Detail-Enhanced Cross-Modality Face Synthesis via Guided Image Filtering

Yunqi Dang, Feng Li, Zhaoxin Li, and Wangmeng Zuo(✉)

Computational Perception and Cognition Center, School of Computer Science and Technology,
Harbin Institute of Technology, Harbin 150001, China
dyqhitcs@126.com, {fengli_hit,cszli}@hotmail.com,
cswmzuo@gmail.com

Abstract. Face images in different modalities are often encountered in many applications, such as face image in photo and sketch style, visible light and near-infrared style. As an active yet challenging task, cross-modality face synthesis aims to transform face images between modalities. Many existing methods successfully recover global features for a given photo, however, fail to capture fine-scale details in the synthesis results. In this paper, we propose a two-step algorithm to tackle this problem. Firstly, KNN is used to select the *K* most similar patches in training set for an input patch centered on each pixel. Then combination of patches is calculated for initial results. In the second step, guided image filtering is used on initial results with test photo as guidance. Fine-scale details can be transferred to the results via local linear transformation. Comparison experiments on public datasets demonstrated the proposed method is superior to the state-of-the-art method in simultaneously keeping global features and enhancing fine-scale details.

Keywords: Photo-sketch synthesis · Guided image filtering · KNN · Local linear transformation

1 Introduction

In many cases, we can obtain face image pairs of the same person in different modalities, such as face image in photo or sketch style, visible light (VIS) or near-infrared (NIR) style, etc. For example, since there are difficulties in obtaining photos of the criminal suspects, sketches of suspects are usually drawn by artists to hunt them. However, drawing face sketches is both time consuming and restricted by painting level of artists. Face images under NIR are on good condition and unaffected by visible lights in the environment. So, face recognition [1] using NIR images is contributive. Thus, automatic cross-modality face synthesis plays an important role in law enforcement. Besides, face sketch can also be applied to digital entertainment [2,3,4]. Studies on cross-modality face synthesis problem has been carried out for several years. And a number of algorithms had been proposed. Among these existing approaches, there are several representative ones. Linear subspace learning-based approaches [5,6,7] are based on the assumption that each output patch can be generated

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 200–209, 2015.
DOI: 10.1007/978-3-662-48558-3_20

by using a linear combination of the selected K nearest neighbors. But the synthesis results tend to be over-smoothed and lose some details. Sparse representation-based method is also an important branch [8,9]. Image patches could be sparsely represented by an over-complete dictionary of atoms. Although it is effective to use sparse coding and dictionary learning to address this problem, it needs excessive time to learn the dictionary and mapping between dictionaries in different modalities.

In this paper, we propose a detail-enhanced approach for cross-modality face synthesis. Our method is composed of two steps. In the first step, we adopt KNN algorithm [10] to select the K most similar patches in training dataset for input patch centered on each pixel of input photo. Then combination of the K patches is calculated for generating the initial synthesis result. Second, we recover some fine-scale details of facial features on initial result by introducing guided image filtering [11]. Input test photo image is regarded as the guidance image, and initial synthesis result constructed by the KNN algorithm is input image. With the help of guidance image, we can keep global features while enhancing fine-scale details.

The remainder of this paper is organized as follows. In Section 2, we review the related work on cross-modality face synthesis, especially photo-sketch synthesis. Section 3 describes the proposed method, including the K-NN-based step and the guided image filtering-based step. Section 4 presents our experiments results. Finally, some conclusion remarks are presented in Section 5.

2 Related Work

In this section, we briefly review several work on recent cross-modality face synthesis. By far, most studies focused on face sketch synthesis, while NIR face synthesis received relatively less attention. The first kind of method is forward method. Tang et al. [5] presented a photo-to-sketch transformation method based on eigentransform by exploiting PCA. And their approach is based on the following assumption: the process of photo-to-sketch can be approximated as linear. However, this assumption is too strong because the mapping between photo and sketch may be nonlinear. Inspired by LLE (locally linear embedding), Liu et al.[12] proposed a nonlinear approach based on local linear preservation of geometry between photo and sketch and achieved better results.

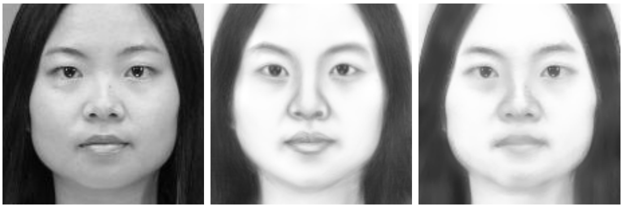
Another kind of method is based on MRF [13], which characterizes the pixel-wise dependency and can generate smooth results. For example, Tang et al. [14] proposed a multi-scale MRF model for face photo-sketch synthesis and recognition. And they also applied face photo-sketch synthesis method to face recognition [15]. Zhou et al. [16] presented a weighted Markov random fields method to overcome the drawbacks that MRF-based method [14] cannot synthesize new sketch patches.

Besides, the sparse representation-based approaches also play an important role in accomplishing image reconstruction. A series of work for photo-sketch synthesis based on sparse representation have been proposed in [8], [17], and [18]. Yang et al. [17] proposed a two-step method based on sparse coding to address the problem of FSR (face super-resolution). Their model was then utilized by Chang et al. [8] for face photo-sketch synthesis. In [18], the multi-dictionary sparse representation model is used to generate details which is lack in the initial sketch by using LLE. While different from the above methods, Wang et al. [19] assumed that the two sparse representations are connected through a linear transformation rather than the same representations. The objective function was separated into three sub-problems: sparse coding for training samples, dictionary updating and linear transformation matrix updating.

For efficient implementation, Song et al. [20] proposed a real-time approach for face sketch synthesis, and the method can be extended to temporal domain. However, the experimental results also tend to lack of some detailed information.

3 Cross-Modality Face Synthesis

In this section, we first present the KNN-based algorithm (baseline approach) and then propose an enhanced algorithm for face sketch synthesis. In [12], Liu et al. proposed a pseudo-sketch synthesis method based on the assumption that corresponding photo and sketch image patches are similar in geometrical structure. And in [20], Song et al. also utilized this model to their method. In order to address the problem of noisy sketch, they proposed a denoising approach called Spatial Sketch Denoising (SSD). However, the results are lack of some fine-scale details such as eyebrows, eyes, and nose. (see Fig. 1).



(a) Input (b) SSD (c) Proposed

Fig. 1. Cross-modality face synthesis results by using SSD method [20] and the proposed method. Comparing with our result, the result of SSD method lacks of some fine-scale details.

3.1 K-NN-Based Algorithm

In the first step of the proposed method, KNN [12] is used for generating the initial sketch. The KNN-based method consists of two stages, namely KNN search and linear combination of the result patches. This method is selected as baseline.

First, the training face photo images and sketch images are both divided into patches for searching the K most similar patches. Given a test face photo, we also divide it into patches. For an image patch centered at a pixel in test photo image, we collect N patches around that patch from each training photo image. Then, its K most similar patches can be searched from the corresponding N sketch patches and the reconstruction coefficients of the linear mapping are calculated by using a conjugate gradient solver [21]. Second, a linear combination of the K most similar sketch patches is used to generate the initial sketch image.

We denote the $\mathbf{T}_\mathrm{p}$ as a patch centered at a pixel p in the test photo and $\mathbf{S}_\mathrm{p}$ as the estimated corresponding sketch patch. The linear combination of searched the K most similar patches from training sketch patch is formulated as,

$$\mathbf{S}_\mathrm{p} = \sum_{k=1}^{K} w_p^k \, \mathbf{I}_\mathrm{p}^\mathrm{k}, \tag{1}$$

where $\mathbf{I}_\mathrm{p}^\mathrm{k}$ is one of the K most similar sketch image patches selected by using KNN search, and w_p^k is the calculated coefficient. The KNN method is summarized in Algorithm-1.

Algorithm-1. KNN-based initial synthesis

```
Input: Training photo image $X$ and sketch image $Y$, test face photo
       $I$, patch size $s$, search radius $r$, number of candidates $K$
1. Divide $X$, $Y$ and $I$ into patches respectively
2. Search for the $K$ most similar patches
3. Calculate the linear coefficients $w_p^k$ in Eq. 1.
4. Get the initial sketch image $I'$ by Eq. 1.
Output: Initial sketch $I'$
```

3.2 Guided Image Filtering-Based Algorithm

The sketch image generated by using KNN-based synthesis is over-smoothed and lack of fine-scale details. To enhance more fine-scale details on the sketch images, in this section, we set sketch image generated by KNN-based synthesis as an initial sketch image, and introduce the guided image filtering to enhance the details based on the original gray-scale image as a guidance image.

Guided Image Filtering. Guided filtering proposed by He et al. [11] is an efficient image filtering method, whose output is based on local linear transform of guidance image. Guidance image can be the input image or another different image. With the help of guidance image, filtering output image can fully obtain details of the guidance image, meanwhile, keep overall characteristics of the input image. Guided image filtering is both effective and efficient in a flurry of applications, such as edge-aware smoothing, detail enhancement, etc.

Denote I as the guidance image, q as output. In [11], the filter output in pixel i is represented by the following formula,

$$q_i = a_k I_i + b_k, \forall i \in \omega_k, \tag{2}$$

where ω_k is a window centered at the pixel k, and the a_k and b_k are the linear coefficients assumed to be constant in ω_k. And a square window of radius r is used. To calculate the linear coefficients a_k and b_k, they define the mapping as the following,

$$q_i = p_i - n_i, \tag{3}$$

where q_i is the output, p_i is the input, and n_i is noise. Therefore, a minimization function can be defined,

$$E(a_k, b_k) = \sum_{i\in\omega_k}((a_k I_i + b_k - p_i)^2 + \epsilon a_k^2), \tag{4}$$

where ϵ is the regularization parameter. The Equation (4) is the linear ridge regression model [22], and a_k and b_k can be calculated by,

$$a_k = \frac{\frac{1}{|\omega|}\sum_{i\in\omega_k} I_i p_i - \mu_k \overline{p_k}}{\sigma_k^2 + \epsilon}, \tag{5}$$

$$b_k = \overline{p_k} - a_k \mu_k, \tag{6}$$

where μ_k and σ_k^2 are the mean and variance of I in the window ω_k, $|\omega|$ is the number of pixels in ω_k. For the reason that one pixel can be calculated repeatedly in all window which it is in, an average values of q_i should be computed. So, the Equation (2) is reformulated as,

$$q_i = \frac{1}{|\omega|}\sum_{k|i\in\omega_k}(a_k I_i + b_k). \tag{7}$$

After computing a_k and b_k for all windows w_k, the linear function can be reformulated as,

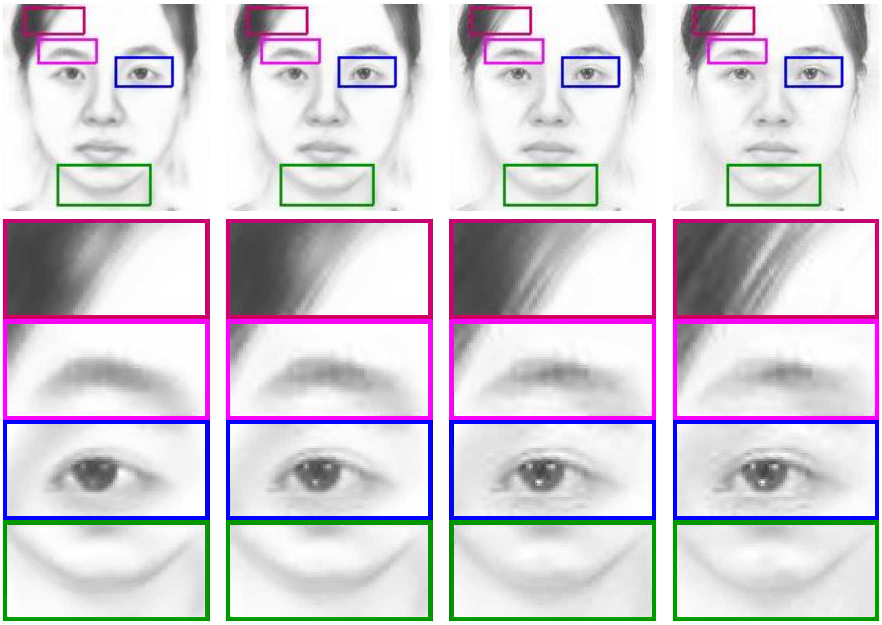
$$q_i = \overline{a_i} I_i + \overline{b_i}, \tag{8}$$

where $\overline{a_i} = \frac{1}{|w|}\sum_{k\in w_i} a_k$ and $\overline{b_i} = \frac{1}{|w|}\sum_{k\in w_i} b_k$ are the average coefficients of all windows overlapping i.

Algorithm-2. Guided filter-based enhancement

```
Input: Initial sketch $I'$, test face photo $I$, window radius $r'$,
       regularization $\epsilon$
1. $mean_I = f_{mean}(I)$
   $mean_{I'} = f_{mean}(I')$
   $corr_I = f_{mean}(I.* I)$
   $corr_{II'} = f_{mean}(I.* I')$
2. $var_I = corr_I - mean_I.* mean_I$
   $cov_{II'} = corr_{II'} - mean_I.* mean_{I'}$
3. $a = cov_{II'}./(var_I + \epsilon)$
4. $b = mean_{I'} - a.* mean_I$
5. $mean_a = f_{mean}(a)$
   $mean_b = f_{mean}(b)$
6. $I'' = mean_a.* I + mean_b$
Output: Final sketch $I''$
```

Enhancement with Guided Image Filtering. To utilize its detail-enhancing capability, guided filtering is adopted to refine the output of KNN algorithm for the



(a) $r' = 1$ (b) $r' = 2$ (c) $r' = 4$ (d) $r' = 8$

Fig. 2. Photo-sketch synthesis results of the proposed method with different r' values.

synthesis of final face sketch image. More concretely, we at first obtain the initial sketch image estimate via KNN in Algorithm-1, then let I denote the test face image as guidance image, we can utilize the Eq. (8) to filter the initial sketch. We summarize the proposed method in Algorithm-2, where f_{mean} is a mean filter. The proposed approach involves two aspects: the KNN approach and the refinement algorithm using guided image filtering. It should be noted that our approach uses the test face photo as guidance for filtering, and the initial sketch generated by KNN as input of guided image filtering. While [20] selects generated initial sketch to guide the test face photo, this makes the results more similar with the face photo.

4 Experimental Results

In this section, we evaluate the proposed algorithms. The KNN algorithm [12] is set as the baseline. Firstly, different parameter settings are tested for evaluating our proposed approach. Then, we compare the results with the baseline algorithm, MRF method [14] and the SSD method [20] on the benchmark datasets. Further, we demonstrate that the proposed method can also be used in other applications such as VIS-NIR image synthesis, and the results are showed.

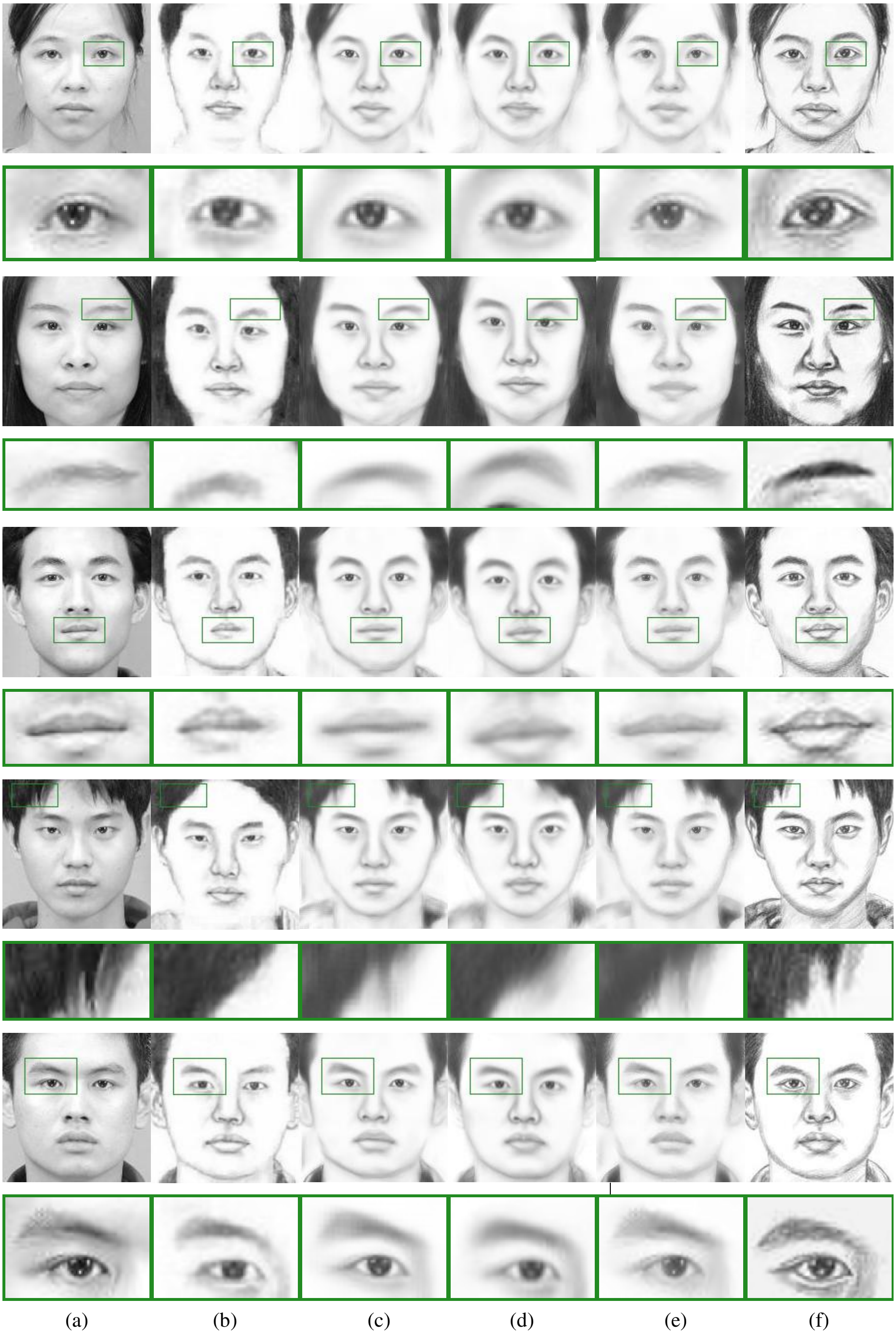


Fig. 3. Photo-sketch synthesis results. (a) photos; (b) results by MRF [14]; (c) results by K-NN search; (d) results by SSD [20]; (e) results by our method; (e) sketches drawn by artists.

4.1 Evaluation on the CUHK Benchmark Dataset

We evaluated our proposed algorithm on CUHK student dataset [14]. The dataset consists of 188 pairs of photo-sketch pair images, in which 88 pairs are used for training and the rest of images are used for testing. In our experiments, the search radius r is set to be 8 pixels around each pixel. K, the number of candidate patches, is set to be 10 and the patch size s is 5×5. The radius of window r' is set to be 2, and the regularization parameter ϵ is set to 0.5 during processing of the guided image filtering. Our algorithm in experiments is coded in MATLAB and run on the computer with Intel(R) Xeon(R) CPU E3-1230 v3 @ 3.30GHz.

Fig. 2 shows synthesis results with different radius of window in guided image filtering. The larger the value of r', the more details kept from the guidance image. However, if value of r' is too big, synthesis results would be too similar with photo. In order to maintain sketch-style while retrieving some details, we adopt $r' = 2$ for guided image filtering in our following experiments. Fig. 3 shows the synthesis results and demonstrates that our method is better on enhancing fine-scale detail information compared with the KNN algorithms, MRF [14] and SSD method [20], e.g. hair, eyes, and other facial parts.

Further, we compared running time of the proposed method with the other two approaches. Among these methods, SSD is the most efficient method in generating sketches on good condition, which takes 4.97 seconds in average for synthesizing one sketch. The average runtime of MRF method is about 103.84 seconds. And for our method, the average running time is 57.46 seconds in KNN, and it is 0.007 seconds in the process of guided image filtering. Table 1 shows the comparison of three methods in runtime. It is worth noting that the algorithm of SSD is implemented with C++ and our program is unoptimized in MATLAB.

Table 1. Speed (*sec.*) comparison of three method

Method	MRF	SSD	Proposed
Time	103.84	4.97	57.47

4.2 Visible-to-Near-Infrared Face Synthesis

Our model can also be used in other applications such as VIS-NIR face synthesis. The experiment was conducted on the CASIA NIR database [1] which contains 400 pairs of faces. 280 pairs are used for training and others are used for testing. Each pairs of faces has one face photo image and one corresponding NIR image. In this experiment, we set the search radius r to 8 pixels around each pixel, the number of candidates K is 10 and the patch size s is 5×5 for baseline method. The radius of window r' is set to be 5, and the regularization parameter ϵ is 0.5 for guided image filtering. Synthesis results are shown in Fig. 4. From the experimental results, we can see that our method can effectively address the problem of losing details and has much less artifacts than the baseline.

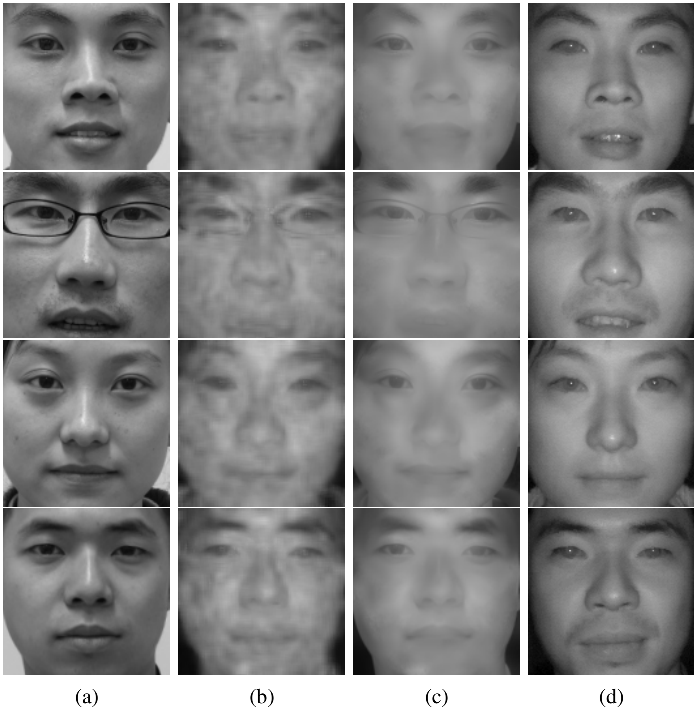


Fig. 4. VIS to NIR face synthesis results. (a) visible lights photos; (b) initial results using baseline method; (c)results by the proposed method; (d) ground truth.

5 Conclusions

In this paper, we propose a detail-enhanced cross-modality face synthesis approach. In order to solve the over-smoothed problem of face synthesis, we proposed a two-step algorithm which composes of a KNN-based method for initialization and the guided image filtering-based method for further refinement. In particular, guided image filtering-based method can enhance fine-scale details from initial synthesis results of KNN algorithm. Experimental results validate the superiority of our method comparing with the state-of-the-art on face sketch synthesis. Moreover, our proposed method can also be applied to VIS-NIR face synthesis.

References

1. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(4), 627–639 (2007)
2. Salisbury, M.P., Anderson, S.E., Barzel, R., Salesin, D.H.: Interactive pen-and-ink illustration. In: 21st annual conference on Computer graphics and interactive techniques, pp. 101–108. ACM, New York (1994)

3. Salisbury, M.P., Wong, M.T., Hughes, J.F., Salesin, D.H.: Orientable textures for image-based pen-and-ink illustration. In: 24th annual conference on Computer graphics and Interactive Techniques, pp. 401–406. ACM Press/Addison-Wesley Publishing Co, New York (1997)
4. Chen, H., Zheng, N.N., Liang, L., Li, Y., Xu, Y. Q., Shum, H.Y.: PicToon: a personalized image-based cartoon system. In: tenth ACM international conference on Multimedia, pp. 171–178. ACM, New York (2002)
5. Tang, X., Wang, X.: Face sketch recognition. IEEE Transactions on Circuits and Systems for Video Technology **14**(1), 50–57 (2004)
6. Tang, X., Wang, X.: Face photo recognition using sketch. In: Proceedings of the 2002 International Conference on Image Processing, vol. 1, pp. I-257–I-260. IEEE (2002)
7. Tang, X., Wang, X.: Face sketch synthesis and recognition. In: Ninth IEEE International Conference on Computer Vision, pp. 687–694. IEEE (2003)
8. Chang, L., Zhou, M., Han, Y., Deng, X.: Face sketch synthesis via sparse representation. In: 20th International Conference on Pattern Recognition (ICPR), pp. 2146–2149. IEEE (2010)
9. Gao, X., Wang, N., Tao, D., Li, X.: Face sketch–photo synthesis and retrieval using sparse representation. IEEE Transactions on Circuits and Systems for Video Technology **22**(8), 1213–1226 (2012)
10. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician **46**(3), 175–185 (1992)
11. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(6), 1397–1409 (2013)
12. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A nonlinear approach for face sketch synthesis and recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 1005–1010. IEEE (2005)
13. Li, S.Z.: Markov random field modeling in image analysis. Springer, Berlin (2010)
14. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(11), 1955–1967 (2009)
15. Wang, X., Tang, X.: Random sampling for subspace face recognition. International Journal of Computer Vision **70**(1), 91–104 (2006)
16. Zhou, H., Kuang, Z., Wong, K.Y.: Markov weight fields for face sketch synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1091–1097. IEEE (2012)
17. Yang, J., Tang, H., Ma, Y., Huang, T.: Face hallucination via sparse coding. In: 15th IEEE International Conference on Image Processing (ICIP), pp. 1264–1267. IEEE (2008)
18. Wang, N., Gao, X., Tao, D., Li, X.: Face sketch-photo synthesis under multi-dictionary sparse representation framework. In: Sixth International Conference on Image and Graphics (ICIG), pp. 82–87. IEEE (2011)
19. Wang, S., Zhang, D., Liang, Y., Pan, Q.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2216–2223. IEEE (2012)
20. Song, Y., Bao, L., Yang, Q., Yang, M.-H.: Real-time exemplar-based face sketch synthesis. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VI. LNCS, vol. 8694, pp. 800–813. Springer, Heidelberg (2014)
21. Paige, C.C., Saunders, M.A.: LSQR: An algorithm for sparse linear equations and sparse least squares. ACM Transactions on Mathematical Software (TOMS) **8**(1), 43–71 (1982)
22. Draper, N.R., Smith, H., Pownell, E.: Applied regression analysis. Wiley, New York (1966)