Face Sketch Synthesis via Multivariate Output Regression

Liang Chang[1], Mingquan Zhou[1], Xiaoming Deng[2], Zhongke Wu[1], and Yanjun Han[3]

[1] College of Information Science and Technology, Beijing Normal University, Beijing 100875, P.R. China
[2] Institute of Software, Chinese Academy of Sciences, Beijing 100190, P.R. China
[3] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.R. China
changliang@bnu.edu.cn

Abstract. This paper presents a multivariate output regression based method to synthesize face sketches from photos. The training photos and sketches are divided into small image patches. For each pairs of photo patch and its corresponding sketch patch in training data, a local regression model is built by multivariate output regression methods such as kernel ridge regression and relevance vector machine (RVM). Compared with commonly used single-output regression, multivariate output regression can enforce the synthesized sketch patches with structure constraints. Experiments are given to show the validity and effectiveness of the approach.

Keywords: Face sketch synthesis, multivariate regression.

1 Introduction

Face sketch synthesis is useful in applications such as face recognition and digital entertainments. In applications such as law enforcement, when the face photo of a suspect is not available, traditional photo based face recognition can not be used. Then a database of face sketches is needed for retrieval and recognition. However, to synthesize a face sketch from a photo is challenging, because the sketch generation mechanism of an artist is difficult to be approximated by rules. Previous methods for face sketch synthesis, most of which are based on global principle component analysis (PCA) or local k-nearest neighbor, can hardly approximate the intrinsic nonlinear mapping between face photo and sketch.

Several studies have been conducted on sketch synthesis. Tang[5] 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[3] presented a method which is similar in the spirit to local linear embedding[4,7]. 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 of the

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2011, LNCS 6761, pp. 555–561, 2011.
© Springer-Verlag Berlin Heidelberg 2011

number of nearest neighbors. Wang[2] proposed a method using a multiscale Markov random field model, which learns the face structure across different scales. As reported in [2], this method is relatively time-consuming due to use an inference procedure with belief propagation.

In this paper, a novel face sketch synthesis method based on multivariate output regression is presented. Multivariate output regression models are popular in computer vision and machine learning. They are widely used in applications such as handwritten digits reconstruction and 3D human pose estimation[8]. In multivariate regression, the input-output pairs of the training set are both vectors.

In our method, the face photos and sketches are divided into small overlapped patches for training and synthesis. For each pair of patches, multivariate output regression model is built to learn the vector-valued function between them. The input of the regression function is the feature vector of the small photo patch, and the output is the feature vector of the corresponding sketch patch. Since the number of variables in the regression function is very large, the regression problem is often ill-posed. We use different regularization methods including the ridge regression [1] and relevance vector regression machine(RVM) [9] to solve the regression problem. In RVM, Gaussian priors are given for each parameter and many basis vectors are of the coefficient of zero, so the regression model has sparse property.

Our method has advantages in two aspects:

(1) Compared with the single-variable regression, in which the output is scalar value, the outputs in multivariate regression are vectors, which can enforce the texture structure constraints of face sketch and approximate the local texture of sketch patches well;
(2) Our method is nonlinear intrinsically. We use kernel bases to obtain the nonlinear mapping between local photo and sketch regions, and it can greatly enhance the sketch synthesis performance.

2 Face Sketch Synthesis Method

In this section, face sketch synthesis method via multivariable output kernel regression is proposed. The method is based on the regression of local image patches, and ridge regression and relevance vector machine are used.

2.1 Multivariate Output Regression

The multivariate regression model maps the input feature vector to an output vector. Suppose the output vector to be $\mathbf{y} \in R^m$ and the input vector to be $\mathbf{x} \in R^d$, the mapping between $\mathbf{x}$ and $\mathbf{y}$ can be approximated with a linear combination of basis functions $\{\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \cdots, \varphi_p(\mathbf{x})\}$ as follows

$$\mathbf{y} = \beta_1 \varphi_1(\mathbf{x}) + \cdots + \beta_p \varphi_p(\mathbf{x}) \tag{1}$$

where $\{\beta_k\}_{k=1}^n \in R^m$ are weight vectors.

We transform $\{\beta_k\}_{k=1}^n$ to a weight matrix $\mathbf{B}_{m\times p} = (\beta_1, \beta_2, \cdots \beta_p)$ and the basis functions into a R^p valued function $f(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \cdots \varphi_p(\mathbf{x}))^T$. Equation (1) can be formulated as

$$\mathbf{y} = \mathbf{B}f(\mathbf{x}) + \varepsilon \tag{2}$$

where ε is residual error vector, and $f(\mathbf{x})$ is the feature vector of input vector $\mathbf{x}$.

For underdetermined system, the solution to equation (2) is not unique[1]. To seek for the unique solution to the weight matrix $\mathbf{B}$, minimization problem of the square residual error with a regularization item is required to be solved:

$$\min_{\mathbf{B}} \sum_{i=1}^{n} \left\| \mathbf{B}f(\mathbf{x}_i) - \mathbf{y}_i \right\|^2 + R(\mathbf{B}) \tag{3}$$

where $R(\mathbf{B})$ is a regularization item on $\mathbf{B}$.

After simple manipulations, equation (3) can be transformed into the following matrix formulation:

$$\min_{\mathbf{B}} \left\| \mathbf{BF} - \mathbf{Y} \right\|^2 + R(\mathbf{B}) \tag{4}$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)$ and $\mathbf{F} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \cdots, f(\mathbf{x}_n))$.

We build the feature vector $f(\mathbf{x})$ by kernel bases as $f(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x}_1), \cdots, K(\mathbf{x}, \mathbf{x}_n))^T$, in which $K(\mathbf{x}_i, \mathbf{x}_j)$ is chosen as Gaussian kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 / \sigma^2)$ with σ estimated from scatter matrix of the training data. Therefore, $\mathbf{F}$ in equation (4) is taken as the kernel matrix of the training data, i.e. $\mathbf{F} = (K(\mathbf{x}_i, \mathbf{x}_j))$.

2.2 Ridge Regression and Relevance Vector Machine

In our problem, we aim to synthesize a face sketch from a face photo. The synthesis problem is transformed into a regression problem(See equation (2)), in which the input vector $\mathbf{x}$ is of a photo patch and the output vector $\mathbf{y}$ is the corresponding sketch patch of $\mathbf{x}$. However, the number of unknown variables in the weight matrix $\mathbf{B}$ is large, the regression is usually ill-posed, and conventional least square regression can no longer be used [1]. In order to get the unique solutions, regression methods with different regularization items are adopted. In this paper, the ridge regression and relevance vector machine are used.

(1) Ridge Regression

In ridge regression, the weight matrix $\mathbf{B}$ is estimated by minimizing the sum of square residual error and a regularization term[1]. The regularization term of the weight matrix takes the form of $R(\mathbf{B}) = \lambda \left\|\mathbf{B}\right\|^2$, thus the ridge regression takes the form as follows:

$$\min_{\mathbf{B}} \|\mathbf{BF}-\mathbf{Y}\|^2 + \lambda \|\mathbf{B}\|^2 \tag{5}$$

where λ is the regularization parameter.

(2) Relevance Vector Machine

Relevance vector machines (RVM) are widely used in classification and regression in computer vision and machine learning [9]. It belongs to sparse Bayesian approaches. The column vector $\mathbf{b}_i$ of $\mathbf{B}$ takes the prior of the form $\|\mathbf{b}_i\|^{-\lambda}$. If λ is large, the regularization term makes many parameters in $\mathbf{B}$ to be zeros, thus the regression model is sparse. With the log likelihood priors, RVM takes the minimization form as follows:

$$\min_{\mathbf{B}} \|\mathbf{BF}-\mathbf{Y}\|^2 + \lambda \sum_k \log \|\mathbf{b}_k\| \tag{6}$$

where λ is the regularization parameter.

2.3 Face Sketch Synthesis Method

In this section, the face sketch synthesis method via multivariate regression is proposed. We divide the training face photos and sketches into a set of overlapped patches. For each pairs of training photo and sketch patch at the same location, the multivariable output kernel regression model is built and the weight matrix $\mathbf{B}$ is estimated (See Section 2.2). Finally, we can reconstruct the sketch image of a test face photo by the estimated weight matrix $\mathbf{B}$. We divide the test face photo into a set of overlapped patches, and denote $f(\mathbf{x}_i)$ to be the feature vector of a photo patch $\mathbf{x}_i$, and then the corresponding sketch patch can be reconstructed by $\mathbf{B}f(\mathbf{x}_i)$.

Algorithm: Face Sketch Synthesis via Multivariable Output Regression
Input: A training set of face photos and the corresponding sketches, and a test face photo.
Output: A synthesized sketch of the test face photo.

(1) All the training photos and sketches and the test photo are divided into overlapped patches.
(2) For each image patch $\mathbf{x}_i$ of the test photo

a) Build the ridge regression or relevance vector machine model with photo and sketch patch pairs of the same location in the training images, and then estimate the weight matrix $\mathbf{B}$ (See Section 2.2).
b) Reconstruct the sketch patch of the photo patch $\mathbf{x}_i$ by multiplying the weight matrix $\mathbf{B}$ and the feature vector $f(\mathbf{x}_i)$ of the photo patch $\mathbf{x}_i$, i.e. $\mathbf{B}f(\mathbf{x}_i)$.

(3) Reconstruct the sketch image by enforcing smoothness between the overlapped regions of the patches.

3 Experiments

The experiments are based on face photo-sketch database from CUHK student database [6]. Since the same face component is roughly in the same region of photos and sketches due to geometry alignment in the preprocessing step, we divide the training face photos and sketches into a set of regions by scanning the whole image area. We use 88 faces for training and 100 faces for testing. For each face photo, 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 with Gaussian kernel bases (See $f(\mathbf{x})$ in equation (2)) by the gray values of the corresponding photo and sketch patches. The size of all the face and sketch images are 160×120. The size of small patches 7×7. For adjacent patches, we keep half of the patch overlapped.

In Fig. 1 and Fig. 2, we demonstrate the experiment results by our method with ridge regression with Gaussian kernel bases, and the regularity parameter λ is set to be 1.0. In Fig. 1, we compare our results with those by eigen-face method [5] and nearest neighbor method[3]. It can be seen that our method can synthesize the hair region well, which is hard for the eigen-face method[5]. Moreover, our method can avoid the selection of neighborhood number in the nearest neighbor method[3]. Experiments show that the results of our method can approximate the sketches drawn by the artist in the style and structures.

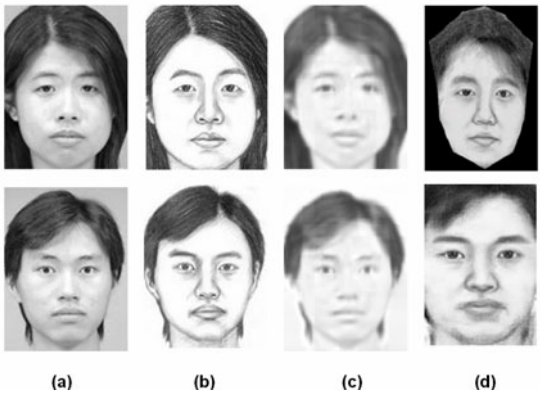

Fig. 1. (a) Photos; (b) sketches drawn by artists; (c) sketches by our method using kernel ridge regression; (d) sketches by the eigen-face method in [5](the first line) and k-nearest neighbor method in [3] (the second line), which are both from [2]

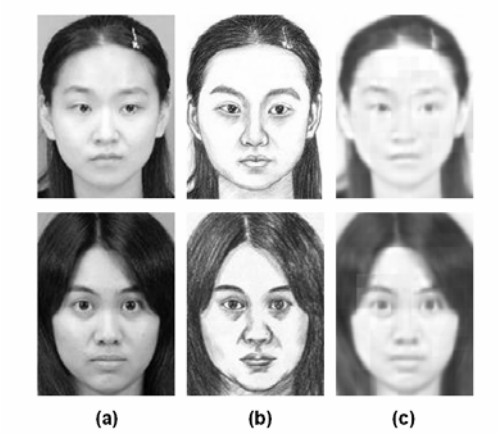

Fig. 2. (a) Photos; (b) sketches drawn by artists; (c) sketches by our method using kernel ridge regression

4 Conclusion

In this paper, we propose a face sketch synthesis method via multivariable output regression. In our method, all the face photos and sketches are divided into overlapped patches, and then multivariable output regression such as kernel ridge regression and relevance vector machine are used to build the regression model between the photo patch and sketch patch. Experiments show that the synthesized sketch can resemble the sketch draw by artists. Our research indicates that multivariable output regression approach is effective in face sketch synthesis.

Acknowledgments. The work is supported by National Natural Science Foundation of China under Grant No. 60736008/F010303 and 61005039, the open projects program of National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Sciences under Grant No. 20090096.

References

1. Bishop, C.: Neural Networks for Pattern Recognition. Oxford Univ. Press, Oxford (1995)
2. Wang, X., Tang, X.: Face Photo Sketch Synthesis and Recognition. IEEE Trans. on PAMI (2009)
3. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A Nonlinear Approach for Face Sketch Synthesis and Recognition. In: Proceedings of CVPR (2005)
4. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through Neighborhood Embedding. In: Proceedings of CVPR (2004)
5. Tang, X., Wang, X.: Face Sketch Recognition. IEEE Trans. on CSVT 14, 50–57 (2004)

6. The Chinese University of Hong Kong, Student sketch database (2009), http://www.mmlab.ie.cuhk.edu.hk/facesketch.html
7. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
8. Agarwal, A., Triggs, B.: Recovering 3D human pose from monocular images. IEEE Trans. on PAMI 28, 44–58 (2006)
9. Tipping, M.: The relevance vector machine. Neural Information Processing Systems (2000)