Learning to sketch human facial portraits using personal styles by case-based reasoning

Bingwen Jin1 · Songhua Xu2 · Weidong Geng1

Abstract This paper employs case-based reasoning (CBR) to capture the personal styles of individual artists and generate the human facial portraits from photos accordingly. For each human artist to be mimicked, a series of cases are firstly built-up from her/his exemplars of source facial photo and hand-drawn sketch, and then its stylization for facial photo is transformed as a style-transferring process of iterative refinement by looking-for and applying best-fit cases in a sense of style optimization. Two models, fitness evaluation model and parameter estimation model, are learned for case retrieval and adaptation respectively from these cases. The fitness evaluation model is to decide which case is best-fitted to the sketching of current interest, and the parameter estimation model is to automate case adaptation. The resultant sketch is synthesized progressively with an iterative loop of retrieval and adaptation of candidate cases until the desired aesthetic style is achieved. To explore the effectiveness and advantages of the novel approach, we experimentally compare the sketch portraits generated by the proposed method with that of a state-of-the-art example-based facial sketch generation algorithm as well as a couple commercial software packages. The comparisons reveal that our CBR based synthesis method for facial portraits is superior both in capturing and reproducing artists’ personal illustration styles to the peer methods.

Keywords Facial portraits · Sketch generation · Style-transferring · Artistic rendering · Personalized exaggeration · Case-based reasoning

Bingwen Jin
jinbw@zju.edu.cn

Weidong Geng
gengwd@zju.edu.cn

Songhua Xu
songhua.xu@njit.edu

1 Department of Computer Science and Technology, Zhejiang University, Hangzhou, China

2 Department of Information Systems, New Jersey Institute of Technology, Newark, NJ, USA
1 Introduction

Image-based artistic rendering (IB-AR) algorithms often rely on manually-encoded heuristics to emulate specific artistic illustration styles [17]. These heuristics can be classified into three categories [17]: stroke-based rendering for image approximation [24], region-based techniques [52], and image processing (filtering) [31, 32]. However, it is difficult for artist to explicitly depict the rules of artistic illustration because these rules are often sub-consciously exercised and are not always expressible verbally or symbolically. Therefore, example-based artistic rendering (EBAR) is proposed to generate stylized sketch, which is more popular and more effective since the styles are implicitly embodied in the examples, as it is natural to specify artistic styles by showing a set of examples [10].

Existing EBAR methods can be broadly classified as model-based and model-free methods. The model-based methods embed prior knowledge about artistic rendering styles into generative models that carry a set of process control parameters. For example, Lu et al. [23] built-up parametric histogram models on the tone distribution inside the sketch example. The resultant sketch was then generated via tone-adjusting by this model according to the tone map in the source image. Tu et al. [45] learned novel direct combined models from the joint distributions of the feature vector pairs of input facial image vs neutral facial shape, neutral vs exaggerated facial shape, and facial sketch vs combination of input facial image and exaggerated facial shape, and then the system optimally synthesized the corresponding output sketch by applying the MMSE criterion [50] within the combined eigenspace. The model-free methods synthesize the resultant style-preserving illustrations directly by the correlation relationship between pixels, patches, or strokes etc., in the exemplars. Wang et al. [48] divided facial region into overlapping patches from the source photo, and for each patch, located a similar photo patch from examples and collected its corresponding sketch for the synthesis of the resultant sketch smoothly. Zhao et al. [53] built a dictionary of stroke templates of oil painting portrait with complete information of artist stroke-by-stroke drawing processes. Their method painterly rendered new portrait by reusing brush strokes from a matched template in the dictionary in terms of facial shape and color in the source image. Model-based methods can generate diverse new illustrations with parametric styles through generalization on exemplars. However subtle and unique artistic characteristics related to individual artist are somewhat lost while modelling the stylization. Model-free methods naturally incorporate the visual correlation in examples into the resultant sketch. But the richness of resulting sketches are usually limited by that in the given examples.

In the community of artificial intelligence, case-based reasoning (CBR) solves a new problem instance by first recalling one or multiple similar, previously solved problem instances and then reapplying the known solutions, often with adaptations according to the new context, to address the new problem instance [1]. From the point of the view of problem solving, the exemplars, a photo and its corresponding artistic illustration, in EBAR are very similar to the cases in CBR. This motivates us to employ case-based reasoning to generate the stylized human facial portraits for an individual artist. Figure 1 shows two facial sketches automatically created by us using the styles learned from two artists respectively. Sketches generated by two commercial software packages are also presented for comparison.

From a technical perspective, the major challenges for CBR-based stylization of portrait lie in case retrieval and adaptation respectively. Case retrieval in stylization is complex, and goes beyond selecting a best example for a given input as that in [48], because portrait sketching practice of human artists is usually a multi-step and progressive painting process, in which the case retrieval may occur in multiple steps, accounting for not only the given
Fig. 1  Comparison between facial portraits in sketching generated by two commercial software packages (Akvis Sketch [2] and a plug-in for Microsoft Outlook 2003 [26]), two human artists, and our method mimicking the illustration styles of the two artists respectively. In this experiment, our method first learns the personal sketching styles of the two human artists and then generates the sketch portraits for the same input facial photo using the two captured illustration styles respectively.

input, but also the current interim sketch and the desired resulting sketch in his/her mind. However, during the phase of looking for the suitable cases, the expected resultant sketch is not available yet for the time being. Case adaptation for stylization is to transfer the style-related correlations in the selected cases into the current sketching. Although existing style-transferring approaches [45, 48, 53] can be utilized, its key-point is how to automate the case adaptation for stylization, especially when multiple cases are applicable, since automatic case adaptation is necessary to enable CBR systems to function autonomously and to serve naive as well as expert users [18].

We address these challenges with an approach specific to personalized stylization for sketching faces. First, we designed an iterative pipeline as an overall algorithmic mechanism for CBR-based stylization, in which the current sketch is refined iteratively by the best-fit case until the desired sketch appears. Second, motivated by human artists practice of forming a desired sketch in mind before drawing, we train-up a predictive model from cases for each artist to hypothetically create them. The model evaluates the fitness of each candidate case in terms of the visual similarity to the presumed “resultant sketch”. The best-fit one is selected to sketch current face. Furthermore, in order to automate the case adaptation, a parameter estimation model is learned for each artist in advance, which will automatically assign the appropriate values to parameters for case adaptation, while getting a new case for refinement of resulting sketch.

To summarize, our paper makes the following contributions:

– A novel case-based reasoning algorithmic pipeline to iteratively stylize the human portraits by exemplars, producing rich portraits and preserving personal styles well.
– A predictive framework integrated with generate-and-evaluate mechanism is newly proposed for best-fit case retrieval with capability of hypothetically creating the desired sketch in the mind of artist, which greatly facilitate the evaluation of best-fit cases.
– Innovatively learning parameter estimation model for each artist which enables an automatic case adaptation for stylized sketching of facial photo.

In this paper, we focus on generating facial sketch drawing without hair, which is still very important and challenge. Since one’s hair style can be changed easily, human face is the key to represent the characteristics of an individual person’s head. Facial sketch synthesis without hair is challenge. For non-exaggerated face sketch, as face are very familiar to us, we notice small deviations from our concept of how a face should appear [33]. For exaggerated facial sketch, although artist actively introduce facial shape deviations in drawings,
many rules still should be carefully followed [36]. Otherwise the exaggerated face may not be recognizable.

2 Related work

A multitude of image-based artistic rendering (IB-AR) techniques have been proposed. Kyprianidis et al. [17] gave an in depth survey of IB-AR techniques. Here we merely focus on example-based artistic rendering (EBAR), which can be roughly categorized into two classes: model-based and model-free methods.

Model-based methods acquire prior knowledge of artistic rendering styles from examples and accordingly built-up the stylization models. Besides the model-based approaches in [23, 45] (see Section 1), Reinhard et al. [37] modeled color style of a source image through the means and standard deviations along each of the three axes in $l\alpha\beta$ color space, and then imposed the means and standard deviations onto the target image, transferring the color style to target image. Liang et al. [21] models facial pexaggeration style through analyzing the correlation between the image caricature pairs using partial least-squares (PLS). The model-based hatching in [14] trained a mapping from the features of input 3D object to hatching properties. A new illustration was generated in target style according to predicted properties. The aforementioned model-based methods can generate diverse new illustrations with parametric styles through generalization on exemplars. However subtle and unique artistic characteristics related to individual artist are somewhat lost while modelling the stylization.

Model-free methods generate new artistic illustrations directly by reusing the correlation relationship provided by all exemplars. Hertzmann et al. [13] proposed image analogies algorithm which reuse a source image A and artistic depiction of that image A’ to synthesize an artistic illustration B’ for a new image B in pixel level. Each new pixel is synthesized by reusing the pixel in exemplar that best matches the pixel being synthesized. Later, an extension of this algorithm incorporated image gradient direction to better preserve object shape of the target image [19]. Example-based stippling [15] proposed a texture similarity metric based on gray-level co-occurrence matrix, aiming at generating stipple textures that are perceptually close to input samples. Their reuse of examples merely accounted for pixels related to stipple primitives. Besides pixel-level reusing of examples, patch-level reusing are also proposed, which often divide original image exemplars into patch exemplars [48]. Liu et al. [22] took the similar approach in [48], and retrieved multiple exemplars for each photo patch of a new face and synthesize a sketch patch by linearly blending sketch patched in the candidate exemplars. Wang et al. [47] improved this approach by defining a probabilistic model to optimize both the reconstruction fidelity of the input photo and the synthesis fidelity of the target output sketch. Song et al. [42] further extended it through a Spatial Sketch Denoising method. Moreover, the face image can also be decomposed into patches in terms of anatomical structure of human face as that in [10, 28, 49, 51]. However, the style of a rendition is mainly embodied in the strokes of a sketch, instead of pixels or patches in an image. Therefore stroke-based reusing in EBAR are also investigated [53]. Berger et al. [5] presented an approach which reuse real artists’ strokes to the image. To retain specific styles, they composed a stroke library for each artist, and “cloned” the relevant strokes that matched to the detected edges in the source image. This approach specialized in line-type sketch. It is difficult for this approach to generate rich texture in results. Our approach is different as we reuse sample sketch image patches to generate new sketch image patches and...
compose new facial sketch drawing. Hence, our approach is able to generate sketch drawings composed by both line and texture, which can be clearly noticed in our experiment results.

3 Method overview

Figure 2 gives an overview of our CBR-based sketch synthesis framework that produces visually superior results than existing synthesis methods. Given exemplars of source facial photo and stylized sketch hand-drawn by a human artist, its preprocessing phase for case-based reasoning proceeds in three steps: 1) we generate sketch Synthesis (STS) cases and construct a library for them; 2) Fitness Evaluation (FE) models for the artist are trained up from the STS cases; 3) Parameter Estimation (PE) models are learned for automatic case adaptation. During the runtime phase, we use STS cases, FE models, and PE models to iteratively synthesize a new facial sketch via CBR for each newly given facial photo.

The best-fit case retrieval is carried out by a predictive framework of generate-and-evaluate mechanism. From the point of view of style imitation, best-fit case should be the one that maximally preserves the human artist’s style in resulting sketch. However, the resulting sketch is to-be-generated, and sketch is not at-hand yet, which arises the problem of unknown “ground truth”. Our solution is to hypothesize there is always a ground truth in the mind of a human artist, while drawing the sketch. In fact a human artist can easily select the best-fit case by the desired sketch in his/her mind. Therefore we propose a predictive framework embedded with generate-and-evaluate mechanism, in which each candidate case
is implicitly applied on the sketch of current interest. FE model is trained to rank its fitness in terms of the similarity between the optimally generated sketch and the one mentally imagined in human artist’s mind.

Our case adaptation for sketch synthesis is implemented through a blending operator, whose parameter setting should maximize the similarity between generated sketch and its hypothesized groundtruth. An explicit parameter searching for optimization is time-consuming, therefore we learn a PE model to automatically configure the blending parameter, i.e. to generate a new sketch, we blend the stylized sketch from the best-fit case and the sketch of current interest with the parameter assigned by PE model automatically.

During the phase of iteratively synthesis, we firstly segment the entire facial region into multiple regions according to the anatomical structure of a human face (see Fig. 3 for facial regions painted with different colors). Then for each segmented region, the resultant sketch is synthesized progressively with an iterative loop of retrieval and adaptation of candidate cases until the desired aesthetic style is achieved. This is motivated by the human artist’s multi-step, progressive portrait sketching practice. FE model guides case retrieval, while PE model automates case adaptation. At last, the overall sketch for the entire human face is composed and formed globally from sketches of these local facial regions according to the relative spatial layout based on facial feature points.

Moreover, a significant characteristics of facial sketch is exaggeration. With our CBR’s framework, simulation of the exaggeration is easily embedded as a post-processing step. After our iterative synthesis process, the exaggeration field is generated from STS cases, and then applied on the synthesized sketch to imitate exaggeration style of a human artist.

4 Building up cases

To build up STS cases (see Fig. 4), we start from a set of paired source photo and its corresponding stylized sketch, $\Pi_l = \{(I_1, S_1), (I_2, S_2), \cdots, (I_i, S_i), \cdots\}$, where $(I_i, S_i)$ is the $i$-th sample composed of a front-view facial photo $I_i$ and its corresponding stylized sketch $S_i$. All sketches in $\Pi_l$ are illustrated using the same style $l$ as denoted by its subscript. The cardinality of $\Pi_l$, $|\Pi_l|$, is the number of sample pairs inside it.
A STS case consists of 6 components: source photo $I_i$, stylized sketch $S_i$, neutral sketch $S'_i$, exaggeration field $V_i$, and two feature point sets, $Q(I_i)$ and $Q(S_i)$, of $I_i$ and $S_i$ respectively. Each source photo merely has one human face in front view. Stylized sketch is hand-drawn by human artist. Neutral sketch is the one neutralized by removing the exaggeration from the stylized sketch. Neutral sketch will be identical to stylized sketch if there is no exaggeration in it. Exaggeration field is a 2D vector field representing the displacement of relevant pixels due to exaggeration. Each feature point set has 67 pre-defined points (see Fig. 3). One is from source photo, and the other one is from stylized sketch. Formally, we denote a STS case $r_i$ as:

$$r_i = (I_i, S_i, S'_i, V_i, Q(I_i), Q(S_i)).$$

(1)

A well-established practice of representing the geometry of a facial image is based on the spatial layout of facial feature points. The human facial portrait generation methods proposed by Chen et al. [10] and Zhao et al. [53] both use feature points to represent facial geometry. By this principle, the geometry of $I_i$ and $S_i$ are respectively represented by $Q(I_i)$ and $Q(S_i)$. $Q(I_i)$ is a set of 67 pre-defined facial feature points extracted from $I_i$ (see Fig. 3). Motivated by the usage of active shape model (ASM) used in the work proposed by Chen et al. [10], we also an active shape model (ASM) [27] to auto-locate the feature point positions in $I_i$. Each feature point is uniquely assigned to one of the nine aforementioned facial regions. Similarly, $Q(S_i)$ is a set of feature points extracted from $S_i$, whose initial positions are also determined by ASM [27]. However, facial elements sketched by artists sometimes deform geometrically and/or positionally from their counterparts in the corresponding facial photo, partly due to exaggeration. Meanwhile, some line-type sketch region also makes it difficult to locate feature points robustly. To get the correct positions for $Q(S_i)$, manual adjustment is often required. The index numbers of feature points in the two sets $Q(I_i)$ and $Q(S_i)$ as the same. That is, points with the same index number presumably depict the same location on a face, assuming a perfect positional alignment between $I_i$ and $S_i$.

Feature point position vector may not be a good choice for representing a contour curve since it is too sparse. However, sketching is a process discarding many details. In the generated and hand-drawn neutral sketch drawings, sketch contour curves usually do not follow real facial contours precisely due to error of our process of generating neutral facial drawing and artist’s estimation error in non-exaggerated sketch drawings. Using low-level dense
representation for contour curve or other methods which can capture curve shape precisely will capture the error at the same time. Nevertheless, feature points only capture some key features of facial shape, which is more intuitively suitable in our case. Meanwhile, feature point vector is widely used in NPR applications for facial shape representation, i.e. [10, 45, 49, 53].

More details about exaggeration field, features, and neutral sketch are given below.

4.1 Exaggeration field

Exaggeration field (EF) is a matrix of two-dimensional vectors. The dimension of the matrix is the same as the size of the EF’s corresponding image. Each pixel in the image has a counterpart two-dimensional vector that represents the pixel’s horizontal and vertical displacements from its position in the neutral sketch to the corresponding position in the deformed sketch with exaggeration.

The EF representing the exaggeration in $S_i$ is denoted as $V_i$. To generate $V_i$, the positions of feature points in $Q(S_i)$ and $Q(I_i)$ are aligned, and the geometric image transformation from $I_i$ to $S_i$ is computed by the image deformation algorithm based on Moving Least Squares (MLS) [39]. The user can also manually modulate the transformation if needed. In MLS algorithm, a denser set of displacement vectors $V_i$ are interpolated by the translations between feature points $Q(I_i)$ and $Q(S_i)$. In our prototype implementation, 50000 two-dimensional vectors will be derived in an EF for an image of 200 by 250 pixels.

4.2 Features

Given a facial image (sketch or photo) and its feature points, each pixel in the image inside the outline of facial feature points and the facial image is segmented into nine regions by its nearest pre-defined facial feature point by the $L_2$ distance. Let $\phi_i$ be one of the nine facial regions in $I_i$; $\psi_i$ be $\phi_i$’s corresponding area in $S_i$; $\text{region}(\cdot)$ be a region identification function whose input and output are a facial region and an index representing region type respectively. Without loss of generation, let $\phi_i$ be the 1st facial region in $I_i$, which means $\text{region}(\phi_i) = 1$. We denote $Q(\phi_i)$ and $Q(\psi_i)$ as the feature point sets of $\phi_i$ and $\psi_i$ respectively.

Let $\tau_{\text{photo}}(\phi_i, Q(\phi_i))$ be a vector of features for $\phi_i$, which includes four types of features, i.e. $\tau_{\text{photo}}(\phi_i, Q(\phi_i)) = (\tau_{\text{surf}}(\phi_i), \tau_{\text{gray}}(\phi_i), \tau_{\text{dir}}(\phi_i), \tau_{\text{context}}(Q(\phi_i)))$. Similarly, we extract a set of features for characterizing $\psi_i$, i.e. $\tau_{\text{sketch}}(\psi_i, Q(\psi_i))$.

The features used to characterize photo region $\phi_i$ are:

- The feature vector $\tau_{\text{surf}}(\phi_i)$ is a SURF descriptor [3]. For a photo patch $\phi_i$, we extract a 64 dimensional SURF feature vector to describe the distribution of Haar-wavelet responses in the photo patch.
- The feature vector $\tau_{\text{gray}}(\phi_i)$ is a normalized histogram on the gray value of pixels in $\phi_i$. We empirically set the histogram dimension to 32 in our implementation.
- The feature vector $\tau_{\text{dir}}(\phi_i)$ is a normalized histogram on the gradient directions of pixels in $\phi_i$ where the gradient direction of each pixel is calculated using the Sobel operator [43]. We also empirically set the histogram dimension to 32.
- The feature vector $\tau_{\text{context}}(Q(\phi_i))$ is composed of simplified shape context descriptors [4], which is a log-polar histogram of the coordinates of the remaining points measured using a reference point as the origin. Before generating shape context descriptors, we sample the outline of $Q(\phi_i)$ with uniform spacing, resulting in 100 vertices. Then, for
each feature point in $Q(\phi_i)$, a shape context descriptor is computed with 1 and 8 bins for $\log r$ and $\theta$ respectively. Therefore, the dimensionality of $\tau_{\text{context}}(\phi_i)$ is $8|Q(\phi_i)|$.

These features depict the high-level visual characteristics because exaggerated sketch illustrations usually focus more on the salient visual features of an object or scene [29] rather than the details.

### 4.3 Neutral sketch

To generate the neutral sketch, we also use MLS algorithm to derive the geometric image transformation from $S_i$ to $I_i$, which is similar with the method introduced in Section 4.1. By applying this transformation to $S_i$, we manage to remove the exaggeration from the sample sketch illustration $S_i$, resulting in a neutral sketch $S'_i$. As a byproduct, the method also produces the new positions of the 67 pre-defined facial feature points in $S'_i$, which are denoted by a new feature point set $Q(S'_i)$. Figure 5 shows two restored sketches from the stylized facial portraits.

It is also worth noticing that there may be artist’s estimation error in non-exaggerated sketch drawings, which causes the variations between the contours on facial image and the contour on hand drawing. This cannot be generated by artist’s style, and should be reduced in style learning. Fortunately, our neutral sketch generation method can also be used to reduce this error, since it aligns the key points in facial drawing to the ones in real facial image. If there is noticeable variations between the contours on facial image and hand drawing, user can also perform neutral sketch generation process.

### 5 Fitness evaluation model for case retrieval

In case retrieval, generate-and-evaluate mechanism is used to compute the fitness of a candidate case, which is not based on the case itself, but based on the result generated by applying the candidate case. More specifically, fitness of a candidate case is evaluated in terms of the visual similarity between generated sketch and groundtruth sketch illustrated by the target human artist for the same input facial photo. During the training phase, groundtruth is

![Fig. 5 Two instances of the recovered sketches after removing exaggeration](image-url)
given, the metric for visual similarity assessment is based on the normalized mutual information [30], partly due to mutual information neither depends on any assumption of the data [44] nor requires the extraction of additional features, such as edges and corners, a process of which may introduce additional geometrical errors [25]. It is worth noticing that the groundtruth sketch is typically not available during the sketch synthesis process, which motivates the introduction of the FE model to predict the fitness of a candidate case with respect to a hypothetically available groundtruth.

For each facial region, we train one FE model. Let the FE model trained for the 1st facial region be \( f^{FE}_1 \). Its output is fitness quality \( \theta_{x,i} \) between a case \( r_i \) from our case set and a case \( r_x \) which could be another case in our case set or a new case. Its input is a composite feature vector computed from \( r_i \) and \( r_x \) which is defined as \( (\mathbf{r}_{\text{photo}}(\phi_x, Q(\phi_x)), \mathbf{r}_{\text{photo}}(\phi_i, Q(\phi_i)), \mathbf{r}_{\text{sketch}}(\psi_x, Q(\psi_x))) \). Then we have

\[
\theta_{x,i} = f^{\psi}_1(\mathbf{r}_{\text{photo}}(\phi_x, Q(\phi_x)), \mathbf{r}_{\text{photo}}(\phi_i, Q(\phi_i)), \mathbf{r}_{\text{sketch}}(\psi_x, Q(\psi_x))).
\] (2)

5.1 Training data generation

To generate training data for \( f^{\psi}_1 \), we start from a case set \( R = \{ r_i | i = 1, 2, \ldots, |\Pi_1| \} \), which is randomly divided into 10 equal-sized sub-sets \( \{ \mathcal{R}_k | k = 1, 2, \ldots, 10 \} \). Sketches in one of the sub-sets will be used as the ground truth while the remaining 9 sub-sets are used to generate the sketch results, which will then be cross-validated using the groundtruth sub-set. This process will be performed 10 folds. Therefore, each case will be used once as a groundtruth case. Without loss of generation, let current groundtruth case set be \( \mathcal{R}_1 \), and remaining case set be \( \hat{\mathcal{R}}_1 = \{ \mathcal{R}_k | k = 2, 3, \ldots, 10 \} \), where \( |\mathcal{R}_1| = 9 |\mathcal{R}_1| = \frac{9}{10} |R| \).

Formally, given a STS case set \( \mathcal{R}_1 \) and a photo region \( \phi_x \) from \( \mathcal{R}_1 \), we identify and adapt one or multiple cases to iteratively synthesize a neutral sketch region to maximally approximate the groundtruth \( \hat{\psi}_x \). Let \( \psi^{\text{final}}_x \) and \( \psi^{(h)}_{x} \) respectively be the final resultant sketch and the interim sketch after the \( h \)-th iteration.

In the first iteration, the algorithm selects a case \( \tilde{r}^{(1)}_x \) whose source photo region \( \tilde{\phi}^{(1)}_x \) appears visually closest to \( \phi_x \) in terms of the normalized mutual information [30]. The sketch in the selected case is then denoted as \( \psi^{(1)}_x \).

For the \( h \)-th iteration, the method searches in \( \hat{\mathcal{R}}_1 \) for a best-fit case \( \tilde{r}^{(h)}_x \). For each candidate case \( r_i \) in \( \hat{\mathcal{R}}_1 \), the adaptation on \( \psi_i \) is carried out by blending \( \psi_{i} \) and \( \psi^{(h-1)}_{x} \). The maximum similarity between resultant sketch patch of blending operation and groundtruth \( \hat{\psi}_x \) is

\[
\theta_{x,i}^{(h)} = \max_{\omega \in [0,1]} \theta \left( \psi_{i,x,\omega}^{(h-1)}, \hat{\psi}_{x} \right),
\] (3)

where \( \theta_{x,i}^{(h)} \) is also represented as the fitness value of \( r_i \), \( \psi_{i,x,\omega}^{(h-1)} \) is the resultant sketch patch of blending \( \psi_i \) and \( \psi^{(h-1)}_{x} \) with blending weight \( \omega \), and \( \theta(\psi_{i,x,\omega}^{(h-1)}, \hat{\psi}_{x}) \) is the visual similarity between \( \psi_{i,x,\omega}^{(h-1)} \) and \( \hat{\psi}_{x} \) in terms of normalized mutual information [30].

The image blending operator to synthesize \( \psi_{i,x,\omega}^{(h-1)} \) is originally proposed in [20]. It generates an inbetween image, \( I_t | t \in [0,1] | p \), of two input images, \( I_0 \) and \( I_1 \), by \( I_t = (1 - t) W_0(t, I_0) + t W_1(t, I_1) \), where \( W_0 \) and \( W_1 \) are two non-linear warping functions built from all pairs of corresponding feature points between \( I_0 \) and \( I_1 \). It is obvious that \( t \) is identical to our \( \omega \). During the process of blending \( I_0 \) and \( I_1 \), their feature point sets, \( Q(I_0) \) and \( Q(I_1) \),
Fig. 6  The sizes of the training sample sets for the nine facial regions. The reported feature dimensionality refers to the reduced dimensionality of the selected feature set with the optimally chosen regression models are also blended to compute the feature point set for \( I_f \). Formally, we define the blending operation as follows:

\[
\psi_{i,x,\omega}(\mathbf{h}^{(h-1)}) = \psi_{i,x,\omega}(\mathbf{h}^{(h-1)}) + \psi_{i,x,\omega}(\mathbf{h}^{(h-1)}) \psi_{i,x,\omega}(\mathbf{h}^{(h-1)}), \quad (4)
\]

where \( \psi_{i,x,\omega}(\mathbf{h}^{(h-1)}) \) is the resulting image of blending \( \psi_i \) and \( \psi_{(h-1)} \) with the blending parameter \( \omega \in [0, 1] \); \( Q(\psi_{i,x,\omega}(\mathbf{h}^{(h-1)})) \) is the resulting set of facial feature points.

A non-linear optimization method [35] is employed as a solver for (3). After computing \( \theta_{x,i}(\mathbf{h}) \) for all candidate cases, the case with maximum \( \theta_{x,i}(\mathbf{h}) \) will be selected as \( \mathbf{r}^{(h)}_x \). We denote the maximum \( \theta_{x,i}(\mathbf{h}) \) as \( \bar{\theta}_{x,i}(\mathbf{h}) \). Resultant sketch patch corresponding to \( \theta_{x,i}(\mathbf{h}) \), \( \psi_{x}(\mathbf{h}) \), will be involved in the next iteration. This iterative synthesis procedure terminates when \( \frac{\bar{\theta}_{x,i}(\mathbf{h}) - \bar{\theta}_{x,i}(\mathbf{h}-1)}{\bar{\theta}_{x,i}(\mathbf{h}-1)} \) is less than a threshold (it is set to 1 % in our implementation).

All \( \theta_{x,i}(\mathbf{h}) \) generated in aforementioned iterative process are collected as sample set \( \{\theta_{x,i}(\mathbf{h})| r_i \in R, h = 1, 2, \cdots, n_{\text{step}}(\phi_x)\} \), where \( n_{\text{step}}(\phi_x) \) is the number of iterations, and all

Fig. 7  The cross-validation errors of candidate regression models of \( f_{\text{FE}}^1 \) for \( \Pi_1 \). The averaged training time for each model is also shown
Table 1  Selected optimal regression models for the FE models for datasets $\Pi_1$, $\Pi_2$, and $\Pi_3$ respectively

| Dataset | Region |
|---------|--------|
| $\Pi_1$ | $M_1$ | $M_9$ | $M_9$ | $M_3$ | $M_9$ | $M_1$ | $M_1$ |
| $\Pi_2$ | $M_1$ | $M_1$ | $M_9$ | $M_1$ | $M_1$ | $M_1$ | $M_1$ |
| $\Pi_3$ | $M_1$ | $M_1$ | $M_7$ | $M_1$ | $M_7$ | $M_1$ | $M_7$ |

combinations of $(\phi_x, Q(\phi_x), \psi(h-1)_x, Q(\psi(h-1)_x))$ are collected as training samples. Hence we generate $|\bar{R}_1|n_{\text{step}}(\phi_x)$ samples for $f_{\text{FE}}^1$ after synthesizing a sketch for $\phi_x$. Supposing that the mean value of $n_{\text{step}}(\phi_x)$, $n_{\text{step}}(\phi_x)$, is 5, and the size of R is 50, we can generate $10|\bar{R}_1|\bar{R}_1|n_{\text{step}}(\phi_x)=0.9|R|^2n_{\text{step}}(\phi_x)=11250$ training samples.

5.2 Learning fitness evaluation model

Given the training samples for $f_{\text{FE}}^1$, we train the fitness evaluation model via regression. To identify optimal regression model for $f_{\text{FE}}^1$, we employ the ten-fold cross validation (CV) technique during the model selection process [40].

The definitions of $\tau_{\text{photo}}(\cdot)$ and $\tau_{\text{sketch}}(\cdot)$ indicate that the input to $f_{\text{FE}}^1$ consists of hundreds of features. The dimensionality of the input features is between 528 and 648 wherein the exact dimensionality depends on the specific type of the facial region. Its feature selection is accomplished using the minimal-redundancy-maximal-relevance (mRMR) criterion [34]. However, the mRMR method can only identify the most important features by a user-specified number. To optimally get this number, we employ the best-first search algorithm [16] to look for it through minimizing the CV error. The best regression model is selected by the minimized CV error derived through performing feature selection for each candidate regression model.

Figure 6 shows the sample number and feature dimensionality in the model selection process based on dataset $\Pi_1$ (see Section 8 for details about $\Pi_1$). Using the Weka toolkit [12], 12 most popular regression models, $\{M_k \mid k = 1, 2, \cdots, 12\}$, are taken into consideration: bagging regression tree ($M_1$), SVM regression ($M_2$), M5P ($M_3$), regression tree ($M_4$),
conjunctive case (M5), M5Cases (M6), isotonic regression (M7), additive regression (M8), KNN (M9), linear regression (M10), pace regression (M11), and decision stump (M12). Let $\epsilon_k$ be the cross-validation errors of the $k$-th candidate regression model $M_k$ for $f_{FE}$, which are applicable for the 1st facial region. Then $\{\epsilon_k | k = 1, 2, \cdots, 12\}$ are used to select an optimal regression model for $f_{FE}$. Figure 7 shows the cross-validation errors of the candidate regression models of $f_{FE}$ for $\Pi_1$, from which we find the optimal one is bagging trees [7]. Table 1 shows the regression models selected for $\Pi_1$, $\Pi_2$, and $\Pi_3$ (see Section 8 for details about $\Pi_2$ and $\Pi_3$), in which we observe that the optimal regression models chosen for FE models may vary across facial regions and datasets.

6 Parameter estimation model for automatic case adaptation

Case adaptation is carried out by blending the sketch from the retrieved case with the interim sketch from previous iteration. To automate the case adaptation, we train PE models to identify the optimal blending parameter, maximizing the similarity between the resultant sketch of blending and the hypothetic groundtruth sketch.

Let $f_{PE}^1$ be the PE model trained for the 1st facial region. In each iteration of sketch synthesis, a new photo region $\phi_x$, a sketch region of current interest $\psi_x$, and a case $r_i$ are given. PE model is to estimate the optimal parameter $\omega_{x,i}$ for blending $\psi_x$ and $\psi_i$. That is:

$$\omega_{x,i} = f_{PE}^1(\tau_{\text{photo}}(\phi_x, Q(\phi_x)), \tau_{\text{photo}}(\phi_i, Q(\phi_i)), \tau_{\text{sketch}}(\psi_x, Q(\psi_x))).$$

(5)

The learning of PE model takes the same training samples for FE model. It is worth noticing that an optimal blending weight is also acquired, after solving (3). Therefore, the training data for $f_{PE}^1$ is also accordingly generated as a by-product while preparing training data for $f_{FE}^1$, which is described in Section 5.1.

We use the same learning method introduced in Section 5.2 to train PE model. Figure 6 also shows the sample number and feature dimensionality in the model selection process for PE models. Figure 8 shows the cross-validation errors of the candidate regression model for $f_{PE}^1$ of $\Pi_1$. Table 2 shows the regression models selected for $\Pi_1$, $\Pi_2$, and $\Pi_3$ (see Section 8 for details about $\Pi_1$, $\Pi_2$ and $\Pi_3$), in which we can also observe that the optimal regression models chosen for PE models vary across facial regions and datasets.

7 Synthesizing the desired sketch via CBR

Given an input facial photo and a set of STS cases of a human artist, the resultant sketch is synthesized progressively with an iterative loop of retrieval and adaptation of candidate cases until the desired aesthetic style is presented. Figures 9 and 10 clearly show that the synthesized sketch using multiple cases is much more close to the ground truth image than

| Dataset | Region | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------|--------|---|---|---|---|---|---|---|---|---|
| $\Pi_1$ |        | $M_1$ | $M_1$ | $M_1$ | $M_1$ | $M_1$ | $M_6$ | $M_6$ | $M_1$ |
| $\Pi_2$ |        | $M_3$ | $M_{11}$ | $M_1$ | $M_1$ | $M_1$ | $M_3$ | $M_1$ | $M_9$ | $M_1$ |
| $\Pi_3$ |        | $M_1$ | $M_1$ | $M_3$ | $M_3$ | $M_1$ | $M_1$ | $M_1$ | $M_1$ | $M_1$ |  |
Fig. 9 Comparison between sketch synthesis results using single case (first column) and multiple cases (second column). The third column is the ground truth. The number of the cases used and the similarity with the ground truth image measured by normalized mutual information [30] is also given for reference.

the sketch produced using single case in a sense of normalized mutual information [30]. Moreover, exaggeration imitation is embedded as a post-processing step.

7.1 Iterative sketching by cases

To synthesize the sketch in a given input facial photo $I_x$, the key point is how to optimally retrieve relevant cases and adaptively fuse them to produce the resultant sketch. Our CBR based synthesis pipeline is illustrated in Fig. 11, which originates from the general CBR framework [1].

Given a new facial region $\phi_x$ from a new photo $I_x$ and a set of STS cases, the process for synthesizing a new sketch $\psi_x$ is very similar to the one in Section 5.1. The major difference is that $f_{FE}^1$ and $f_{PE}^1$ are already known now. Therefore, $\theta_{x,i}^{(h)}$ and optimal $\omega$ of (3) can be directly calculated by $f_{FE}^1$ and $f_{PE}^1$ respectively, ignoring the non-linear optimization for (3).

The most important parameter of our method is max number of cases used to generate a sketch patch in the iterative sketching process. In all following experiments, it is not configured in order to maximize result quality. However, in some situations sensitive to processing time, user may configure this parameter for acceleration. Typically speaking, smaller max case number leads to worse result, which can also be noticed in Figs. 9 and 10.
Fig. 10 Comparison between the synthesized facial sketches using single case (first column) and multiple cases (second column) for each facial region as compared with the groundtruth facial sketches (third column). “Sim” is its visual similarity with the groundtruth image measured by normalized mutual information [30]

We are aware of that Liu et al. [22] introduced a similar method for finding optimal example sketch patches and their blending parameters during a facial texture synthesis procedure. Their method is based on the local linearity assumption, which searches the most visually similar examples for an input photo region. The optimal blending parameters are identified through minimizing the reconstruction error in terms of the visual similarity between the input photo and the blended result of the selected example photos. Instead of calculating blending parameters directly from examples, a large amount of training data from a limited

Fig. 11 The pipeline of synthesize sketches for individual facial regions iteratively through case-based reasoning. Feature points on images are represented by yellow dots
number of sketch synthesis cases are generated to train PE models, which allows us to fully utilize available cases for parameter estimation.

The blended image will appear more blurry than the source one, and the post-processing is usually required. We extend the image analogies algorithm [13], as an example-based image sharpening procedure with a multi-scale autoregression process, which can learn from multiple pairs of example images. More concretely, we use \( \{ \psi_i | i = 1, 2, \ldots, |\Pi_l| \} \) as the “filtered” examples and apply a Gaussian kernel, whose radius and standard deviation are empirically set to 3 and 1 pixels respectively, over each \( \psi_i \) to generate its “unfiltered” version \( \psi_{\text{smoothed}}^i \). According to the exemplified mapping relation \( \{ (\psi_{\text{smoothed}}^i, \psi_i) | i = 1, 2, \ldots, |\Pi_l| \} \), the sharpened one of \( \psi_{\text{final}}^x \) is synthesized by image analogy. Figure 12 shows that its sharpened sketch is much better than the one processed by the high-pass filter-based method [11] in terms of the visual appearance of sketching.

7.2 Exaggeration

Exaggeration is popular in facial sketch illustrations, which is often captured as a 2D transformation from \( S_i \) to \( S'_i \), represented by exaggeration field \( V_i \).

In our CBR framework, a new EF for exaggeration imitation, \( V_x \), is generated by a linear system \( H \) as \( H (Q(S'_x)) = V_x \), where \( S'_x \) is a new facial sketch and \( Q(S'_x) \) is the set of feature points on \( S'_x \).

The linear system has the properties of superposition and homogeneity of degree 1 [9], therefore \( H \left( \sum_{i=1}^{||\Pi_l||} \omega_i Q(S'_i) \right) = \sum_{i=1}^{||\Pi_l||} \omega_i H (Q(S'_i)) \), where \( \{ Q(S'_i) | i = 1, 2, \ldots, |\Pi_l| \} \) is a set of sample inputs of \( H \), and \( \omega_i \) is the salience weight associated with \( S'_i \). Given a new input \( Q(S'_x) \), we can optimally find a linear combination of sample inputs to \( H \) that best reconstructs \( Q(S'_x) \). That is:

\[
\tilde{\Omega} = \arg \min_{\Omega} \, D \left( Q(S'_x), \sum_{i=1}^{||\Pi_l||} \omega_i Q(S'_i) \right), \tag{6}
\]

s.t. \( \omega_i \in [0, 1], \sum_{i=1}^{||\Pi_l||} \omega_i = 1 \),

\[
H (Q(S'_x)) \approx H \left( \sum_{i=1}^{||\Pi_l||} \tilde{\omega}_i Q(S'_i) \right) = \sum_{i=1}^{||\Pi_l||} \tilde{\omega}_i V_i, \tag{7}
\]

where \( \Omega = \{ \omega_i | i = 1, 2, \ldots, |\Pi_l| \} \), \( \tilde{\Omega} = \{ \tilde{\omega}_i | i = 1, 2, \ldots, |\Pi_l| \} \), and \( D(\cdot, \cdot) \) is defined as the sum of Squared Euclidean Distances between pairs of corresponding points in two input feature point sets, and \( V_i \) is the sample output of \( H \) corresponding to \( Q(S'_i) \) (see Section 4.1.
Deforming a sketch using the exaggeration field and feature point displacement vectors sampled from the exaggeration field. a is an initial sketch. b is generated by a standard image remapping algorithm [6]. c is generated by our revised approach. Notice: b and c are best viewed at 400 % zoom for the method of generating $V_i$). According to the index number specified for each feature point (see Fig. 3), we can naturally establish a pairwise point-to-point correspondence between the two sets of feature points.

We cannot theoretically prove the linear assumption of our method. However, experimentally, the resultant exaggeration by the predicted EF appears highly close to the ones illustrated by the target human artist.

Once $V_x$ is generated, a straightforward approach for exaggerating $S'_x$ is to warp $S'_x$ following the guidance of $V_x$ through a standard image remapping function [6]. However, $V_x$ is usually noisy, which would cause unreliable deformations (see Fig. 13b). Therefore we firstly get rid of the noise from $V_x$ by interpolating the control point displacements [39]. And then the smoothed $V_x$ is used to warp $S'_x$ through image remapping [6], which generates the exaggerated sketch $S_x$ accordingly (see Fig. 13c).

8 Experimental results

In our experiments, we prepared 3 sets of facial sketch illustrations, respectively denoted as $\Pi_1$, $\Pi_2$, $\Pi_3$ for three human artists. Each set consists of facial sketch portraits illustrated by one human artist, as differentiated through the subscript of $\Pi_i$’s. Every set contains 49 front-view facial photos of 49 different people where each photo is accompanied by its sketch illustrated by the artist. The resolutions of the images are 200 by 250 pixels in $\Pi_1$, 191 by 235 pixels in $\Pi_2$, and 194 by 247 pixels in $\Pi_3$. In particular, for $\Pi_1$, all its facial photos and their sketch illustrations are from the CUHK student dataset [48]. For $\Pi_2$ and $\Pi_3$, 39 photos are from $\Pi_1$ and the remaining 10 new human facial photos are newly taken, which are both included in $\Pi_2$ and $\Pi_3$. For the 49 photos respectively in $\Pi_2$ and $\Pi_3$, we hired two artists to draw 49 facial sketches for each set respectively. The artist for $\Pi_2$ has more than 20 years of professional experiences in creating human facial portraits. The artist for $\Pi_3$ is a PhD student of digital art and design. He has been a freelance illustrator for 7 years and showed his caricature art pieces in a national culture and art expo event. In this manual sketch illustration process, we used a 21-inch LCD to display the human facial photo one by one and asked artists to draw sketch illustrations for each of the displayed facial photo on their A4 paper canvas. Our artists were asked to spend as long as they want in creating these sketch illustrations. After that, all hand-drawn sketches are digitized by a scanner. The original resolutions of the images in $\Pi_2$ and $\Pi_3$ are 2480 by 3508 pixels. As the image resolution of the released CUHK student dataset is 200 by 250 pixels, we downsample the
images in $\Pi_2$ and $\Pi_3$ to make our experimental conditions comparable among the three datasets.

To compare the aesthetic stylization between our approach with that of a facial sketch synthesis approach in [48] and two commercial packages [2, 26], we perform the leave-one-out test on $\Pi_1$. It is noted that since the sketch illustrations in $\Pi_1$ do not present the exaggerative style, the procedure of removing exaggeration from example sketches (Section 4.3) is not performed. That is, for $\Pi_1$, $S_i$ and $S'_i$ are the same. Figure 14 shows the results by our method and the two peer commercial software packages, which clearly demonstrate that the sketch illustrations generated by us appear visually much closer to the target artist’s hand-drawn ones. Figure 15 gives more sketching instances by our approach and the state-of-the-art peer algorithmic approach [48], which also intuitively leads to the same qualitative comparison conclusion. In this experiment, as we perform the leave-one-out test on $\Pi_1$, the number of training samples used for sketch generation is 48, which is significantly less than the number of sample sketches, at least 88, in [48]. In all experiments in this paper, less than 50 training samples are used to capture an artist’s personal facial sketch illustration style.

To make more assessment on the stylization of the sketches produced by our approach and the peer algorithm, we further conducted a user survey on the sketches generated in the

![Fig. 14](image)

Comparing our approach with two commercial software packages. Column (a) and (b) are generated by the two commercial software packages [26] and [2] respectively. Column (c) is generated by our approach with the training examples from $\Pi_1$. The sketch illustrations created by artist I are the groundtruths, which are not involved in the training process.
Fig. 15 Generating sketch illustrations of facial portraits from input human facial photos in personal illustration styles of three artists. Column (a) are generated by the peer algorithm proposed in [48] while column (b) are produced by our algorithm.

above experiment. We selected 10 artist-drawn sketches from $\Pi_1$ and their corresponding sketches generated by our approach and [48] respectively. We then created an online questionnaire consisting of 10 questions. Each question presents three images side by side. The hand-drawn sketch illustration is always shown in the middle. The corresponding sketch illustrations generated by the two algorithms respectively are randomly placed on the left and right. Each question asks a human viewer to select among the two images displayed on the left and right positions of the image triplet. The subjects are asked to answer which one appears visually more similar to the middle image. We released the online questionnaire on

Fig. 16 The user study comparing the sketch generation quality by our approach and [48]. Data collected from 2624 completed questionnaires by answering 10 questions.
a twitter-like social media community dedicated to comic fans [41], who are generally very familiar with artistic facial drawing styles. 2624 completed questionnaires were collected during 3 days. None of the participants was compensated monetarily, who took part in the online survey due to their curiosity and interest in facial sketch illustrations. Through a dynamic webpage tracking feature, it is shown each questionnaire takes around 2.5 minutes on average to be answered. Among the ten pairs of sketches in our online survey, seven of them get more votes for our sketches by the subjects, and the remaining three get more votes for the sketches from the peer algorithm. Overall, 62 % of all the 26240 answers to the ten questions favor resulting sketches by our method (one-sample t-test, \( p \)-value < 0.001; two-sample t-test, \( p \)-value < 0.001). Figure 16 presents more details about our assessment. The voting clearly shows the superiority of our method in terms of the visual appearance of sketching.

Figure 17 shows our two failure cases from user survey. Users’ votes show that the two sketches in Fig. 17 generated by Wang et al. [48] are more similar with artist’s hand-drawing than ours.

Besides the evaluation procedure based on user survey, there are objective evaluation procedures. In fact, the mutual information based method we adopted in our FE model training phase is an objective evaluation method for image similarity. Nevertheless, according to the NPR Turing Test proposed by Salesin [38], computer-generated imagery should be indistinguishable from imagery produced by humans, which means user evaluation of imagery should be conducted as the final test.

We also perform leave-one-out tests on the data sets \( \Pi_2 \) and \( \Pi_3 \). Figure 15 gives more synthesized sketches, which demonstrate that our approach is indeed capable of imitating multiple personal facial illustration styles, with or without exaggerations. The first artist sketches human face portraits more or less following the visual characteristics in the original facial photo. The other two artists present significant exaggeration in their facial illustrations.

In Fig. 15, it is obvious that the peer algorithm [48] fails to capture the sketch illustration styles of artists II and III as the generated sketch illustrations have significant differences to the ones created by the target artists. This is partly because their algorithm is designed upon the assumption that each facial region in a photo occupies the same image area as its counterpart region in the corresponding sketch illustration does, which is not always true in reality, particularly for sketch illustrations with significant exaggeration.

---

**Fig. 17** Failure cases from online user survey. For the above two photos, our method failed to generate better sketch drawings than Wang et al. [48], since users’ votes show the above two sketches generated by Wang et al. [48] are more similar with artist’s hand-drawing than ours.
Fig. 18 Comparing our approach with the method proposed by [47]. The faces in column (c) and (e) are generated by our algorithm and the peer algorithm proposed in [47] respectively.

We also compare our approach with the transductive learning algorithm proposed by Wang et al. [47]. Figure 18 shows our resultant sketches in the leave-one-out test on $\Pi_1$ and sketches from Wang’s website of results on CUHK face sketch database [46]. The nasolabial folds [8] is one of the essential elements of facial stylization and often appears.

Fig. 19 Human facial sketch portraits generated in two illustration styles for the same set of input facial photos. Images in column (a) and (b) are generated by our approach. The training data for generating the images in columns (a) and (b) come from $\Pi_1$ and $\Pi_2$ respectively.
in face portraits drawn by human artist. They are well preserved and can be significantly observed in our resultant sketches. Unfortunately, nasolabial folds are little presented in sketches generated by the transductive learning algorithm. Hence, our resultant sketches more closely resembles the groundtruth sketches by the target human artists.

Figure 19 shows more resulting sketches in the leave-one-out tests on $\Pi_2$ and $\Pi_3$. It demonstrates that our method can successfully learn personal sketching styles of multiple artists and accordingly generate the individual stylized sketches for the same input photos.

Regarding the time performance, it takes about 1 minute to generate a facial sketch from an input photo of the resolution of 200 by 250 pixels using the unoptimized, single computing core prototype implementation of our algorithm executed on a PC equipped with Intel i5-3450 3.1GHz CPU and 3.2G memory, while the peer algorithm [48] takes 3 minutes on the same hardware platform.

Originally our training takes about 3 days to finish. However, we observe that it is straightforward to implement our training algorithm concurrently since FE and PE models for each facial region can be trained independently. We evaluated the parallel computing implementation on a PC with Intel i7 CPU (8 cores) and 6G memory. It took about 8 hours to complete all training tasks. We are thus optimistic about the computational efficiency of our algorithm in practice with a parallel implementation.

9 Conclusion and discussions

We propose a new CBR-based sketch synthesis algorithm that can produce visually superior results than existing synthesis methods. To the best of our knowledge, it is the first CBR framework that is explicitly introduced to personally stylize human facial portrait.

For each human artist to be mimicked, a series of cases are firstly built-up from her/his exemplars of source facial photo and hand-drawn sketch, and then its stylization for facial photo is transformed as a style-transferring process of iterative refinement by looking for and applying a series of best-sit cases in a sense of style optimization. The presented experimental results demonstrate that in comparison with a state-of-the-art method and a couple commercially available software packages, the new approach is capable of generating visually more appealing portraits from the point of view of personal style, which also more closely resemble the groundtruth sketches by the target human artists.

Our work still has some limitations. The first limitation is the lack of overall statistics of the experiment results, which requires a large-scale user survey. With current online survey method, each subject should answer all questions. This makes it difficult to collect completed questionnaires, when the question number is large. In future, we will develop a tool which can randomly pick up 10 questions from our dataset for each online subject, collect online answers, and do statistics for this large-scale user survey. The second limitation is that our approach has not supported the synthesize sketch illustrations for human hair yet. This limitation comes from the difficulty in establishing correspondences between strokes or texture areas that depict hairs and the counterpart hair regions on human facial photos. A related future research direction of this problem is Cross-Modal Face Matching [54]. Besides, if we focus on the hair of a specific group of people, it is possible to represent the correspondences through manually indicated key points, like that by Chen et al. [10].

Acknowledgments This work was supported by a grant from National Program on Key Basic Research Project of China (973 Program, 2013CB329504), National High Technology Research and Development Program of China (863 Program, 2013AA013705), National Natural Science Foundation of China (N0.
References

1. Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun 7(1):39–59
2. AKVIS: Akvis sketch v.13.0 (2012) http://akvis.com/en/sketch/index.php
3. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. ECCV. Springer, pp 404–417
4. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. PAMI 24(4):509–522
5. Berger I, Shamir A, Mahler M, Carter E, Hodgins J (2013) Style and abstraction in portrait sketching. TOG 32(4):55
6. Bradski G (2000) The opencv library. Dr Dobb’s Journal of Software Tools
7. Breiman L (1996) Bagging predictors. Mach Learn 24:123–140
8. Burgess CM (2005) Cosmetic dermatology. Springer
9. Chen H, Liu Z, Rose C, Xu Y, Shum HY, Salesin D (2004) Example-based composite sketching of human portraits. In: NPAR. ACM, pp 95–153
10. Gonzalez RC, Woods RE (2002) Digital image processing, 2nd edn. Prentice Hall, SL
11. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: an update. SIGKDD Explor News1 11(1):10–18
12. Hertzmann A, Jacobs CE, Oliver N, Curless B, Salesin DH (2001) Image analogies. In: SIGGRAPH. ACM, pp 327–340
13. Kalogerakis E, Nowrouzezahrai D, Breslav S, Hertzmann A (2012) Learning hatching for pen-and-ink illustration of surfaces. ACM Trans Graph 31(1):1:1–1:17
14. Kim SY, Maciejewski R, Isenberg T, Andrews WM, Chen W, Sousa MC, Ebert DS (2009) Stippling by example. In: NPAR. ACM, pp 41–50
15. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1):273–324
16. Kyprianidis J, Collomosse J, Wang T, Isenberg T (2013) State of the art: a taxonomy of artistic stylization techniques for images and video. TVCG 19(5):866–885
17. Leake DB, Kinley A, Wilson D (1995) Learning to improve case adaptation by introspective reasoning and cbr. In: Case-based reasoning research and development. Springer, pp 229–240
18. Lee H, Seo S, Ryoo S, Yoon K (2010) Directional texture transfer. In: NPAR. ACM, pp 43–48
19. Lee SY, Chwa KY, Hahn J, Shin SY (1996) Image morphing using deformation techniques. JVCA 7(1):3–23
20. Liang L, Chen H, Xu YQ, Shum HY (2002) Example-based caricature generation with exaggeration. In: PG. IEEE Computer Society, p 386
21. Liu Q, Tang X, Jin H, Lu H, Ma S (2005) A nonlinear approach for face sketch synthesis and recognition. In: CVPR. IEEE Computer Society, pp 1005–1010
22. Lu C, Xu L, Jia J (2012) Combining sketch and tone for pencil drawing production. In: NPAR. Eurographics Association, pp 65–73
23. Lu J, Sander PV, Finkelstein A (2010) Interactive painterly stylization of images, videos and 3d animations. In: Proceedings of the 2010 ACM SIGGRAPH symposium on interactive 3d graphics and games. ACM, pp 127–134
24. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997) Multimodality image registration by maximization of mutual information. T-MI 16(2):187–198
25. Microsoft: Microsoft office homestyle+ trial edition (2003). http://www.microsoft.com/zh-tw/download/details.aspx?id=4851
26. Milborrow S, Nicolls F (2008) Locating facial features with an extended active shape model ECCV
27. Min F, Suo JL, Zhu SC, Sang N (2007) An automatic portrait system based on and-or graph representation. In: EMMCVPR. Springer, pp 184–197
28. Mish FC (1994) Merriam webster’s collegiate dictionary. Merriam Webster Inc
29. Modat M, Ridgway GR, Taylor ZA, Lehmnn M, Barnes J, Hawkes DI, Fox NC, Ourselin S (2010) Fast free-form deformation using graphics processing units. Comput Methods Prog Biomed 98:278–284
30. Papari G, Petkov N (2009) Continuous glass patterns for painterly rendering. IEEE Trans Image Process 18(3):652–664
32. Papari G, Petkov N, Campisi P (2007) Artistic edge and corner enhancing smoothing. IEEE Trans Image Process 16(10):2449–2462
33. Parke FI (1972) Computer generated animation of faces. In: Proceedings of the ACM annual conference, vol 1. ACM, pp 451–457
34. Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. PAMI 27(8):1226–1238
35. Powell MJD (1998) Direct search algorithms for optimization calculations. Acta Numerica:287–336
36. Redman L (2012) How to draw caricatures. McGraw Hill Professional
37. Reinhard E, Ashikhmin M, Gooch B, Shirley P (2001) Color transfer between images. IEEE Comput Graph Appl 21(5):34–41
38. Salesin DH (2002) Non-Photorealistic animation & rendering: 7 grand challenges. In: Keynote talk at second international symposium on non-photorealistic animation and rendering (NPAR 2002, Annecy, France, June 3–5, 2002). http://www.research.microsoft.com/salesin/NPAR.ppt
39. Schaefer S, McPhail T, Warren J (2006) Image deformation using moving least squares. TOG 25:533–540
40. Schaffer C (1993) Selecting a classification method by cross-validation. Mach Learn 13(1):135–143
41. Semillon (2013) Semillon’s homepage on sina weibo. http://weibo.com/semillon
42. Song Y, Bao L, Yang Q, Yang MH (2014) Real-time exemplar-based face sketch synthesis. Springer, pp 800–813
43. Sonka M, Hlavac V, Boyle R (1993) Image processing, analysis and machine vision. Chapman and Hall, London
44. Thévenaz P, Unser M (2000) Optimization of mutual information for multiresolution image registration. IEEE Trans Image Process 9(12):2083–2099
45. Tu CT, Lien JJJ (2010) Automatic location of facial feature points and synthesis of facial sketches using direct combined model. Trans Sys Man Cyber Part B 40:1158–1169
46. Wang N (2013) Results on cuhk face sketch database. https://nannanwang.github.io/Result_CUFS.htm
47. Wang N, Tao D, Gao X, Li X, Li J (2013) Transductive face sketch-photo synthesis. IEEE Transactions on Neural Networks and Learning Systems 24(9):1364–1376
48. Wang X, Tang X (2009) Face photo-sketch synthesis and recognition. PAMI 31:1955–1967
49. Wei G, Rui M, Lei W, Yi Z, Zhenyun P, Yaohui Z (2009) Template-based portrait caricature generation with facial components analysis. In: ICIS, pp 219–223
50. Wu C, Liu C, Shum HY, Xy YQ, Zhang Z (2004) Automatic eyeglasses removal from face images. IEEE Trans Pattern Anal Mach Intell 26(3):322–336
51. Zhang Y, Dong W, Deussen O, Huang F, Li K, Hu BG (2014) Data-driven face cartoon stylization. In: SIGGRAPH Asia 2014 technical briefs. ACM, pp 14:1–14:4
52. Zhao M, Zhu SC (2010) Sisley the abstract painter. In: NPAR. ACM, pp 99–107
53. Zhao M, Zhu SC (2011) Portrait painting using active templates. In: NPAR. ACM, pp 117–124
54. Zhou H, Kuang Z, Wong KY (2012) Markov weight fields for face sketch synthesis. In: IEEE conference on computer vision and pattern recognition (CVPR), 2012, pp 1091–1097

Bingwen Jin received his B.A. degree in optoelectronic information engineering from Zhejiang University, China in 2008. From September 2008, he is studying as a Ph.D. student in computer application at Zhejiang University, China. His primary research interests include Non-photorealistic Rendering, Cartoon Animation, and Machine Learning.
Songhua Xu is an assistant professor at the Information Systems Department of New Jersey Institute of Technology. He received his Ph.D., M.Phil., M.S., all in computer science from Yale University. His primary research interests include digital arts, information retrieval and management, web search and data mining, innovative applications of AI, and intelligent systems for biomedical applications.

Weidong Geng received the BSc degree from the Computer Science Department in Nanjing University, China, in 1989, and an MSc degree from the Computer Science Department of National University of Defense Technology in 1992. In 1995, he received the PhD degree from the Computer Science and Engineering Department of Zhejiang University, China. Currently, is a professor of College of Computer Science, Zhejiang University, China. From 1995 to 2000, he was in Zhejiang University, where he took charge of a number of projects about CAD/CG, and intelligent systems. He joined Fraunhofer Institute for Media Communication (former GMD.IMK), Germany, as a research scientist in 2000. In 2002, he worked in Multimedia Innovation Center, The Hong Kong Polytechnic University, Hong Kong. Since 2003, he has been working in State Key Laboratory of CAD/CG, Zhejiang University, and his current research focuses on computer aided design, computer animation, perceptual user interface, interactive media, and digital entertainment.