Pattern Recognition Letters 34 (2013) 77–84

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Heterogeneous image transformation

Nannan Wang [a,b], Jie Li [a], Dacheng Tao [b], Xuelong Li [c], Xinbo Gao [a,*]

[a] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[b] Faculty of Engineering and Information Technology, University of Technology, Sydney 2007, Australia
[c] 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, China

ARTICLE INFO

Article history:
Available online 21 April 2012

Keywords:
Heterogeneous image transformation
Near infrared image
Sketch-photo synthesis
Sparse representation
Support vector regression

ABSTRACT

Heterogeneous image transformation (HIT) plays an important role in both law enforcements and digital entertainment. Some available popular transformation methods, like locally linear embedding based, usually generate images with lower definition and blurred details mainly due to two defects: (1) these approaches use a fixed number of nearest neighbors (NN) to model the transformation process, i.e., K-NN-based methods; (2) with overlapping areas averaged, the transformed image is approximately equivalent to be filtered by a low pass filter, which filters the high frequency or detail information. These drawbacks reduce the visual quality and the recognition rate across heterogeneous images. In order to overcome these two disadvantages, a two step framework is constructed based on sparse feature selection (SFS) and support vector regression (SVR). In the proposed model, SFS selects nearest neighbors adaptively based on sparse representation to implement an initial transformation, and subsequently the SVR model is applied to estimate the lost high frequency information or detail information. Finally, by linear superimposing these two parts, the ultimate transformed image is obtained. Extensive experiments on both sketch-photo database and near infrared–visible image database illustrates the effectiveness of the proposed heterogeneous image transformation method.

1. Introduction

In real world applications, there are diverse imaging modes and correspondingly different image modalities for different purpose, such as near infrared (NIR) image applied for illumination invariant face recognition (Li et al., 2007), complicated sketch applied for law enforcement or digital entertainment (Wang and Tang, 2009; Gao et al., 2008a), thermal-infrared image applied for life detection, a high resolution image reconstructed from one or more frames of surveillance video which lies in a much lower resolution (Chang et al., 2004; Gao et al., 2012; Zhang et al., 2011c, in press), and so on. We call these different image modalities heterogeneous images.

In some cases, available query image is in a different image modality with the image gallery. Taking NIR images and sketches for example, a photo of the suspect is not always available in many scenarios which can be substituted by a sketch drawn from the cooperation of an artist and eyewitnesses, and mug-shot nevertheless are always in a photo image modality; similar situations are encountered in invariant illumination face recognition, since illumination condition affects face recognition a lot but NIR image is not sensitive to illumination invariant. Recent research on heterogeneous image based face recognition (Tang and Wang, 2004; Gao et al., 2008a; Chen et al., 2009) illustrated that conventional photo-photo matching algorithms (Phillips et al., 2005; Zhao et al., 2003) work poorly due to the great difference in appearance of heterogeneous images. One way to improve the face recognition performance is to transform heterogeneous images into homogeneous ones. Therefore, heterogeneous image transformation is crucial for face recognition or face retrieval. Furthermore, heterogeneous image transformation may favor to the development of other applications, such as animation industry.

Recent research on heterogeneous image transformation focuses on face sketch-photo synthesis and face synthesis between visible images and near infrared images. Zhang et al. (2011a) proposed a coupled information-theoretic encoding algorithm to extract common features to perform face recognition. Klar et al. (2011) transformed photos and forensic sketches into features in the same representation space and then carried out face recognition. Since forensic sketches are drawn according to the description of eyewitnesses rather than a static photo image, it is of great difficulty to design effective approaches. Developments on sketch-photo synthesis could be mainly grouped into two classes: linear methods (Li et al., 2006; Tang and Wang, 2004) and nonlinear methods (Gao et al., 2008b; Liu et al., 2005, 2007; Wang and Tang, 2009; Xiao et al., 2009; Zhang et al., 2010). Tang and Wang proposed the eigen-trans-

* Corresponding author.
E-mail address: xbgao.vipsl@gmail.com (X. Gao).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.04.005

form algorithm utilizing the idea of principal component analysis. Li et al. (2006) extended the eigen-transform to a hybrid space consisting of sketch space and photo space. Since the mapping from sketch to photo is not a simple linear relation, vice versa, these two algorithms perform not very well, especially under the condition of including hair regions. Liu presented a locally linear embedding (LLE) (Roweis and Saul, 2000) based face sketch synthesis approach (Liu et al., 2005), with the idea of locally linear approximating global nonlinear. This method works at patch-level which can depict much more detail information than the global scheme like eigen-transform. Wang and Tang proposed a multi-scale Markov random field based face sketch-photo synthesis approach, with every image as a Markov random field and each image patch as a node (Wang and Tang, 2009). Then they extended this idea to the lighting and pose variant face sketch-photo synthesis. Gao et al. proposed some methods using embedded hidden Markov model and selective ensemble strategy (Gao et al., 2008a,b; Xiao et al., 2009). Chen et al. first used the transformation idea to perform heterogeneous face recognition between visible images and near infrared images (Chen et al., 2009). The above nonlinear face sketch-photo synthesis methods (Gao et al., 2008b; Liu et al., 2005, 2007; Wang and Tang, 2009; Xiao et al., 2009; Zhang et al., 2010) and Chen's method (Chen et al., 2009) share either one or both of the following two defects: (1) A fixed number of nearest neighbors (NN) is used to model the transformation process, i.e., K-NN-based method; (2) With overlapping areas averaged, the transformed image are approximately equivalent to be filtered by a low pass filter, losing the high frequency or detail information. These two drawbacks will result in low definition and blurring effect.

In order to overcome the above two disadvantages correspondingly, a novel two-step framework is proposed. First, SFS is introduced to sparsely select the fewest but closely related neighbors to construct the model (Gao et al., in press). Then, SVR-based image enhancement is to compensate the lost high frequency information.

The rest of this paper is organized as follows. Section 2 describes the SFS model. SVR based image enhancement is presented in Section 3. Experimental results and analysis are given in Section 4 and Section 5 concludes the paper.

2. HIT based on sparse feature selection

2.1. Sparse feature selection (SFS)

Sparse representation of a signal could be described as decomposing a signal into products of an over-completed dictionary and a coefficient vector with few nonzero entries. It can be mathematically formulated as

$$x_{optimal} = \arg\min_x \|x\|_0 \quad \text{s.t.} \|y - Ax\|_2 \leqslant \varepsilon, \tag{1}$$

where $y \in \mathbb{R}^m$ is a signal, $A \in \mathbb{R}^{m\times n}$ is an over-completed dictionary, and $x \in \mathbb{R}^n$ is the coefficient vector. The fidelity term $\|y - Ax\| \leqslant \varepsilon$ constrains the energy of the noise no more than ε. Since the problem in (1) is an NP-hard problem, it is difficult to solve directly. Recent research (Donoho, 2006a,b) claimed that for most underdetermined systems of linear equations, the minimal ℓ^1-norm solution is also the sparest solution. Then Eq. (1) can be transformed into the following optimal equation.

$$x_{optimal} = \arg\min_x \lambda\|x\|_1 + \|y - Ax\|_2. \tag{2}$$

Conventional K-NN method finds K nearest neighbors under Euclidean distance or other distance metrics. One defect of this method is that the number of K nearest neighbors is fixed to a constant which may be not appropriate in some cases in need of adaptively selecting neighbors. This can be easily understood by the following example: if K = 6 and there are actually only 5 nearest neighbors most related to the testing image patch, then K-NN based method may still choose another image patch which is in fact a mismatched patch. For the proposed sparse feature selection (SFS), this would be avoided because it will adaptively determine the number of images patches to minimize the reconstruction error. Due to this fact, this paper introduces a SFS method to find closely related nearest neighbors adaptively based on our previous work (Gao et al., in press). Supposing A is composed of some samples and y is an input query example, by formula (2), we can obtain the corresponding coefficient vector $x_{optimal}$. Then we can obtain the candidate neighbors by selecting those samples whose corresponding coefficient's absolute value is larger than a predefined threshold value σ (σ is a small positive real number). Therefore, the closely related nearest neighbors are found. One should be noticed that σ should be larger than zero in order to cancel out noises raised by small absolute value coefficients near zero and some less related patches. The similar idea has been explored by Ji et al. (2011). However, they preserved all sparse representation coefficients and lack a normalization process while with a threshold σ, we could control a proper number of patches selected to synthesize the output image patch.

2.2. SFS-based HIT algorithm

For heterogeneous image transformation in this paper, y denotes a probe image patch and A consists of columns of training image patches. Here each image patch is concatenated into a column. Subsequently we will use superscript 1 to denote an image modality and superscript 2 to denote corresponding image modality. Given a query image i^1, it is first divided into M even patches with some overlapping, i.e., $i^1 = \{i_1^1, i_2^1, \ldots i_M^1\}$, where i_j^1 means the jth patch of image i^1. Let A^1 denote a dictionary whose columns consist of patches sampled from the image training set I^1 and A^2 indicates the dictionary consisting of the image patches from training set I^2 corresponding to A^1. Then the sparse representation of i_j^1 can be represented as

$$w_j = \arg\min_{w_j} \lambda\|w_j\|_1 + \left\|i_j^1 - A^1 w_j\right\|_2, \tag{3}$$

where $w_j = (w_{j1}, w_{j2}, \ldots w_{jn})^T$ indicates the coefficient vector. We can solve the above optimization problem for w_j. Afterwards the neighborhood of image patch i_j^1 is achieved according to the following criterion.

$$N\left(i_j^1\right) = \{k|\delta(w_{jk}) \neq 0, 1 \leqslant k \leqslant n\}, \tag{4}$$

where $N\left(i_j^1\right)$ denotes the neighborhood for i_j^1 and $\delta(\cdot)$ is a neighbor selection function.

$$\delta(w_{jk}) = \begin{cases} w_{jk}, & |w_{jk}| \geqslant \sigma, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where σ is a small positive real number, which is set to 0.001 in our experiments (it can be also set to a much larger value such as 0.01, 0.05, and so on, but better for not more than 0.1). Once the neighborhood is determined, the weight for every neighbor in $N(i_j^1)$ is calculated as follows.

$$W_{jk} = \delta(w_{jk})/sum(\delta(w_j)), \tag{6}$$

where $sum(\delta(w_j))$ denotes the summation of $\delta(w_{jk})$, $1 \leqslant k \leqslant n$ and W_{jk} is the weight of the kth neighbor to jth input image patch. Finally, the corresponding initial estimate image patch i_j^2 can be synthesized as follows.

$$i_j^2 = A^2 w_j. \tag{7}$$

For each patch in the query image i^1, we iterate the above steps and then fuse all these synthesized patches into a whole image i^2 by averaging the overlapping areas. The SFS-based HIT algorithm is concluded as follows:

SFS-based HIT algorithm

Input: heterogeneous image training pairs (sketch-photo pairs or photo-near infrared image pairs), the query image
Output: the synthesized image corresponding to the input image
Step 1: Divide each image (both training and query images) into M patches with some overlapping;
Step 2: For j = 1:M
(1) Solve the sparse representation problem to obtain the coefficient vector:

$$w_j = \arg\min_{w_j} \lambda \|w_j\|_1 + \left\| i_j^1 - A^1 w_j \right\|_2.$$

(2) Select nearest neighbors according to the following criteria:

$$N(i_j^1) = \{k | \delta(w_{jk}) \neq 0, 1 \leqslant k \leqslant n\}, \quad \text{where } \delta(w_{jk}) = \begin{cases} w_{jk}, & |w_{jk}| \geqslant \sigma, \\ 0, & \text{otherwise}, \end{cases}$$

(3) Normalize the weight vector: $W_{jk} = \delta(w_{jk})/sum(\delta(w_j))$
(4) Synthesize the jth patch: $i_j^2 = A^2 w_j$;
End
Step 3: Fuse all M synthesized patches into a whole image with overlapping regions averaged.

3. HIT based on SFS–SVR

3.1. Support vector regression (SVR)

Support vector machine (Vapnik, 1995) has been successfully used in pattern recognition, and support vector regression (SVR) as one of its variants has been attracted more attentions. It specifically favors to the small-sample size problem with good generalization ability. A brief introduction to SVR (Vapnik, 1995) is given as follows: Consider a set of training data

$$\{(x_i, y_i) | x_i \in \mathbb{R}^n,\ y_i \in \mathbb{R}, \quad i = 1, \ldots, N\}, \tag{8}$$

where x_i is a sample from the input space $\mathbb{R}^n$ and y_i is the corresponding output (y_i is the label of the input sample in classification problem). The object of regression problem is to find a function to minimize the deviation between $f(x)$ and y_i, 1,...,N. Suppose that $f(x)$ takes the following form.

$$f(x) = w \cdot \phi(x) + b, \tag{9}$$

where $w \in \mathbb{R}^n$, $b \in \mathbb{R}$, and ϕ is a nonlinear function to transform x to a high-dimensional space. One has to find the value of w and b to minimize the regression risk error

$$\min E(w, b) = \min \left\{ C \sum_{i=0}^{N} L(y, f(x)) + \frac{1}{2} \|w\|_2 \right\}. \tag{10}$$

$E(w,b)$ is the objective function of SVR model using Lagrange multiplier method. The term $L(y,f(x))$ takes the form

$$L(y, f(x)) = \begin{cases} 0, & |y - f(x)| \leqslant \varepsilon, \\ |y - f(x)| - \varepsilon, & |y - f(x)| > \varepsilon. \end{cases} \tag{11}$$

The optimization problem of Eq. (10) is convex and its optimal solution can be formulated as

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)(\phi(x_i) \cdot \phi(x)) + b = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b, \tag{12}$$

where $K(\cdot)$ is a kernel function, α and α^* are Lagrange multipliers. In all our experiments, $K(\cdot)$ took the form of radial basis function (RBF). For SVR implementation, we use LIBSVM[1] and the default parameters are utilized as in the software.

3.2. SFS–SVR-based HIT algorithm

In Section 2, we have introduced the proposed SFS method which can adaptively select closely related nearest neighbors. Thus SFS method may have some advantages over conventional K-NN-based synthesis algorithm. However, the average process is just like a low-pass filter which filters some high frequency crucial for face recognition and other applications. Thus, in this subsection, we will endeavor to compensate those lost high frequency. In our previous work (Zhang et al., 2011b), we have used SVR to learn the mapping relation of the high frequency information between sketches and photos. And the experimental results illustrated the powerful capacity of SVR. As a result, we will use SVR model to regress the high frequency lost in the SFS synthesis process.

For SVR applied to high frequency synthesis, it consists of two stages: training stage and regression stage as introduced in Section 3.1. In the training stage, the input of the SVR model is the mean value of patch intensity subtracted by patch intensity and the output is the mean value of patch intensity subtracted by the central intensity of the corresponding image patch. In the testing stage, the input is extracted from the input image in the same way as in the training stage and the output is expected to be the lost high frequency. Fig. 1 gives the details about the input and output of the SVR model.

The proposed method combining SFS and SVR is named after SFS–SVR method. The framework is shown in Fig. 2 and SFS–SVR method is summed up as follows:

SFS–SVR algorithm

Input: heterogeneous image training pairs (sketch-photo pairs or photo-near infrared image pairs), the query image
Output: the synthesized image corresponding to the input image
Phase 1 Generate the initial estimate i_l^2 (the subscript l represents low and mid frequency)
Use the algorithm described in Section 2.2 to generate an initial estimate i_l^2;
Phase 2 Synthesize the high frequency
Step 1: Extract features both from training image patches and query image patches;
Step 2: Training.
Taking all extracted features to the SVR training function:

$$\min E(w, b) = \min \left\{ C \sum_{i=0}^{N} L(y, f(x)) + \frac{1}{2} \|w\|_2 \right\},$$

$L(y,f(x))$ takes the form $L(y, f(x)) = \begin{cases} 0, & |y - f(x)| \leqslant \varepsilon \\ |y - f(x)| - \varepsilon, & |y - f(x)| > \varepsilon \end{cases}$. The final regression function can be written as $f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b$ where all parameters in this function can be solved using training features.
Step 3: For j = 1: M
Input jth patch feature into the regression function to obtain the output value:

(*continued on next page*)

[1] Chang C. and Lin C., LIBSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

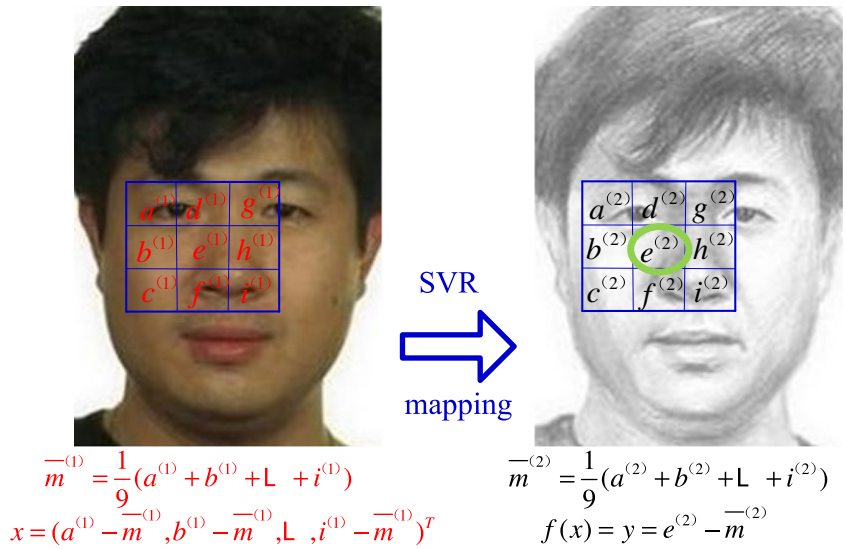

Fig. 1. Input and output of SVR model: here the patch size is 3 × 3 as an example. The intensity of each patch is shown and denoted by $a^{(t)}, b^{(t)}, \ldots, i^{(t)}$ (t = 1 or 2). It should be noticed that the superscript 1 and 2 in the patches represents the image modality.

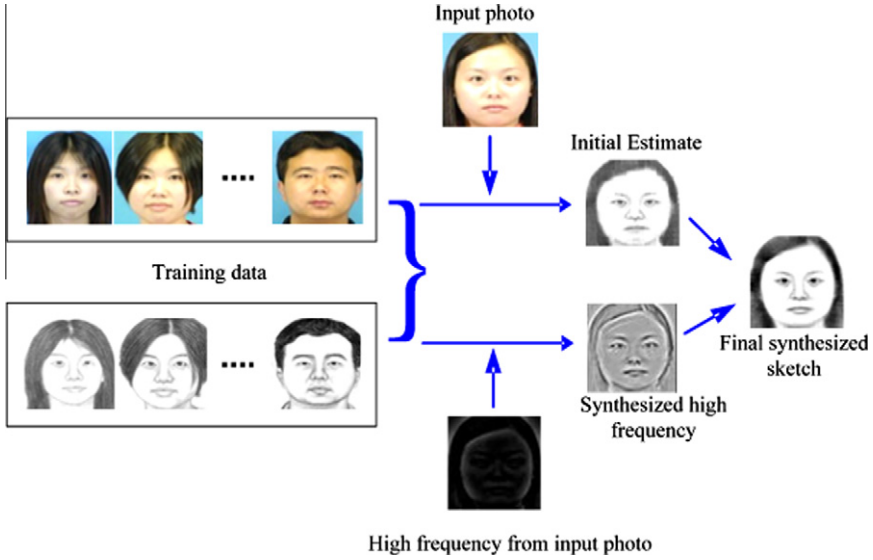

Fig. 2. Framework of proposed SFS–SVR method.

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b,$$

End

Step 4: Fuse M high frequency patch into a whole one i_h^2 (the subscript h denotes high frequency)

The final output $i = i_h^2 + i_h^2$

4. Experimental results and analysis

We first introduce several databases used in our experiments: CUHK student sketch-photo database (Tang and Wang, 2004; Wang and Tang, 2009), VIPSL sketch-photo database (Gao et al., 2011), visible–NIR image database. The CUHK student sketch-photo database is constructed by Multimedia Lab of the Chinese University of Hong Kong, consisting of 188 sketch-photo pairs. The photos in CUHK student database are all taken in neutral expression, frontal pose, from the same race and have the same skin color. The VIPSL database is constructed by Video & Image Processing System Laboratory of Xidian University, consisting of 200 photos and 1000 sketches drawn by 5 different artists with each photo corresponding to 5 sketches. The photos therein come from different benchmark face databases: FERET (Phillips et al., 1998, 2000), FRAV2D (Serrano et al., 2007), Indian Face database.[2] The associated people are from different area of the world and have different skin colors, so the VIPSL database is of much more challenge than the CUHK student database. In our experiments, we only chose 200 sketches drawn by the same artist and corresponding 200 photos. The visible–NIR image database is composed of 100 visible–NIR image pairs. Some experimental examples of these three databases are shown in Fig. 3.

[2] http://vis-:www.cs.umass.edu/~vidit/IndianFaceDatabase/2002.

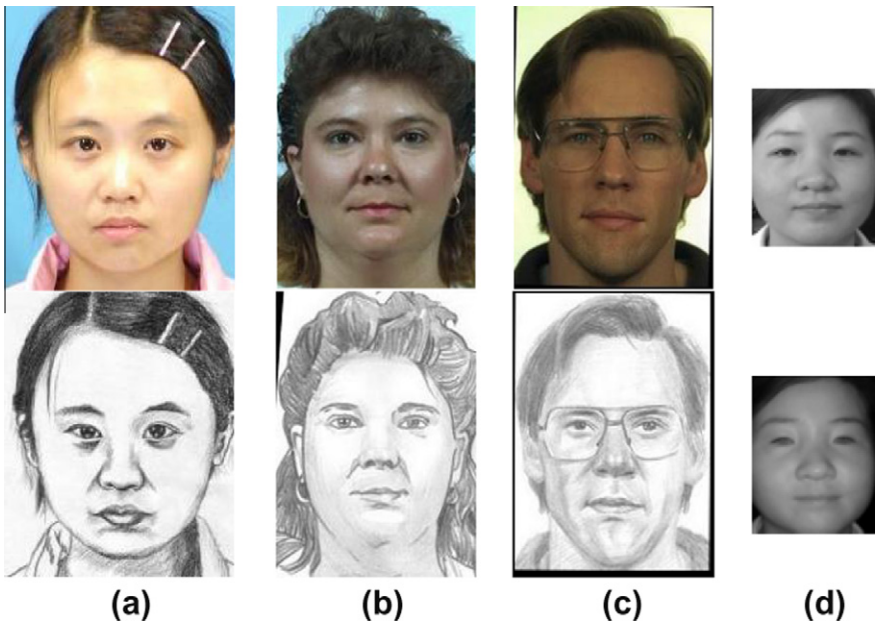

Fig. 3. Some heterogeneous image examples from different databases: (a) sketch-photo pairs from CUHK student database; (b,c) two sketch-photo pairs from VIPSL database; (d) visible–NIR image pairs.

4.1. Image transformation

In this subsection, we will compare several state-of-the-art approaches (Liu et al., 2005; Gao et al., 2008; Xiao et al., 2009; Chen et al., 2009) with the proposed method. For sketch-photo synthesis, all sketches and photos are cropped into the size of 163×200 and the patch size of 36×36 with 28×28 area overlapping. For visible and near infrared images, the image size, patch size and overlapping size are 64×80, 16×16, and 12×12 respectively. For the CUHK student database, 88 sketch-photo pairs are used as the training set and the other 100 pairs as the testing set. For the VIPSL database, 100 pairs serve as the training set and the testing set each. For visible–NIR image database, we utilize the leave-one-out strategy since there are only 95 image pairs. For methods in comparison,

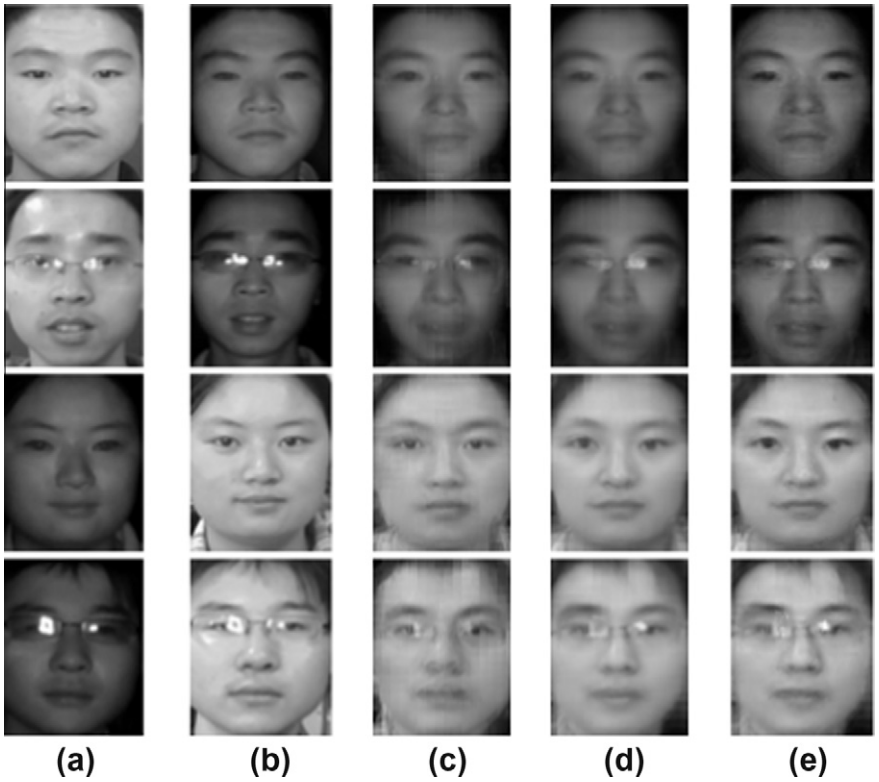

Fig. 4. Synthesized visible images and near infrared images. (a) Input images; (b) ground truth corresponding to input images; (c) synthesized images using Chen's methods (2009); (d) synthesized images using the proposed SFS method; (e) synthesized images using the proposed SFS–SVR method.

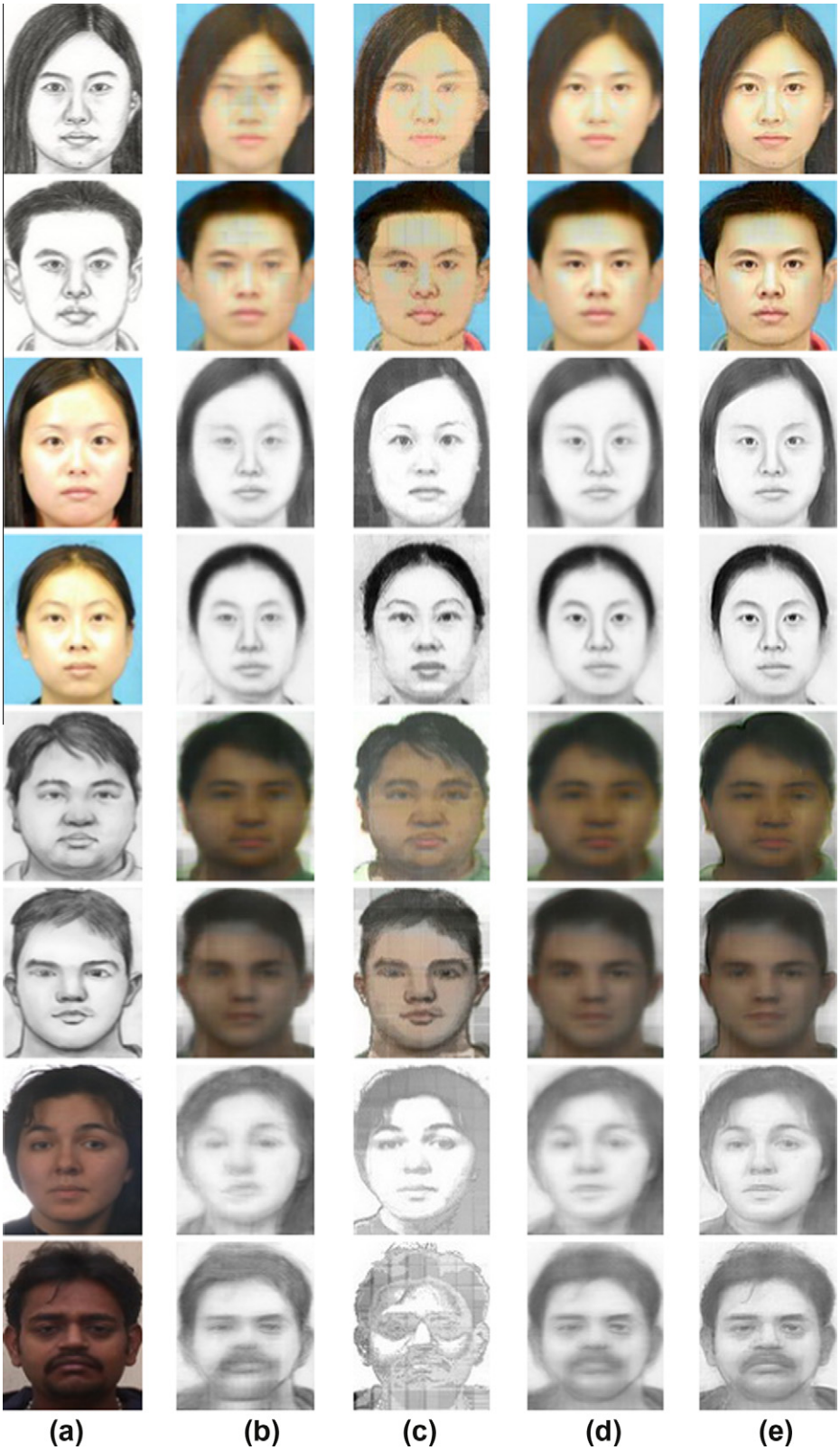

Fig. 5. Synthesized sketches and photos. (a) input images; (b) synthesized images using Liu et al.'s method (2005); (c) synthesized sketches using Gao et al.'s method (2008) and Synthesized photos using Xiao et al.'s method (2009); (d) synthesized images using the proposed SFS method; (e) synthesized images using the proposed SFS–SVR method.

the parameters are set as those in the corresponding literatures. In order to reduce the time cost in testing phase, we utilize clustering strategy before implementing SFS and SVR based image enhancement. The *K*-means clustering method is used to cluster image patches in SFS stage into 20 categories and cluster image feature patches in SVR based image enhancement stage into 25 categories. Then for the training phase, 20 SFS model and 25 SVR model are constructed in corresponding category respectively. For testing

Table 1
VIF values of corresponding category of synthesized images.

Synthesized images	Chen's method (2009)	Liu's method (2005)	Gao's method (2008)	Xiao's method (2009)	The proposed SFS method	The proposed SFS–SVR method
CUHK-sketch	–	0.0921	0.0948	–	0.0956	**0.0972**
CUHK-photo	–	0.1224	–	0.1306	0.1438	**0.1447**
VIPSL-sketch	–	0.0547	0.0594	–	0.0584	**0.0657**
VIPSL-photo	–	0.0881	–	0.1471	0.1538	**0.1746**
Synthesized visible	0.1515	–	–	–	0.1714	**0.1833**
Synthesized NIR	0.2076	–	–	–	0.2247	**0.2373**

It should be noticed that "–" denotes that the method is not used to synthesize corresponding images.

Table 2
Face recognition results on CUHK student and VIPSL database (Eigenface performance is in the parenthesis).

Recognition rate (%)	LLE	EHMM	SFS	SFS–SVR
CUHK sketch	100 (84)	100 (87)	100 (91)	**100** (**93**)
VIPSL sketch	91 (47)	93 (47)	96 (42)	**98** (**51**)
CUHK synthesized photos	100 (88)	100 (90)	100 (91)	**100** (**95**)
VIPSL synthesized photos	86 (26)	90 (25)	91 (23)	**95** (**28**)

Table 3
Face recognition results on visible–NIR image database (Eigenface performance is in the parenthesis).

Recognition rate (%)	LLE	SFS	SFS–SVR
Synthesized visible image	49.47 (31.58)	58.95 (27.37)	**64.21** (**32.63**)
NIR image	53.68 (36.84)	61.05 (34.74)	**69.47** (**42.11**)

phase, the nearest category could be found in Euclidian distance for each testing image patch or image feature patch. Then the standard SFS process or SVR process can be implemented in this category. Some synthesized images are shown in Figs. 4 and 5.

From Figs. 4 and 5, it can be found that the proposed SFS method achieve better results than most available approaches and the enhanced SFS, SFS–SVR, can obtain even better performances from a perceptual manner. The above methods in comparison used fixed number of nearest neighbors to estimate the synthesized image patch corresponding to the input image patch and then averaged the overlapping area, which led to blurring effect and bringing in some noise. The proposed SFS method adaptively selects the nearest neighbors based on sparse representation and the proposed image enhancement strategy based on SVR compensates the lost high frequency information due to average.

4.2. Synthesized image quality assessment

In Section 4.1, we have illustrated the proposed method performed better than other approaches from the subjective view. This subsection will further illustrate the effectiveness of the proposed method from the objective perspective using visual information fidelity (VIF) (Sheikh and Bovik, 2006). VIF is one of the most effective full reference image quality assessment metrics that quantifies the information that is presented in the reference image and how much of this reference information can be extracted from the distorted image. In this paper, the reference image is the original images either drawn by artist or taken by near infrared imaging system or by generic color imaging system. The synthesized images can be viewed as distorted images. The larger the VIF value, the higher the synthesized image quality. Table 1 shows the mean VIF values of corresponding category of synthesized images. The largest scores are marked in bold values and the similar meaning for bold values in the following tables.

From Table 1, though the value is small overall, the proposed two methods scored larger than other state-of-the-art methods. Specially, the proposed SFS–SVR achieved higher score than the proposed SFS method, claiming that the image enhancement based on SVR is effective.

4.3. Face recognition

Since face recognition is one of the most important applications of heterogeneous image transformation, we can evaluate the synthesized image quality in terms of the pattern recognition rate. Recently, Wright et al., 2009 has proposed a sparse representation classification algorithm (SRC) which worked well under well-controlled condition (frontal pose, neutral expression) (2009). Here, we will use this algorithm to illustrate the superiority of the proposed method. Furthermore, in order to illustrate that not only the sparse representation classification but also the proposed transformation method favors to the performance of synthesized face image recognition, we also employ the eigenface method (Turk and Pentland, 1991) to conduct face recognition on the visible–NIR database and the experimental results are listed in the parenthesis in Tables 2 and 3. The results are shown in Tables 2 and 3. In these two tables, "LLE" denotes Liu's method (2005) for sketch-photo synthesis and Chen's method (2009) for visible–NIR image transformation, and "EHMM" represents Gao's method (2008) for sketch synthesis and Xiao's method for photo synthesis (2009).

From Table 2, we can see that the face recognition rate on CUHK student database is as high as 100 percent using SRC method. This is due to the fact that the sketches and photos in this database are simple in structure, i.e., all these photos are taken in the same background and all people are yellow skin color. Therefore, it is easy to learn the mapping function between them. However, since VIPSL database is composed of photos taken in several different backgrounds and in different skin colors, it is more difficult to perform face recognition than CUHK student database. Nevertheless, the proposed SFS–SVR method can reach 95% for synthesized photo based and 98% for sketch based face recognition, exceeding current state-of-the-art method. In Chen's paper (2009), the face recognition rate can achieve a rate as high as 94.2% for homogeneous illumination because there are 6 images for each person in database yet only one image in our experiment. Though the recognition rate for visible and NIR image is low overall, the superiority of the proposed SFS and SFS–SVR method can be seen from Table 3. From the Tables 2 and 3, it can be found that the SRC performs much better than eigenface while eigenface can still obtain consistent tendency with SRC's. This further verified the effectiveness of the proposed heterogeneous image transformation method.

5. Conclusions

In this paper, we claimed that there existed two disadvantages for most available heterogeneous image transformation methods: (1) the number of nearest neighbors is fixed which incurs blurring

effect or brings in noise; (2) some important detail information or high frequency information loses due to average of overlapping areas. We correspondingly proposed two strategies to overcome or improve these two defects: SFS and SVR based image enhancement. Extensive experimental results reveal a remarkable improvement over LLE-based methods and other competing approaches. In future, we will discuss the effect of sketches drawn by different artists and the heterogeneous image transformation problem under different poses and views.

Acknowledgements

The authors would like to thank the helpful comments and suggestions from the anonymous reviewers. The authors would also like to thank Professor Stan Z. Li of Institute of Automation, Chinese Academy of Sciences for providing the visible-NIR face database. This research was supported partially by the National Basic Research Program of China (973 Program) (Grant No. 2012CB316400), the National Natural Science Foundation of China (Grant Nos. 61125204, 61125106, 91120302, 61172146 and 60832005), the Fundamental Research Funds for the Central Universities, Open Project Program of the State Key Lab of CAD&CG, Zhejiang University (Grant No. A1006), the PhD Programs Foundation of Ministry of Education of China (Grant No. 20090203110002), and the State Administration of STIND (B1320110042).

References

Chang, H., Yeung, D., Xiong, Y., 2004. Super-resolution through neighbor embedding. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 947–954.

Chen, J., Yi, D., Yang, J., Zhao, G., Li, S., Pietikainen, M., 2009. Learning mappings for face synthesis from near infrared to visual light images. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 156–163.

Donoho, D., 2006a. For most large underdetermined systems of linear equations, the minimal ℓ^1-norm solution is also the sparsest solution. Comm. Pure Appl. Math. 59 (6), 797–829.

Donoho, D., 2006b. For most large underdetermined systems of linear equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution. Comm. Pure Appl. Math. 59 (7), 907–934.

Gao, X., Zhong, J., Li, J., Tian, C., 2008a. Face sketch synthesis algorithm based on E-HMM and selective ensemble. IEEE Trans. Circuits Syst. Video Technol. 18 (4), 487–496.

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

Gao, X., Wang, N., Tao, D., Li, X., in press. Face sketch-photo synthesis and retrieval using sparse representation. In: IEEE Trans. Circuits Syst. Video Technol.

Gao, X., Zhang, K., Tao, D., Li, X., 2012. Joint learning for single image super-resolution via a coupled constraint. IEEE Trans. Image Process. 21 (2), 469–480.

Ji, N., Chai, X., Shan, S., Chen, X., 2011. Local regression model for automatic face sketch generation. In: Proc. Internat. Conf. on Image and Graphics, pp. 412–417.

Klar, B., Li, Z., Jain, A., 2011. Matching forensic sketches to mug shot photos. IEEE Trans. Pattern Anal. Machine Intell. 33 (3), 639–646.

Li, Y., Savvides, M., Bhagavatula, V., 2006. Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters. In: Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, pp. 357–360.

Li, S., Chu, R., Liao, S., Zhang, L., 2007. Illumination invariant face recognition using near-infrared images. IEEE Trans. Pattern Anal. Machine Intell. 29 (4), 627–639.

Liu, Q., Tang, X., Jin, H., Liu, J., 2005. A nonlinear approach for face sketch synthesis and recognition. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1005–1010.

Liu, W., Tang, X., Liu, J., 2007. Bayesian tensor inference for sketch-based facial photo hallucination. In: Proc. Internat. Joint Conf. on Artificial Intelligence, pp. 2141–2146.

Phillips, P., Wechsler, H., Huang, J., Rauss, P., 1998. The FERET database and evaluation procedure for face recognition algorithms. Image Vision Comput. 16 (5), 295–306.

Phillips, P., Moon, H., Rizvi, S., Rauss, P., 2000. The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Machine Intell. 22 (10), 1090–1104.

Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W., 2005. Overview of the face recognition grand challenge. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 947–954.

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

Serrano, I. Diego, Conde, C., Cabello, E., Shen, L., Bai, L., 2007. Influence of wavelet frequency and orientation in an SVM-based parallel Gabor PCA face verification system. In: Proc. Conf. on Intelligent Data Engineering and Automated Learning, pp. 219–228.

Sheikh, H., Bovik, A., 2006. Image information and visual quality. IEEE Trans. Pattern Anal. Machine Intell. 15 (2), 430–444.

Tang, X., Wang, X., 2004. Face sketch recognition. IEEE Trans. Circuit Syst. Video Technol. 14 (1), 50–57.

Turk, M., Pentland, A., 1991, Face Recognition Using Eigenfaces. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586–591.

Wang, X., Tang, X., 2009. Face photo-sketch synthesis and recognition. IEEE Trans. Pattern Anal. Machine Intell. 31 (11), 1955–1967.

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

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag, New York.

Xiao, B., Gao, X., Li, X., Tao, D., 2009. A new approach for face recognition by sketches in photos. Signal Process. 89 (8), 1531–1539.

Zhang, W., Wang, X., Tang, X., 2010. Lighting and pose robust face sketch synthesis. In: Proc. European Conf. on Computer Vision (6), pp. 420–433.

Zhang, W., Wang, X., Tang, X., 2011a. Coupled information-theoretic encoding for face photo-sketch recognition. In: Proc. IEEE Internat. Conf. on Computer Vision and Pattern Recognition, pp. 513–520.

Zhang, J., Wang, N., Gao, X., Tao, D., Li, X., 2011b. Face sketch-photo synthesis based on support vector regression. In: Proc. Internat. Conf. on Image Processing, pp. 1149–1152.

Zhang, K., Gao, X., Tao, D., Li, X., 2011c. Partially supervised neighbor embedding for example-based image super-resolution. IEEE J. Selected Topics Signal Process. 5 (2), 230–239.

Zhang, K., Gao, X., Tao, D., Li, X., in press. Multi-scale dictionary for single image super-resolution. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition.

Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P., 2003. Face recognition: A literature survey. ACM Comput. Survey 34 (4), 399–458.