Markov Weight Fields for Face Sketch Synthesis

Hao Zhou, Zhanghui Kuang and Kwan-Yee K. Wong
Department of Computer Science,
The University of Hong Kong,
Pokfulam Road, Hong Kong.
{hzhou, zhkuang, kykwong}@cs.hku.hk

Abstract

Great progress has been made in face sketch synthesis in recent years. State-of-the-art methods commonly apply a Markov Random Fields (MRF) model to select local sketch patches from a set of training data. Such methods, however, have two major drawbacks. Firstly, the MRF model used cannot synthesize new sketch patches. Secondly, the optimization problem in solving the MRF is NP-hard. In this paper, we propose a novel Markov Weight Fields (MWF) model that is capable of synthesizing new sketch patches. We formulate our model into a convex quadratic programming (QP) problem to which the optimal solution is guaranteed. Based on the Markov property of our model, we further propose a cascade decomposition method (CDM) for solving such a large scale QP problem efficiently. Experimental results on the CUHK face sketch database and celebrity photos show that our model outperforms the common MRF model used in other state-of-the-art methods.

1. Introduction

Face sketch synthesis, being an important branch of face style transformation [1, 10, 11, 20], has received much attention in recent years, and has many important applications in both digital entertainment and law enforcement. For instance, some people prefer using face sketches to using real photos as their profile pictures in facebook and personal websites. Artists may employ face sketch synthesis technique to simplify animation production. Police officers need to match a sketch of a suspect drawn by an artist against a mug-short database to help identifying the suspect. Face sketch synthesis can help narrow down the gap between face photos and face sketches [16]. Despite its wide application, early attempts to synthesize face sketches from photos were not very successful due to the fundamental difference between face photos and face sketches. Recently, Markov Random Fields (MRF) model was introduced to solve this problem [16], and this brought about great progress to face
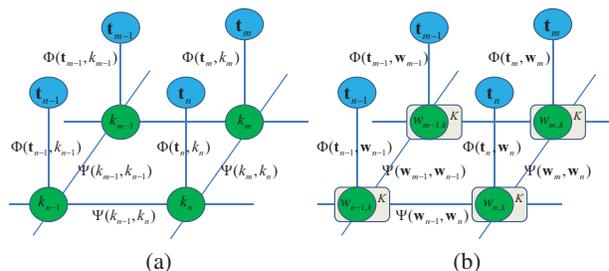

Figure 1. Graphical representations of the MRF and MWF models. (a) MRF model [16, 19]: each node in the lower sketch layer corresponds to a single variable (i.e., the index of a single candidate sketch patch). (b) Our proposed MWF model: each node in the sketch layer corresponds to a list of variables (i.e., a list of weights for the candidate sketch patches).

sketch synthesis. Commonly, face photos and sketches are divided into patches, and the "best" sketch patch for representing a test photo patch will be selected via solving a MRF. Such methods [16, 19], however, cannot synthesize new sketch patches as they only represent a target patch by the best sketch patch from the training data. As a result, they sometimes cannot produce good results for features like eyes and mouths which vary a lot from person to person, especially when the number of subjects in the training data is small. Moreover, the optimization problem involved in solving the MRF is NP-hard, and the commonly used Belief Propagation (BP) [7, 18] algorithm cannot guarantee to give an optimal solution.

In this paper, a new face sketch synthesis method is proposed to overcome the aforementioned limitations. We first propose a novel MRF model that is capable of synthesizing new sketch patches. Unlike the commonly used MRF model [16, 19] in which each node in the sketch layer corresponds to a single variable (i.e., a single candidate sketch patch), each node in the sketch layer in our model corresponds to a list of variables (i.e., a list of weights for the candidate sketch patches), and a target sketch patch is represented by a linear combination of some candidate s-

ketch patches (see Figure 1). We hence call our model the Markov Weight Fields (MWF) model. MWF model is superior to the commonly used MRF model in that it can synthesize new sketch patches and can be formulated into a convex Quadratic Programming (QP) problem to which the optimal solution is guaranteed. Note that, being a large scale QP problem, MWF model still cannot be solved easily by off-the-shelf optimization algorithms. By exploiting the Markov property of our model, we propose a cascade decomposition method (CDM) to decompose the original large scale QP problem into a number of small conditionally independent QP problems, each of which can be solved by some common optimization algorithms, resulting in a highly parallelizable computing framework.

The contributions of this paper are: (1) Proposing a MWF model which is capable of synthesizing new sketch patches, and can be formulated into a convex QP problem to which the optimal solution is guaranteed; (2) Proposing a cascade decomposition method to solve the large scale QP problem based on the Markov property of MWF.

2. Related Work

Most of the research studies in face sketch synthesis focus on two kinds of sketches, namely profile sketches [4, 17] and shading sketches [3, 8, 12, 15, 16, 19]. Compared with profile sketches, shading sketches are more expressive and popular. By assuming a linear transformation between face photos and face sketches, Tang and Wang [15] proposed to compute a global eigentransformation for synthesizing face sketches from face photos. Face photos and face sketches are, however, two different modalities, and therefore cannot be simply explained by a linear transformation. To account for the non-linearity, Liu *et al*. [12] adopted a patch-based approach. Their method divides a face photo into some overlapping patches and represents each target sketch patch by a linear combination of some candidate sketch patches. The drawback of [12] is that it synthesizes each sketch patch independently, and therefore large scale structures spanning multiple patches cannot be synthesized well. Wang and Tang [16] introduced the MRF model to tackle the problem of preserving large scale structures across sketch patches. Their method synthesizes a face sketch by selecting the "best" candidate patches that maximize the a posteriori estimation of their MRF model. Zhang *et al*. [19] further extended the work of Wang and Tang [16] by introducing shape priors specific to facial components and using SIFT feature [13] as a descriptor of each photo and sketch patch. Their method was proven to be robust against illumination changes and pose variations. Computing the maximum a posteriori of their MRF model, however, can only produce an approximate solution, and selecting the "best" candidate sketch patch cannot synthesize new patches that do not exist in the training data. In contrast, our MWF model is capable of preserving large scale features as well as synthesizing new sketch patches. Besides, it can also be formulated into a convex QP problem to which the optimal solution is guaranteed.

Large scale QP problems have been well studied in computer vision because of its important applications [9, 14]. A popular approach is decomposing the original large scale problem into smaller sub-problems. Osuna *et al*. [14] proposed an iterative method in which variables are decomposed into a working set and a fixed set, and only those in the working set are being optimized in each iteration. Their method determines the working set arbitrarily, and only one variable in the working set will be replaced after each iteration. Joachims [9] proposed an algorithm to select a good working set by finding the deepest feasible direction. As their objective function was too complicated, they used a first order approximation to formulate the problem [2]. Different from [9, 14], we propose a cascade decomposition method based on the Markov property of our model. In this method, variables are divided into two working sets whose memberships are fixed throughout the whole optimization process, and are optimized alternatively. By carefully selecting these two sets, each set can be further decomposed into many smaller independent QP problems, each of which can be optimized efficiently.

3. Markov Weight Fields

Consider a training data set consisting of M face photo sketch pairs. We divide each photo into N overlapping $l \times l$ patches and represent each of them by a L-vector where $L = l^2$. Likewise, we divide each sketch into N patches and represent each of them by a L-vector. Now given a test face photo, we divide it into N patches in exactly the same way we have done for the training data. Let $\mathbf{t}_i$ denote the L-vector of the ith test patch, where $i \in 1 \ldots N$. For each test patch, we find K candidate photo patches from the training data that are "closest" to it in terms of Euclidean distance between their L-vectors. Let $\mathbf{p}_{i,k}$ and $\mathbf{s}_{i,k}$ denote the L-vectors of the kth candidate photo patch and its corresponding sketch patch, respectively, for the ith test patch. We assume a test photo patch and its target sketch patch can be represented by the same linear combination of the K candidate photo patches and sketch patches respectively. The rationale behind this has already been discussed in [12].

Figure 1 (b) shows a graphical representation of the proposed MWF model. Similar to the MRF model in [16, 19], our MWF model consists of two layers, namely the upper photo layer and the lower sketch layer. As in [16, 19], each node in the photo layer corresponds to a test photo patch. Unlike [16, 19] where each node in the sketch layer corresponds to (the index of) a single candidate sketch patch, each node in our sketch layer corresponds to a list of weights for the K candidate sketch patches, denoted by the K-

vector $\mathbf{w}_i$ with elements $w_{i,k}$ being the weight of the kth candidate sketch patch for the ith test patch. The joint probability of $\mathbf{t}_i$ and $\mathbf{w}_i$, $\forall i \in \{1, \ldots, N\}$, is given by

$$p(\mathbf{t}_1, ..., \mathbf{t}_N, \mathbf{w}_1, ..., \mathbf{w}_N)$$
$$\propto \prod_{i=1}^{N} \Phi(\mathbf{t}_i, \mathbf{w}_i) \prod_{(i,j)\in\Xi} \Psi(\mathbf{w}_i, \mathbf{w}_j), \quad (1)$$

where

$$\Phi(\mathbf{t}_i, \mathbf{w}_i) = \exp\{-||\mathbf{t}_i - \sum_{k=1}^{K} w_{i,k}\mathbf{p}_{i,k}||^2/2\sigma_D{}^2\} \quad (2)$$

and

$$\Psi(\mathbf{w}_i, \mathbf{w}_j)$$
$$= \exp\{-||\sum_{k=1}^{K} w_{i,k}\mathbf{o}_{i,k}^j - \sum_{k=1}^{K} w_{j,k}\mathbf{o}_{j,k}^i||^2/2\sigma_S{}^2\} (3)$$

Here $(i, j) \in \Xi$ means the ith and jth patches are neighbors. $\mathbf{o}_{i,k}^j$ denotes the overlapping area of the kth candidate for the ith sketch patch with the jth patch. The posterior probability can be written as

$$p(\mathbf{w}_1, ..., \mathbf{w}_N | \mathbf{t}_1, ..., \mathbf{t}_N)$$
$$= \frac{1}{Z} p(\mathbf{t}_1, ..., \mathbf{t}_N, \mathbf{w}_1, ..., \mathbf{w}_N), \quad (4)$$

where $Z = p(\mathbf{t}_1, ..., \mathbf{t}_N)$ is a normalization term. By maximizing the posterior probability, the best weight for each of the candidate sketch patch can be achieved. This is equivalent to minimizing the following cost function

$$\min_{\mathbf{W}} \sum_{i=1}^{N} ||\mathbf{t}_i - \mathbf{P}_i\mathbf{W}||^2 + \alpha \sum_{(i,j)\in\Xi} ||\mathbf{O}_i^j\mathbf{W} - \mathbf{O}_j^i\mathbf{W}||^2. \quad (5)$$

Here $\alpha = \sigma_D{}^2/\sigma_S{}^2$. $\mathbf{P}_i$ and $\mathbf{O}_i^j$ are two matrices, with the $((i-1)K+k)$th column being $\mathbf{p}_{i,k}$ and $\mathbf{o}_{i,k}^j$, respectively, and all other columns being zero vectors. $\mathbf{W}$ is a NK-vector, with the $((i-1)K+k)$th element being $w_{i,k}$. (5) can be formulated into a standard QP problem:

$$\min_{\mathbf{W}} \mathbf{W}^{\mathrm{T}}\mathbf{Q}\mathbf{W} - 2\mathbf{W}^{\mathrm{T}}\mathbf{H} + \mathbf{V}. \quad (6)$$

where

$$\mathbf{Q} = \sum_{i=1}^{N} \mathbf{P}_i^{\mathrm{T}}\mathbf{P}_i$$
$$+ \alpha \sum_{(i,j)\in\Xi} (\mathbf{O}_i^j - \mathbf{O}_j^i)^{\mathrm{T}}(\mathbf{O}_i^j - \mathbf{O}_j^i), \quad (7)$$

$$\mathbf{H} = \sum_{i=1}^{N} \mathbf{P}_i^{\mathrm{T}}\mathbf{t}_i, \quad (8)$$

$$\mathbf{V} = \sum_{i=1}^{N} \mathbf{t}_i^{\mathrm{T}}\mathbf{t}_i. \quad (9)$$

In order to make this optimization trackable, we enforce $\mathbf{W}$ to satisfy the constraint $\sum_{k=1}^{K} w_{i,k} = 1$ and $w_{i,k} \geq 0$ for all i, k. As the term $\mathbf{V}$ in (6) has no effect on the minimization, we can ignore this term in the final formulation:

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathbf{W}^{\mathrm{T}}\mathbf{Q}\mathbf{W} - 2\mathbf{W}^{\mathrm{T}}\mathbf{H} \\ s.t. \quad & \mathbf{A}\mathbf{W} = \mathbf{b} \\ & w_{i,k} \geq 0, \quad \forall i \in \{1, \ldots, N\}, k \in \{1, \ldots, K\}. \end{aligned}$$
$$(10)$$

Here $\mathbf{A}$ is a $N \times NK$ matrix. For the ith row of $\mathbf{A}$, the elements from $((i-1)K+1)$ to $(i \times K)$ are 1 and all others are 0. $\mathbf{b}$ is a N-vector with all elements being 1. (10) is a standard QP problem and as $\alpha > 0$,(5), (6) and (10) are convex. Our model, as a result, has a very good property that the global optimum can be achieved.

4. Cascade Decomposition Method

Directly solving (10) using off-the-shelf optimization algorithms is intractable due to the extremely high dimension of $\mathbf{W}$. For instance, the number of variables will be more than $19,000$ in our problem which is too large to be optimized simultaneously. Inspired by the work of [9, 14] and considering the structure of MWF, we propose an efficient way to solve this large scale optimization problem which we call cascade decomposition method.

In our proposed method, variables are divided into two sets, namely $\mathbf{W}_B$ and $\mathbf{W}_F$, and $\mathbf{Q}$, $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{b}$ are divided accordingly:

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_B \\ \mathbf{W}_F \end{pmatrix},$$
$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{BB} & \mathbf{Q}_{BF} \\ \mathbf{Q}_{FB} & \mathbf{Q}_{FF} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_B \\ \mathbf{H}_F \end{pmatrix},$$
$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_B & 0 \\ 0 & \mathbf{A}_F \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_B \\ \mathbf{b}_F \end{pmatrix}. \quad (11)$$

Since $\mathbf{Q}$ is symmetric, $\mathbf{Q}_{FB}{}^{\mathrm{T}} = \mathbf{Q}_{BF}$, and the original optimization problem can be written as

$$\begin{aligned} \min_{\mathbf{W}_B,\mathbf{W}_F} \quad & \mathbf{W}_B^{\mathrm{T}}\mathbf{Q}_{BB}\mathbf{W}_B - 2\mathbf{W}_B^{\mathrm{T}}\mathbf{H}_B + 2\mathbf{W}_B^{\mathrm{T}}\mathbf{Q}_{BF}\mathbf{W}_F \\ & + \mathbf{W}_F^{\mathrm{T}}\mathbf{Q}_{FF}\mathbf{W}_F - 2\mathbf{W}_F^{\mathrm{T}}\mathbf{H}_F \\ s.t. \quad & \mathbf{A}_B\mathbf{W}_B = \mathbf{b}_B \\ & \mathbf{A}_F\mathbf{W}_F = \mathbf{b}_F \\ & w_{i,k} \geq 0, \quad \forall i \in \{1, 2, ..., N\}, k \in \{1, 2, ..., K\}. \end{aligned}$$
$$(12)$$

Notice that if we fix one set of the variables and optimize (12) with respect to the other, the problem is still convex. We therefore propose to solve this problem by optimizing $\mathbf{W}_B$ and $\mathbf{W}_F$ alternatively.

Nonetheless, we still need to solve two large scale optimization problems. Fortunately, by taking the advantage of the structure of MWF, these two sub-problems can be solved by optimizing many independent smaller problems if $\mathbf{W}_B$ and $\mathbf{W}_F$ are chosen carefully. According to the Markov property of the MWF model, if all the neighbors of $\mathbf{w}_i$ are known, $\mathbf{w}_i$ will be independent of all the other nodes in the sketch layer. Based on this observation we choose $\mathbf{W}_B$ and $\mathbf{W}_F$ as follows:

$$\mathbf{W}_B = \{\mathbf{w}_{i(p,q)}|(p+q)\%2 = 0\}, \qquad (13)$$
$$\mathbf{W}_F = \{\mathbf{w}_{i(p,q)}|(p+q)\%2 = 1\}. \qquad (14)$$

Here $i(p,q)$ denotes the index of the node located at the pth row and qth column of the MWF. By dividing the nodes in this way, every node in $\mathbf{W}_B$ will be independent of each other when $\mathbf{W}_F$ is fixed, and vice verse. This allows a further decomposition of $\mathbf{W}_B$ and $\mathbf{W}_F$ into many smaller sets, each of which only contains one single node (i.e., $\mathbf{w}_i$). The number of variables in each optimization problem is then reduced to K. The final optimization problem we need to solve becomes:

$$\min_{\mathbf{w}_i} \quad \mathbf{w}_i{}^T \mathbf{Q}_{ii} \mathbf{w}_i + 2\mathbf{w}_i{}^T (\sum_{a=1}^{4} \mathbf{Q}_{ia} \mathbf{w}_{ia} - \mathbf{H}_i)$$
$$s.t. \quad \mathbf{A}_i \mathbf{w}_i = 1$$
$$w_{i,k} \geq 0, \ \forall k \in \{1, 2, ..., K\}. \qquad (15)$$

Here $\mathbf{w}_{ia}$ is the neighbor node of $\mathbf{w}_i$, and $\mathbf{Q}_{ii}$, $\mathbf{Q}_{ia}$, $\mathbf{H}_i$ and $\mathbf{A}_i$ are the decomposition of $\mathbf{Q}$, $\mathbf{H}$ and $\mathbf{A}$ according to $\mathbf{w}_i$ and $\mathbf{w}_{ia}$ respectively.

Algorithm 1 Cascade Decomposition Method

Initialize $\mathbf{W}$ according to the distances between the test and candidate photo patch vectors.
repeat

Step 1   For every node $\mathbf{w}_i \in \mathbf{W}_B$
           compute $\mathbf{Q}_{ii}, \mathbf{A}_i, \mathbf{Q}_{ia}, \mathbf{H}_i$ and $\mathbf{w}_{ia}$,
           optimize (15) with respective to $\mathbf{w}_i$.

Step 2   For every node $\mathbf{w}_i \in \mathbf{W}_F$
           compute $\mathbf{Q}_{ii}, \mathbf{A}_i, \mathbf{Q}_{ia}, \mathbf{H}_i$ and $\mathbf{w}_{ia}$,
           optimize (15) with respective to $\mathbf{w}_i$.
until $\mathbf{W}$ does not change any more.

The process of cascade decomposition method is summarized in **Algorithm 1**. Each optimization procedure within Step 1 and Step 2 of **Algorithm 1** can actually be computed in parallel as they are independent of each other. This makes the algorithm very efficient. We find that our algorithm will normally converge within 10 iterations.

# 5. Experiments

5.1. Implementation Details

A coarse to fine approach is adopted. Firstly in the coarse step, face photos are divided into patches with a size of 20×20, and adjacent patches have a 5-pixel-wide overlapping area. Coarse face sketches are synthesized using the proposed MWF method. Next in the fine step, face photos are divided into smaller patches with a size of 10×10, and again adjacent patches have a 5-pixel-wide overlapping area. An additional smoothing term is included in the objective function of the fine step to enforce the synthesized sketch patches being consistent with the corresponding patches synthesized in the coarse step. For each test photo patch, $K = 10$ candidates are selected from the training data. Since photos have been aligned by the positions of the eyes, to save computation time, candidates of a test patch are selected from within a 30×30 local region around it in the coarse step, and a 26×26 local region in the fine step. In both steps, α is set to 0.25. In the fine step, the weight of the smoothing term for ensuring the consistency of the sketches synthesized in the two steps is set to 0.08. A minimum error boundary method [6] is used to find the cutting path between the overlapping area of two adjacent patches. Like [16], we use *Luv* color to form the vectors representing the photo and sketch patches.

5.2. Experimental Results

We validated our method using the CUHK database [16] which contains 188 photo sketch pairs. In all the experiments, 88 subjects from the database were selected as the training data, and the rest 100 were used as test cases. Besides, we tested our method using photos with pose variations and photos of some Chinese celebrities. We compared our results with those produced by other state-of-the-art methods to demonstrate the high quality of our results.

In Figure 2, we compare our results with those published in [12]. As pointed out in [16], the results of [12] cannot preserve large scale features, whereas our results do not suffer from such a problem.

Figure 3 shows the comparison of our results with those synthesized by the MRF method proposed in [16]. It can be seen that their method cannot synthesize the eyes and mouths very well. One reason is that eyes and mouths are two distinctive features of human face, and they vary a lot from person to person. Therefore, there is a high chance that the MRF method cannot find a suitable patch in the training data for some target patches containing eyes or mouths. Another reason is that the formulation of their energy function is NP-hard, and the BP algorithm used in their method cannot guarantee to give the optimal solution. This implies that even if there exists a suitable patch in the training data for a target patch, there is still a chance that their method
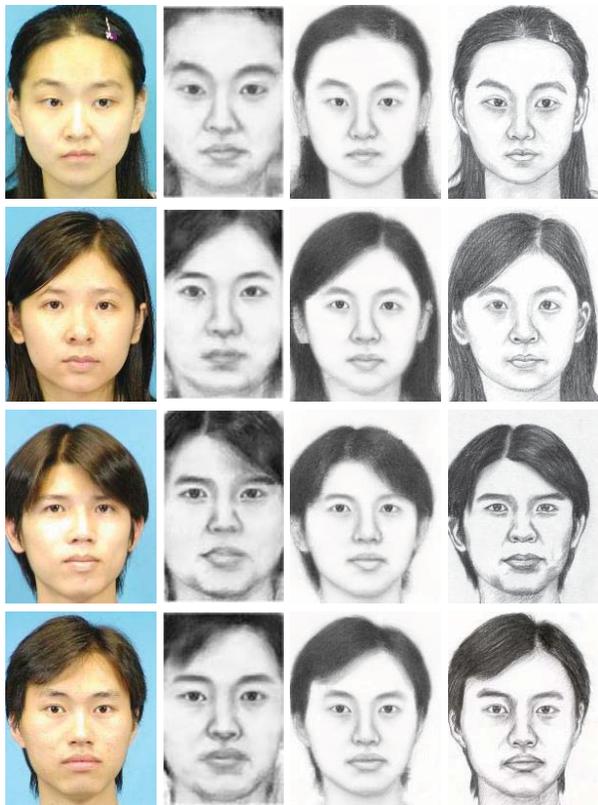
| (a) Photos | (b) Liu *et al.* [12] | (c) Ours | (d) Artist |

Figure 2. Face sketch synthesis results.

| (a) Photos | (b) Wang and Tang [16] | (c) Ours | (d) Artist |

Figure 3. Face sketch synthesis results.

may not be able to find that patch. Unlike the MRF method, our MWF model uses a linear combination of 10 candidate sketch patches to synthesize the target patch, and can generate new patches that do not exist in the training data. To give a quantitative evaluation, we compare the sketches synthesized by these two methods using the rank-1 and rank-10 recognition rates[1] on the 100 test cases. We simply used P-CA [5] as the recognition method, and the sketches drawn by the artist were used as the training data for computing the projection matrix. Figure 4 shows the recognition results for the sketches produced by these two methods. We can see that the recognition rates for the sketches generated by our MWF model are higher than those by the MRF model.

Although our method spares no special effort in dealing with pose variation, we also tested it on the pose variation data set. For these photos, aligning the positions of the eyes cannot align the faces very well. Hence, the candidate sketch patches of a target patch were selected from within a larger local region (60×60 and 50×50 in the coarse and

---
[1]Although recognition rates cannot reflect the visual quality of synthesized sketches completely, it is the only way we know to evaluate the quality of synthesized sketches quantitatively [19].

fine steps respectively). Figure 5 shows some synthesized face sketches under pose variation. We find that our results are comparable to or even better than those produced by the method in [19], which is specially designed to tackle pose variation. The method proposed in [19] uses the same MRF model as in [16], and thus it suffers from the same problem of not being capable of synthesizing new patches. This can be seen in the sketches of the first and second subjects in Figure 5 (the eyes synthesized by their method are not as good as ours).

We also tested our method on some photos of Chinese celebrities. As the colors in these photos are quite different from those in the training data, we simply converted all the photos into gray-scale, and directly used the gray level intensity to form the vectors representing the photo patches. From Figure 6, it can be seen that our results look more natural than those produced by [16] and [19]. The reason behind this is that these photos are quite different from those in the training data, and it is very hard to find suitable patches for the target patches. By using a linear combination of candidate sketch patches, our method can synthesize some new patches which can better represent the target patches.

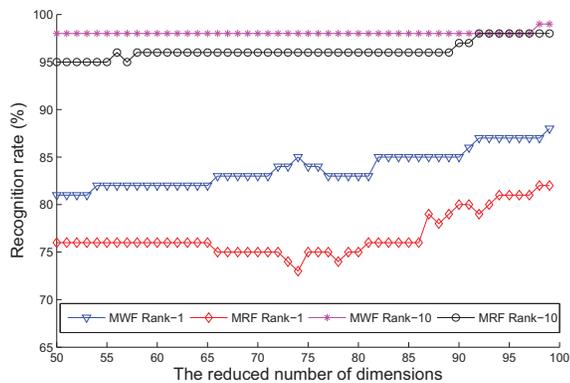By interchanging the roles of photos and sketches, our

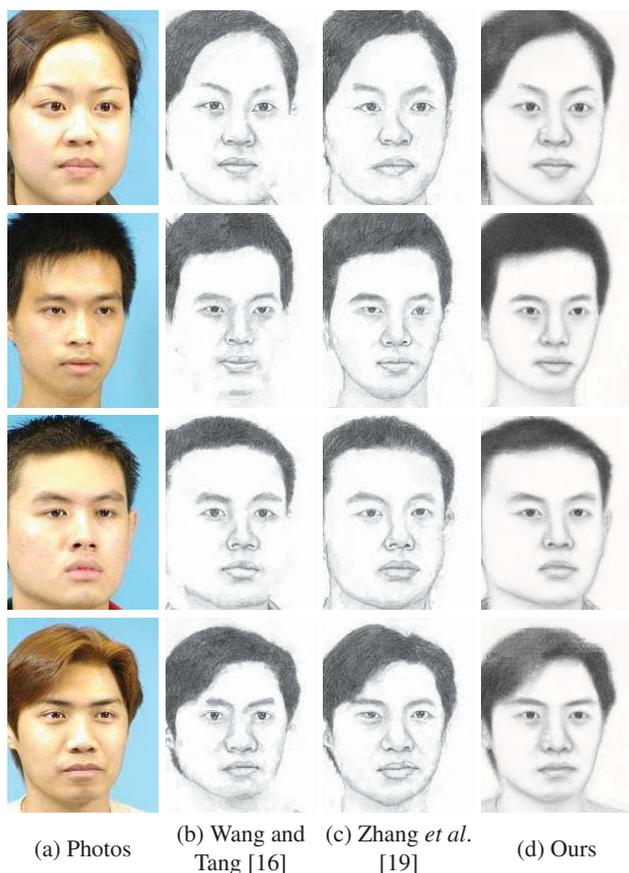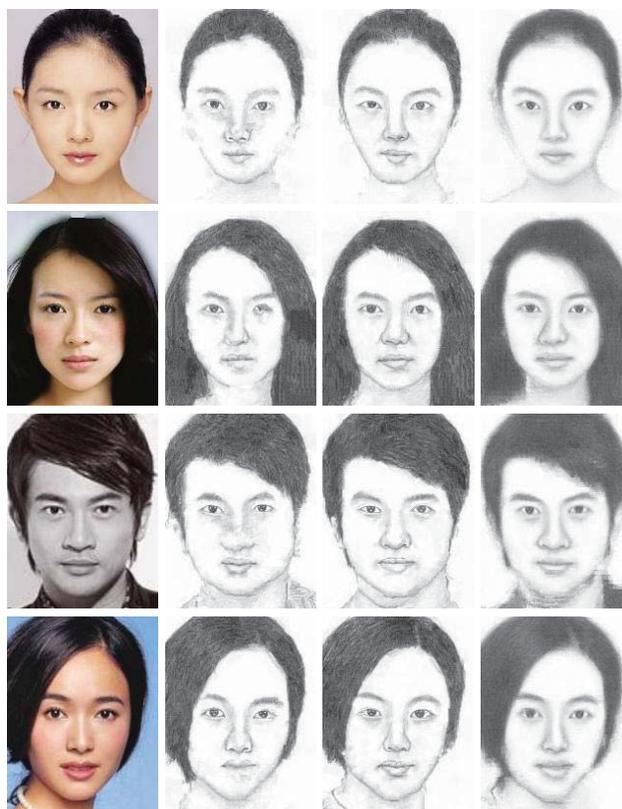Figure 4. Rank-1 and Rank-10 recognition rates using PCA.

(a) Photos   (b) Wang and   (c) Zhang *et al*.   (d) Ours
Tang [16]   [19]

Figure 5. Synthesized sketches with pose variation.

(a) Photos   (b) Wang and   (c) Zhang *et al*.   (d) Ours
Tang [16]   [19]

Figure 6. Synthesized sketches for Chinese celebrities.

method can also be used to synthesize photos from sketches. Figure 7 shows the photos generated by our method using the sketches drawn by artist as input. It can be seen that the synthesized photos are very similar to the original photos.

6. Discussion

In this paper, we proposed a new MRF model which we call Markov Weight Fields for face sketch synthesis. Different from the commonly used MRF model, our model is capable of synthesizing new sketch patches as a linear combination of candidate sketch patches. Besides, our model can be formulated into a convex quadratic programming problem which is guaranteed to have an optimal solution. An efficient cascade decomposition method was proposed to solve this large scale QP problem which can often converge within 10 iterations. A large number of experiments proved that our method outperforms other state-of-the-art methods, especially when there is no suitable sketch patch in the training data for a target patch.

Acknowledgments

References

[1] S. Baker and T. Kanade. Hallucinating faces. In *Proc. of Automatic Face and Gesture Recognition*, pages 83–88, 2000.

[2] L. Bottou and C.-J. Lin. Support vector machine solvers. In L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, editors, *Large Scale Kernel Machines*, pages 301–320. MIT Press, Cambridge, MA., 2007.
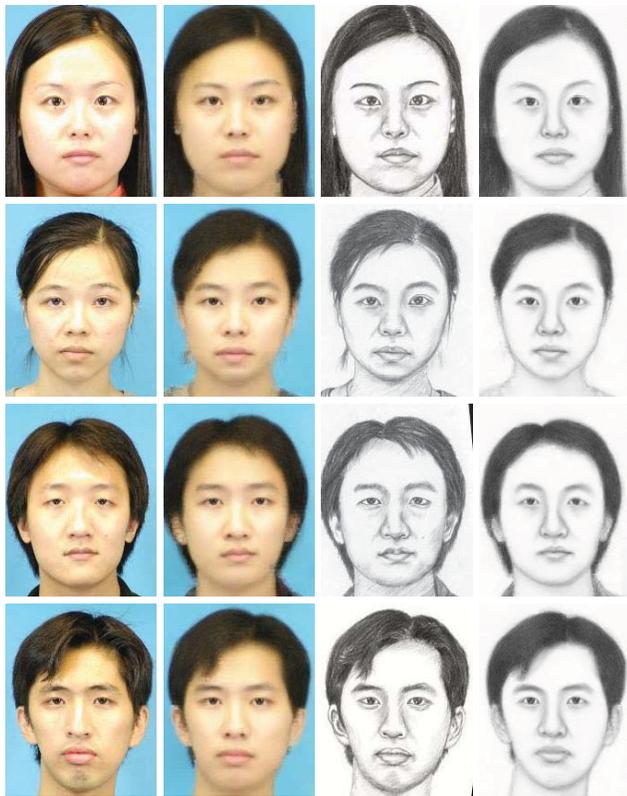
| | | | |
| (a) Photos | (b) Synthesized photos | (c) Sketches drawn by artist | (d) Synthesized sketches |

Figure 7. Synthesized sketches and photos.

[3] L. Chang, M. Zhou, Y. Han, and X. Deng. Face sketch synthesis via sparse representation. In *Proc. of International Conference on Pattern Recognition*, pages 2146 –2149, 2010.

[4] H. Chen, Y.-Q. Xu, H.-Y. Shum, S.-C. Zhu, and N.-N. Zheng. Example-based facial sketch generation with non-parametric sampling. In *Proc. of International Conference on Computer Vision*, volume 2, pages 433 –438, 2001.

[5] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, 15:252–260, 2005.

[6] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *Proc. of SIGGRAPH*, pages 341–346, 2001.

[7] W. Freeman and E. Pasztor. Learning low-level vision. In *Proc. of International Conference on Computer Vision*, volume 2, pages 1182 –1189, 1999.

[8] X. Gao, J. Zhong, J. Li, and C. Tian. Face sketch synthesis algorithm based on E-HMM and selective ensemble. *IEEE Transactions on Circuits and Systems for Video Technology*, 18:487 –496, April 2008.

[9] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999.

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

[11] C. Liu, H.-Y. Shum, and C.-S. Zhang. A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages 192 – 198, 2001.

[12] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages 1005 – 1010, 2005.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[14] E. Osuna, R. Freund, and F. Girosi. An improved training algorithm for support vector machines. In *Proc. of Neural Networks for Signal Processing VII, IEEE Workshop*, pages 276 –285, Sep. 1997.

[15] X. Tang and X. Wang. Face sketch synthesis and recognition. In *Proc. of Internatinoal Conference on Computer Vision*, volume 1, pages 687 –694, 2003.

[16] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1955 –1967, 2009.

[17] Z. Xu, H. Chen, S.-C. Zhu, and J. Luo. A hierarchical compositional model for face representation and sketching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:955 –969, 2008.

[18] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, pages 239–269, 2003.

[19] W. Zhang, X. Wang, and X. Tang. Lighting and pose robust face sketch synthesis. In *Proc. of European Conference on Computer Vision*, pages 420–433, 2010.

[20] M. Zhao and S.-C. Zhu. Portrait painting using active templates. In *Proc. of the ACM SIGGRAPH/Eurographics*, pages 117–124, 2011.