Example-based Automatic Portraiture

Hong Chen[1*], Lin Liang[1], Ying-Qing Xu[1], Heung-Yeung Shum[1] and Nan-Ning Zheng[2]

[1] Microsoft Research, China
[2] Xi'an Jiaotong University, China

Abstract

In this paper, we present an example-based approach for automatically generating a life-like portrait from a frontal face image. Based on an inhomogeneous Markov Random Field Model, an inhomogeneous non-parametric sampling scheme is used to capture the complex statistical characteristics of face image and corresponding portrait. In our approach, only those pixels corresponding to a portrait point are sampled. Such a strategy is crucial for maintaining facial structure and guaranteeing coherence of portrait lines. Experimental results demonstrate the effectiveness and lifelikeness of our approach.

1 Introduction

A portrait is a visual representation of individual person, especially of the face. Only well-trained artists are capable of exhibiting great skills in drawing portraits. It is a challenging and difficult work to use computer to automatically generate a portrait from a given image.

There have been few attempts to automatically generate a stylistic facial sketch by observing images drawn by artists. A few template-based facial caricature systems were developed, for instance, by Murakami et al. [6], and Li et al. [7]; these systems simply link facial feature points using image processing methods. But these approaches did not attempt to observe and learn from an artist's products, and thus produced stiff sketches.

Inspired by recent development in texture synthesis, texture transfer [4, 9, 5] and face hallucination [1], Chen et al. [2] developed an example-based automatic stylistic facial sketch generating system. They used inhomogeneous non-parametric sampling to capture the statistical likelihood between the sketch and original facial image, and fit a flexible template that models the statistical prior of the sketch. However, this method is limited to sketch generation.

In this paper, we present a new example-based approach that can generate portraits with varying styles. We adopt an inhomogeneous Markov Random Field model, which can model not only the likelihood of a portrait given the original image, but also the prior statistical characteristics of the portrait. Based on this statistical model, we propose two sampling strategies: iterative sampling which is simple and efficient, and simulated annealing which is more robust.

The rest of this paper is organized as follows. We present our example-based learning framework in Section 2. The statistical model for portraiture is described in Section 3. The detailed algorithms are presented in Section 4. Results are shown in Section 5. In Section 6, we summarize our work and address future research topics.

2 Example-based Portraiture Generation

As shown in Figure 1, given a face image I, we aim to automatically generate its portrait S with artistic styles such as pencil, line-drawing. Since there are no clear rules to formulate the artist's drawing style and intention, we have chosen to take an example-based approach under the Bayesian inference framework.

2.1 Statistic Learning Approach

Based on the Bayes rule, the posterior probability $P(S|I)$ can be represented as:

$$P(S|I) = \frac{P(I|S)P(S)}{P(I)} \tag{1}$$

The prior $P(S)$ represents the statistical property of S. The likelihood $P(I|S)$ is the probability of the observed image I given the portrait S. To estimate S, we adopt MAP criterion: finding the optimal one to maximize the the posterior probability $P(S|I)$. Since the evidence $P(I)$ can be treated as a normalization constant, MAP actually maximizes the product of likelihood $P(I|S)$ and prior $P(S)$,

$$S^* = \arg\max_S P(I|S)P(S) \tag{2}$$

In our approach, an example-based approach is employed to learn the prior $P(S)$ and the likelihood $P(I|S)$ from a set of training examples.

[*] Visiting from Artificial Intelligence and Robotics Lab, Xi'an Jiaotong University, China

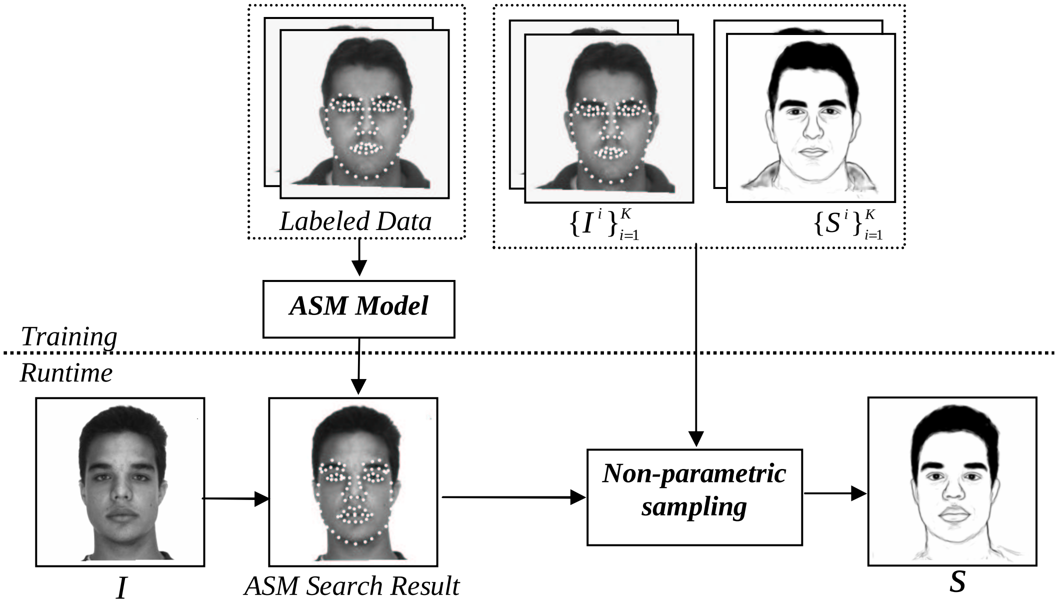

Figure 1. System Framework

2.2 Training Data

The examples in our training data include a set of frontal face images $\{I^i\}_{i=1}^K$ taken from the AR data set [8] and corresponding artist drawings $\{S^i\}_{i=1}^K$ as shown in Figure 1. We select and prepare the training data to satisfy following important prerequisites:

- Frontal view only (no hats and glasses)
- Each pair of image and portrait matches perfectly
- Portraits are drawn with a consistent style

To get matched training pairs, we have asked the artist to draw the portrait on the top of the original image(e.g., with a different layer in PhotoShop). As we shall see later, matched image and portrait will make our learning process much simpler.

2.3 The System Framework

Because human faces are highly structural, we assume inhomogeneous Markov assumption to model the complicated spatial probability $P(S|I)$: the conditional probability of a portrait point q is determined by its neighborhood and position. To construct such a probability, an inhomogeneous non-parametric sampling strategy is employed: only the points at the same facial location with q from different training images are sampled.

As shown in Figure 1, our system consists of a training phase and a runtime phase. In the training phase, we start with the set of face images with manually labeled facial feature points on them.

- An ASM model is trained to automatically locate the facial feature points in any input image

At runtime, for a given image I, we generate a portrait S by the following steps:

- Apply the ASM to extract the facial feature points
- Employ inhomogeneous non-parametric sampling to obtain the MAP solution of portrait S
  - Construct points correspondence between I and each training image
  - Find the MAP solution by iterative local search or simulated annealing

Our statistical model and sampling strategy are explained in the following sections.

3 Statistical Model for Portraiture

3.1 Inhomogeneous Markov Random Field Model

In our system, an inhomogeneous Markov Random Field (MRF) model is assumed to describe the inhomogeneity of facial features. Let $S(q)$ denote the grey value of pixel q in S. Then under the MRF assumption, the probability distribution of $S(q)$ will depend on the small neighbor regions both in I and S.

$$P_q(S(q)|I;\{S(q')\},\forall q'\neq q) = P_q(S(q)|N_I(q),N_S(q)) \quad (3)$$

where $N_I(q)$ and $N_S(q)$ denote the neighborhood regions of q in I and S, respectively.

For a homogeneous MRF, $P_q(S(q)|N_I(q), N_S(q))$ is independent of the relative location of q in the image lattice. But in our inhomogeneous MRF model, the conditional density functions of pixels at different positions are not identical. Thus, for two pixels $q \neq v$, even if the neighborhood conditions are the same, their local conditional probabilities may be different.

Under MRF assumption and Bayes rule, the conditional (posterior) probability can be represented as:

$$P_q(S(q)|N_I(q),N_S(q)) \propto P_q(N_I(q)|S(q))P_q(S(q)|N_S(q)) \quad (4)$$

We call $P_q(N_I(q)|S(q))$ the local likelihood and $P_q(S(q)|N_S(q))$ the local prior and assume them to be exponential distributions:

$$P_q(S(q)|N_S(q))=\frac{1}{Z_{q,P}}exp\{-E_{q,P}(S(q),N_S(q))/T_P\} \quad (5)$$

$$P_q(N_I(q)|S(q))=\frac{1}{Z_{q,L}}exp\{-E_{q,L}(S(q),N_I(q))/T_L\} \quad (6)$$

where $E_{q,L}(S(q), N_I(q))$ is the local likelihood energy, $E_{q,P}(S(q), N_S(q)))$ the local prior energy, $Z_{q,P}$, $Z_{q,L}$ the normalizing constant, and temperature T_P and T_L are used to control the smoothness of the distribution.

Let $Z_q = Z_{q,P} \times Z_{q,L}$, $T = T_L$. The posterior probability becomes

$$P_q(S(q)|N_I(q),N_S(q))=\frac{1}{Z_q}exp\{-E_q(S(q),N_I(q),N_S(q))/T\} \quad (7)$$

where

$$\begin{aligned} E_q(S(q), N_I(q), N_S(q)) &= E_{q,L}(S(q), N_I(q)) \\ &\quad +\lambda E_{q,P}(S(q), N_S(q)) \quad (8) \end{aligned}$$

The posterior energy is represented as the weighted sum of local likelihood energy and local prior energy. The weighting coefficient λ adjusts the statistical constraints coming from $N_I(q)$ and $N_S(q)$.

3.2 Non-parametric Probability Representation

Since learning the above MRF model parameters [12] is very complex, inspired by a non-parametric sampling method successfully used in texture synthesis [4], we construct the discrete probability distribution of a portrait point evaluated only at training examplars. Unlike homogeneous non-parametric sampling [4, 9, 5], we only use the training examplars at the corresponding facial position of q to construct its distribution.

Mathematically, suppose Ω_q contains M examplars whose facial positions correspond to q: $\Omega_q = \{z_q^j\}_{j=1}^M$, where j is the pixel's index in Ω_q , z_q^j the pixel's position, $\{I(z_q^j), S(z_q^j)\}$ the pixel values of these sample points, and $N_I(z_q^j)$ and $N_S(z_q^j)$ denote the neighborhoods of pixel z_q^j in I and S, respectively. For non-parametric sampling, the pixels with neighborhood close to q's neighborhood are most possible to be chosen as q, then we can define the local likelihood energy $E_{q,L}(S(q), N_I(q))$ and local prior energy $E_{q,P}(S(q), N_S(q))$ as:

$$E_{q,L}(S(q), N_I(q)) = \sum_{j=1}^{M} \delta(k_q - j)d(N_I(q), N_I(z_q^j)) \quad (9)$$

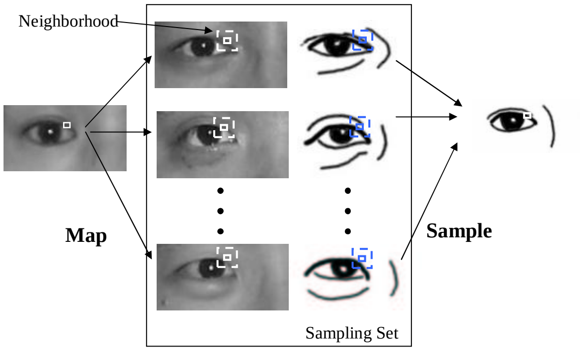

Figure 2. Inhomogeneous non-parametric sampling. On the training set, the smaller solid white squares represent corresponding pixels, while the larger dashed windows are sampled neighborhoods.

$$E_{q,P}(S(q),N_S(q))=\sum_{j=1}^{M} \delta(k_q-j)d(N_S(q),N_S(z_q^j)) \quad (10)$$

where k_q is the pixel's index in Ω_q where the portrait point is sampled from, $\delta(.)$ is the Dirac function, and $d(.)$ is the similarity metric between neighborhood regions. Then the local posterior energy will be

$$\begin{aligned} &E_q(S(q), N_I(q), N_S(q)) = \\ &\sum_{j=1}^{M} \delta(k_q-j)[d(N_I(q),N_I(z_q^j))+\lambda d(N_S(q), N_S(z_q^j))] \quad (11) \end{aligned}$$

The similarity metric $d(.)$ plays a crucial role for sampling. For different tasks, different similarity metrics can be defined. We will discuss it as implementation details in next section.

4 Inhomogeneous Non-parametric Sampling

Our inhomogeneous sampling strategy is important for guaranteeing global facial structure, the coherence of facial drawing lines. Detailed algorithms are discussed in following subsections.

4.1 Sampling Set Construction

For a given point q, its sampling set Ω_q only contains the training examplars at the corresponding positions (as shown in Figure 2). Since the training images are usually not aligned with the new input image I, we have to establish the correspondence first. To do this automatically, the ASM model [3] is employed to locate facial feature points of I (as shown in Figure 1). Then the corresponding pixel position v_q^i of q in the ith training image is determined.

4.2 Sampling Strategy

To get the MAP solution of S, we have designed two different sampling strategies. Iterative local search is simple and efficient, but easy to get stuck at the local minima.

4.2.1 Iterative Local Search

We use an iterative local conditional distribution maximization procedure, sequentially updating each pixel's value $S^{(n)}(q)$ to $S^{(n+1)}(q)$ by maximizing $P_q(S(q)|N_I(q),N_S(q))$, i.e. minimizing $E_q(S(q),N_I(q),N_S(q))$. Based on the posterior energy expression (11), this can be done by non-parametric sampling: compare the neighborhood region of the output pixel $\{N_I(q), N_S(q)\}$ with the neighborhood regions of sampling set $\{N_I(z_q^j, N_S(z_q^j)\}_{j=1}^M$, then the pixel value $S(z_q^j)$ whose neighborhood region best matches q is set as $S(q)$. The sampling process is summarized in the following pseudo code:

```
Function $Sampling(\{I^i, S^i\}_{i=1}^M, I)$
  $Initialize(S^{(0)})$
  For each iteration $n$, from 1 to $N$ do:
    For each pixel $q \in S^{(n)}$ do:
      Construct sampling set $\Omega_q$
      find pixel $z_q^r \in \Omega_q$ to minimize energy (11)
      $S^{(n)}(q) \leftarrow S(z_q^r)$
    end
    $S^{(n+1)} = S^{(n)}$
    $n = n + 1$
  end
  $S = S^{(N)}$
  return $S$
```

We synthesize pixels of S in scan-line order and use square neighborhood during iteration. The neighborhood contains synthesized pixel values at this iteration and the initial pixel values determined from last iteration. To get initialized MAP solution $S^{(0)}$, L-shape neighborhood [9] is used to include prior constraints.

By iteratively updating each portrait point from propagating its neighboring points, lines generated will be smoother than by sampling only once, especially for the line drawing style, as shown in Figure 3.

4.2.2 Simulated Annealing

The local search method often falls into local minima. The problem is more severe for generating line drawing style portraits because unexpected disconnected lines will appear. To deal with this problem, simulated annealing can be incorporated into the above local search method. The temperature T in the distribution (7) is decreased from a high value to a low value during the iterative minimization. At a fixed T, the sampling is done according to the distribution (7). When T is close to zero, the global minimum solution is nearly found. The sampling process with simulated annealing is shown in the following pseudo code:

```
Function SamplingWithSA($\{I^i, S^i\}_{i=1}^M$,$I$)
  Initialize $T$
  While $T > \varepsilon$ ($\varepsilon$ is a small positive number) do:
    For each pixel $q \in S$ do:
      Construct sampling set $\Omega_q$
      Set $S^{(n)} = S(z_q^j)$ with probability (7)
  end
  decrease $T$
end
return $S$
```

Simulated annealing allows occasional energy ascent instead of always performing energy descent. This helps the algorithm to escape from local minima. Some results are shown in Figure 3.

4.3 Implementation Details

4.3.1 Similarity Metric

The similarity metric $d(.)$ in (11) plays a very crucial role in nonparametric sampling. Many kinds of image features can be selected to define $d(.)$, such as grey values, directional filter responses, and pyramid structures [4, 9, 5]. We have used a multi-scale grey level metric in our system, which has proven to be effective in our experiments. The image feature vector $F(q)$ of pixel q contains grey values of neighborhoods at two pyramid levels, $d(.)$ is defined as:

$$d(N(q), N(z_q^j)) = \|F(q) - F(z_q^j)\|/L$$

where L is the length of the feature vector. Including neighborhood on low resolution level is good for capturing long scale structure, such as drawing lines.

4.3.2 Acceleration Strategy

Similar to patch-pasting texture synthesis approaches [10, 11], we extend the synthesis from a pixel to a small $\omega \times \omega$ square patch to accelerate the generation speed. The results shown in this paper are all generated using such an acceleration scheme.

5 Experiments

We have selected 40 frontal images from the AR data set to test our algorithm. Figure 6 shows some examples generated by our method. The results are generated using a two-level pyramid, where the neighborhood size is set to 5×5 at the coarse level and 7×7 at the fine level. For a point, all examples within a 5×5 window are used to construct the sampling set. The parameter λ of the distance metric is set to 0.2 for both styles. The number of iterations is 3, and the patch size is 3×3. It takes about 50 seconds to generate a 256×256 portrait based on 100 training images of size 256×256.

Figure 4 shows the influence of the distance metric parameter λ in (11) which controls the tradeoff between the

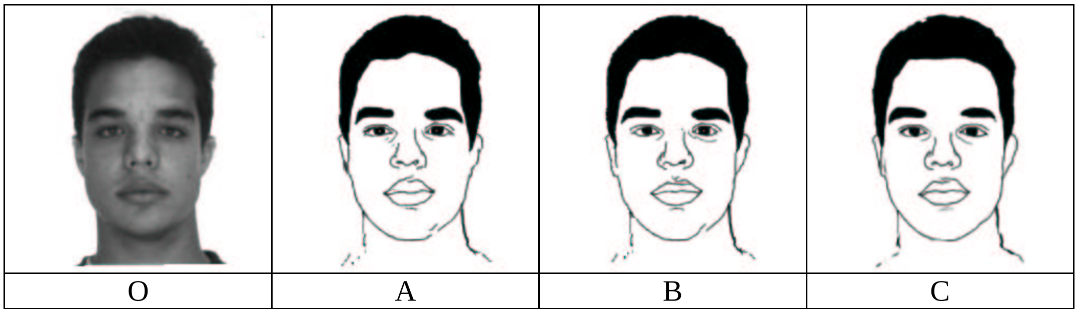

Figure 3. The effect of iteration numbers and simulated annealing. (O) the original image. (A) 1 iteration; (B) 3 iterations; (C) simulated annealing.

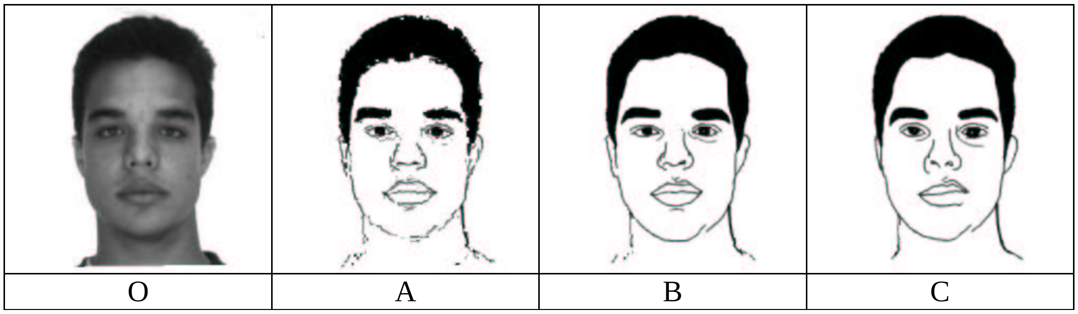

Figure 4. The effect of λ. (O) the original image. The λ is set to (A) 0; (B) 0.2; (C) 1.0.

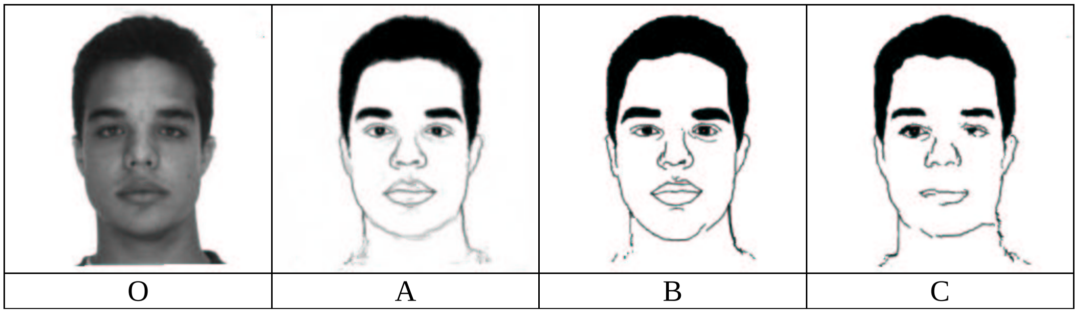

Figure 5. Comparison between our method and homogeneous non-parametric sampling. (O) source image; (A) pencil style portrait of our method; (B) line drawing style portrait of our method;(C)line drawing generated by homogeneous non-parametric sampling.

local likelihood and local prior. When λ increases, the results tend to keep the drawing style more strictly, but will lose some similarity with the original image, and vice versa. In our experiments, we found 0.2 to be a good choice. We have also compared our results with those using homogeneous sampling method similar to Image Analogies [5] in Figure 5. Obviously, our method can generate better results.

6 Summary and Future Works

In this paper, we have presented an example-based approach to automatically generate a life-like portrait from a given frontal face image. Based on a set of training examples drawn by an artist, the Inhomogeneous Markov Random Field model is employed as the statistical model and a non-parametric sampling scheme is used to capture the complex statistical characteristics of a face image. Such a strategy is crucial for maintaining facial structure and guaranteeing coherence of portrait lines. Our method can automatically generate portraits with different styles.

Our approach of portraiture is, up to now, limited to frontal face images. Extending the current method to different views and even to video is necessary for some applications. There are also a number of interesting topics that we can address in the future: speeding up the non-parametric sampling; dealing with source images with complex backgrounds; and finding better features and similarity metrics.

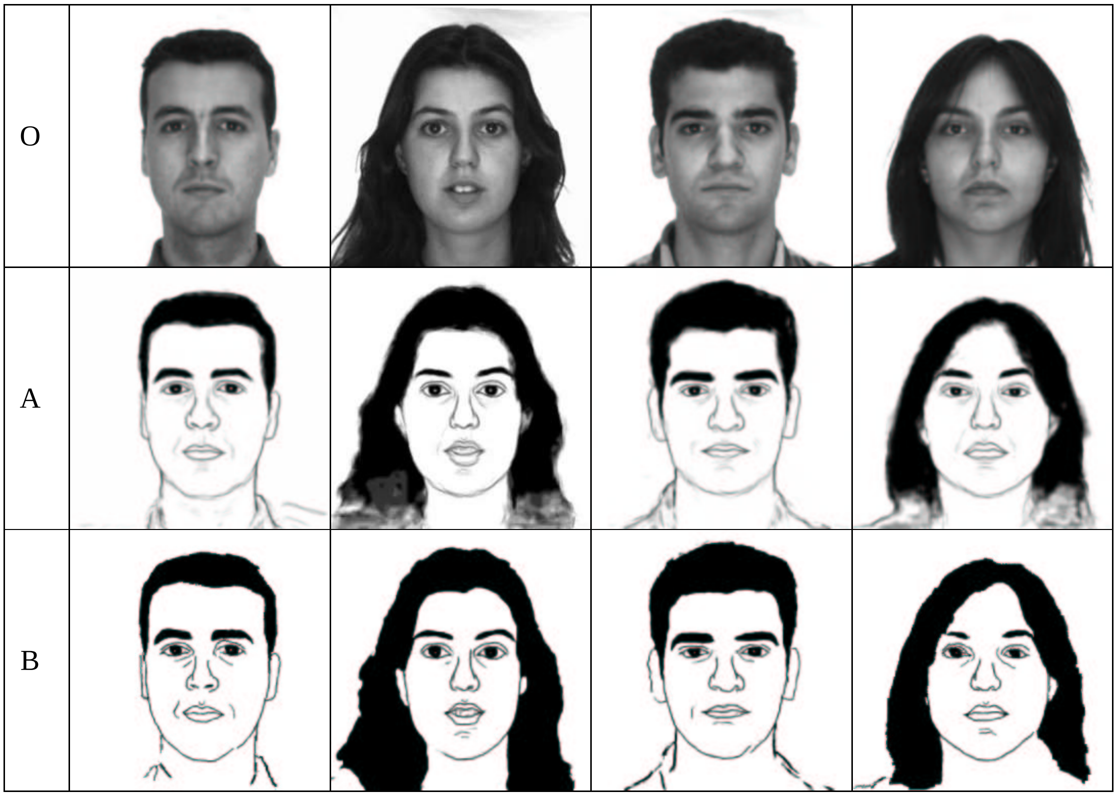

Figure 6. More results of pencil style and line drawing style portraits generated by our method. (O) source image; (A) pencil style portrait; (B) line drawing style portrait.

References

[1] S. Baker and T. Kanade. Hallucinating face. In *The fourth International Conference on Automatic Face and Gesture Recogntion*, 2000.

[2] H. Chen, Y. Q. Xu, H. Shum, S. C. Zhu, and N. N. Zheng. Example-based facial sketch generation with non-parametric sampling. In *the Eighth International Conference on Computer Vision*, 2001.

[3] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer version. Technical report, University of Manchester, Manchester M13 9PT, U.K., 2000.

[4] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *the Seventh International Conference on Computer Version*, pages 20–27, 1999.

[5] A. Hertzmann, C. E. Hacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In *SIGGRAPH'2001*, 2001.

[6] H. Koshimizu, M. Tominaga, T. Fujiwara, and K. Murakami. On kansei facial processing for computerized facial caricaturing system picasso. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 6, pages 294 –299, 1999.

[7] Y. Li and H. Kobatake. Extraction of facial sketch based on morphological processing. In *IEEE international conference on image processing*, volume 3, pages 316–319, 1997.

[8] A. Martinez and R. Benavente. The ar face database. Technical Report 24, CVC, 1998.

[9] L. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *SIGGRAPH'2000*, pages 479–488, 2000.

[10] Y. Q. Xu, B. N. Guo, and H. Shum. Chaos mosaic:fase and memory efficient texture synthesis. Technical Report 32, Microsoft Research Technical Report, Apr. 2000.

[11] Y. Q. Xu, S. C. Zhu, B. N. Guo, and H. Shum. Asymptotically admissible texture synthesis. In *Proc. of 2nd Int'l Workshop on Statistical and Computational Theories of Vision*, page 12, 2001.

[12] S. C. Zhu, Y. N. Wu, and D. Mumford. Filters,random fields and maximum entropy:towards a unified theory for texture modeling. *International Journal of Computer Vision*, 12(2), 1998.