 22nd British Machine Vision Conference (BMVC), Aug. 2011 (2011)

91. Arandjelovic, O., Cipolla, R.: A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In: Proc. IEEE 11th Int. Conf. on Computer Vision (ICCV), Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8 (2007)

92. Huang, H., He, H.T.: Super-resolution method for face recognition using nonlinear mappings on coherent features. IEEE Trans. Neural Netw. 22(1), 121–130 (2011)

93. Li, B., Chang, H., Shan, S.G., Chen, X.L., Gao, W.: Hallucinating facial images and features. In: Proc. IAPR/IEEE 19th Int. Conf. on Pattern Recognition (ICPR), Tampa, Florida, USA, Dec. 2008, pp. 1–4 (2008)

94. Tax, D., Duin, R.: Support vector domain description. Pattern Recognit. Lett. 15, 1191–1199 (1999)

95. Arandjelović, O., Zisserman, A.: Automatic face recognition for film character retrieval in feature length films. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, June 2005, vol. 1, pp. 581–588 (2005)

96. Wang, Z.F., Miao, Z.J.: Scale invariant face recognition using probabilistic similarity measure. In: Proc. IAPR/IEEE 19th Int. Conf. on Pattern Recognition (ICPR), Tampa, Florida, USA, Dec. 2008, pp. 3735–3738 (2008)

97. Zhuang, L.S., Wang, M.L., Yu, W., Yu, N.H., Qian, Y.C.: Low-resolution face recognition via sparse representation of patches. In: Proc. IEEE 5th Int. Conf. on Image and Graphics (ICIG), Xi'an, Shanxi, China, Sep. 2009, pp. 200–204 (2009)

98. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31(2), 210–227 (2009)

99. Shekhar, S., Patel, V.M., Chellappa, R.: Synthesis-based recognition of low resolution faces. In: Proc. Int. Joint Conf. on Biometrics (IJCB), Washington, DC, USA, Oct. 2011, pp. 1–6 (2011)

100. Zhou, C.T., Zhang, Z.W., Yi, D., Lei, Z., Li, S.Z.: Low-resolution face recognition via simultaneous discriminant analysis. In: Proc. Int. Joint Conf. on Biometrics (IJCB), Washington DC, USA, Oct. 2011, pp. 1–6 (2011)

101. Wurm, L.H., Legge, G.E., Isenberg, L.M., Luebker, A.: Color improves object recognition in normal and low vision. J. Exp. Psychol. Hum. Percept. Perform. 19(4), 899–911 (1993)

102. Yang, J., Liu, C.J., Zhang, L.: Color space normalization: enhancing the discriminating power of color spaces for face recognition. Pattern Recognit. 43(4), 1454–1466 (2010)

103. Vijaya Kumar, B.V.K., Savvides, M., Xie, C.Y.: Correlation pattern recognition for face recognition. Proc. IEEE 94(11), 1963–1976 (2006)

104. Yeom, S.: A linear discriminant analysis for low resolution face recognition. In: Proc. IEEE 2nd Int. Conf. on Future Generation Communication and Networking Symposia (FGCN), Sanya, Hainan, China, Dec. 2008, vol. 3, pp. 230–233 (2008)

105. Hadid, A., Pietikainen, M., Ahonen, T.: A discriminative feature space for detecting and recognizing faces. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, June 2004, vol. 2, pp. 797–804 (2004)

106. Ren, C.X., Dai, D.Q.: Piecewise regularized canonical correlation discrimination for low-resolution face recognition. In: Proc. Chinese Conf. on Pattern Recognition (CCPR), Chongqing, China, Oct. 2010, pp. 1–5 (2010)

107. Peng, Y., Zhang, D.Q., Zhang, J.C.: A new canonical correlation analysis algorithm with local discrimination. Neural Process. Lett. 31, 1–15 (2010)

108. Ben, X.Y., Jiang, M.Y., Wu, Y.J., Meng, W.X.: Gait feature coupling for low-resolution face recognition. IET Journal on Electronics Letters **48**(9) (2012)

109. Deng, Z.X., Dai, D.Q., Li, X.X.: Low-resolution face recognition via color information and regularized coupled mappings. In: Proc. Chinese Conf. on Pattern Recognition (CCPR), Chongqing, China, Oct. 2010, pp. 1–5 (2010)

110. France, S.L., Carroll, J.D.: Two-way multidimensional scaling: a review. IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. **41**(5), 644–661 (2011)

111. Yang, J.C., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)

112. Lin, D.H., Tang, X.O.: Inter-modality face recognition. In: Proc. 9th European Conf. on Computer Vision (ECCV), Graz, Austria, May 2006, vol. 4, pp. 13–26 (2006)

113. Lei, Z., Li, S.Z.: Coupled spectral regression for matching heterogeneous faces. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Miami Beach, Florida, USA, June 2009, vol. 2, pp. 1123–1128 (2009)

114. Lei, Z., Zhou, C.T., Yi, D., Jain, A.K., Li, S.Z.: An improved coupled spectral regression for heterogeneous face recognition. In: Proc. IAPR/IEEE 5th Int. Conf. on Biometrics (ICB), New Delhi, India, Mar. 2012 (2012)

115. Tome, P., Fierrez, J., Alonso-Fernandez, F., Ortega-Garcia, J.: Scenario-based score fusion for face recognition at a distance. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, June 2010, pp. 67–73 (2010)

116. Ebrahimpour, R., Sadeghnejad, N., Amiri, A., Moshtagh, A.: Low resolution face recognition using combination of diverse classifiers. In: Proc. Int. Conf. on Soft Computing and Pattern Recognition (SoCPaR), Dec. 2010, pp. 265–268 (2010)

117. Rara, H.M., Ali, A.A., Elhabian, S.Y., Starr, T.L., Farag, A.A.: Face recognition at-a-distance using texture, dense- and sparse-stereo reconstruction. In: Proc. IAPR/IEEE 20th Int. Conf. on Pattern Recognition (ICPR), Istanbul, Turkey, Aug. 2010, pp. 1221–1224 (2010)

118. Yang, A.Y., Wright, J., Ma, Y., Sastry, S.: Feature selection in face recognition: A sparse representation perspective. Tech Report UCB/EECS-2007-99, UC Berkeley (2007)

119. Martinez, A.M., Benavente, R.: The AR-face database. CVC Technical Report 24 (June 1998)

120. Yale University: The Yale Face Database. http://cvc.yale.edu/projects/yalefaces/yalefaces.html

121. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 643–660 (2001)

122. Gao, W., Cao, B., Shan, S.G., Zhou, D.L., Zhang, X.H., Zhao, D.B.: The CAS-PEAL large-scale Chinese face database and baseline evaluations. Technical Report, JDL-TR-04-FR-001, ICT-ISVISION, Chinese Academy of Sciences (May 2004)

123. Philips, P., Moon, H., Pauss, P., Rivzvi, S.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)

124. Color FERET Face Database. http://www.nist.gov/humanid/colorferet

125. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination and expression database. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1615–1618 (2003)

126. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image Vis. Comput. **28**(5), 807–813 (2010)

127. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, June 2005, vol. 1, pp. 947–954 (2005)

128. Grgic, M., Delac, K., Grgic, S.: SCface-surveillance cameras face database. Multimed. Tools Appl. **51**(3), 863–879 (2011)

129. Messer, K., Mastas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: the extended M2VTS database. In: Proc. 2nd Int. Conf. on Audio and Video-Based Biometric Person Authentication (AVBPA), Washington, DC, USA, pp. 72–77 (1999)

130. The Sheffield (previously UMIST) Face Database. http://www.sheffield.ac.uk/eee/research/iel/research/face

131. Cambridge and U.K. AT Research, Laboratories: The ORL Face Database. http://www.uk.research.att.com/facedatabase.html

132. Hwang, B.-W., Byun, H., Roh, M.-C., Lee, S.-W.: Performance evaluation of face recognition algorithms on the Asian face database, KFDB. In: Proc. 4th Int. Conf. on Audio and Video-Based Biometric Person Authentication (AVBPA), Guildford, UK, June 2003, pp. 557–565 (2003)

133. Gross, R., Shi, J.: The CMU Motion of Body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University (June 2001)

134. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Video-based face recognition using probabilistic appearance manifolds. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Madison, Wisconsin, USA, June 2003, vol. 1, pp. 313–320 (2003)

135. Stiefelhagen, R., Bernardin, K., Bowers, R., Garofolo, J., Mostefa, D., Soundararajan, P.: The CLEAR 2006 evaluation. In: Stiefelhagen, R., Garofolo, J. (eds.) CLEAR 2006. Lecture Notes in Computer Science, vol. 4122, pp. 218–229. Springer, Berlin (2006)

136. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (Oct. 2007)

137. Pinto, N., DiCarlo, J.J., Cox, D.D.: How far can you get with a modern face recognition test set using only simple features? In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Miami Beach, Florida, USA, June 2009, pp. 2591–2598 (2009)

138. Yao, Y., Abidi, B.R., Kalka, N.D., Schmid, N.A., Abidi, M.A.: Improving long range and high magnification face recognition: database acquisition, evaluation, and enhancement. Comput. Vis. Image Underst. **111**(2), 111–125 (2008)

139. Ni, J., Chellappa, R.: Evaluation of state-of-the-art algorithms for remote face recognition. In: Proc. IEEE 17th Int. Conf. on Image Processing (ICIP), Hong Kong, China, Sep. 2010, pp. 1581–1584 (2010)

140. Chang, J.-M., Kirby, M., Kley, H., Peterson, C., Draper, B., Beveridge, J.R.: Recognition of digital images of the human face at ultra low resolution via illumination spaces. In: Proc. 8th Asian Conf. on Computer Vision (ACCV), Tokyo, Japan, Nov. 2007, vol. 2, pp. 733–743 (2007)

141. Medioni, G., Choi, J., Kuo, C.-H., Fidaleo, D.: Identifying non-cooperative subjects at a distance using face images and inferred three-dimensional face models. IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. **39**(1), 12–24 (2009)

142. Hadid, A., Pietikainen, M.: From still image to video-based face recognition: an experimental analysis. In: Proc. IEEE 6th Int. Conf. on Automatic Face and Gesture Recognition (FG), Seoul, Korea, May 2004, pp. 813–818 (2004)

143. Wheeler, F.W., Liu, X.M., Tu, P.H.: Multi-frame super-resolution for face recognition. In: Proc. IEEE 1st Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS), Crystal City, VA, USA, Sep. 2007, pp. 1–6 (2007)

144. Sharma, A., Jacobs, D.W.: Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. In: Proc. IEEE

Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, USA, June 2011, pp. 593–600 (2011)

145. Razzak, M.I., Khan, M.K., Alghathbar, K., Yusof, R.: Multimodal biometric recognition based on recognition based on fusion of low resolution face and finger veins. Int. J. Innov. Comput. Inf. Control **7**(8), 4679–4689 (2011)

146. Zhou, X.L., Bhanu, B.: Integrating face and gait for human recognition at a distance in video. IEEE Trans. Syst. Man Cybern., Part B, Cybern. **37**(5), 1119–1137 (2007)

Zhifei Wang received his B.E. degree in biomedical engineering from Beijing Jiaotong University, Beijing, China, in 2006 and he is now a Ph.D. candidate in human–machine interaction at Institute of Information Science, Beijing Jiaotong University. His research interests include computer vision, face recognition, and pattern recognition.

Zhenjiang Miao (M'11) received the B.E. degree from Tsinghua University, Beijing, China, in 1987, and the M.E. and Ph.D. degrees from Northern Jiaotong University, Beijing, in 1990 and 1994, respectively. From 1995 to 1998, he was a Post-Doctoral Fellow with the École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications, Institut National Polytechnique de Toulouse, Toulouse, France, and was a Researcher with the Institute National de la Recherche Agronomique, Sophia Antipolis, France. From 1998 to 2004, he was with the Institute of Information Technology, National Research Council Canada, Nortel Networks, Ottawa, ON, Canada. He joined Beijing Jiaotong University, Beijing, in 2004. He is currently a Professor with Beijing Jiaotong University. His research interests include image and video processing, multimedia processing, and intelligent human–machine interaction.

Q.M. Jonathan Wu (M'92–SM'09) received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990. He was with the National Research Council of Canada for ten years from 1995, where he became a Senior Research Officer and a Group Leader. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He has published more than 250 peer-reviewed papers in computer vision, image processing, intelligent systems, robotics, and integrated microsystems. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks. Dr. Wu holds the Tier 1 Canada Research Chair in Automotive Sensors and Information Systems. He is an Associate Editor for the IEEE Transactions on Systems, Man, and Cybernetics Part A, and the International Journal of Robotics and Automation. He has served on technical program committees and international advisory committees for many prestigious conferences.

Yanli Wan received her B.E. degree in computer science and technology from Jinan University in 2003, her M.E. degree in computer application technology from China University of Mining and Technology in 2006, and her Ph.D. degree from Beijing Jiaotong University in 2012. Her research interests include computer vision, image processing, and pattern recognition.

Zhen Tang received his B.E. degree from China University of Petroleum in 2006 and his Ph.D. degree from Beijing Jiaotong University in 2012. His research interests include computer vision, image and video editing, and pattern recognition.