# FACE SKETCH SYNTHESIS USING NON-LOCAL MEANS AND PATCH-BASED SEAMING

*Liang Chang*[1,2], *Yves Rozenholc*[3], *Xiaoming Deng*[4], *Fuqing Duan*[1,2], *Mingquan Zhou*[1,2]

[1]College of Information Science and Technology, Beijing Normal University, Beijing, China
[2]Engineering Research Center of Virtual Reality and Applications, Ministry of Education, China
[3]MAP5 - UMR 8145 Université Paris Descartes, and, INRIA Saclay Ile-de-France équipe SELECT, France
[4]Beijing Key Lab of HCI, Institute of Software, Chinese Academy of Sciences, Beijing, China

## ABSTRACT

This paper proposed a face sketch synthesis method by using non-local means (NL-Means), which takes the advantage of the non-local self-similarity of face photo and sketch patches. With a learning database of individuals described by one face photo and one face sketch, we assume that, for a given individual, the NL-Means coefficient of a given face photo patch is the same as its corresponding sketch patch. In order to handle the visible seam due to intensity difference of neighbor overlapping patches, we use patch based optimal seam to enforce the consistency of synthesized overlapping sketch patches. Experimental results on CUHK Face Sketch Database illustrate that our method has the advantage of easy implementation and much less required training samples, meanwhile our method can achieve fairly competitive synthesis results.

***Index Terms*—** Face sketch synthesis, Non-local means, Patch-based seaming

## 1. INTRODUCTION

Sketches are often considered to be more discriminative for face recognition than a photo in law enforcement [1] [2]. The sketch can also be the only information available on a suspect if it its obtained from the description of witness(es). In this context, both to improve discrimination and to provide similar information to face recognition systems, it appears of importance to derive a sketch from a photo when this information is available. This process is called "Face photo-sketch synthesis". Naturally, in other applications the reverse operation called "Face sketch-photo synthesis" may also be of interest [13]. Despite that our presentation below focuses only on face photo-sketch synthesis, our method can be also applied to the second problem by exchanging photos and sketches.

Photo-sketch synthesis methods can be classified into three categories, all of which aimed to learn the transformation of "Face photo-sketch synthesis".

1. Estimation of the transformation between photos and sketches using, either, an eigenvalue decomposition assuming the transformation to be linear [3], or, a locally linear embedding which bypasses the linear assumption but requires a careful choice of the number of nearest neighbors [4].
2. Markov random field (MRF) based methods. The whole photo is divided into $N$ photo-patches defining the nodes of a network each attached to a sketch-patch. MRF are used to optimize the link between the patch-sketch collection and the patch-photo network by computing a maximum a posteriori or a minimum mean-square error estimator on the network. Such approaches are known to be NP hard and can provide only an approximation of the expected optimum. They also suffer from the decomposition of the face into non-overlapping patches, technique known to reduce the potential of generalization when applied in small databases. In this spirit, [1] uses a multi-scale Markov random field model to overcome this last limitation with belief propagation, and [5] proposes to use Markov weight fields formulation, and it can be turned into a large-scale convex quadratic programming problem, the optimal solution of which is guaranteed.
3. Sparse representation based methods, which rely on the construction of two dictionaries: one for the photos and the other for the sketches. In this context, Chang et. al. [6] assume that both the face photo patch and its corresponding sketch patch have the same sparse representation coefficients or, in other words, that, in these sparse representations, the corresponding photo-to-sketch transformation is simply the identity. Wang et. al. [7] propose a semi-coupled dictionary learning method which does not assume the sparse codings to be same, however relies on a linear assumption for the link between photo descriptors and sketch descriptors.

Non-local means is an effective method in image denoising. The main difference between NL-means and other local filters or frequency domain filters is the systematic use of all possible self-predictions the image can provide [8, 9]. The method is based on a non-local averaging of all pixels in the image. Non-local means has shown advantage in the image denoising field due to the fact that it creates less artifacts than traditional wavelet based methods. NL-means has been successfully applied for super-resolution [15], however, it has never been used in the context of face sketch synthesis.

In this paper, we propose a face sketch synthesis approach based on NL-means idea in the spirit of [15]. Following the NL-means construction, our approach does not rely, either,

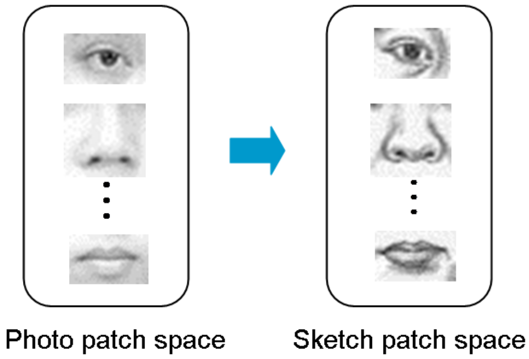


**Fig. 1**. Face sketch synthesis with face photo-sketch pairs.

on an expensive optimization procedure, or, on any linear assumption of the transformation“face to sketch”. It also do not need careful choice of meta-parameters. In order to handle the visible seam problem in the synthesized sketch due to the intensity difference of overlapping patches, we use patch based optimal seaming in the sprite of “image quilting” [10], which can keep the sketch texture consistency and overcome blurring effects by averaging overlapping patches in previous works.

Our contributions lie in the following aspects:

1. We propose a face photo-sketch synthesis method using NL-means, which takes the advantage of the nonlocal self-similarity of face photo and sketch patches. We select the intensity value, the first and second-order gradient as features for synthesis, we can get synthesized sketches with local image smoothness and edge sharpness.

2. We use image seaming to keep the texture consistency of the overlapping sketch patches. It can avoid the blurring effects due to averaging the overlapping sketch patches in previous methods.

## 2. FACE SKETCH SYNTHESIS USING NON-LOCAL MEANS

We use the database constructed by face photo and sketch pairs for training. Given a test photo, we construct its corresponding sketch. Each face photo has its sketch drawn by the same artist. All the photos and sketches are preprocessed, i.e. translated, rotated, and scaled such that the two eye centers of all the face images are at fixed position. Geometric normalization aligns the same face components in different images roughly to the same region. Then, we adopt non-local means for training.

### 2.1. Non-local means

Non-local means was firstly proposed in patch based image denoising[8]. Given an observed noisy image, the method reconstructs a denoised image. Suppose the observed image is the noisy image $\tilde{u}$, the reconstructed image is the denoised image $\hat{u}$. Denote $u(i)$ to be a square image patch centered at the pixels $i$. In NL-means, for each pixel $i$, we select a square reference patch $\tilde{P}$ around $i$. We compute a denoised patch $\hat{P}$ by computing a weighted average of patches $\tilde{Q}$ in a square neighborhood around $i$. The weight is proportional to $w(\tilde{P}) = e^{-\frac{d^2(\tilde{P},\tilde{Q})}{\sigma^2}}$ where $d(\tilde{P},\tilde{Q})$ is the Euclidean distance between $\tilde{P}$ and $\tilde{Q}$, and $\sigma$ is the noise standard deviation. We recover a final denoised value $\hat{u}(i)$ at each pixel $i$ by averaging all values at $i$ of all denoised patches containing $i$. Compared with other local filters, NL-means has advantage in that it systematic use of all possible self-predictions the image can provide[8].

We use NL-means for face sketch synthesis. Firstly, photo and sketch database are built. Then we divide each image into small overlapping patches. For each face photo patch, we compute its NL-Means coefficient. We assume that the NL-Means coefficient of each face photo patch is the same as its corresponding sketch patch. Thus, the face sketch patch of a testing face photo can be synthesized by using the NL-Means coefficient and sketch patches.

### 2.2. Algorithm

We divide the photo and sketch pairs in the training set into small overlapping patches. Denote $\bar{\mathbf{P}} = \{\bar{\mathbf{P}}_i, i = 1, \cdots, n\}$, $\bar{\mathbf{S}} = \{\bar{\mathbf{S}}_i, i = 1, \cdots, n\}$ to be the training sets composed by photo and corresponding sketch patches. The sketch synthesis proceeds patch by patch. Given a test photo $\mathbf{P}$, we aim to synthesize its corresponding sketch $\mathbf{S}$. We divide $\mathbf{P}$ into overlapping patches $\{\mathbf{P}_i, i = 1, \cdots, N\}$ in the same way. We synthesize the corresponding sketch patches $\{\mathbf{S}_i, i = 1, \cdots, N\}$ in an order from top to bottom and from left to right of the whole image. We use intensity value combined with gradient value feature as a descriptor for each image patch.

For photo patch $\mathbf{P}_i$, we synthesize $\mathbf{S}_i$ by taking weighted average of all similar sketch patches in the training set. Denote $\hat{\mathbf{S}}_i$ to be the estimation of $\mathbf{S}_i$, it is computed as follows:

$$\hat{\mathbf{S}}_i = \sum_j w(\mathbf{S}_i, \bar{\mathbf{S}}_j)\bar{\mathbf{S}}_j / \sum_j w(\mathbf{S}_i, \bar{\mathbf{S}}_j) \tag{1}$$

where the NL-means coefficient $w(\mathbf{S}_i, \bar{\mathbf{S}}_j)$ measures the similarity between $\mathbf{S}_i$ and $\bar{\mathbf{S}}_j$, it is defined as:

$$w(\mathbf{S}_i, \bar{\mathbf{S}}_j) = \exp\left(-\frac{\left\|\mathbf{G}_{\sigma^2} * (\mathbf{S}_i - \bar{\mathbf{S}}_j)\right\|^2}{2\lambda^2}\right) \tag{2}$$

where $\mathbf{G}_{\sigma^2}$ is the Gaussian kernel with standard deviation $\sigma^2$. By using Gaussian kernel,we give bigger weight to central pixels in each image patch, and smaller weight to neighboring pixels. The parameter $\lambda$ acts as a degree of filtering. It controls the decay of the exponential function and therefore the decay of the weights as a function of the Euclidean distances.

We assume the testing photo patch $\mathbf{P}_i$ and the desired sketch patch $\mathbf{S}_i$ has the same NL-means coefficient, i.e.

$$w(\mathbf{P}_i, \bar{\mathbf{P}}_j) / \sum_j w(\mathbf{P}_i, \bar{\mathbf{P}}_j) = w(\mathbf{S}_i, \bar{\mathbf{S}}_j) / \sum_j w(\mathbf{S}_i, \bar{\mathbf{S}}_j)$$

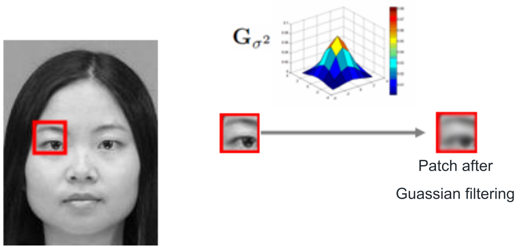


**Fig. 2**. Image patch after Gaussian filtering. $G_{\sigma^2}$ is the gaussian kernel with standard deviation $\sigma^2$.

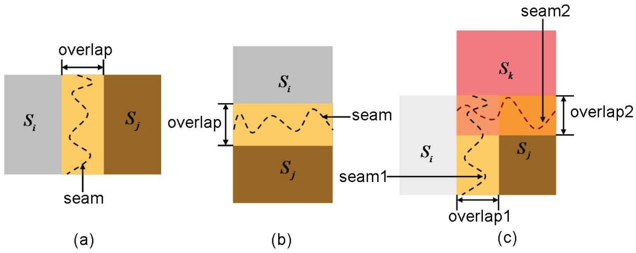


**Fig. 3**. Optimal patch seam. (a) vertical overlap; (b) horizontal overlap; (c) both vertical and horizontal overlap.

then the sketch patch can be synthesized by:

$$\hat{\mathbf{S}}_i = \sum_j w(\mathbf{P}_i, \bar{\mathbf{P}}_j)\bar{\mathbf{S}}_j / \sum_j w(\mathbf{P}_i, \bar{\mathbf{P}}_j) \tag{3}$$

In Eq. (3), $w(\mathbf{P}_i, \bar{\mathbf{P}}_j)$ can be easily computed with the test photo patch, and then we can get the synthesized sketch patch.

By exchanging the role of $\bar{\mathbf{P}}$ and $\bar{\mathbf{S}}$, we can also synthesize face photo patch from sketch patches as follows:

$$\hat{\mathbf{P}}_i = \sum_j w(\mathbf{S}_i, \bar{\mathbf{S}}_j)\bar{\mathbf{P}}_j / \sum_j w(\mathbf{S}_i, \bar{\mathbf{S}}_j) \tag{4}$$

In the image synthesis literature, researchers have suggested extracting different features for image patch to boost the prediction accuracy. The used features include high-pass filter to extract the edge information [11], a set of Gaussian derivative filters to extract the contours [12], the first and second-order gradients [14]. Here, we also use the first, second-order derivatives and intensity value as the feature for image patches due to their simplicity and effectiveness. Denote $\mathbf{P}_{x,i}, \mathbf{P}_{y,i}, \mathbf{P}_{xx,i}, \mathbf{P}_{yy,i}$ to be the first and second gradient of $\mathbf{P}_i$ along $x$ axis and $y$ axis, then the weight in Eq. (3) can be transformed as follows:

$$w(\mathbf{P}_i, \bar{\mathbf{P}}_j) = \exp(-\frac{||\mathbf{G}_{\sigma^2} * (\mathbf{P}'_i - \bar{\mathbf{P}}'_j)||^2}{2\lambda^2}) \tag{5}$$

where $\mathbf{P}'_i = [\mathbf{P}_i, \mathbf{P}_{x,i}, \mathbf{P}_{y,i}, \mathbf{P}_{xx,i}, \mathbf{P}_{yy,i}]^T$. By using intensity value, the first and second order gradient features and the weight computed by Eq.(5), we can get better synthesis results.

## 3. PATCH BASED SKETCH IMAGE SEAMING

Two overlapped patches usually generate visible seams in the overlapped region because of intensity difference(as illustrated in Fig.3). A commonly used method is to take average of the neighboring patches. However, this will lead to blurring effect. Another method is to find the seam between two overlapping patches on the pixels where they match best (i.e., where the overlap error is lowest). The minimum cost path through the error matrix is computed with dynamic programming [10].

The minimal cost path through the error matrix is computed in the following manner. As shown in Fig. 3 (a), denote $\mathbf{S}_i$ and $\mathbf{S}_j$ to be two patches that overlap along their vertical edge with the overlapped regions $\mathbf{P}$ and $\mathbf{Q}$, respectively, then the cost matrix is defined as $e_{\mathbf{S}_i,\mathbf{S}_j} = \|\mathbf{P} - \mathbf{Q}\|^2$. Let $(u, v)$ to be a pixel in $\mathbf{P}$ and $\mathbf{Q}$, then $e_{u,v} = \|\mathbf{P}(u, v) - \mathbf{Q}(u, v)\|^2$. To find the minimal vertical seam through this cost matrix, we traverse $e$ and compute the cumulative minimum error $E$ for all paths as follows:

$$E_{u,v} = e_{u,v} + \min(E_{u-1,v-1}, E_{u-1,v}, E_{u-1,v+1}) \tag{6}$$

The minimum value of the last row in $E$ indicates the end of the minimal vertical path though the cost matrix and one can trace back and find the path of the best seam. The optimal seam for other overlapped directions as shown in Fig. 3 (b)(c) can be found in similar approaches.

After obtaining the optimal seam, the pixels on the optimal seam are computed by averaging, and the other pixels in the target region are filled by the source pixels to avoid blurring. Denote $\hat{\mathbf{Q}}$ to be the blended patch at the same location as $\mathbf{Q}$ with the optimal seam $S_{PQ}$, and we can compute its intensity values as follows:

$$\hat{\mathbf{Q}}(u, v) = w_{u,v}\mathbf{P}(u, v) + (1 - w_{u,v})\mathbf{Q}(u, v) \tag{7}$$

where the weight mask $w$ is defined in the following:

$$w_{u,v} = \begin{cases} 1, & (u, v) \in R_P \\ 0, & (u, v) \in R_Q \\ 0.5, & (u, v) \in S_{PQ} \end{cases}$$

where $R_P$ is the region on the left and top of the optimal seam $S_{PQ}$ between $\mathbf{P}$ and $\mathbf{Q}$, $R_Q$ is the region on the right and downside of the optimal seam $S_{PQ}$ in the region $\mathbf{Q}$.

## 4. EXPERIMENTS

Our experiments are based on CUHK Face Sketch Database [1]. The database contains 88 faces for training and 101 faces for testing. For each face, there is a photo taken in a frontal pose and a sketch drawn by the artist. The photos are taken in frontal pose. In our experiments, we use 20 training samples. All the photo and sketch images are divided into $7 \times 7$ patches, and neighboring patches have an overlapping area of $7 \times 4$ pixels . The Gaussian filtering parameter is $\lambda = 10$,


**Fig. 4**. Comparison between the face sketch synthesis resul (a) Face photo; (b) Sketches synthesized by NL-means a seaming; (c) Sketches with the method in [1]; (d) Sketch drawn by the artist.

the Gaussian smoothing parameter is chosen as $\sigma^2 = 8$. In all our experiments, the parameter settings are fixed.

In Fig.4, we compare our method with multiscale MRF model [1]. Although the MRF based method has more clear contours, our method is more similar to the sketch drawn by the artist especially in the eye region. Moreover, our method uses only 20 training photo-sketch pairs for synthesis, while the MRF based method used 88 image pairs.

In Fig.5, we compare the results with our method, MRF method [1], and semi-coupled dictionary learning method [7]. The proposed methods are competitive in the highlighted regions, and our method with seaming can get much clearer sketch contours than results with the other methods. The results with MRF based method [1] has artifacts, and the semi-coupled dictionary learning method [7] tends to get over-smoothed sketches .

In Fig.6, we compare the NL-means method with seaming and without seaming. The results with NL-means with seaming are better in the hair contour around the face, while the results without seaming have mosaic effects in this area [1].

[1]The mosaic effects can be more easily noticed if Fig. 6 is magnified twice or more larger.

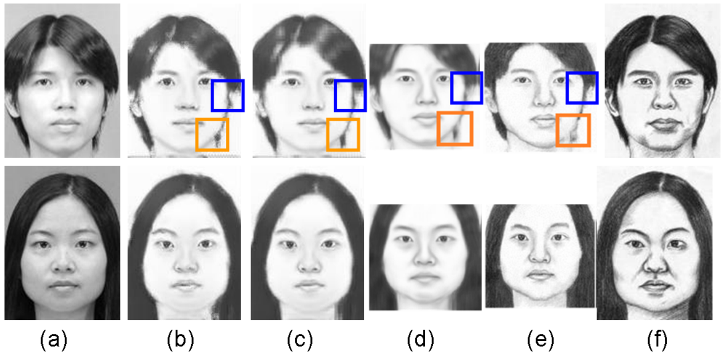


**Fig. 5**. Comparison between the face sketch synthesis results. (a) Face photo; (b) Sketches synthesized by NL-means with seaming; (c) Sketches synthesized by NL-means without seaming; (d) Results with the method in [7]; (e) Results with the method in [1]; (f) Sketches drawn by the artist.

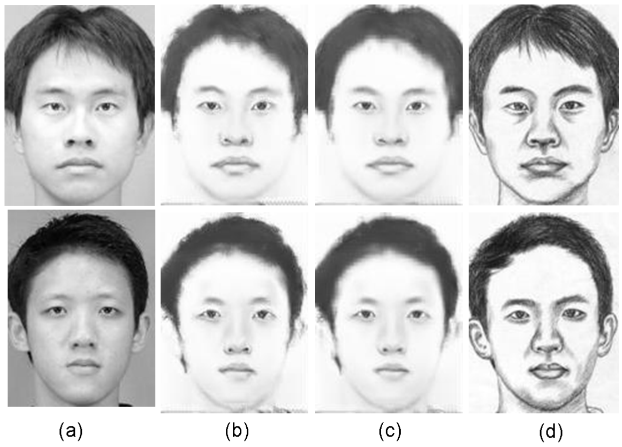


**Fig. 6**. Comparison between the sketch synthesis results. (a) Face photo; (b)Sketches synthesized by NL-means and seaming; (c) Sketches by NL-means without seaming; (d) Sketches drawn by the artist.

A limitation of the method lies in that the results of synthesized hair sketch is a little different in the line stroke style than those by artists. We hope to improve the hair results in the future works by using more line or curve discriminative features, such as wavelet features.

## 5. CONCLUSION

In this paper, we propose a face sketch synthesis method using NL-means and patch based seaming. Given a face photo, its sketch can be synthesized by using a non-local means, which learns the face texture transformation from photos to sketches. Experimental results on CUHK Face Sketch Database illustrate that our method has the advantage of easy implementation and much less required training samples, meanwhile our method can achieve fairly competitive synthesis results.

## 6. ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China(GrantNo.61402040, 61473276, 61272363, 61170203), Beijing Higher Education Young Elite Teacher Project(No.YETP0231).

## 7. REFERENCES

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

[2] Hu Han, Brendan F. Klare, Kathryn Bonnen, Anil K. Jain, Matching composite sketches to face photos: a component-based approach, IEEE Transactions on Information Forensics and Security, 8(1): pp.191–204, 2013.

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

[4] Qingshan Liu, Xiaoou Tang, Hongliang Jin, Hanqing Lu, Songde Ma, A nonlinear approach for face sketch synthesis and recognition, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1005–1010, 2005.

[5] Hao Zhou, Zhanghui Kuang, Kwan-Yee K. Wong, Markov weight fields for face sketch synthesis, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1091–1097, 2012.

[6] Liang Chang, Xiaoming Deng, Mingquan Zhou, Fuqing Duan, Zhongke Wu, Smoothness-constrained face photo-sketch synthesis using sparse representation, Proceedings of International Conference on Pattern Recognition, pp.3025–3029, 2012.

[7] Shenlong Wang, Lei Zhang, Yan Liang, Quan Pan, Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch image synthesis, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2216–2223, 2012.

[8] Antoni Buades, Bartomeu Coll, Jean-Michel Morel, A non-local algorithm for image denoising, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.60–65, 2005.

[9] Antoni Buades, Bartomeu Coll, Jean-Michel Morel, Non-local image and movie denoising, International Journal of Computer Vision, 76(2): 123–139, 2008.

[10] Alexei A. Efros, William T. Freeman, Quilting for texture synthesis and transfer, Proceedings of ACM Conference Computer Graphics and Interactive Techniques(SIGGRAPH), pp.341–346,2001.

[11] William T. Freeman, Egon C. Pasztor, Owen T. Carmichael, Learning low-level vision, International Journal of Computer Vision, 40(1):25–47, 2000.

[12] Jian Sun, Nanning Zheng, Hai Tao, Heung-Yeung Shum, Image hallucination with primal sketch priors, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 729–736, 2003.

[13] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, Jie Li, A comprehensive survey to face hallucination, International Journal of Computer Vision, 106(1): 9–30, 2014.

[14] Hong Chang, Dit-Yan Yeung, Yimin Xiong, Super-resolution through neighbor embedding, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 275–282, 2004.

[15] Matan Protter, Michael Elad, Hiroyuki Takeda, Peyman Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction, IEEE Transactions on Image Processing, 18(1): 36–51, 2009.