International Journal of Computer Vision manuscript No.
(will be inserted by the editor)

A Comprehensive Survey to Face Hallucination

Nannan Wang · Dacheng Tao · Xinbo Gao · Xuelong Li · Jie Li

Received: date / Accepted: date

Abstract This paper comprehensively surveys the development of face hallucination (FH), including both face super-resolution (FSR) and face sketch-photo synthesis (FSPS) techniques. Indeed, these two techniques share the same objective of inferring a target face image (e.g. high-resolution face image, face sketch and face photo) from a corresponding source input (e.g. low-resolution face image, face photo and face sketch). Considering the critical role of image interpretation in modern intelligent systems for authentication, surveillance, law enforcement, security control, and entertainment, FH has attracted growing attention in recent years. Existing FH methods can be grouped into four categories: Bayesian inference approaches, subspace learning approaches, a combination of Bayesian inference and subspace learning approaches, and sparse representation-based approaches. In spite of achieving a certain level of development, FH is limited in its success by complex application conditions such as variant illuminations, poses, or views. This paper provides a holistic understanding and deep insight into FH, and presents a comparative analysis of representative methods and promising future directions.

Keywords Face hallucination · face sketch-photo synthesis · face super-resolution · heterogeneous image transformation

1 Introduction

Face images, compared to other kinds of biometrics such as fingerprint, iris, and retina, can be acquired in a more convenient, natural, and direct way because they are collected in a non-intrusive manner (Jain et al, 2000). Consequently, a growing number of face image-based applications have been developed and investigated. These include face detection (Zhang and Zhang, 2010), alignment (Liu, 2009), tracking (Ong and Bowden, 2011), modeling (Tao et al, 2008), and recognition (Chellappa et al, 1995; Zhao et al, 2003) for security control, surveillance monitoring, authentication, biometrics, digital entertainment and rendered services for a legitimate user only, and age synthesis and estimation (Fu et al, 2010) for explosively emerging real-world applications such as forensic art, electronic customer relationship management, and cosmetology.

The intrinsic fluidity of face imaging and uncontrollable extrinsic imaging conditions (such as an intended target deliberately concealing his/her identity) means

N. Wang
VIPS Lab, School of Electronic Engineering, Xidian University, 710071, Xi'an, P. R. China
E-mail: nannanwang.xidian@gamil.com

D. Tao
Centre for Quantum Computation & Intelligent Systems, Faculty of Engineering & Information Technology, University of Technology, Sydney, 235 Jones Street, Ultimo, NSW 2007, Australia
E-mail: Dacheng.Tao@uts.edu.au

X. Gao
VIPS Lab, School of Electronic Engineering, Xidian University, 710071, Xi'an, P. R. China
E-mail: xbgao@mail.xidian.edu.cn

X. Li
Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China
E-mail: xuelong_li@opt.ac.cn

J. Li
VIPS Lab, School of Electronic Engineering, Xidian University, 710071, Xi'an, China
E-mail: leejie@mail.xidian.edu.cn

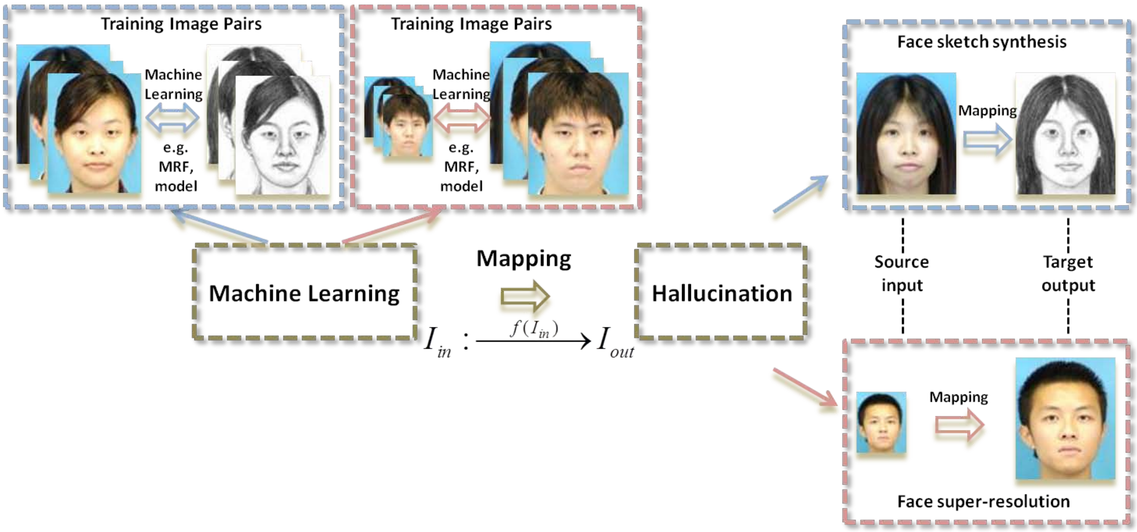

Fig. 1 Diagram of face hallucination

that suitable face images for processing and identifying a person cannot always be obtained. In cases where low-resolution face images are acquired by live surveillance cameras at a distance or face sketches are drawn by an artist, however, face hallucination (FH) techniques can be used to enhance low-resolution images and transform sketches to photos and photos to sketches for the subsequent utilizations.

It has been widely acknowledged that FH can be used to generate imagery or information from an input source face image with different modalities (resolution, style, or imaging modes) (Baker and Kanade, 2000a). In this paper, FH refers to both face super-resolution (FSR) and face sketch-photo synthesis (FSPS) because they share the similar intrinsic mathematical model; that is, they infer an image lying in an image space from its corresponding counterpart lying in another space. A brief introduction to both techniques is given below.

Low-resolution images impose hard restrictions on real world applications dealing with face recognition and high-resolution display. We intend to approximate high-resolution images from low-resolution images, using the super-resolution technique. Available super-resolution techniques can be grouped into two categories: reconstruction-based approaches and learning-based approaches. Reconstruction-based methods estimate a high-resolution image from a sequence of blurred and down-sampled low-resolution images (Elad and Feuer, 1997, 1999; Hardie et al, 1997) and there are inherent limitations in relation to increasing the magnification factor (Baker and Kanade, 2002). In recent years, learning-based approaches have been proposed and obtained competitive results for various low-level vision tasks (Fan and Yeung, 2007; Freeman and Pasztor, 1999; Freeman et al, 2000, 2002), including image hallucination (Sun et al, 2003; Xiong et al, 2009), image analogy (Hertzmann et al, 2001), image stitching (Brown and Lowe, 2007), cartoon character synthesis (Yu et al, 2012b,a), and texture synthesis (Efros and Leung, 1999; Efros and Freeman, 2001; Zalesny et al, 2005). Learning-based methods explore mapping relations between high- and low-resolution image pairs to infer high-resolution images from their low-resolution counterparts. Compared to reconstruction-based methods, learning-based methods achieve higher magnification factors and better visual quality, especially for single-image super-resolution (Lin et al, 2007, 2008). This is also the main reason underlying the proposal of FSR algorithms. The application scenario therefore needs to be constrained in such a way that more specific prior knowledge, e.g. human skin color, the strong structure of faces, and gender information, can be further exploited to improve the estimation.

In searching for criminal suspects, a photo of a criminal suspect is not always available and thus the best substitute may be a sketch drawn by an artist with the aid of eyewitnesses. However, because of the great difference between face sketches and face photos in both geometry and texture, using direct face recognition to identify a criminal suspect performs poorly when a sketch is compared with an existing photo gallery (Gao et al, 2008b; Tang and Wang, 2004). To reduce the visual difference between sketches and photos, sketches and photos can be transformed to the same modality. There are two ways to accomplish this: transformation of the sketches to photos, or transformation of the photos to sketches (FSPS for short). Note that an FSPS algorithm is not constrained to face recognition but can also be

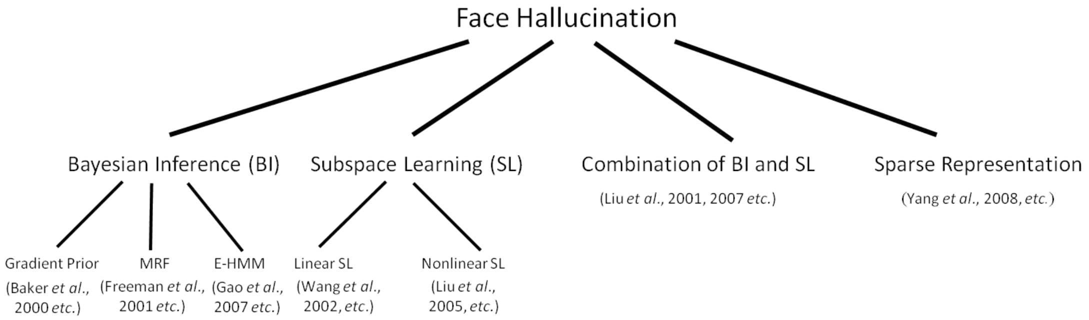

Fig. 2 Tree diagram for different categories of face hallucination algorithms

Table 1 Notations

Symbols	Descriptions
$\boldsymbol{I}_{in}$	Input of FH, used when FSR is not distinguished from FSPS
$\boldsymbol{I}_{out}$	Output of FH corresponding to $\boldsymbol{I}_{in}$
$\boldsymbol{I}_H$	High-resolution image for FSR
$\boldsymbol{I}_L$	Low-resolution image corresponding to $\boldsymbol{I}_H$
$\boldsymbol{I}_H^g$	Global high-resolution face image used in some methods for FSR
$\boldsymbol{I}_H^l$	Local high-resolution face image corresponding to $\boldsymbol{I}_H^g$ used for FSR
$\boldsymbol{I}_S$	Face sketch for FSPS
$\boldsymbol{I}_P$	Face photo for FSPS
$\boldsymbol{x}, \boldsymbol{x}_i$	sketch patch or high-resolution image patch, or observation feature of a pixel on a sketch
$\boldsymbol{y}, \boldsymbol{y}_i$	photo patch or low-resolution image patch, or observation feature of a pixel on a photo
K	Number of nearest neighbors

applied to digital entertainment (Iwashita et al, 1999; Koshimizu and Tominaga, 1999; Wang and Tang, 2009; Yu et al, 2012b).

Both learning-based FSR and FSPS generate a target image from a corresponding input source image by using training image pairs (e.g. low- and high-resolution image pairs and sketch-photo pairs) based on various machine learning algorithms. In the learning stage, learning-based FSR and FSPS learn the underlying relation between training image pairs and in the inference stage, the output target image corresponding to the input source image is predicted via the learned mapping relations. Fig. 1 illustrates the framework for FSPS and FSR, from which we can see that the main difference between the two techniques is that the transformation between sketches and photos (FSPS) is invertible while this reversibility is not required in FSR. The mapping obtained in the learning stage is similar for these two different applications. It is symmetric for sketch synthesis and photo synthesis, and a synthesis process can be completed by switching the roles of sketches and photos of another synthesis process. Thus, an FSPS model can be constructed from a learning-based FSR model by adjusting the training image pairs and features used as the input to the model. In this paper, we therefore prefer not to differentiate between FSR and FSPS algorithms when they are categorized, except as noted.

As shown in Fig. 1, the mapping learned from the training image pairs using a machine learning algorithm is critical to the FH algorithm. This mapping may be explicit, such as a function mapping from input to output, or implicit, in which it is hidden in the model and relies on various approaches to construct the output model. Based on the approaches applied to the model construction, FH methods can be divided into four categories: Bayesian inference framework, subspace learning framework, combination of Bayesian inference and subspace learning methods, and sparse representation methods. FH techniques in each of these four categories may be further classified in a much more detailed manner. Fig. 2 shows these different classes of FH algorithms in a tree diagram.

Table 1 summarizes frequently used notations in this paper. The rest of this paper is organized as follows. Methods under Bayesian inference framework are described and comprehensively analyzed in Section 2, and a description of the subspace learning-based methods follows in Section 3. A compound framework that combines Bayesian inference and subspace learning-based methods is presented in Section 4. Section 5 discusses several methods for FH in the realm of sparse representation. A comparative analysis of these four categories and their performance are given in Section 6. Finally, insights on recent trends and promising future directions in this field are given in Section 7, and concluding remarks are made in Section 8.

2 The Bayesian Inference Framework

Bayesian inference exploits evidence to update the state of the uncertainty over competing probability models. Bayes' theorem is critically important in Bayesian inference, and is written as $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, where A and B represent two events in the event space (Gelman et al, 2003). Given that $\boldsymbol{I}_{in}$ and $\boldsymbol{I}_{out}$ denote the input (observation) and output image (to be estimated) for FH, respectively, the maximum a posteriori (MAP) decision rule in Bayesian statistics for FH is written as

$$\begin{aligned} \boldsymbol{I}^*_{out} &= \underset{\boldsymbol{I}_{out}}{\operatorname{argmax}}\, P(\boldsymbol{I}_{out}|\boldsymbol{I}_{in}) \\ &= \underset{\boldsymbol{I}_{out}}{\operatorname{argmax}}\, P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out})P(\boldsymbol{I}_{out}). \end{aligned} \tag{1}$$

Since $\boldsymbol{I}_{in}$ is an observation, $P(\boldsymbol{I}_{in})$ is a constant and it can be ignored in Eq. (1). In the above equation, $P(\boldsymbol{I}_{out})$ is known as the prior, which is learned from training images pairs, and $P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out})$ is the likelihood and can also be taken as a Gaussian form under the assumption that each pixel on $\boldsymbol{I}_{in}$ is identically treated. For different methods under this framework, $P(\boldsymbol{I}_{in})$ and $P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out})$ take different concrete forms which are discussed below. Fig. 3 shows a diagram of the Bayesian inference framework using face sketch synthesis as an example, and the following figures show the face sketch synthesis process only except for special explanation. In Fig. 3, the partition mask is applied to divide images into patches. Holistic methods such as Tang and Wang (2002, 2003, 2004) synthesized a sketch as a whole, thus the partition mask might degenerate to an identical transformation which actually preserve a whole image as itself.

2.1 Gradient-based Prior for Data Modeling

Baker and Kanade (2000a) proposed the first FH algorithm. By treating FSR as predicting the lowest level of the Gaussian Pyramid (Burt, 1981; Burt and Adelson, 1983), this method processes in a pixel-wise manner and aims to improve the face recognition performance.

The likelihood term $P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out})$ in Eq. (1) is given by:

$$\begin{aligned} &P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out}) \propto exp\Big\{ \\ &-\frac{1}{2\sigma^2}\sum_{m,n}\big[\boldsymbol{G}_k(m,n) - \sum_{p,q}\boldsymbol{W}(m,n,p,q)\boldsymbol{G}_0(p,q)\big]^2\Big\}, \end{aligned} \tag{2}$$

where $\boldsymbol{G}_k, k = 0, 1, \cdots, N$ is the k-th level Gaussian pyramid and the level 0 pyramid is the high-resolution image. The subscripts to the sum index the corresponding pixel on a specific Gaussian pyramid. The weight $\boldsymbol{W}(\cdot)$ is a function of down-sample factor which measures the number of overlapped low-resolution pixels and high-resolution pixels. σ^2 is the variance. The likelihood mainly considers the fidelity between the low-resolution image and the down-sample version of the high-resolution image to be predicted.

The prior $P(\boldsymbol{I}_{out})$ in Eq. (1) is learned from the spatial distribution of image gradient vectors. The gradient vector is given by the concatenation of Laplacian pyramids, the horizontal and vertical first- and second-order derivatives of Gaussian pyramids. The predicted gradient vector of the high-resolution image corresponding to the low-resolution input is copied from the gradient vector of a training high-resolution image. This high-resolution image is identified by searching the nearest gradient vector of the input low-resolution image through all training low-resolution images. Then the prior is modeled by the errors between the gradients of the target high-resolution image and the above predicted gradients. These errors are assumed to be i.i.d. and the prior $P(\boldsymbol{I}_{out})$ can be modeled by a Gaussian distribution with variance σ^2_∇,

$$\begin{aligned} &P(\boldsymbol{I}_{out}) \propto exp\Big\{ \\ &-\frac{1}{2\sigma^2_\nabla}\sum_{m,n}\big[\boldsymbol{H}_0(\boldsymbol{G}_0)(m,n) - \overline{\boldsymbol{H}}_0(m,n)\big]^2 \\ &-\frac{1}{2\sigma^2_\nabla}\sum_{m,n}\big[\boldsymbol{V}_0(\boldsymbol{G}_0(m,n)) - \overline{\boldsymbol{V}}_0(m,n)\big]^2\Big\}, \end{aligned} \tag{3}$$

where $\boldsymbol{H}_0(\cdot)$ and $\boldsymbol{V}_0(\cdot)$ denote the actual horizontal and vertical first order derivative of the Gaussian pyramids, and $\overline{\boldsymbol{H}}_0$ and $\overline{\boldsymbol{V}}_0$ are the corresponding predicted derivatives, respectively.

Finally, the target high-resolution image is resolved from the objective function, a combination of the likelihood model $P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out})$ and the gradient prior model $P(\boldsymbol{I}_{out})$, by the gradient descent method. The authors reported that this algorithm enhanced face images by a factor of 8 (e.g. from 12 × 16 to 96 × 128). This method was further investigated in their subsequent

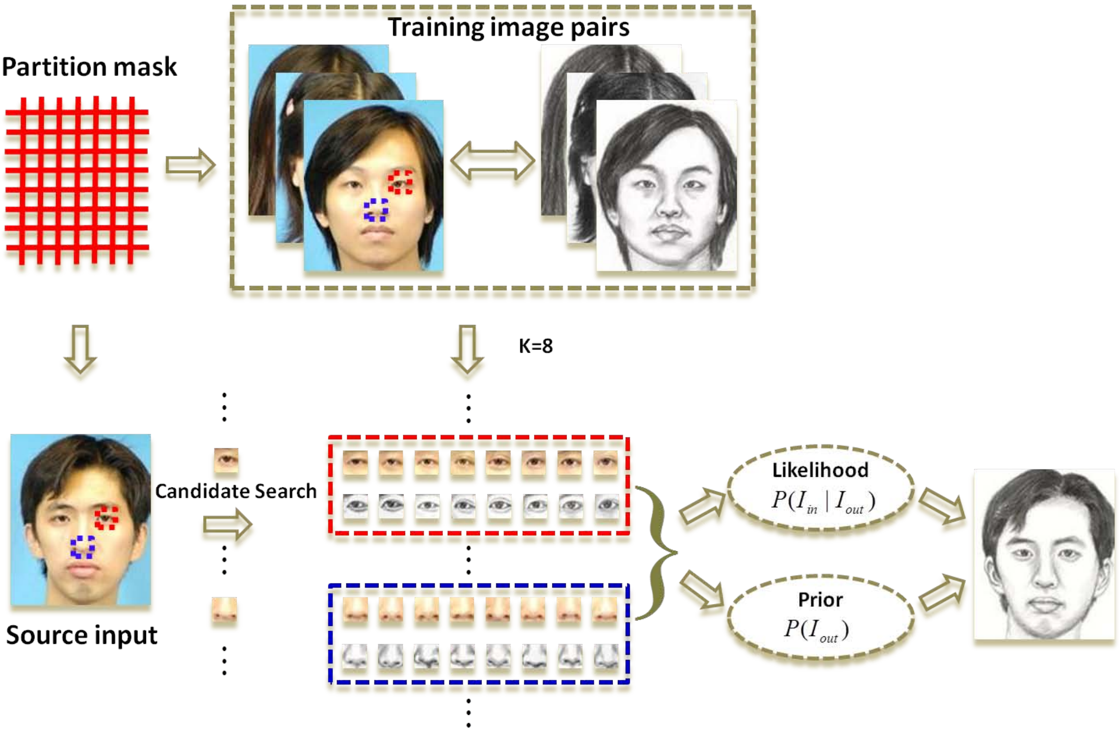

Fig. 3 Bayesian inference framework for sketch synthesis

works (Baker and Kanade, 2000b), (Baker and Kanade, 2002) that demonstrated useful information provided by the reconstruction constraints (i.e. the prior information) reduces with the increase of the magnification factor.

Inspired by Baker and Kanade (2000a), the gradient-based prior was carefully explored. Dedeoglu et al (2004) explored a similar idea for video hallucination and demonstrated the resolution of a human face video by a factor of 16 from 8 × 6 to 128 × 96 . Since these methods search for the nearest neighbor pixel by pixel, they are time-consuming; furthermore, the pixel-based strategy is susceptible to noise. Unlike the gradient feature extracted by Baker and Kanade (2000a), Su et al (2005) proposed the exploitation of a steerable pyramid to model the prior generated by oriented steerable filters to extract multi-orientation and multi-scale information of local low-level face features. With regard to the feature of each pixel of the source input, its nearest neighbor was chosen in a different way from the strategy in (Baker and Kanade, 2000a,b; Dedeoglu et al, 2004). Baker and Kanade (2000a) searched the nearest neighbor of an input pixel from the feature of pixels in the same location on the training images. Su et al (2005) found its nearest neighbor from the feature of pixels around the location on the training images, which alleviates the requirements for exact face alignment. However, this method is still subject to high computation cost due to the high dimension of extracted features. Their experimental results showed that they could enhance a 24 × 32 face image into its 96 × 128 counterpart.

2.2 Markov Random Fields-based Method

Markov random fields (MRF) (Li, 2010) characterize the dependency relationship between neighboring pixels or features. The principal consideration is given by

$$P(\boldsymbol{f}_i|\boldsymbol{f}_1, \cdots, \boldsymbol{f}_N) = P(\boldsymbol{f}_i|\mathbf{N}(i)) \tag{4}$$

where $\boldsymbol{f}_i, i = 1, \cdots, N$ is the i-th feature and $\mathbf{N}(i)$ denotes the neighborhood. An image can be modeled by MRF; for example, given that the intensity of each grid on images is the variable, then the probability of an image intensity configuration is usually the product of a data constraint term and a smooth constraint term. The data constraint term models the fidelity between the observation and the target output and the smooth constraint models the local neighborhood relationship of the target output.

Freeman et al. (1999; 2000; 2002) proposed an example-based learning framework for the low-level vision problem and took super-resolution as one of its applications. In their seminal works, images (low-resolution images) and scenes (high-resolution images) were modeled by Markov Random Fields. Both a scene $\boldsymbol{I}_{out}$

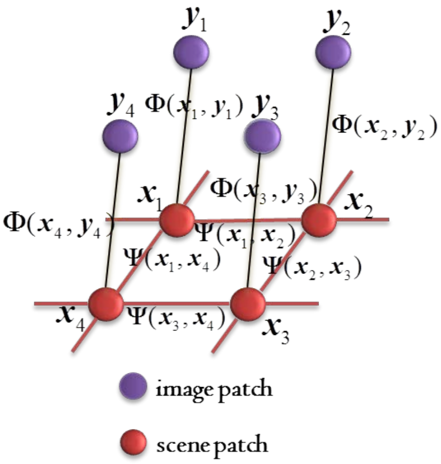

Fig. 4 Illustration of the Markov network utilized in Freeman and Pasztor (1999); Freeman et al (2000).

and its corresponding image $\boldsymbol{I}_{in}$ are first divided into patches $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ and $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\}$, respectively. Each of these patches is represented as a node in the Markov network as shown in Fig. 4. For any input image patch, K nearest neighbors are searched from the training image patches to construct the compatibility matrix $\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})$ between image and scene nodes. Simultaneously, K target scene candidate patches are collected from the training scene patches corresponding to the selected training image patches. Then the neighborhood relationship (smooth constraint) is constructed from the compatibility matrix $\boldsymbol{\Psi}(\boldsymbol{x}, \boldsymbol{y})$ between neighboring scene nodes. The joint probability over a scene $\boldsymbol{I}_{out}$ and its corresponding image $\boldsymbol{I}_{in}$ can be written as

$$\begin{aligned} P(\boldsymbol{I}_{in}, \boldsymbol{I}_{out}) &= P(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_N) \\ &\propto \prod_{(i,j)} \boldsymbol{\Psi}(\boldsymbol{x}_i, \boldsymbol{x}_j) \prod_k \boldsymbol{\Phi}(\boldsymbol{x}_k, \boldsymbol{y}_k), \end{aligned} \tag{5}$$

where (i, j) indexs a pair of neighboring scene nodes i and j. The compatibility functions $\boldsymbol{\Psi}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{\Phi}(\boldsymbol{x}_k, \boldsymbol{y}_k)$ are defined by

$$\begin{aligned} \boldsymbol{\Psi}(\boldsymbol{x}_i^l, \boldsymbol{x}_j^m) &= exp^{-\|\boldsymbol{d}_{ji}^l - \boldsymbol{d}_{ij}^m\|^2 / 2\sigma_s^2}, \\ \boldsymbol{\Phi}(\boldsymbol{x}_k^l, \boldsymbol{y}_k) &= exp^{-\|\boldsymbol{y}_k^l - \boldsymbol{y}_k\|^2 / 2\sigma_p^2}, \end{aligned} \tag{6}$$

where $\mathbf{d}_{ji}^l (l = 1, \cdots, K)$ is a vector of the pixel intensities of the l-th possible candidate for the scene patch $\boldsymbol{x}_i$ and lies in the overlap region with the patch $\boldsymbol{x}_j$. $\boldsymbol{y}_k^l (l = 1, \cdots, K)$ is the l-th nearest neighbor of the image patch $\boldsymbol{y}_k$. σ_s and σ_p are two predefined parameters. Eq. (1) and Eq. (5) indicate that maximizing a posterior is equivalent to maximizing the joint probability $P(\boldsymbol{I}_{out}, \boldsymbol{I}_{in})$ and then we have

$$\begin{aligned} &P(\boldsymbol{I}_{in} | \boldsymbol{I}_{out}) \propto \prod_k \boldsymbol{\Phi}(\boldsymbol{x}_k, \boldsymbol{y}_k), \\ &P(\boldsymbol{I}_{out}) \propto \prod_{(i,j)} \boldsymbol{\Psi}(\boldsymbol{x}_i, \boldsymbol{x}_j). \end{aligned} \tag{7}$$

Bayesian belief propagation (Pear, 1988; Yedidia et al, 2001) is used to find a local maximum of the posterior probability for the target scene node. The integrated scene is obtained by merging these patches with an average of the overlapping regions. In (Bishop et al, 2003), the model was applied to video sequences, but introduces severe video artifacts. To reduce the number of artifacts and to obtain coherent resultant videos, an adhoc solution that re-uses the high-resolution solutions is adopted.

Inspired by the promising results obtained by the patch-based nonparametric sampling used in texture synthesis (Bonet, 1997; Chen et al, 2001; Efros and Leung, 1999; Efros and Freeman, 2001; Liang et al, 2001), Liu et al. (2001; 2007a) proposed a nonparametric MRF-based FSR method. This two-step global and local modeling framework assumes that a high-resolution face image is naturally a composition of two parts- a global face image corresponding to the low frequency and a local face image corresponding to middle and high frequencies,

$$\boldsymbol{I}_{out} = \boldsymbol{I}_H = \boldsymbol{I}_H^l + \boldsymbol{I}_H^g. \tag{8}$$

Under this assumption, the objective function (1) can be rewritten as

$$\boldsymbol{I}_{out}^* = \boldsymbol{I}_H^* = \underset{\boldsymbol{I}_H^g, \boldsymbol{I}_H^l}{\operatorname{argmax}} P(\boldsymbol{I}_L | \boldsymbol{I}_H^g + \boldsymbol{I}_H^l) P(\boldsymbol{I}_H^l | \boldsymbol{I}_H^g) P(\boldsymbol{I}_H^g). \tag{9}$$

Since $\boldsymbol{I}_L$ mainly consists of the low-frequency part of $\boldsymbol{I}_H$, then $P(\boldsymbol{I}_L | \boldsymbol{I}_H^g + \boldsymbol{I}_H^l) = P(\boldsymbol{I}_L | \boldsymbol{I}_H^g)$. The likelihood $P(\boldsymbol{I}_{in} | \boldsymbol{I}_{out})$ and the prior $P(\boldsymbol{I}_{out})$ are

$$\begin{aligned} &P(\boldsymbol{I}_{in} | \boldsymbol{I}_{out}) = P(\boldsymbol{I}_L | \boldsymbol{I}_H^g), \\ &P(\boldsymbol{I}_{out}) = P(\boldsymbol{I}_H^l | \boldsymbol{I}_H^g) P(\boldsymbol{I}_H^g). \end{aligned} \tag{10}$$

In contrast to the aforementioned methods, Liu et al. (2001; 2007a) did not model the likelihood and prior, respectively. They constructed a global model for the terms $P(\boldsymbol{I}_L | \boldsymbol{I}_H^g) P(\boldsymbol{I}_H^g)$ by using PCA. Given the global face image $\boldsymbol{I}_H^g$, a patch-based nonparametric Markov network similar to the MRF model in Freeman et al. (1999; 2000; 2002) is built to model the residual local face image $\boldsymbol{I}_H^l$ (i.e. the residual term $P(\boldsymbol{I}_H^l | \boldsymbol{I}_H^g)$). By combining $\boldsymbol{I}_H^g$ and $\boldsymbol{I}_H^l$, the target high-resolution image can be obtained. Related works include (Fan and

Yeung, 2007; Hsu et al, 2009; Jia and Gong, 2006, 2008; Kumar and Aravind, 2008b; Liang et al, 2010; Liu et al, 2005b,c,d, 2007b; Tanveer and Iqbal, 2010; Wang et al, 2011; Zhang et al, 2008, 2011a; Zhuang et al, 2007).

The aforementioned methods construct the same pairwise edge-based compatibility functions for all patches on a face image. In contrast to this, Stephenson and Chen (2006) proposed a method that structured several different pairwise compatibility functions, in which patches lying on the same region or the same group shared the same compatibility function. This method improves the probability of incorporating more relevant information between a query image patch and the selected nearest neighbors. Subsequently, similar approximation procedures (Freeman et al, 2000) are applied to the estimation of the target high-resolution image.

Considering the strong structural property of face images, the uniform scale of MRF has limited ability to address the long range dependency among local patches; thus, Wang and Tang proposed a multi-scale MRF model for FSPS (Wang and Tang, 2009). Their method constructs the pairwise compatibility functions through the nearest neighbors searched from a training set across different scales. Under the MAP rule, the best matched neighbor patch is then taken as the target patch corresponding to the input image patch. This method uses the image quilting (Efros and Freeman, 2001) technique to stitch the overlapping areas, which reduces both the blurring effect due to the strategy of averaging the overlapping areas and the blocking artifacts because of the incompatible nearest neighbor patches. The authors also performed subspace face recognition (Wang and Tang, 2006) by using synthesized sketches and photos. They extended this work to lighting and pose robust FSPS (Zhang et al, 2010) by taking photo-to-photo patch matching, photo-to-sketch patch matching, shape priors, intensity compatibility, and gradient compatibility into account. The experimental results show that their proposed method achieved a better visual effect than the results reported in (Wang and Tang, 2009).

Zhou et al (2012) claimed that above MRF-based sketch-photo synthesis method (Wang and Tang, 2009) had two major drawbacks: cannot synthesize new sketch patches (i.e. each patch of final target output is from the training set) and NP-hard for the optimization problem in solving the MRF model. Then they proposed a weighted Markov random fields method (Zhou et al, 2012) to model the relation between sketch and photo patches. By a linear combination of selected K candidate sketch patches, their method could synthesize new sketch patches existing not in the training sketch set. Furthermore, the objective function is a convex optimization problem which has the unique optimal solution. Experimental results illustrated they indeed improved some deformation yet not as clear as that generated by (Wang and Tang, 2009).

Aforementioned methods are based on inductive learning which may result in high loss for test samples. This is because inductive learning minimizes the empirical loss for training examples. Wang et al (2013b) proposed a transductive face sketch-photo synthesis method that took the given test samples into the learning process to minimize the loss on these test samples. The generative process of both photos and sketches could be modeled by Bayesian inference. The relation between sketch and photo patches are modeled by a graphical model similarly as weighted Markov random fields method (Zhou et al, 2012). Experimental results illustrate this method achieves state-of-the-art performance both from subjective (synthesized examples) and objective (face recognition accuracy) manner.

2.3 Embedded Hidden Markov Model

Hidden Markov models (HMM) track the time-varying stochastic process through probability statistics, and have been widely applied to acoustic speech signal processing (Rabiner, 1989). Samaria (1994) first constructed a one-dimensional HMM on a face partitioned into five regions (hair, forehead, eye, nose, and mouth), each region corresponding to a hidden state. The intensities of each region are taken as the observation. Three basic problems in HMM-based methods illustrate the backbone of this class of methods: (1) how can one efficiently compute the probability $P(\boldsymbol{O}|\boldsymbol{\lambda})$ of the observation sequence $\boldsymbol{O} = (\boldsymbol{o}_1, \cdots, \boldsymbol{o}_T)$ (T denotes the number of observations) given the HMM model $\boldsymbol{\lambda}$ (model parameters); (2) how does one choose a corresponding state sequence $\boldsymbol{Q} = (q_1, \cdots, q_T)$ that is optimal in some meaningful sense (e.g. $\max_{\boldsymbol{I}_H^g, \boldsymbol{I}_H^l}$); and (3) how can the model $\boldsymbol{\lambda}$ be adjusted to maximize the probability in problem (1) $P(\boldsymbol{O}|\boldsymbol{\lambda})$. These three problems can be solved with the help of the backward-forward algorithm, the Viterbi decoding algorithm, and the Baum-Welch algorithm, respectively. A detailed discussion and analysis of these three problems and HMM can be found in Rabiner (1989). Owing to the fact that a face image contains two-dimensional spatial information, conventional HMM is challenged by two problems: the loss of some spatial information and high computation cost. Later, the use of embedded hidden Markov models (E-HMM) (Nefian and Hayes, 1999) was proposed to model the face image at a moderate computation cost. Gao et al. (2008b; 2008c; 2009; 2010; 2007) employed E-HMM to

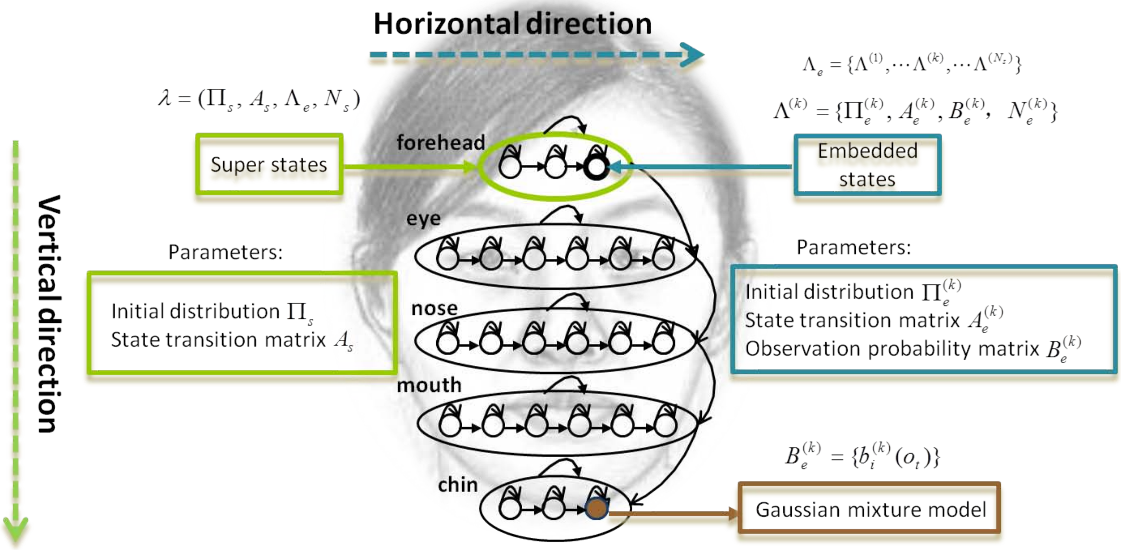
Horizontal direction
$\lambda = (\Pi_s, A_s, \Lambda_e, N_s)$
$\Lambda_e = \{\Lambda^{(1)}, \cdots \Lambda^{(k)}, \cdots \Lambda^{(N_s)}\}$
$\Lambda^{(k)} = \{\Pi_e^{(k)}, A_e^{(k)}, B_e^{(k)}, \ N_e^{(k)}\}$
Vertical direction
Super states
forehead
Embedded states
eye
Parameters:
Initial distribution Π_s
State transition matrix A_s
nose
Parameters:
Initial distribution $\Pi_e^{(k)}$
State transition matrix $A_e^{(k)}$
Observation probability matrix $B_e^{(k)}$
mouth
$B_e^{(k)} = \{b_i^{(k)}(o_t)\}$
chin
Gaussian mixture model

Fig. 5 E-HMM structure and parameters for face image.

learn the nonlinear relationship between sketches and their counterpart photo images.

Before discussing these methods, the construction of the E-HMM for a holistic face image should be introduced. In this model, E-HMM consists of $N_s = 5$ super-states (corresponding to 5 different sections: forehead, eye, nose, mouth, and chin) that model the face information in the vertical direction. Each super-state can be decomposed into embedded-states that describe the face information from the horizontal direction. Each super-state and its embedded-states can be regarded as a one dimensional HMM, where each observation (each pixel has an observation (vector)) in an image corresponding to one hidden state, i.e., embedded state. The following parameters support the E-HMM model: initial super-state distribution $\boldsymbol{\Pi}_s$, super-state probability transition matrix $\boldsymbol{A}_s$, initial embedded state distribution $\boldsymbol{\Pi}_s^{(k)}$, and embedded-state probability transition matrix $\boldsymbol{A}_e^{(k)}$. In addition, the distribution $\boldsymbol{b}_i^{(k)}(\boldsymbol{o}_t)$ of each observation $\boldsymbol{o}_t$ (t indexes the pixel) under the hidden embedded-state s_i^k (super-state and embedded-state are indexed by k and i respectively) is represented by a Gaussian mixture density function parameterized by mixture weights, mean vector, and covariance. The observation vector of each pixel is the concatenation of five vectors extracted from the image by five operators: grayscale value-extracting operator, Gaussian operator, Laplace operator, horizontal and vertical derivative operator (see Fig. 5).

Gao et al. (2008b; 2007) generated sketches from an input test photo by using E-HMM. Fig. 6 shows the generation of the sketch-photo pairs from groups of hidden variables. In comparison to the model defined by (1), this method does not model the likelihood $P(\boldsymbol{I}_{in}|\boldsymbol{I}_{out})$ and the prior $P(\boldsymbol{I}_{out})$ directly. Instead, the hidden variables $\boldsymbol{z} = \{z_1, \cdots, z_N\}$ are taken into account

$$\begin{aligned} \boldsymbol{I}_{out}^* &= \operatorname*{argmax}_{\boldsymbol{I}_{out}, \boldsymbol{z}} P(\boldsymbol{I}_{out}, \boldsymbol{z}|\boldsymbol{I}_{in}) \\ &= \operatorname*{argmax}_{\boldsymbol{I}_{out}, \boldsymbol{z}} P(\boldsymbol{I}_{out}, \boldsymbol{z}, \boldsymbol{I}_{in}) \\ &= \operatorname*{argmax}_{\boldsymbol{I}_{out}, \boldsymbol{z}} P(\boldsymbol{I}_{in}, \boldsymbol{z}) P(\boldsymbol{I}_{out}|\boldsymbol{I}_{in}, \boldsymbol{z}) \\ &= \operatorname*{argmax}_{\boldsymbol{I}_{out}, \boldsymbol{z}} P(\boldsymbol{I}_{in}, \boldsymbol{z}) P(\boldsymbol{I}_{out}|\boldsymbol{z}). \end{aligned} \tag{11}$$

To obtain $\boldsymbol{I}_{out}^*$, a coupled E-HMM model $\boldsymbol{\lambda}$ is jointly trained to maximize the likelihood $P(\boldsymbol{O}|\boldsymbol{\lambda})$ by using the Baum-Welch algorithm under the assumption that a sketch and the counterpart photo share the same supper-state and embedded-state transition probability matrix. Here the observation sequence $\boldsymbol{O}$ is the concatenation of features extracted from one sketch-photo pair $(\boldsymbol{I}_P, \boldsymbol{I}_S)$ by the aforementioned five operators. Afterward, two sub-E-HMM models $\boldsymbol{\lambda}_P$ and $\boldsymbol{\lambda}_S$ are obtained by uncoupling the E-HMM model as $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_P; \boldsymbol{\lambda}_S]$. In the synthesis stage, K E-HMM models are selected with respect to the similarity between the source input photo and the trianing photos. The similarity is measured by $P(\boldsymbol{O}_{in}|\boldsymbol{\lambda}_P)$ calculated by the forward-backward algorithm, where $\boldsymbol{O}_{in}$ denotes the observation sequence extracted from the input photo $\boldsymbol{I}_{in}$ and $\boldsymbol{\lambda}_P$ is the E-HMM model of a training photo image. With regard to each selected $\boldsymbol{\lambda}_{P_i} (i = 1, \cdots, K)$, the probability $P(\boldsymbol{I}_{in}, \boldsymbol{z})$ under this model can be further represented as $P(\boldsymbol{O}_{in}, \boldsymbol{z}|\boldsymbol{\lambda}_{P_i})$. Gao et al (2008b) solved the above

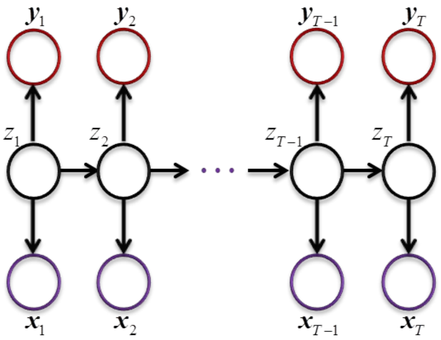

Fig. 6 Graphical illustration of the model in (Gao et al, 2008b; Zhong et al, 2007). Here $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T$ and $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T$ denote the observations extracted from a sketch-photo pair respectively, i.e. $\boldsymbol{o}_i = [\boldsymbol{x}_i; \boldsymbol{y}_i]$, $i = 1, \cdots, T$. $z_1, \cdots, z_N$ are hidden variables.

Eq. (11) in three steps. First, the optimal state sequence $\boldsymbol{z}$ is decoded from the observation sequence $\boldsymbol{O}_{in}$ by $\boldsymbol{\lambda}_{P_i}$ exploiting the Viterbi algorithm

$$\boldsymbol{z}^* = \underset{\boldsymbol{z}}{\operatorname{argmax}}\, P(\boldsymbol{O}_{in}, \boldsymbol{z}|\boldsymbol{\lambda}_{P_i}). \tag{12}$$

Then, the observation sequence $\boldsymbol{O}_{out}$ corresponding to the target sketch is then reconstructed according to the computed optimal state sequence $\boldsymbol{z}^*$ under the E-HMM $\boldsymbol{\lambda}_{S_i}$

$$\boldsymbol{O}^*_{out} = \underset{\boldsymbol{O}_{out}}{\operatorname{argmax}}\, P(\boldsymbol{O}_{out}|\boldsymbol{z}^*, \boldsymbol{\lambda}_{S_i}) \tag{13}$$

The above optimization problem can be resolved by assigning the mode of a special Gaussian component of the Gaussian mixture model, where the index of the special component is determined by the corresponding state value in the optimal state sequence $\boldsymbol{z}^*$. Subsequently a sketch can be rearranged from grayscale values extracted from the observation sequence $\boldsymbol{O}^*_{out}$. Finally, the target sketch is synthesized by a linear combination of these K sketches weighted by the sum-normalized similarity measure $P(\boldsymbol{O}_{in}|\boldsymbol{\lambda}_P)$.

Since the method in (Gao et al, 2008b; Zhong et al, 2007) was conducted on a holistic face image, certain fine local features such as those associated with eyes, nose, and mouth, could not be learned. Furthermore, some noise exists in the synthesized sketches. To overcome these defects, Gao et al (2008c) extended the aforementioned method to local patch-based sketch synthesis in a subsequent work. Here, all the images were divided into even patches with some overlap. For each patch in the source input image, the corresponding target image patch is synthesized using the approach introduced above (Gao et al, 2008b; Zhong et al, 2007). The approach in (Xiao et al, 2009) extended the algorithm (Gao et al, 2008c) to face photo synthesis, employing a similar idea to Gao et al (2008c). Several of the above E-HMM-based methods average the overlapping regions which may result in blurring effect. In consideration of this, the image quilting technique (Efros and Freeman, 2001) was exploited to stitch the neighbor patches both for sketch synthesis and photo synthesis in Xiao et al (2010).

2.4 Discussion

Gradient-based prior for data modeling-based methods find only neighbors related to pixels in the same location, which may result in low compatibility between neighboring patches and sensitive to small misalignment of face images. MRF-based methods compensate for this shortcoming by defining two compatibility functions between the low-resolution patch (or sketch/photo patch) and the corresponding high-resolution patch (photo/sketch patch), and among high-resolution neighboring patches (photo/sketch patches), respectively. However, MRF-based methods always adopt the MAP criterion to select the most appropriate neighboring patch to hallucinate the target patch. This requires that there are sufficient examples in the training dataset to contain every possible patch state; otherwise the MAP strategy may lead to deformation as a result of its neighbor selection limitation. E-HMM-based methods enforce compatibility between neighboring states by a transition probability matrix from one state to other neighboring states. From the analyses carried out in subsections 2.1 to 2.3, we found that all three subcategory methods share the same drawbacks of high computation cost and heavy memory load. Gradient-based prior for data modeling-based methods are subject to this defect because of pixel-based feature extraction and computation. MRF-based methods may avoid this curse by adopting the neighbor search strategy in (Wang and Tang, 2009). E-HMM-based methods tolerate this shortcoming as a result of both the pixel-based feature extraction strategy and iterative Viterbi decoding estimation.

3 The Subspace Learning Framework

Subspace learning refers to the technique of finding a subspace $\Re^m$ embedded in a high dimensional space $\Re^n (n > m)$. Linear subspace learning (e.g. principal component analysis, locality preserving projection (He, 2005)) is mainly achieved by a projection matrix $\boldsymbol{U} \in \Re^{n \times m}$, which is learned from training examples. The matrix $\boldsymbol{U}$ can always be calculated by solving a standard eigenvalue decomposition problem (Zhang et al,

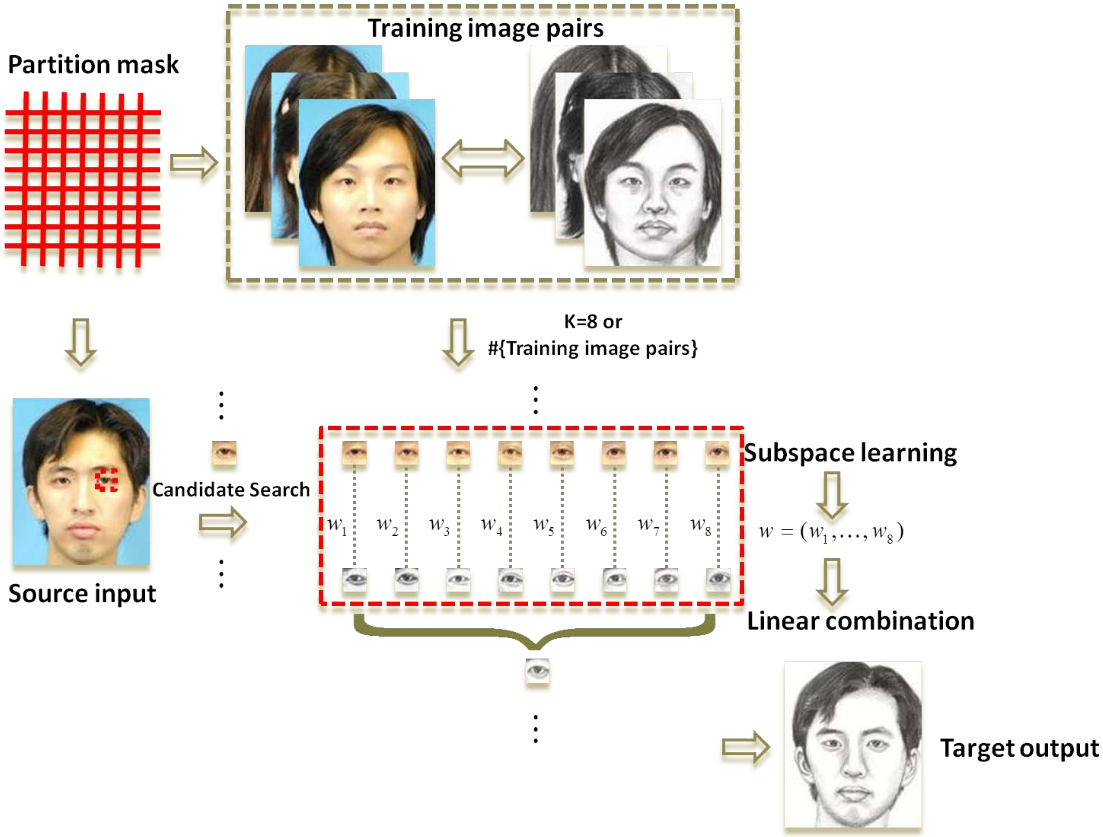

Fig. 7 Diagram of the subspace learning framework.

2009) or generalized eigenvalue decomposition problem (He, 2005),

$$\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{B}\boldsymbol{u}_i \tag{14}$$

where $\boldsymbol{A}$ and where $\boldsymbol{B}$ denote various meanings for different subspace learning methods, $\boldsymbol{u}_i$ is the eigenvector corresponding to eigenvalue λ_i, and $\boldsymbol{U}$ is composed of columns of $\boldsymbol{u}_i$. Given an input image or image feature $\boldsymbol{f} \in \Re^n$, we can find its projection on subspace $\Re^m$ from $\boldsymbol{f}_proc = \boldsymbol{U}^T\boldsymbol{f}$. In addition to above vector- and matrix-based subspace learning, it could be similarly extended to multilinear analysis, i.e. tensor analysis (Tao et al, 2007a,b). Nonlinear subspace learning mainly refers to nonlinear manifold learning (e.g. locally linear embedding (Roweis and Saul, 2000)). The concept of constructing a local neighborhood has been explored since the methods of such a sub-category have no explicit mapping function. When the subspace-learning framework is applied to FH, most methods assume that both sides of the face hallucination share the same linear combination of weights. An illustration of the subspace-learning framework is shown in Fig. 7; for an the holistic method such as eigentransformation, the partition mask denotes the identity map which preserves the whole face image and K equals to the number of training image pairs. The patch should be substituted by the holistic face image.

3.1 Linear Subspace Learning-based Approaches

Tang and Wang (2002; 2003; 2004; 2003; 2005) proposed an eigentransform method for face sketch synthesis and FSR by exploiting principal component analysis (PCA). This method assumes that the sketch and corresponding photo share the same linear combination coefficients (Tang and Wang, 2002, 2004). The input photo $\boldsymbol{p}_r$ is first projected on the photo training set $\boldsymbol{P}$ to obtain the linear combination coefficients $\boldsymbol{c}_p$

$$\boldsymbol{p}_r = \boldsymbol{P}\boldsymbol{c}_p = \sum_{i=1}^{M} c_{p_i}\boldsymbol{P}_i. \tag{15}$$

The target sketch $\boldsymbol{s}_r$ is then synthesized by linearly combining sketches $\boldsymbol{S}$ in the training set with aforementioned coefficients $\boldsymbol{c}_p$

$$\boldsymbol{s}_r = \boldsymbol{S}\boldsymbol{c}_p = \sum_{i=1}^{M} c_{p_i}\boldsymbol{S}_i. \tag{16}$$

In Tang and Wang (2003), the shape was first separated from the texture and then eigentransform was applied to shape and texture. Finally, the shape and texture are fused to obtain the sketch corresponding to the input photo. The idea of eigentransform was then applied to FSR (Wang and Tang, 2003, 2005).

Considering the fact that less information is provided in the sketch than in an original face image, which may affect face recognition performance, Li et al (2006) proposed an algorithm for synthesizing a photo from its sketch counterpart. In the proposed algorithm, they performed eigen-analysis (Turk and Pentland, 1991) on a hybrid space consisting of training sketches and training photos instead of on the photo space, as in the methods discussed previously. By separating the hybrid projection matrix obtained from the eigen-analysis into two coupled matrixes-an eigenphoto matrix and an eigensketch matrix, the projective coefficients are obtained by projecting the query sketch on the sketch space spanned by the columns of the eigensketch matrix. Finally, the pseudo-photo is synthesized from the linear combination of eigenphotos weighted by the obtained coefficients.

Park and Lee proposed a method for FSR (Park and Lee, 2003) that was similar to Tang and Wang's work (Tang and Wang, 2003). The method is based on a face morphable model under the framework of top-down learning (Hwang and Lee, 2003; Jones et al, 1997). The shape and texture information are first separated using backward warping and then decomposed into a linear combination of shapes and textures collected from the low-resolution training images. The reconstruction weight is found by solving a least square problem, after which the high-resolution image could be hallucinated by combining the estimated high-resolution shape with the estimated high-resolution texture using forward warping. They extended this idea further to a two-step method by incorporating a recursive error back-projection procedure (Park and Lee, 2008).

In contrast to the method in (Park and Lee, 2008), which utilized the prototype faces trained from the raw training images rather than from the residual images themselves, Hsu et al (2009) presented a two-step method. In the method, both high and low-resolution training images are divided into subsets and then, for each input low-resolution image, the closest training subset is selected by finding the index of the nearest cluster under Euclidian distance metric.

Different from the aforementioned PCA-based methods all performed on a whole face image, which may result in some blurring effect and the loss of some critical fine detail information, Liu et al (2005c) proposed a patch-based method that utilizes multilinear analysis techniques and a coupled residue compensation strategy. An initial estimate is obtained via tensorpatch super-resolution. By performing PCA on both low and high-resolution training images, the final high-resolution image is reached by settling a least square problem. The experimental results indicates that the high-resolution face image generated by their proposed method resulted in improvements, especially for some detail parts, compared to other global methods.

In addition to PCA, locality preserving projection (LPP) (He, 2005) is explored to compute the projection weights. Zhuang et al (2007) proposed a two step FSR method: locality preserving hallucination for the initial estimate and neighbor reconstruction for residue compensation. LPP is first utilized to extract the embedding features from training images and the low-resolution input image. Next, the radial basis function (RBF) regression model is learned from the training image features obtained and the training image intensities, characterized by the regression coefficients. The whole image is then output from the RBF regression model after inputting the feature of the source input image. At the residual compensation stage, neighbor embedding (Chang and Xiong, 2004) is explored to hallucinate the high frequency information. Finally, the high-resolution image is fused by adding the estimated holistic face image and the high frequency information. Zhang et al (2008) proposed an adaptive learning method based on LPP. Given an input low-resolution photo, they first interpolated it to obtain its high-resolution counterpart and then filtered it using a low-pass filter to generate the low frequency face image. A similar process is also applied to the low-resolution training images. The LPP procedure is used on the low-resolution training images to obtain the basis and the mapped data matrixes. Residual faces are then obtained from the high-resolution faces by removing the corresponding low frequency parts. By projecting the input low frequency image patch on the basis matrix, the low dimensional feature is obtained. Similar features are selected and utilized in the obtained mapped data matrix under the metric of Euclidean distance, and the high frequency residual image is hallucinated using an eigentransformation-like method from the training residual images whose indexes are determined in the feature selection step. The final high-resolution image is synthesized by adding the low frequency face image to the residual image.

3.2 Nonlinear Manifold Learning-based Approaches

Inspired by locally linear embedding (LLE) (Roweis and Saul, 2000), Chang and Xiong (2004) proposed a super-resolution algorithm. This method assumes that the low-resolution face images and their corresponding high-resolution counterparts are sampled from two manifolds share a similar geometrical structure. The proposed method works at patch-level and all referred images are divided into patches at the outset. Given an input low-resolution image patch $\boldsymbol{y}$, K nearest neighbors $\boldsymbol{y}^i (i = 1, \cdots, K)$ are first found in the patches extracted from low-resolution training images. The reconstruction weight vector $\boldsymbol{w}$ is calculated by solving a least squares problem

$$\min_{\boldsymbol{w}} \parallel y - \sum_{i=1}^{K} w_i \boldsymbol{y}^i \parallel^2, s.t. \sum_{i=1}^{K} w_i = 1. \tag{17}$$

By linearly combining high frequency information $\boldsymbol{x}^i$ of the K high-resolution image patches corresponding to the selected K low-resolution candidates, $\sum_{i=1}^{K} w_i \boldsymbol{x}^i$, the target high-resolution image patch is generated by adding the corresponding low frequency information transferred from the low-resolution input image. All the obtained image patches are put together to obtain the final target.

Aware that global FH methods might lose critical fine detail information, the local neighborhood construction idea of LLE was subsequently explored in FH methods. Liu et al (2005a) applied a similar idea to that in (Chang and Xiong, 2004) to face sketch synthesis, taking the image intensities as the input and the output directly rather than the high frequency feature. Their experimental results showed that this nonlinear method achieved improvements over the global linear method like (Tang and Wang, 2002, 2003, 2004). Liu et al (2005d) utilized the neighbor embedding method to generate an initial high-resolution image from its low-resolution counterpart and then explored generalized singular value decomposition to hallucinate the high frequency information to compensate the residual for the initial estimate. They applied singular value decomposition to obtain two projection matrixes and then, given an input low-resolution image patch, synthesized the initial estimate by addressing a least square problem, thereby exploring their previous work (Liu et al, 2005b). The residual image is generated using a similar procedure. (Fan and Yeung, 2007) proposed a two-step image hallucination method using neighbor embedding over visual primitive features. In the first stage, neighbor embedding is also used to obtain an initial estimate. In the second stage, the residual error is compensated for by averaging the residual error of K nearest neighbors that are stored in the training phase. Chen et al (2009) applied neighbor embedding to visible image-near infrared image synthesis with LBP features as the input. Through their face recognition experimental results, great improvements were made for illumination variation cases.

Though patch-based methods improved the detail, a global search strategy among all training image patches is time-consuming. Ma et al. (2009; 2010a; 2010b) proposed a position-based FH method that borrowed the idea of neighbor embedding. After dividing all images into patches, the high-resolution counterpart of a given input low-resolution image patch is estimated by applying neighbor embedding to those training image patches located in the same position as the test patch. The authors also applied the position-based FH method to multi-view FH in which a multi-view face synthesis procedure was conducted before hallucination by utilizing a method similar to neighbor embedding (Ma et al, 2010a). They further investigated whether residue compensation was a necessary step for FH and declared that it was not indispensable if the FH algorithm did not incorporate dimension reduction methods such as PCA, or LPP which incur the loss of non-feature information (Ma et al, 2010b). Liang et al (2010) also utilized a method similar to neighbor embedding to compensate for the initial high-resolution estimate obtained from an image decomposition perspective.

To further improve the quality of local detail, pixel-structure is explored to refine the local characteristics. Also inspired by LLE, Hu et al (2010, 2011) proposed a method from local pixel structure to global image FSR. This method assumes that two face images belonging to the same person should have similar local pixel structures and that each pixel could be generated by a linear combination of its neighborhoods weighted by coefficients. They conducted their method in three steps as follows: (1) K example faces are searched from the low-resolution image training set containing images that are most similar to the input and K corresponding high-resolution example images are warped to the input face using optical flow (Brox et al, 2004); (2) the local pixel structures for the target high-resolution face image are learned from the warped high-resolution example faces; and (3) the target high-resolution face image is estimated by addressing a constrained least square problem by means of an iterative procedure. When the peak signal to noise ratio (PSNR) and structured similarity index metric (SSIM) (Wang and Bovik, 2004) values were compared to other methods, the proposed method was found to be superior.

Li et al (2009) claimed that the assumption adopted by many learning-based super-resolution methods that the low-resolution representation manifold and the corresponding high-resolution representation manifold share similar local geometry might not hold due to the non-isometric one-to-multiple mappings from low-resolution image patches to high-resolution image patches. They proposed a manifold alignment method for FH that projected the two manifolds to a common hidden manifold. An input low-resolution image is first projected onto the common manifold and the FH task is then conducted among the common manifold and the high-resolution manifold by way of the neighbor embedding.

3.3 Discussion

PCA-based FH methods preserve the holistic property but ignore the neighboring relations for encoding some local facial features. Manifold learning-based FH methods make up for this imperfection by enforcing certain constraints: for example, LLE assumes that each patch of a face can be reconstructed by a linear combination of its nearest neighbors and this relation is preserved for both low-resolution image (photo) and high-resolution image (sketch); LPP is a linear approximation of a Laplacian eigenmap (Belkin and Niyogi, 2001), which preserves the local geometry by constructing a graph of neighboring nodes connected by edges. Then LPP-based FH methods also take local facial features into account when they are applied to model the hallucination process. Though the local neighboring relations of manifold learning-based methods may be well preserved, the global shape information of a face might be not easily modeled by these methods.

4 Combination of Bayesian Inference and Subspace Learning Framework

Some works have explored both Bayesian inference (Section 2) and subspace learning methods (Section 3). Methods in this category have mostly applied subspace analysis (Fig. 6) to the prior model (Fig. 3) or explored subspace learning methods to generate an initial estimate, in which case the two step framework (Liu et al, 2001, 2007a) is adopted.

Although the method (Liu et al, 2001, 2007a) is introduced in Section 2.2 (under the Bayesian inference framework) for the convenience of introduction, indeed, it can be deemed as a representative approach in the contexts of Bayesian inference and subspace learning. In this approach (Liu et al, 2001, 2007a), principal component analysis is first applied to obtain an initial global face image and subsequently an MAP-MRF is exploited to calculate the local face image.

Liu et al (2007b) applied a two-step procedure to photo synthesis from an input sketch. In the first step, a method similar to (Liu et al, 2005a) (LLE-based) is used to generate an initial estimate. Then, by exploiting the proposed tensor model whose modes consisted of people identity, patch position, patch style (sketch or photo) and patch features, the high frequency residual error is inferred under the Bayesian MAP framework on the assumption that a sketch-photo patch pair shares the same tensor representation parameter. By adding these two parts, a photo with much more detailed information could be synthesized from the input sketch.

Zhang and Cham (2008, 2011) proposed a FSR approach in the DCT domain under the MAP framework. In this method, the high frequency DCT coefficients are abandoned due to their weak energy. The DC coefficient is calculated by an interpolation-based method, while the AC coefficients are estimated by their corresponding K nearest neighbors exploring the idea of LLE under a simplified MRF model that assumes there were no dependency relations between neighboring AC coefficients. Prefiltering procedures are executed along the boundaries block-wise locally before performing DCT and postfiltering procedures are carried out after applying IDCT.

Unlike many LLE-based methods that assumed the low-resolution patch and high-resolution patch had the same reconstruction weights or coefficients, Park and Savvides (2007) proposed a LPP-based FSR, which inferred the LPP projection coefficients of each high-resolution patch via Bayesian MAP criterion from an input low-resolution image patch. Together with the LPP projection matrix learned from the training of high-resolution image patches, the high-resolution image patch can be synthesized from their linear combination; hence, a final high-resolution image can be fused from the patches obtained.

Following Park and Savvides's procedure for projection coefficients (Park and Savvides, 2007), several similar methods are proposed. Kumar and Aravind (2008b) proposed a two-step method using orthogonal locality preserving projections (OLPP) (Cai et al, 2006) and kernel ridge regression (KRR). OLPP is utilized to estimate the coefficients of each high-resolution image patch, as in (Park and Savvides, 2007), while KRR is used to estimate the high frequency needed to compensate for the residual error. Kumar and Aravind (2008a) also proposed a similar idea that combined 2D-PCA (Zhang and Zhou, 2005) and KRR to hallucinate input low-resolution images, in which 2D-PCA was explored to estimate the 2D-PCA projection features (coefficients) of the high-resolution image and KRR was applied to estimate the residue. Subsequently, Cai et al (2006) substituted OLPP into the direct locality preserving projections (DLPP) method (Ahmed et al, 2008) to estimate the projection coefficients while using KRR to compensate for the residual image.

Several works investigate the learning procedure applied in multi-frame or video sequence FSR. Capel and Zisserman (2001) proposed a FSR method that could work either by constraining the solution to a restricted subspace or by defining a prior via subspace analysis where both subspaces were spanned by PCA components. An image is divided into four regions: the eyes (a pair), nose, mouth, and cheek (two sides) areas and some PCA components are trained separately from the corresponding training image regions. A maximum likelihood (ML) estimator is obtained by restricting the solution lying on the PCA subspace and a MAP estimator is produced by extending this ML estimator through adding a prior defined on the coefficients of the principal components. Another MAP estimator is formed by encouraging the estimated image to lie near to the PCA subspace as a prior. Similar work was also carried out by Chakrabarti et al (2007) who proposed a multi-frame face image super-resolution method via kernel PCA and took the prior in the form of a Gibbs function. The energy function reflects the energy of the high-resolution image outside the principal subspace that could be written in a least square distance of high-resolution image from its projection on the principal subspace. The high-resolution image is computed using the gradient-descent approach by solving a constrained least squares problem.

Considering the fact that one application of face image super-resolution is face recognition and that dimensionality reduction is frequently used in state-of-the-art face recognition systems, Gunturk et al (2003) proposed an eigenface-domain super-resolution method especially for face recognition using a sequence of images extracted from surveillance videos. FSR is usually seen as a preprocessing procedure for generic FSR-based recognition systems; however, in their proposed method, they first extracted feature vectors from sequential low-resolution images by PCA and then estimated the feature vector of the corresponding high-resolution image by exploring eigenface analysis under the MAP framework. The feature vector could then be used both for face recognition and high-resolution image reconstruction. After multiplying the feature vectors by the PCA projection matrix calculated from the high-resolution training images, the super resolved face image is obtained. This method has the advantage of a reduction in computational complexity but suffers from poor visual quality.

Most of the above methods are dedicated to frontal FSR or frontal FSPS, although some insight into the pose and view variation issue has been provided. Li and Lin (2004) proposed a FSR method with pose variation in which they first utilized a SVM classifier to estimate the pose label of the input low-resolution face image. The frontal low-resolution face image is then estimated by solving a least square problem based on the corresponding training images with the same pose label. Using the estimated frontal face image, they applied the hallucination method in (Gunturk et al, 2003) to hallucinate the high-resolution image. Following a similar idea to Gunturk et al.'s method (Gunturk et al, 2003), Jia and Gong (2005) proposed a multi-view, multi-illumination tensor face-based FSR approach. The tensor is composed of four modes: identities, views, illuminations, and pixels. This method derives a model for the reconstruction of identity parameter vectors in the high-resolution tensor space from the corresponding identity parameter vectors of low-resolution space. By substituting the principal component analysis in (Gunturk et al, 2003) with tensor analysis under the maximum likelihood estimation framework, the identity parameter is calculated and both face recognition and FSR are carried out. The tensor model was subsequently extended to the multi-resolution patch tensor for face expression hallucination under the MAP framework (Jia and Gong, 2006, 2008). This method could hallucinate several high-resolution images with different expressions given a low-resolution image. The tensor consists of modes of identities, expressions, resolutions, patches (patch location), and pixels. An image is decomposed into two parts: the low and middle frequency information part, and the high frequency information part, which results in the MAP objective function being solved by a two-step sequential solution. In the first step, the low and middle frequency information is evaluated by solving a least square problem. The high frequency information is then compensated by exploiting a nonparametric patch learning process in the second step. Combining these two parts, the target high-resolution image with some expression is computed.

5 The Sparse Representation-based Approaches

Sparse representation accounts for a decomposition that represents a signal $\boldsymbol{y}_{sig} \in \Re^n$ into a linear combination of basis signals $\boldsymbol{D}_i \in \Re^n (i = 1, \cdots, k)$, which are often called atoms, weighted by few nonzero coefficients. Given that $\boldsymbol{D} = [\boldsymbol{D}_1, \cdots, \boldsymbol{D}_k]$ denotes an over-complete dictionary ($k > n$), the sparse representation of the signal $\boldsymbol{y}_{sig}$ is represented as follows:

$$\underset{\boldsymbol{x}_{coe}}{\operatorname{argmin}} \|\boldsymbol{y}_{sig} - \boldsymbol{D}\boldsymbol{x}_{coe}\|_2 + \lambda \|\boldsymbol{x}_{coe}\|_0 \tag{18}$$

However, solving this "L_0-norm" (which is actually not a norm since it does not satisfy the three necessary conditions of the definition of norm) regularized problem is NP-hard and is computationally prohibitive. Nevertheless, Donoho (2006) recently proved that the minimal L_1-norm solution approximates the sparsest solution under mild conditions. Thus, the optimization problem in Eq. (18) is reformulated:

$$\operatorname*{argmin}_{\boldsymbol{x}_{coe}} \|\boldsymbol{y}_{sig} - \boldsymbol{D}\boldsymbol{x}_{coe}\|_2 + \lambda\|\boldsymbol{x}_{coe}\|_1 \tag{19}$$

which is known in statistical literature as the Lasso, essentially a linear regression regularized with L_1-norm on the coefficients (Tibshirani, 1996). In some applications, such as image denoising (Elad and Aharon, 2006) and image restoration (Mairal et al, 2008), the dictionary is always learned from training examples by alternatively optimizing $\boldsymbol{D}$ and $\boldsymbol{x}_{coe}$ respectively, while in applications such as face recognition (Wright et al, 2009), $\boldsymbol{D}$ is predefined as a set of either patches or features extracted from images.

Yang et al (2008a) proposed a two-step method for FSR based on sparse coding. In the first step, non-negative matrix factorization (Lee and Seung, 1999) is used to obtain a non-negative basis matrix $\boldsymbol{B}$ which spans a face subspace. An MAP problem is defined to reconstruct an initial estimiation to the target high-resolution image

$$\begin{aligned} &\max_{\boldsymbol{I}_H} P(\boldsymbol{I}_L|\boldsymbol{I}_H)P(\boldsymbol{I}_H) \Leftrightarrow \\ &\boldsymbol{c}^* = \operatorname*{argmin}_{\boldsymbol{c}} \| \boldsymbol{MBc} - \boldsymbol{I}_L \|_2^2 + \lambda\| \boldsymbol{\Gamma Bc} \|_2, \quad s.t. \boldsymbol{c} \succeq 0, \end{aligned} \tag{20}$$

where $\boldsymbol{M}$ is a blurring and dow-sampling matrix, $\boldsymbol{c}$ is the non-negative reconstruction coefficient vector, λ is a trade-off factor between the reconstruction term and the prior term, and $\boldsymbol{\Gamma}$ is a high-pass filtering matrix. Finally the initial high-resolution image is generated by $\boldsymbol{Bc}^*$. Since the regularization in Eq. (20) requires the result to be smooth, some critical high-frequency information can be filtered.

In the second step of this method, sparse representation on both low-resolution image patches (features) and high-resolution image patches (features) is applied to obtain the residual image to compensate the missing detailed information. Before performing sparse representation, two dictionaries $\boldsymbol{D}_L$ and $\boldsymbol{D}_H$ are constructed from some patch pairs randomly sampled from the training images (both low-resolution and hihg-resolution images). The target high-frequency information is computed by $\boldsymbol{D}_h\boldsymbol{\alpha}$, where the coefficient vector $\boldsymbol{\alpha}$ is given by

$$\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_1 + \frac{\eta}{2}\| \widetilde{\boldsymbol{D}}\boldsymbol{\alpha} - \widetilde{\boldsymbol{y}} \|_2^2 \tag{21}$$

where $\widetilde{\boldsymbol{D}} = \begin{bmatrix} \boldsymbol{FD}_L \\ \beta\boldsymbol{ED}_H \end{bmatrix}$ and $\widetilde{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{Fy} \\ \beta\boldsymbol{\omega} \end{bmatrix}$. The parameter η controls the tradeoff between the sparsity of the coefficient and the fidelity of the data term and β balances the low-resolution reconstruction and the compatibility among neighboring patches. $\boldsymbol{F}$ extracts the gradients of patches. $\boldsymbol{E}$ extracts the overlapped region between the current target patch and the neighboring reconstructed patch. $\boldsymbol{\omega}$ consists of the intensities of a neighboring patch on the overlapped region. $\boldsymbol{y}$ is the input low-resolution image patch. Finally the target output high-resolution image is obtained by superimposing $\boldsymbol{D}_h\boldsymbol{\alpha}$ on $\boldsymbol{Bc}^*$.

A similar sparse representation model is borrowed in several other works. In (Chang et al, 2010), Yang's model (2008a) was applied to face sketch-photo synthesis by substituting the low- and high-resolution images with sketches and photos. Considering that different areas of the face might have their own characteristics, Wang et al (2011) proposed a multi-dictionary based sparse representation framework in which a sub-dictionary was learned from a cluster of training image patches. A simple version of this model was also borrowed to perform FSR and FSPS with the dictionary predefined as a collection of image patches (Ji et al, 2011; Jung et al, 2011). Now that the number of nearest neighbors is fixed in most existing methods, which might introduce some deformation and noise into the result, Gao and Wang et al (2012; 2013a) utilized sparse representation to conduct feature selection for FSPS and heterogeneous image transformation respectively. The principal motivation is to adaptively select closely related features whose sparse representation coefficient is larger than a threshold value. By substituting the L_0-norm or L_1-norm regularized prior term in Eq. (18) or (19) with another regression regularized prior such as ridge regression prior (Komarek, 2004) and relevance vector machine prior (Tipping, 1991) (which belongs to the sparse Bayesian approaches), Chang et al (2011) proposed a multivariate output regression-based face sketch-photo synthesis method. Zhang et al (2011a) proposed a support vector regression-based two-step method for a face sketch-photo synthesis method.

Different from above methods all assumed that the source input and the target output had the same sparse representations, Wang et al (2012) relaxed this assumption supposing they had their respective sparse representations. These two sparse representations are connected through a linear transformation. Then the objective function is composed of two sparse representation parts, one fidelity term between the sparse representation coefficients, and the regularization term on the linear transformation matrix, under some scale constraints to each atom of dictionaries. They separated the objective function into 3 sub-problems: sparse coding for training samples, dictionary updating and linear transformation matrix updating.Experimental results seem to be over-smoothed.

6 Performance Evaluations

The evaluation for FH can be subjective quality assessment or objective quality assessment. Subjective quality assessment can be applied by visual perception or mean opinion score (MOS), which has been used in ITU-T p.910, a standard in multimedia services. Visual perception is predicated on the observers' perception without a numerical quantification. MOS is defined as the average of the quality values ranging from 1 to 5 that are obtained from observers. Although the subjective image quality assessment is the most direct and most accurate metric to reflect a person's perception, it is always subject to the defects of costs and expensive manpower. As a result, objective quality assessment metrics that operate in an automatic manner have been proposed. These include classical PSNR, mean square error (MSE) or root mean square error (RMSE), cross-correlation, the recently proposed SSIM (Wang and Bovik, 2004), and the universal image quality index (UIQI) (Wang and Bovik, 2002) (a special case of SSIM) method for generic image quality assessment. To some extent, face recognition rate can also be seen as an objective image quality assessment metric because it measures the similarity of the query image to images in the gallery. Table 2 summarizes and compares the evaluation metrics of a number of representative methods.

Although FSR is important for improving the performance of face recognition, there are limited results to explain how FSR quantitatively affecting the face recognition performance. Gunturk et al (2003) performed eigenface (Turk and Pentland, 1991) recognition experiments on some real video sequences containing 68 people, collected from the CMU PIE database (Sim et al, 2002). They achieved an accuracy of 44% by utilizing low-resolution images in comparison to 74% by exploring their hallucinated high-resolution face images. Park and Lee (2008) performed eigenface (Turk and Pentland, 1991) recognition experiments on three databases: MPI (Vetter and Troje, 1997), XM2VTS (Messer et al, 1999), and KF (Roh and Lee, 2007). Their results show that recognition performance can be significantly improved by utilizing their hallucinated high-resolution face images compared with exploiting the interpolated high-resolution images. Wang and Tang (2005) conducted direct correlation-based face recognition on 490 face images of 295 subjects in the XM2VTS database (each subject has two images from two different sessions). They found that the recognition accuracies fluctuate slightly when the down-sample factor is not too large (not larger than 5 in the paper). When the down-sample factor is reduced further, the hallucinated high-resolution face images improve the face recognition performance compared to directly utilizing the low-resolution images. They also pointed out that the improvement on face recognition accuracy is not as significant as that in the visual quality. Further studies in psychology and human visual system are valuable to examine how FSR help improve face recognition and verification performance.

Most existing FSPS methods perform synthesis and recognition experiments on the public database: CUFS (Wang and Tang, 2009). This database contains 606 face sketch-photo pairs consisting of three sub-databases: CUHK Student (188 pairs), Purdue AR (123 pairs), XM2VTS (295 pairs). Face photos of this database are generally in neutral expression, normal lighting, and frontal view. In experiments, 306 pairs are usually utilized for model training and the remaining 300 pairs are for model test. Tang and Wang (2003) reported an accuracy of 81.3% by exploring a Bayesian classifier (Moghaddam et al, 2000) in comparison to 25% by applying eigenface (Turk and Pentland, 1991) method on sketches without any synthesis process. Subsequently Liu et al (2005a) improved the accuracy to 88% by adopting the kernel based nonlinear discriminant analysis (Mika et al, 1999) as the dimension reduction algorithm. Wang and Tang (2009) then achieved an accuracy of 96.3% classified by random sampling linear discriminant analysis (Wang and Tang, 2006). Considering the sketches of above database are in a relative simple structure, the multimedia lab of Chinese University of Hong Kong further released the Face sketch FERET database (CUFSF) (Zhang et al, 2011c), which includes 1,194 persons in the FERET database (Phillips et al, 2000). Each person in the CUFSF database has a photo with lighting variation and a sketch with shape exaggeration drawn by the artist.

7 Promising Future Directions and Tasks

In Section 6, we saw that when each method is evaluated by visual perception in a subjective image quality assessment manner, it is expensive and may easily become tedious. Thus, an automatic objective image quality assessment metric is essential in evaluating the performance of the FH algorithm. Classical full reference metrics such as PSNR, MSE, and RMSE are

Table 2 Evaluation Summary and Comparison of Different FH Methods

Method	Database	Category	Subjective Metric	Objective Metric
Baker and Kanade (2000a,b, 2002)	FERET (Philips et al, 1997)	BI(GP)	VP	RMSE
Su et al (2005)	FERET, AR (Martinez and Benavente, 1998), Cohn Kanade (Kanade et al, 2000), PIE (Sim et al, 2002)	BI(GP)	VP	N/A
Wang and Tang (2009)	CUFS (Wang and Tang, 2009)	BI(MRF)	VP	FR
Liu et al (2001)	FERET, AR	BI(MRF)	VP	N/A
Gao et al (2008b,c)	(Gao et al, 2008b,c)	BI(E-HMM)	VP	UIQI, FR
Park and Lee (2008)	KF (Roh and Lee, 2007), XM2VTS (Messer et al, 1999), MPI (Vetter and Troje, 1997)	SL(LSL)	VP	SSIM, FR
Wang and Tang (2005)	CUFS Student (Wang and Tang, 2003)	SL(LSL)	VP	RMSE, FR
Liu et al (2005c)	FERET	SL(LSL)	VP	N/A
Ma et al (2010b)	CAS-PEAL (Gao et al, 2008a), FERET, CMU (Rowley et al, 1998), Stereo-pair (Fransens et al, 2005)	SL(NML)	VP	PSNR
Zhuang et al (2007)	Asian Face (Dong and Gu, 2001)	SL(NML)	VP	PSNR
Hu et al (2011)	FERET, AR, GA (Nefian, 1997)	SL(NML)	VP	PSNR, SSIM
Liu et al (2007b)	CUFS	C-BI-SL	VP	RMSE
Chakrabarti et al (2007)	FERET YALE (Georghiades et al, 2001)	C-BI-SL	VP	SSIM, MSE
Zhang and Cham (2011)	FERET	C-BI-SL	VP	SSIM, MSE
Gunturk et al (2003)	YALE, CMU, AR, HRL (Hallinan, 1994)	C-BI-SL	VP	FR
Wang et al (2011)	CUFS Student (Tang and Wang, 2002), VIPSL (Wang et al, 2011)	SR	MOS, VP	FR
Yang et al (2008a)	FRGC 1.0 (Philips et al, 2005)	SR	VP	N/A
Chang et al (2010)	CUFS Student	SR	VP	N/A

Note: BI-**B**ayesian **I**nference, the notation in the parentheses denotes the sub-category, GP-**G**radient-based **P**rior for data modeling, VP-**V**isual **P**erception, N/A-**N**ot **A**vailable, FR-**F**ace **R**ecognition Rate, SL-**S**ubspace **L**earning, LSL-**L**inear **S**ubspace **L**earning, NML-**N**onlinear **M**anifold **L**earning, C-BI-SL-**C**ombination of **B**ayesian **I**nference and **S**ubspace **L**earning, SR-**S**parse **R**epresentation.

holistic and cannot yet reflect the detailed information that is needed to assess image quality. This point is discussed in detail by Wang and Bovik (2009). Therefore, an effective, objective image quality assessment metric that has much better correlation with subjective visual perception needs to be developed. Several metrics such as UIQI, SSIM, VIF (Sheikh and Bovik, 2006), and FSIM (Zhang et al, 2011b) have been proposed; however, none of them is specialized for hallucinated face images, which have their own unique characteristics due to both the structure of the face and the property of the hallucinated image. Hence, synthesized face image quality assessment may be a promising and helpful research direction.

Recently, sparse representation has achieved great progress in computer vision (Wright et al, 2010) and data analysis (Zhou and Tao, 2013). In particular, methods have been proposed for image reconstruction and state-of-the-art results have been obtained (Mairal et al, 2008; Marial et al, 2008). Yang et al (2008b) applied the idea of the sparse representation model with a coupled learning process to face image super-resolution and achieved good results. Yang et al.'s method (2008b) is not the end of the application of sparse representation to FH, since the method considers less prior knowledge of the face image than the face images provide, and the effective exploration of the sparsity of face images is therefore an interesting problem to resolve.

From Table 2, we find that most image databases used for face image super-resolution were not sampled from surveillance camera videos, since the main application of face image super-resolution is face recognition or face retrieval from a monitor. Therefore, an image database extracted from surveillance videos should be constructed that incorporates pose, illumination, expression, and view variant images. Although the CUFS database has been constructed, there is only one sketch with neutral expression and front view corresponding to each photo in the database for face sketch-photo synthesis; therefore, constructing a database containing several sketches corresponding to each photo across multiple modalities is essential. Furthermore, these two databases will stimulate the progress of study on multi-modality FH and recognition.

Though FSR and FSPS share a similar mathematical form, they are intrinsically different. The first difference comes from how much the face alignment precision affects the hallucination. Face alignment is a critical preprocessing phase before FH, because imprecise localization of the facial features (landmarks) degrades the subsequent processes. Experiments (Liu et al, 2007a; Jia and Gong, 2008; Luo et al, 2012) indicate accurate face alignment is more important for FSR than for FSPS. Because face sketches and corresponding photo counterparts are generally in high or moderate resolution, their alignment is relatively easier. Even a small amount of misalignment can dramatically degenerate the FSR performance. Low-resolution images usually have blurring effect and contain limited structure information, and so many ambiguities exist for facial landmark localization which raises the alignment of low-resolution face images a challenging problem. Another difference lies in whether they need to handle the problem of shape exaggeration. Artists usually exaggerate some distinctive facial features when they draw sketches, which results in some deformation. Wang and Tang (2009) explained that "if a face has a big nose in a photo, the nose drawn in the sketch will be even bigger". Consequently, in contrast to FSR, FSPS needs to handle the problem of shape exaggeration.

From above analysis, precisely detecting facial landmarks on low-resolution images to perform face alignment is still a challenging problem. Moreover, the shape exaggeration causes nonlinear transformation between sketches and photos. Existing FSPS approaches rarely consider the nonlinear mapping resulted by shape exaggeration. Although Tang and Wang (2003) considered the global shape information in their work, the mapping relationship between sketches and photos were assumed to be linear. Thus, effective face alignment on low-resolution images and appropriately modeling the nonlinear relationship between sketches and photos resulted by shape exaggeration are two promising research directions.

Besides learning-based face sketch synthesis methods surveyed in this paper, some sketch synthesis algorithms are not learning-based (Kang et al, 2005; Wen et al, 2006). Whatever these methods are, they are applicable to general images. However, they can hardly handle the styles by different artists. This is because different artists may have different representation and exaggeration styles for many parts of a face. For example, different artists may render the nose, eye, mouth and other parts of a face differently. It may be even more difficult to model these different artistic styles than model the shape exaggeration. To learn these different styles, some discriminative information among them may favor the synthesis process since it can assist to choose a sketch part (here face part can be a face patch or a holistic face) from sketches of desired styles.

Patch-based methods have been widely applied to face hallucination due to their ability to represent the local facial features. However, these methods neglect the global shape information describing the holistically geometric relationships between the individual facial features. Especially in face sketch-synthesis, state-of-the-art methods adopt the patch-based strategies which actually loss some important information about global shape exaggeration. Tang and Wang (2002, 2003, 2004) proposed an eigentransform scheme to take the global shape exaggeration into account. Nevertheless, local facial features were lost and this strategy could hardly distinguish subtle individual facial feature variations. Therefore, designing an approach to integrate both the information of local neighborhood and the configuration of global shape exaggeration is important for face sketch synthesis.

As is shown in (Liu et al, 2007b; Wang et al, 2011; Zhang et al, 2011a), most available methods for face sketch-photo synthesis averaged the overlapping regions, which may result in over-smoothing, and the residual image could therefore be learned from the training images to compensate for the lost high frequency information. However, residual images were learned from the sketch-photo pairs rather than the training synthesized images and corresponding truth images. Thus, a two-step face sketch-photo synthesis framework such as this needs be explored further.

In recent years, many FH methods have been developed and obtained promising performance for face recognition under well-controlled conditions. In particular, FSPS can significantly improve the face recognition accuracy comparing with direct recognition using sketch under well-controlled condition (Wang and Tang, 2009). However, this does not suggest that FH is a solved problem. The recognition performance degenerates when encountering faces collected from uncontrolled conditions such as faces with non-frontal views, expression and lighting variations due to the intrinsic non-rigidness of faces and extrinsic uncontrollable environment conditions. Though some valuable results for face hallucination have been obtained to handle one or two types of the aforementioned variations (Li and Lin, 2004; Jia and Gong, 2005, 2006, 2008; Ma et al, 2010a; Zhang et al, 2010), so much effort is required to face the real challenges when attempting to handle multiple variations simultaneously in practice. Especially, designing effective approaches for modeling these variations is essential to apply FH in various real-world tasks and is a focus of the future research.

8 Conclusion

In this paper, we reviewed the topic of face hallucination incorporating face image super-resolution and face sketch-photo synthesis. The methods utilized are classified into four categories according to the framework under which they fall: Bayesian inference framework, subspace learning framework, combination of Bayesian inference and subspace learning framework, and sparse representation-based methods. Bayesian inference framework-based approaches can be grouped into three subcategories: gradient-based gradient prior model-based methods, MRF-based methods, and E-HMM-based methods. Subspace learning-based algorithms are divided into linear subspace learning-based methods and nonlinear subspace learning-based methods. By means of a comprehensive analysis and comparison of these methods, we found that Bayesian inference methods have the disadvantage of high computation cost and heavy memory load, although neighbor compatibility reduces the boundary noise (except for the gradient-based prior for data modeling-based methods). We also found that subspace learning-based methods make strict assumptions about the geometric structure of two image spaces and low computation cost. The combination of these two frameworks may result in a more accurate method (except for the LLE-based methods in this category). Although different from subspace learning-based methods, sparse representation-based methods also assume that two image spaces share a similar geometric structure; however, this assumption is constrained on two sparse spaces. This relaxes the original, much more restrictive assumption to some extent. Finally, we proposed several promising future directions and tasks, and we believe this survey will help readers to gain a thorough understanding of the face hallucination research landscape. Although face super-resolution and face sketch-photo synthesis share the similar framework, this does not mean that methods which work well for face super-resolution also work well on face sketch-photo synthesis and vice versa. This indicates that applying face super-resolution techniques directly to face sketch-photo synthesis may not always achieve good performance and vice versa. This may be due to the fact that though down-sampling and blurring effect are the mian factors of difference between low-resolution and high-resolution images, they have the similar texture or intensity expressions. However, sketches and photos are in quite different texture expressions.

Acknowledgements We want to thank the helpful comments and suggestions from the anonymous reviewers. This work was supported in part by the National Basic Research Program of Chinas 973 Program under Grant 2012CB316400, the National Natural Science Foundation of China under Grant 61125204, Grant 61125106, Grant 61172146, Grant 91120302, and Grant 61072093, the Australia Research Council Discovery Project under Grant ARC DP-120103730, the Fundamental Research Funds for the Central Universities under Grant K5051202048, and Grant K50513100009, and the Shaanxi Innovative Research Team for Key Science and Technology under Grant 2012KCT-02.

References

Ahmed S, Ghafoor A, Sheri A (2008) Direct hallucination: direct locality preserving projections (dlpp) for face super-resolution. In: Proceedings of International Conference on Advanced Computer Theory and Engineering, pp 105–110

Baker S, Kanade T (2000a) Hallucinating faces. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp 83–88

Baker S, Kanade T (2000b) Limits on super-resolution and how to break them. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 372–379

Baker S, Kanade T (2002) Limits on super-resolution and how to break them. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9):1167–1183

Belkin M, Niyogi P (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of Advances in Neural Information Processing Systems, pp 585–591

Bishop C, Blake A, Marthi B (2003) super-resolution enhancement of video. In: Proceedings of IEEE Workshop on Artificial Intelligence and Statistics

Bonet J (1997) Multiresolution sampling procedure for analysis and synthesis of texture images. In: Proceedings of SIGGRAPH, pp 361–368

Brown M, Lowe D (2007) Automatic panoramic image stitching using invariant features. International Journal of Computer Vision 74(1):59–73

Brox T, Bruhn A, Papenberg N, Weicket J (2004) High accuracy optical flow estimation based on a theory for warping. In: Proceedings of European Conference on Computer Vision, pp 25–36

Burt P (1981) Fast filter transforms for image processing. Computer Graphics and Image Processing 16(1):20–51

Burt P, Adelson E (1983) The laplacian pyramid as a compact image code. IEEE Transactions on Communications 31(4):532–540

Cai D, He X, Han J, Zhang H (2006) Orthogonal laplacianfaces for face recognition. IEEE Transactions on Image Processing 15(11):3608–3614

Capel D, Zisserman A (2001) Super-resolution from multiple views using learnt image models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 627–634

Chakrabarti A, Rajagopaian A, Chellappa R (2007) Super-resolution of face images using kernel-pca-based prior. IEEE Transactions on Multimedia 9(4):888–892

Chang D H Yeung, Xiong Y (2004) Super-resolution through neighbor embedding. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 275–282

Chang M L Zhou, Han Y, Deng X (2010) Face sketch synthesis via sparse representation. In: Proceedings of International Conference on Pattern Recognition, pp 2146–2149

Chang M L Zhou, Deng X, Han Y (2011) Face sketch synthesis via multivariate output regression. In: Proceedings of International Conference on Human-Computer Interaction, pp 555–561

Chellappa R, Wilson C, Sirohey S (1995) Human and machine recognition of faces: a survey. Proceedings of the IEEE 83(5):705–740

Chen H, Xu Y, Shum H, Zhu S, Zheng N (2001) Example-based face sketch generation with nonparametric sampling. In: Proceedings of IEEE International Conference on Computer Vision, pp 433–438

Chen J, Yi D, Yang J, Zhao G (2009) Learning mappings for face synthesis from near infrared to visual light images. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp 156–163

Dedeoglu G, Kanade T, August J (2004) High-zoom video hallucination by exploiting spatial-temporal regularities. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp 151–158

Dong H, Gu N (2001) Asian face image database pf01. Tech. rep., Pohang University of Science and Technology

Donoho D (2006) For most large underdetermined systems of linear equations, the minimal $l1$-norm near-solution approximates the sparsest near-solution. Communications on Pure and Applied Mathematics 59(7):907–934

Efros A, Freeman W (2001) Image quilting for texture synthesis and transfer. In: Proceedings of SIGGRAPH, pp 341–346

Efros A, Leung T (1999) Texture synthesis by non-parametric sampling. In: Proceedings of IEEE International Conference on Computer Vision, pp 1033–1038

Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing 15(12):3736–3745

Elad M, Feuer A (1997) Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images. IEEE Transactions on Image Processing 6(12):1646–1658

Elad M, Feuer A (1999) Super-resolution reconstruction of image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(9):817–834

Fan W, Yeung D (2007) Image hallucination using neighbor embedding over visual primitive manifolds. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp 1–7

Fransens R, Strecha C, Gool L (2005) Parametric stereo for multi-pose face recognition and 3d-face modeling. In: Proceedings of IEEE International Conference on Computer Vision Workshop Analysis and Modeling of Faces and Gestures, pp 109–124

Freeman W, Pasztor E (1999) Learning low-level vision. In: Proceedings of IEEE International Conference on Computer Vision, pp 1182–1189

Freeman W, Pasztor E, Carmichael O (2000) Learning low-level vision. International Journal of Computer Vision 40(1):25–47

Freeman W, Jones T, Pasztor E (2002) Example-based super-resolution. IEEE Computer Graphics and Applications 22(2):56–65

Fu Y, Guo G, Huang T (2010) Age synthesis and estimation via faces: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(11):1955–1976

Gao W, Cao B, Shan S, Chen X, Zhou D, Zhang X, Zhao D (2008a) The cas-peal large-scale chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics: Part A 38(1):149–161

Gao X, Zhong J, Li J, Tian C (2008b) Face sketch synthesis using e-hmm and selective ensemble. IEEE Transactions on Circuits and Systems for Video Technology 18(4):487–496

Gao X, Zhong J, Tao D, Li X (2008c) Local face sketch synthesis learning. Neurocomputing 71(10-12):1921–1930

Gao X, Wang N, Tao D, Li X (2012) Face sketch-photo synthesis and retrieval using sparse representation. IEEE Transactions on Circuits and Systems for Video Technology 22(8):1213–1226

Gelman A, Carlin H J Stern, Rubin D (2003) Bayesian Data Analysis. Chapman & Hall/CRC

Georghiades A, Belhumeur P, Kriegman D (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6):643–660

Gunturk B, Batur A, Altunbasak Y (2003) Eigenface-domain super-resolution for face recognition. IEEE Transactions on Image Processing 12(5):597–606

Hallinan P (1994) A low dimensional representation of human faces for arbitrary lighting conditions. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 995–999

Hardie R, Barnard K, Armstrong E (1997) Joint map registration and high-resolution image estimation using a sequence of undersampled images. IEEE Transactions on Image Processing 6(12):1621–1633

He X (2005) Locality preserving projections. Tech. rep., PhD Thesis, University of Chicago

Hertzmann A, Jacobs C, Oliver N, Curless B, Salesin D (2001) Image analogies. In: Proceedings of SIGGRAPH, pp 327–340

Hsu C, Lin C, Liao H (2009) Cooperative face hallucination using multiple references. In: Proceedings of IEEE International Conference on Multimedia & Expo, pp 818–821

Hu Y, Lam K, Qiu G, Shen T (2010) Learning local pixel structure for face hallucination. In: Proceedings of IEEE International Conference on Image Processing, pp 26–29

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

Hwang B, Lee S (2003) Reconstruction of partially damaged face images based on a morphable face model. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(3):365–372

Iwashita S, Takeda Y, Onisawa T (1999) Expressive face caricature drawing. In: Proceedings of IEEE International Conference on Fuzzy Systems, pp 1597–1602

Jain A, Duin R, Mao J (2000) Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1):4–37

Ji N, Chai X, Shan S, Chen X (2011) Local regression model for automatic face sketch generation. In: Proceedings of International Conference on Image and Graphics, pp 412–417

Jia K, Gong S (2005) Multi-modal tensor face for simultaneous super-resolution and recognition. In: Proceedings of IEEE International Conference on Computer Vision, pp 1683–1690

Jia K, Gong S (2006) Multi-resolution patch tensor for face expression hallucination. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 395–402

Jia K, Gong S (2008) Generalized face super-resolution. IEEE Transactions on Image Processing 17(6):873–886

Jones M, Sinha P, Vetter T, Poggio T (1997) Top-down learning of low-level vision tasks. Current Biology 7(12):991–994

Jung C, Jiao L, Liu B, Gong M (2011) Position-patch based face hallucination using convex optimization. IEEE Signal Processing Letters 18(6):367–370

Kanade T, Cohn J, Tian Y (2000) Comprehensive database for face expression analysis. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp 46–53

Kang H, He W, Chui C, Chakraborty U (2005) Interactive sketch generation. The Visual Computer 21(8-10):821–830

Komarek P (2004) Logistic regression for data mining and high-dimensional classification. Tech. rep., PhD Thesis, Carnegie Mellon University

Koshimizu H, Tominaga M (1999) On kanse face processing for computerized face caricaturing system picasso. In: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, pp 294–299

Kumar B, Aravind R (2008a) A 2d model for face superresolution. In: Proceedings of International Conference on Pattern Recognition, pp 1–4

Kumar B, Aravind R (2008b) Face hallucination using olpp and kernel ridge regression. In: Proceedings of IEEE International Conference on Image Processing, pp 353–356

Lee D, Seung H (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

Li B, Chang H, Shan S, Chen X (2009) Aligning coupled manifolds for face hallucination. IEEE Signal Processing Letters 16(11):957–960

Li S (2010) Markov random field modeling in image analysis. Springer

Li Y, Lin X (2004) Face hallucination with pose variation. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, pp 723–728

Li Y, Savvides M, Bhagavatula V (2006) Illumination tolearn face recognition using a novel face from sketch synthesis approach and advanced correlation filters. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp 357–360

Liang L, Liu C, Xu Y, Guo B (2001) Real-time texture synthesis by patch-based sampling. ACM Transactions on Graphics pp 127–150

Liang Y, Lai J, Xie X, Liu W (2010) Face hallucination under an image decomposition perspective. In: Proceedings of International Conference on Pattern Recognition, pp 2158–2161

Lin Z, He J, Tang X, Tang C (2007) Limits of learning-based superresolution algorithms. In: Proceedings of IEEE International Conference on Computer Vision, pp 1–8

Lin Z, He J, Tang X, Tang C (2008) Limits of learning-based superresolution alogrithms. International Journal of Computer Vision 80(3):406–420

Liu C, Shum H, Zhang C (2001) A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 192–198

Liu C, Shum H, Freeman W (2007a) Face hallucination: theory and practice. International Journal of Computer Vision 75(1):115–134

Liu Q, Tang X, Jin H, Lu H, Ma S (2005a) A nonlinear approach for face sketch synthesis and recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1005–1010

Liu W, Lin D, Tang X (2005b) Face hallucination through dual associative learning. In: Proceedings of IEEE International Conference on Image Processing, pp 873–876

Liu W, Lin D, Tang X (2005c) Hallucinating faces: tensorpatch super-resolution and coupled residue compensation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 478–484

Liu W, Lin D, Tang X (2005d) Neighbor combination and transformation for hallucinating faces. In: Proceedings of IEEE International Conference on Multimedia & Expo, pp 145–148

Liu W, Tang X, Liu J (2007b) Bayesian tensor inference for sketch-based face photo hallucination. In: Proceedings of International Joint Conference on Artificial Intelligence, pp 2141–2146

Liu X (2009) Discriminative face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(11):1941–1954

Luo P, Wang X, Tang X (2012) Hierarchical face parsing via deep learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2480–2487

Ma H X ang Huang, Wang S, Qi C (2010a) A simple approach to multiview face hallucination. IEEE Signal Processing Letters 17(6):579–582

Ma X, Zhang J, Qi C (2009) Position-based face hallucination method. In: Proceedings of IEEE International Conference on Multimedia & Expo, pp 290–293

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

Mairal J, Sapiro G, Elad M (2008) Learning multiscale sparse representations for image and video restoration. SIAM Multiscale Modeling and Simulation 17:214–241

Marial J, Elad M, Sapiro G (2008) Sparse representation for color image restoration. IEEE Transactions on Image Processing 17(1):53–69

Martinez A, Benavente R (1998) The ar face database. Tech. rep., CVC Technical Report #24

Messer K, Matas J, Kittler J, Luettin J, Maitre G (1999) Xm2vtsdb: the extended m2vts database. In: Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication, pp 72–77

Mika S, Ratsch G, Weston J (1999) Fisher discriminant analysis with kernels. In: Proceedings of IEEE Workshop on Neural Networks for Signal Processing, pp 41–48

Moghaddam B, Jebara T, Pentland A (2000) Bayesian face recognition. Pattern Recognition 33(11):1771–1782

Nefian A (1997) Georgia tech face database. http://www.anefian.com/research/face_reco.htm

Nefian A, Hayes M (1999) Face recognition using an embedded hmm. In: Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication, pp 19–24

Ong E, Bowden R (2011) Robust face feature tracking using shape-constrained multiresolution-selected linear predictors. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(1):1–16

Park J, Lee S (2003) Resolution enhancement of face image based on top-down learning. In: Proceedings of SIGMM Workshop on Video Surveillance, pp 59–64

Park J, Lee S (2008) An example-based face hallucination method for single-frame, low-resolution face images. IEEE Transactions on Image Processing 17(10):1806–1816

Park S, Savvides M (2007) Breaking the limitation of manifold analysis for super-resoluton of face images. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp 573–576

Pear J (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann San Francisco, CA

Philips P, Moon H, Rauss P, Rizvi S (1997) The feret evaluation methodology for face-recognition algorithms. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 137–143

Philips P, Flynn P, Scruggs T, Bowyer K, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of face recognition grand challenge. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 947–954

Phillips P, Moon H, Rauss P, Rizvi S (2000) The feret evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10):1090–1104

Rabiner L (1989) A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2):257–286

Roh M, Lee S (2007) Performance analysis of face recognition alogrithms on korean face database. International Journal of Pattern Recognition and Artificial Intelligence 21(6):1017–1033

Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326

Rowley H, Baluja S, Kanade T (1998) Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1):137–143

Samaria F (1994) Face recognition using hidden markov models. Tech. rep., PhD Thesis, University of Cambridge

Sheikh H, Bovik A (2006) Image information and visual quality. IEEE Transactions on Image Processing 15(2):430–444

Sim T, Baker S, Bsat M (2002) The cmu pose, illumination, and expression (pie) database. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp 46–51

Stephenson T, Chen T (2006) Adaptive markov random fields for example-based super-resolution of faces. EURASIP Jouranl on Applied Signal Processing 2006:1–11

Su C, Zhuang Y, Huang L, Wu F (2005) Steerable pyramid-based face hallucination. Pattern Recognition 38(6):813–824

Sun J, Zheng N, Tao H, Shum H (2003) Image hallucination with primal sketch priors. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 729–736

Tang X, Wang X (2002) Face photo recognition using sketches. In: Proceedings of IEEE International Conference on Image Processing, pp 257–260

Tang X, Wang X (2003) Face sketch synthesis and recognition. In: Proceedings of IEEE International Conference on Computer Vision, pp 687–694

Tang X, Wang X (2004) Face sketch recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(1):1–7

Tanveer M, Iqbal N (2010) A bayesian approach to face hallucination using dlpp and krr. In: Proceedings of International confernece on Pattern Recognition, pp 2154–2157

Tao D, Li X, Wu X, Hu W, Maybank S (2007a) Supervised tensor learning. Knowledge and Information Systems 13(1):1–42

Tao D, Li X, Wu X, Maybank S (2007b) General tensor discriminant analysis and Gabor features for gait recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(10):1700–1715

Tao D, Song M, Li X, Shen J, Sun J, Wu X, Faloutsos C, Maybank S (2008) Bayesian tensor approach for 3-D face modeling. IEEE Transactions on Circuits and Systems for Video Technology 18(10):1397–1410

Tibshirani R (1996) Regression shrinkge and selection via the lasso. Journal of Royal Statistics Society, Series B 58(1):267–288

Tipping M (1991) Sparse bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1:586–591

Turk M, Pentland A (1991) Face recognition using eigenfaces. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 586–591

Vetter T, Troje N (1997) Separation of texture and shape in images of faces for image coding and synthesis. Journal of Optical Society of America 14(9):2152–2161

Wang N, Gao X, Tao D, Li X (2011) Face sketch-photo synthesis under multi-dictionary sparse representation framework. In: Proceedings of International Conference on Image and Graphics, pp 82–87

Wang N, Li J, Tao D, Li X, Gao X (2013a) Heterogeneous image transformation. Pattern Recognition Letters 34(1):77–84

Wang N, Tao D, Gao X, Li X, Li J (2013b) Transductive face sketch-photo synthesis. IEEE Transactions on Neural Networks and Learning Systems 24(9):1–13

Wang S, Zhang L, Liang Y, Pan Q (2012) Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2216–2223

Wang X, Tang X (2003) Face hallucination and recognition. In: Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication, pp 486–494

Wang X, Tang X (2005) Hallucinating face by eigentransformation. IEEE Transactions on Systems, Man, and Cybernetics-Part C 35(3):425–434

Wang X, Tang X (2006) Random sampling for subspace face recognition. International Journal of Computer Vision 70(1):91–104

Wang X, Tang X (2009) Face photo-sketch synthesis and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(11):1955–1967

Wang Z, Bovik A (2002) A universal image quality index. IEEE Signal Processing Letters 9(3):81–84

Wang Z, Bovik A (2004) Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4):600–612

Wang Z, Bovik A (2009) Mean squared error: love it or leave it? a new look at signal fidelity measures. IEEE Signal Processing Magazine 26(1):98–117

Wen F, Luan Q, Liang L, Xu Y, Shum H (2006) Color sketch generation. In: Proceedings of International Symposium on Non-photorealistic animation and rendering, pp 47–54

Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse represnetation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2):210–227

Wright J, Ma Y, Mairal J, Sapiro G, Huang T, Yan S (2010) Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE 98(6):1031–1044

Xiao B, Gao X, Tao D, Li X (2009) A new approach for face recognition by sketches in photos. Signal Processing 89(8):1576–1588

Xiao B, Gao X, Tao D, Yuan Y, Li J (2010) Photo-sketch synthesis and recognition based on subspace learning. Neurocomputing 73(4-6):840–852

Xiong Z, Sun X, Wu F (2009) Image hallucination with feature enhancement. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2704–2081

Yang J, Tang H, Ma Y, Huang T (2008a) Face hallucination via sparse coding. In: Proceedings of IEEE International Conference on Image Processing, pp 1264–1267

Yang J, Wright J, Huang T, Ma Y (2008b) Image super-resolution as sparse representation of raw image patches. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1–8

Yedidia J, Freeman W, Weiss Y (2001) Generalized belief propagation. In: Proceedings of Advances in Neural Information Processing Systems, pp 689–695

Yu J, Liu D, Tao D, Seah H (2012a) On combining multiple features for cartoon character retrieval and clip synthesis. IEEE Transactions on Systems, Man, and Cybernetics, Part B 42(5):1413–1427

Yu J, Wang M, Tao D (2012b) Semisupervised multiview distance metric learning for cartoon synthesis. IEEE Transactions on Image Processing 21(11):4636–4648

Zalesny A, Ferrari V, Caenen G, Gool L (2005) Composite texture synthesis. International Journal of Computer Vision 62(1-2):161–176

Zhang C, Zhang Z (2010) A survey of recent advances in face detection

Zhang D, Zhou Z (2005) $(2d)^2$ pca: 2-directional 2-dimensional pca for efficient face representation and recognition. Neurocomputing 69(1-3):224–231

Zhang J, Wang N, Gao X, Tao D, Li X (2011a) Face sketch-photo synthesis based on support vector regression. In: Proceedings of IEEE International Conference on Image Processing, pp 1149–1152

Zhang L, Zhang L, Mou X, Zhang D (2011b) Fsim: a feature similarity index for image quality assessment. IEEE Transactions on Image Processing 20(8):2378–2386

Zhang T, Tao D, Li X, Yang J (2009) Patch alignment for dimensionality reduction. IEEE Transactions on Knowledge and Data Engineering 21(9):1299–1313

Zhang W, Cham W (2008) Learning-based face hallucination in dct domain. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1–8

Zhang W, Cham W (2011) Hallucinating face in the dct domain. IEEE Transactions on Image Processing 20(10):2769–2779

Zhang W, Wang X, Tang X (2010) Lighting and pose robust face sketch synthesis. In: Proceedings of European Conference on Computer Vision, pp 420–423

Zhang W, Wang X, Tang X (2011c) Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 513–520

Zhang X, Peng S, Jiang J (2008) An adaptive learning method for face hallucination using locality preserving projections. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp 1–8

Zhao W, Chellappa R, Phillips P, Rosenfeld A (2003) Face recognition: a literature survey. ACM Computing Surveys 35(4):399–458

Zhong J, Gao X, Tian C (2007) Face sketch synthesis using e-hmm and selective ensemble. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp 485–488

Zhou H, Kuang Z, Wong K (2012) Markov weight fields for face sketch synthesis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1091–1097

Zhou T, Tao D (2013) Double shrinking sparse dimension reduction. IEEE Transactions on Image Processing 22(1):244–257

Zhuang Y, Zhang J, Wu F (2007) Hallucinating faces: Lph super-resolution and neighbor reconstruction for residue compensation. Pattern Recognition 40(11):3178–3194