Face Sketch Synthesis via Sparse Representation

Liang Chang, Mingquan Zhou
College of Information Science and Technology
Beijing Normal University
Beijing 100875, P.R.China
{*changliang, mqzhou*}*@bnu.edu.cn*

Yanjun Han
Institute of Automation
Chinese Academy of Sciences
Beijing 100190, P.R.China
yanjun.han@ia.ac.cn

Xiaoming Deng
Institute of Software
Chinese Academy of Sciences
Beijing 100190, P.R.China
idengxm@gmail.com

***Abstract*—Face sketch synthesis with a photo is challenging due to that the psychological mechanism of sketch generation is difficult to be expressed precisely by rules. Current learning-based sketch synthesis methods concentrate on learning the rules by optimizing cost functions with low-level image features. In this paper, a new face sketch synthesis method is presented, which is inspired by recent advances in sparse signal representation and neuroscience that human brain probably perceives images using high-level features which are sparse. Sparse representations are desired in sketch synthesis due to that sparseness can adaptively selects the most relevant samples which give best representations of the input photo. We assume that the face photo patch and its corresponding sketch patch follow the same sparse representation. In the feature extraction, we select succinct high-level features by using the sparse coding technique, and in the sketch synthesis process each sketch patch is synthesized with respect to high-level features by solving an l_1-norm optimization. Experiments have been given on CUHK database to show that our method can resemble the true sketch fairly well.**

***Keywords*-face sketch; image synthesis; sparse representation;**

I. Introduction

Face sketch is useful in applications such as face recognition and digital entertainment[4], [8]. Compared with face photos, face sketches have the following advantages.

1) In many cases, especially in law enforcement, when the face photo of a suspect is not available, photo-based face recognition methods can not be used. An artist's drawing based on the depiction of the witness is needed and vital[4].
2) Face sketches are more concisely and discriminatively than face photos[8]. Usually, artists can capture the most distinctive characteristics of human faces by observation and human being can recognize their depictions very well.

In face sketch-based applications, sketch synthesis is usually used to construct the sketch database. In the last decade, several studies have been conducted on sketch synthesis. Tang[6] developed an eigentransform based algorithm, in which the transformation between photos and sketches is assumed to be linear. However, the assumption is hard to be satisfied in practice. Liu[4] presented a method which is similar to the spirit in local linear embedding[5]. For each input photo patch, the nearest neighbors are found, and reconstruction weights are computed. Preserving the neighbor relationships and weights, the desired sketch patch can be obtained. This method needs a carefully chosen number of nearest neighbors. Wang[8] proposed a method using a multiscale Markov random field model, which learns the face structure across different scales. As reported in [8], this method is relatively time-consuming due to using an inference procedure with belief propagation.

The theory of sparse representation has been regarded as a breakthrough in signal processing[1]. Sparse representations are desired due to that sparseness can be traced back to the parsimony principle which works as a guiding principle for inference. Researchers have also derived that the role of parsimony exists in human perception and human vision. Hence, it is suitable applying parsimony rule in computer vision tasks, such as face recognition[9], and image recovery[10]. In signal representation, among all subsets of base vectors, sparse representation selects the subset which can express the input image compactly and rejects all other possible but less compact representations. Therefore, the representation is both succinct and discriminative. Although significant advances have been made in sparse representation, yet little studies have been investigated on the sparsity in sketch synthesis.

In this paper, a face sketch synthesis method with sparse representation is proposed. Given a test face photo, we obtain its corresponding sketch image by solving an l_1- norm minimization problem with Lasso[2]. Firstly, the training photos and sketches are divided into overlapped regions. Since the used face photo and sketch images have been aligned that the same face components in different images roughly at the same region, photo and sketch patch pairs within the corresponding photo and sketch regions are used to built a coupled dictionary with much succinct elements using sparse coding [3]; Secondly, for each image patch in the test photo, one can compute its sparse representation coefficient with respect to the photo elements in coupled dictionary. The sketch patch can be recovered with the same coefficient and the sketch elements in coupled dictionary. Finally, the face sketch can be constructed with the obtained

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.526

IEEE computer society

sketch patches.

Compared with previous approaches, our method has advantages in three aspects. Firstly, the computed sparse representation adaptively selects the most relevant patches which give best representations to the test face image. It can avoid the parameter selection of traditional neighbor-based methods[4]; Secondly, former methods require dictionaries containing all the sampling image patches from the training set, while our method requires a much smaller dictionary using sparse coding; Thirdly, instead of adopting a global dictionary for training, our method is performed on local dictionaries over small regions, which makes good use of the alignment of face and sketch images. The sketch synthesis with local dictionaries is efficient and effective both theoretically and numerically. Experiment results on public data set[7] show the superiority of our method.

The organization of the paper is as follows. Section 2 gives the preliminary of sparse representation. In Section 3, face sketch synthesis approach based on sparse representation is proposed. In Section 4, experimental results are given. Finally, we conclude this paper in Section 5.

II. Preliminary of Sparse Representation

Given a set of training samples $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and a test sample $\mathbf{y}$, in which $\mathbf{x}_i \in R^d (\forall 1 \leq i \leq n)$ and $\mathbf{y} \in R^d$. If the test sample can be modeled as the linear combination of the training samples, then we have

$$\mathbf{y} = \sum_{i=1}^{n} w_i \mathbf{x}_i = \underbrace{(\mathbf{x}_1, \cdots, \mathbf{x}_n)}_{\mathbf{A}} \underbrace{\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}}_{\mathbf{w}} \tag{1}$$

where $\mathbf{A} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is the concatenation of the entire training set, and $\mathbf{w} = (w_1, \cdots, w_n)^T$ is called the representation coefficient. Thus, the linear representation (1) can be rewritten as $\mathbf{y} = \mathbf{A}\mathbf{w}$.

In order to get the solution of $\mathbf{w}$ in equation (1), the relationship of the dimensionality d and the cardinality of training set n is discussed.

In case of $d > n$, equation (1) is overdetermined, and it has unique solution.

In case of $d < n$, equation (1) is underdetermined, and its solution is not unique. Conventionally, approximate solution to (1) can be computed by minimizing an l_2-norm optimization problem as follows:

$$(l_2): \ \mathbf{w} = \arg\min \|\mathbf{w}\|_2 \quad s.t. \ \mathbf{A}\mathbf{w} = \mathbf{y}. \tag{2}$$

The disadvantage of l_2-norm optimization is that the solution is usually dense, thereby losing the discriminative ability to select the most relevant training samples.

Given a test sample $\mathbf{y}$, it admits a sparse representation over $\mathbf{A}$, if the following conditions are satisfied: $\mathbf{y}$ can be represented as a linear combination of column vectors in $\mathbf{A}$; the combination just relies on a few column vectors in $\mathbf{A}$.

To seek for the sparest solution, the following l_0-norm optimization is intended to be solved,

$$(l_0): \ \mathbf{w} = \arg\min \|\mathbf{w}\|_0 \quad s.t. \ \mathbf{A}\mathbf{w} = \mathbf{y}, \tag{3}$$

where the nonzero entries in $\mathbf{w}$ are counted with $\|\mathbf{w}\|_0$. However, solving the l_0-norm minimization is both numerically unstable and NP-hard.

Recent theories from compressive sensing[1] suggest that if the solution of $\mathbf{w}$ is sparse enough, then the sparest solution can be exactly recovered via the following l_1-norm optimization,

$$(l_1): \ \mathbf{w} = \arg\min \|\mathbf{w}\|_1 \quad s.t. \ \mathbf{A}\mathbf{w} = \mathbf{y}, \tag{4}$$

where the l_1-norm is defined as $\|\mathbf{w}\|_1 = \sum_{i=1}^{n} |w_i|$. This is a convex optimization problem that can be solved by linear programming methods.

Taking noise into consideration, equation (4) can be transformed into the following l_1-norm optimization,

$$\mathbf{w} = \arg\min \|\mathbf{w}\|_1 \quad s.t. \ \|\mathbf{A}\mathbf{w} - \mathbf{y}\|_2^2 \leq \varepsilon, \tag{5}$$

where ε is the upper bound of the noise term.

By using the Lagrange multiplier method, equation (5) can be transformed into an equivalent problem as follows:

$$\mathbf{w} = \arg\min \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1, \tag{6}$$

where λ is the regularization parameter.

The formulation (6) is a regularized l_1-norm optimization problem, which is also called Lasso in statistics and can be efficiently solved[2].

III. Face Sketch Synthesis

Given a face photo P as input, we will reconstruct the sketch S with a training set of face photos and corresponding sketches. In this section, succinct coupled dictionaries for local regions of face photo and sketch are constructed with sparse coding[3], and then face sketch synthesis method with sparse representation is presented.

A. Locally Coupled Dictionaries

Noticing that the same face component is roughly in the same region of photos and sketches due to geometry alignment, we divide the training face photos and sketches into a set of overlapping regions $\{R_i\}_{i=1}^N$ by scanning the whole image area (See Fig. 1). For each image region, a coupled dictionary is built using image patches within the local region of the training photo and sketch set.

Denote the coupled dictionary for training photo and sketch set within a local region R_i to be $\{\mathbf{D}_P^i, \mathbf{D}_S^i\}$. As shown in Fig.2, a local dictionary for regions in the red

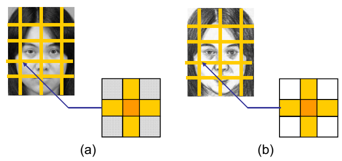

Figure 1. Face photo in (a) and its corresponding sketch in (b) are divided into regions with neighboring overlapped areas.

square is built, which is based on small overlapped patches sampled within the region of the training set.

In the proposed method, instead of building $\{\mathbf{D}_P^i, \mathbf{D}_S^i\}$ on raw image patches, we use the sparse coding[3] from $\{\mathbf{D}_P^i, \mathbf{D}_S^i\}$ to get a succinct dictionary. Sparse coding provides a class of algorithms for finding succinct representations of input stimuli or pattern. Given an over-complete data set, it learns basis elements that capture high-level features. Applying sparse coding to natural images, people can get bases which resemble the receptive fields of neurons in visual cortex. Mathematically, sparse coding aims to represent input vectors approximately as a weighted linear combination of a small number of unknown basis vectors. A number of practical methods have been developed for solving the problem, among which Lee[3] proposed an efficient strategy based on iteratively solving two convex optimization problems, including an l_1 regularized least squares problem and an l_2 constrained least square problem. In this paper, we build the coupled dictionaries succinctly using Lee's method. The number of elements in the dictionary is chose as almost $\frac{1}{10}$ of the number of sampling patches. Therefore, different from previous method in [4] which use the all sampling patches in sketch synthesis, our method is based on a fairly small dictionary constructed by sparse coding.

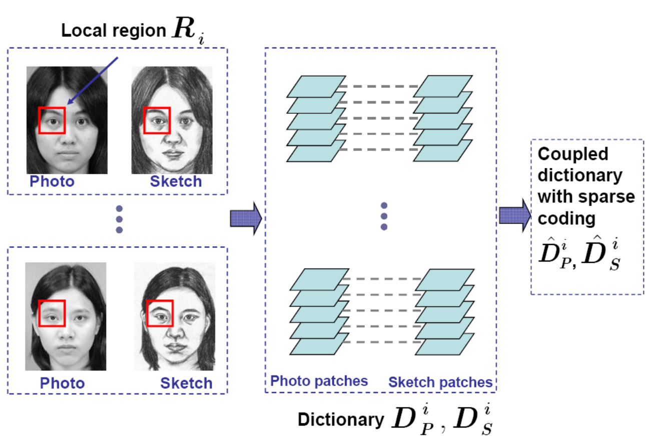

Figure 2. Construction of coupled dictionary for a local region R_i with the training photo and sketch set.

B. Sketch Synthesis Algorithm with Sparse Representation

In our method, we assume that the sparse linear relationship of a given photo patch with respect to photo bases of the coupled dictionary are maintained for the corresponding sketch patch with respect to sketch bases.

Denote the coupled dictionary within a local region R_i of training photo and sketch set as $\hat{\mathbf{D}}_P^i$ and $\hat{\mathbf{D}}_S^i$. Let a photo patch within R_i of a test photo P to be $\tilde{\mathbf{x}}$, and denote its sketch patch as $\tilde{\mathbf{y}}$. If a sparse representation of $\tilde{\mathbf{x}}$ with respect to $\hat{\mathbf{D}}_P^i$ is obtained (i. e. $\tilde{\mathbf{x}} = \hat{\mathbf{D}}_P^i\alpha$), the corresponding sketch patches $\tilde{\mathbf{y}}$ can be computed by multiplying $\hat{\mathbf{D}}_S^i$ and the representation coefficient α with $\tilde{\mathbf{y}} = \hat{\mathbf{D}}_S^i\alpha$.

The sparse representation of $\tilde{\mathbf{x}}$ can be formulated as:

$$\alpha = \arg\min \|\alpha\|_1 \quad s.t.\, \hat{\mathbf{D}}_P^i\alpha = \tilde{\mathbf{x}}. \tag{7}$$

The minimization problem (7) can be relaxed in its constraint, and then can be solved efficiently (See Section 2) ([2], [10]).

Algorithm: Face Sketch Synthesis with Sparse Representation

1) For each local image region R_i in the training photo and sketch set, construct the local coupled dictionary $\{\hat{\mathbf{D}}_P^i, \hat{\mathbf{D}}_S^i\}$ by using sparse coding.
2) For each local region R_i in a test photo
   a) For each patch in R_i;
      i) Compute its sparse representation coefficient α by solving the l_1-norm optimization with respect to $\hat{\mathbf{D}}_P^i$ (See formula (7)).
      ii) Compute the sketch patch using $\hat{\mathbf{D}}_S^i$ and the sparse representation coefficient α by $\hat{\mathbf{D}}_S^i\alpha$.
   b) Construct the sketch region by enforcing local smoothness constraint between overlapped patches.
3) Reconstruct the sketch image by enforcing local smoothness constraint between overlapped regions.

Remark: A simple method is adopted to enforce inter-patch or inter-region relationships by averaging the gray values in the overlapped area between adjacent patches and regions.

IV. Experiments

In our experiments, a face photo-sketch database from the CUHK student database was used[7]. The database contains 88 faces for training and 100 faces for testing. For each face, a sketch by an artist and a photo taken in the front pose are given. The feature vectors of the photos and sketches are represented by the gray values inside the corresponding photo and sketch patches.

In our experiments, the size of all the face and sketch images are 160×120. The size of local image region is set to be 30×30, and the size of small patches in the local region is 7×7. For adjacent regions and patches, we keep $\frac{1}{2}$ area

(a) (b) (c) (d)

Figure 3. Comparison between the sketch synthesis results of the method [6] and our method. (a) photos; (b) sketches synthesized by the method [6], copied from [8]; (c) sketches synthesized by our method; (d) sketches drawn by the artist.

(a) (b) (c) (d)

Figure 4. Comparison between the sketch synthesis results of the method[4] and our method. (a) photos; (b) sketches synthesized by the method[4], copied from [8]; (c) sketches synthesized by our method; (d) sketches drawn by the artist.

overlapped. The regularity parameter λ in l_1 minimization is set to be 0.1, and the synthesized sketches are not sensitive to the parameter λ in practice.

In Fig. 3 and Fig. 4, we compare our method with previous methods in [4], [6]. The results by our method are close to the sketches drawn by the artist in the style, and the face structure as well as details in sketches are synthesized well. Although the synthesis of human hair is challenging due to the variation in the hair style, our method based on local dictionaries can synthesize the human hair region well. In addition, since the method only need to solve l_1-norm optimization problems[1], the sketch synthesis method is efficient. In our implementation with Matlab, the sketch synthesis process of a photo takes less than a minute on a computer with 2.99GHz CPU.

V. Conclusion

In this paper, we proposed a face sketch synthesis method that exploits the nature of sparse representation in face photo and sketch. The work is based on the assumption that the face photo patch takes the same sparse representation coefficient with the corresponding sketch patch. Efforts have been devoted to build succinct coupled dictionaries for local regions of training set and compute the sparse representation coefficient with l_1 optimizations. Experiments show that the obtained sketch resembles the sketch from the artist fairly well. Our research indicates that the sketch synthesis from photo can be effectively synthesized via sparse representation.

Acknowledgment

The work is supported by the National Natural Science Foundation of China under Grant No.60736008/F010303, the open projects program of National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Sciences(20090096), China Postdoctoral Science Foundation(20080440526), and the K. C. Wong Education Foundation, Hong Kong.

References

[1] D. Donoho, *For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution*, Comm. Pure and Applied Math, Vol. 59, no. 6, pp. 797-826, 2006.

[2] B. Efron, T. Hastie, I. M. Johnstone and R. Tibshirani, *Least angle regression*, The Annals of Statistics, Vol.32, no. 2, pp. 407-499, 2004.

[3] H. Lee, A. Battle, R. Raina and A. Y. Ng, *Efficient sparse coding algorithms*, NIPS, 2006.

[4] Q. Liu, X. Tang, H. Jin, H. Lu and S. Ma, *A nonlinear approach for face sketch synthesis and recognition*, Proceedings of CVPR, 2005.

[5] S. T. Roweis and L. K. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science, Vol. 290, pp. 2323-2326, 2000.

[6] X. Tang and X. Wang, *Face Sketch Recognition*, IEEE Trans. on CVST, Vol. 14, pp. 50-57, 2004.

[7] the Chinese University of Hong Kong. Student sketch database. *http://mmlab.ie.cuhk.edu.hk/facesketch.html*, 2009.

[8] X. Wang and X. Tang, *Face Photo-sketch Synthesis and Recognition*, IEEE Trans. on PAMI, 2009.

[9] J. Wright, Y. Yang, A. Ganesh, S. Sastry and Y. Ma, *Robust face recognition via sparse representation*, IEEE Trans. PAMI, 2009.

[10] J. Yang, J. Wright, Y. Ma and T. Huang, *Image super-resolution as sparse representation of raw image patches*, Proceedings of CVPR, 2008.