Advances in Computer Science and Engineering
Volume 5, Number 1, 2010, Pages 49-71
Published Online: August 29, 2010
This paper is available online at http://pphmj.com/journals/acse.htm

EXAMPLE BASED CARICATURE SYNTHESIS

HONGCHUAN YU, WENJUAN CHEN[*] and JIAN J. ZHANG

National Centre of Computer Animation
Bournemouth University
U. K.

[*]Animation School
Communication University of China
P. R. China

Abstract

The likeness of a synthesized caricature and the original face image is an essential and often overlooked part of caricature production. In this paper, we present an example based caricature synthesis technique, consisting of shape exaggeration, relationship exaggeration, and optimization for likeness. Rather than relying on a large training set of caricature face pairs, our shape exaggeration step is based on only one or a small number of examples of facial elements (e.g., eyes and nose). The relationship exaggeration step introduces two definitions which facilitate global facial feature synthesis. The first is the T-shape rule, which describes the relative relationship between the facial elements in an intuitive manner. The second is the so-called proportions, which characterizes the facial features in a proportion form. Finally, we introduce a similarity metric as the likeness metric based on the Modified Hausdorff Distance (MHD) which allows us to optimize the configuration of facial elements, maximizing the likeness while satisfying a number of constraints. The effectiveness of our algorithm is demonstrated with experimental results.

Keywords and phrases: caricature synthesis, exaggeration, likeness, modified Hausdorff distance, texture style transferring.

Received April 7, 2010

1. Introduction

This paper presents a new technique for the synthesis of human face caricatures learning from existing examples. The purpose is twofold. The first is to facilitate caricaturists to produce caricatures efficiently allowing them to concentrate on their creative work. The second is to enable a novice to learn and produce caricatures for entertaining purposes by mimicking one or more existing caricature styles.

Caricature is a face representation where some distinctive features or peculiarities are exaggerated deliberately. Caricatures are prevalent in most forms of media, from newspapers and magazines to cartoons, with themes ranging from political satire to entertainment. The legendary animator Walt Disney equated his animation to caricature. It differs from portrait drawing, since a portrait must preserve the recognizable features rather than exaggerate them. A good caricature should differ from a real face image but should remain recognizable as the caricatured person. The exaggerated features help to convey the comedic aspects of the figurer to the viewer, which can be both funny and critical.

There are three elements essential to caricatures: exaggeration, likeness, and statement [16]. A caricaturist must decide *which features* to exaggerate, and the *scale* of the exaggeration. The likeness emphasizes the visual similarity of the caricature to the subject. Statement allows the artist to add some personality to the subject by editorializing the caricature. Statement is an artistic process and cannot be emulated by a computer. In this paper, we address exaggeration and likeness with the aim to create exciting caricatures by learning from available examples.

Example based learning methods usually need a large training set from a particular artistic tradition, such as [6, 7, 13]. In practice, however, it is impossible to get a large training set of caricatures that have the same style or from the same artist. Commonly, only a small number of caricatures from the same caricaturist or the same artistic tradition are available, making these conventional example based learning approaches ineffective.

Facial features (e.g., facial contour, eyes and nose, etc.) are essential elements of a caricature. Different caricaturists and artistic traditions draw them differently which give caricatures a distinct style. Therefore, a new caricature can be created by taking these individual elements from several caricature examples. For instance, we may want to exaggerate a face with a narrow facial contour and short nose. If both features are present in different examples, then the solution is to pick up the necessary features from the respective example caricatures. However, because the facial features are from different examples, harmonious arrangement of these features is essential. Therefore, we can see two challenging problems:

- How do we generate the desired exaggerated facial elements by using one available example as reference, and

- How do we arrange all facial features (elements) harmoniously with the face?

In this paper, we address these two challenges by presenting a new algorithm. Our contributions can be summarized as following:

- Shape exaggeration. The shape exaggeration of individual face elements is computed based on only one or a small number of examples;

- Relationship exaggeration. The T-shape rule [16] is introduced and the proportional description of the features are utilized to exaggerate the relationships between the facial features. It proves both simple and intuitive;

- Likeness. In existing methods, “likeness” is seldom considered for caricature synthesis due to lack of a “likeness” metric. We introduce the Modified Hausdorff Distance (MHD) [10] to measure the visual similarity. Based on this metric, the likeness is incorporated into the integral caricature by optimizing the configuration of the facial elements, ensuring the resulting caricature resembles the original subject.

Our work mimics the practice of caricature production. The user can choose the styles of the target caricature, and our method semi-automatically merges all exaggerated and non-exaggerated elements into a caricature, while maximizing the resemblance to the original face.

2. Related Works

The relevant approaches to caricature generation can be categorized into three groups: The first is the template based morphing where the user manually deforms a template to produce a new caricature. In the system described in [8, 11], a face image was fitted by a set of pre-designed templates, which was then exaggerated interactively. Akleman first employed the morphing technique to a caricature in [1]. In [2], Akleman et al. further developed a set of new morphing functions with the Implicit Free-form Deformation method. Their technique is capable of, interactively, producing extreme exaggerations of facial features. However, these methods usually require expert knowledge and detailed involvement of experienced artists. For an untrained user, it is not easy to decide which and how the features should be exaggerated.

The second can be summarized as the "Exaggerating the Difference From the Mean" (EDFM). Brennan [4] first presented the idea of EDFM and developed an interactive caricature generator. Koshimizu et al. [12] applied the same idea to their caricature system of PICASSO, where they focused on how to extract the facial features. However, different opinions exist regarding the effectiveness of EDFM. The central question is whether we can equate "the difference from the mean" to the distinctiveness of the facial features. Mo et al. [15] stated that "the distinctiveness of a displaced feature not only depends on its distance from the mean, but also on its variance". They first decomposed a face into a high dimensional space by a non-negative matrix factorization, and then scaled each dimension with different factors. The factors are chosen with a threshold. Chiang et al. [5] dealt with facial elements individually, while Xu et al. [20] applied principal component analysis to each facial element to determine the exaggeration and synthesized them in a caricature. Each exaggeration factor was determined in terms of eigenvalues. Tseng et al. [19] also applied a similar idea to their caricature generation system. These approaches essentially formulate some semi-regular rules to exaggerate the difference. As there are many undetermined parameters, particularly with extreme exaggerations, merging all features into a caricature remains a challenging task. One of the goals of this paper is to address how to harmoniously and simply synthesize all of emphasized and non-emphasized features into a caricature.

The third group includes the example based learning methods. This kind of approach usually needs a training database containing a large number of *caricature face pairs* (each pair is made up of a natural face image and its corresponding caricatured face image). Chen et al. [7] developed an example based sketch generation technique by learning a particular style of an artist. A similar approach was also applied to their caricature system for capturing a particular style of an artist in [13]. In [6], Chen et al. further decomposed the data into components that are structurally related to each other such as eyes or mouth, which were dealt with independently. Each component therefore must possess its individual training subset. Shet et al. [18] presented a cascade correlation neural network based caricature synthesis method to capture the drawing style of an artist or a particular artistic tradition. Their training examples were not further classified into different prototypes, meaning that the influence of the different example prototypes was not taken into account. Although it is relatively easy to acquire a dataset of natural face images, it is nontrivial to construct a training dataset of caricature face pairs. If the training dataset is further divided into several different stylistic subsets, each subset becomes even smaller, making the example based approach ineffective. To avoid this predicament, we believe it is important to explore the feasibility of creating a caricature in the style of a given caricature without the need of a training dataset of caricature face pairs. In this paper, we will address this challenge.

The rest of this paper is organized as follows: We first present our exaggeration approach in Section 3. Section 4 addresses how to implement the "likeness" of a caricature. Experiments and analysis are given in Section 5. Section 6 concludes the paper by looking into the areas of improvement for our future work.

3. Exaggeration

Our goal is to synthesize a caricature based on one or a small number of given caricature face pairs. We will not rely on a training dataset of caricatures due to the practical difficulties mentioned above. A human face can be decomposed semantically into seven facial elements, i.e., facial contour, left and right eyebrows, left and right eyes, a nose and a mouth. Each facial element may be further divided into several prototypes (e.g., the eyebrow element has two prototypes, thick and thin) based on their appearance in the individual training datasets. A caricature is usually represented by two types of exaggerations, shape exaggeration of individual facial elements (which we call *shape exaggeration*); and the exaggeration of the relationship between these facial elements (which we call *relationship exaggeration*). The latter includes position, size and angle of the facial elements [16]. For example, eyebrows are exaggerated in the shape of a thin curve (shape exaggeration) while their locations may be moved apart from each other (relationship exaggeration). Shape exaggeration can be implemented by studying the style of a given shape, while relationship exaggeration usually depends on a global model, which handles the overall arrangements. In general, these two kinds of exaggerations might be handled independently in a drawing.

Capturing the global model implicitly needs a very large training example set such as in [13], since the seven facial elements and their parameters (including scaling, position and orientation) lead to a number of combinations. Most caricaturists employ some semi-regular rules to deal with relationship exaggeration. Previous approaches tried to formulate the rules for relationship exaggeration [1, 4, 19]. Our approach of relationship exaggeration is also based on these semi-regular rules. However, synthesizing all the facial elements into a caricature still remains extremely challenging. This is due to the fact that too many motion parameters of elements need to be tuned. In this section, we present a method to merge the facial elements in a simple manner.

We first address a learning approach for shape exaggeration based on the Principal Component Analysis (PCA) technique, and then explore the relationship exaggeration of facial elements.

3.1. Shape exaggeration

Suppose a given original face training set $\{X_i, i = 1, ..., n\}$, where each X_i consists of a set of feature point coordinates of the seven facial elements, i.e., $X_i = (X_i^{(1)}, ..., X_i^{(7)})^T$. For convenience, we ignore the superscript thereafter, since the same processing procedure is applied to the different facial elements in our algorithm. Let $\{X_0, X_0^*\}$ be a given face image-caricature pair, where X_0 denotes the original natural face while X_0^* denotes the caricatured one. We aim at creating a caricature in the style of this given X_0^* here.

In terms of the training set $\{X_i\}$, we can build an eigenface space as follows:

$$Y = U_k^T (X - \overline{X}) U_k, \tag{1}$$

where Y is the projection of X onto the eigenface space, $\overline{X}$ is the mean of the training dataset, and U_k is the collection of the first k eigenvectors. Our objective is to make all examples of $\{X_i\}$ approximate X_0^*, so that the exaggerated $\{X_i^*\}$ is in the style of X_0^*. It is therefore expected to build a mapping between $X_i^* - X_0^*$ and $X_i - X_0$. To this end, a new eigenspace is first generated with the mean as X_0 (instead of $\overline{X}$) in the same manner of equation (1). Then the mapping between the differences is described as follows:

$$X_i^* - X_0^* = U_k [\lambda_k] U_k^T (X_i - X_0), \tag{2}$$

where $[\lambda_k]$ denotes the approximation coefficients in a diagonal matrix form. If we let $X_i^* = 0,\ X_i = 0,$ then we have a unique λ by solving

$$X_0^* = U[\lambda]U^T X_0. \tag{3}$$

It is impossible that the resulting λ applies to all other face images X_i. For robustness, we formulate this problem of seeking λ over the training set $\{X_i\}$ as a minimization problem with respect to λ as follows:

$$\min_{\lambda} \sum_{i=0}^{n} \| X_0^* - U[\lambda]U^T X_i \|^2. \tag{4}$$

This is because all exaggerated shapes are expected to be closest to X_0^*. The optimal λ can be obtained by solving a linear system. Once λ is yielded, we can select the first *k* principal components $(\text{i.e., } \lambda_k, U_k)$ to compute the deformed X_i^* with equation (2). Usually, *k* can be determined empirically.

Figure 1(a) shows the results of facial contour exaggeration only. The other facial elements (e.g., the eyes and mouth) are merged into it by keeping their individual original shapes and proportions unchanged. With equations (2)-(4), we can create different styles of exaggerated facial contour X_i^* starting from the original X_i using different examples $\{X_0,\ X_0^*\}$. It allows the user to produce the exaggeration style without changing the training set. Figure 1(b) shows the results of two exaggerated facial elements, facial contour and nose shape. The examples $\{X_0,\ X_0^*\}$, which are used for exaggerating the facial contour and nose shape, are from two different caricature face pairs. The other non-exaggerated elements keep their original shapes and proportions. It can be observed that apart from the facial contour, little shape change arises with the other elements, i.e., the exaggerated nose does not obviously stand out in the caricature. This leaves many degrees of freedom for relationship exaggeration.

For a good effect, we should choose a natural face (or facial elements) training set $\{X_i,\ i = 1,\ ...,\ n\}$ in the prototype of the given example X_0 in advance. For example, the shape of facial contour can be categorized into several different groups such as oval, square, long, diamond and heart, etc. According to a specified prototype of facial contour, we need to build a corresponding training set. This is also a basic demand of the PCA technique.

3.2. Relationship exaggeration

For a caricature, the relationships of facial elements plays a dominant role. The relationships include position (e.g., the relative distances between facial elements), size (each element is scalable and the absolute size is treated as a part of the relationship exaggeration) and angle (e.g., relative to the central axis of a face) [16]. The relationships between facial contour and the other six facial elements and also among these six facial elements, are of a powerful mechanism for the production of exciting caricatures. In our algorithm, the facial contour is assumed to be fixed while the other six elements are placed into it. Our basic idea is to adopt the so-called T-shape rule [16] and emphasize a small number of facial features rather than all features. Any given subject might have several different interpretations with respect to the exaggeration of the relationships of its features and each as the other [16]. Our presented approach attempts to mimic the way by which caricaturists exaggerate distinctive features. We will also look into how exaggerated and non-exaggerated features can be arranged automatically in a caricature.

The T-shape rule can be stated as follows: (1) both eyes determine a horizontal axis while the nose and mouth symmetrically distribute around a vertical axis; (2) moving both eyes relative to each other along the horizontal axis results in the nose moving along the vertical axis; (3) if the eyes move apart from each other, then the nose should be shortening; whereas, if the nose is lengthening, then the eyes should move closer to each other. This is shown in Figure 2(b). The T-shape rule is used for emphasizing one or a few distinctive features in drawing a caricature. However, the third item is too restrictive to be employed, in practice. In our algorithm, it is not required for certain exaggeration effects.

To implement the T-shape rule in our technique, we first determine the horizontal axis by connecting points A and B of the facial contour, and the vertical axis by drawing a line between the midpoint O of line $\overline{AB}$ and the bottom point C of the chin as shown in Figure 2(a). Let the top point of the nose share the same point with the midpoint O of $\overline{AB}$. Based on this T-shape, we present two methods for relationship exaggeration.

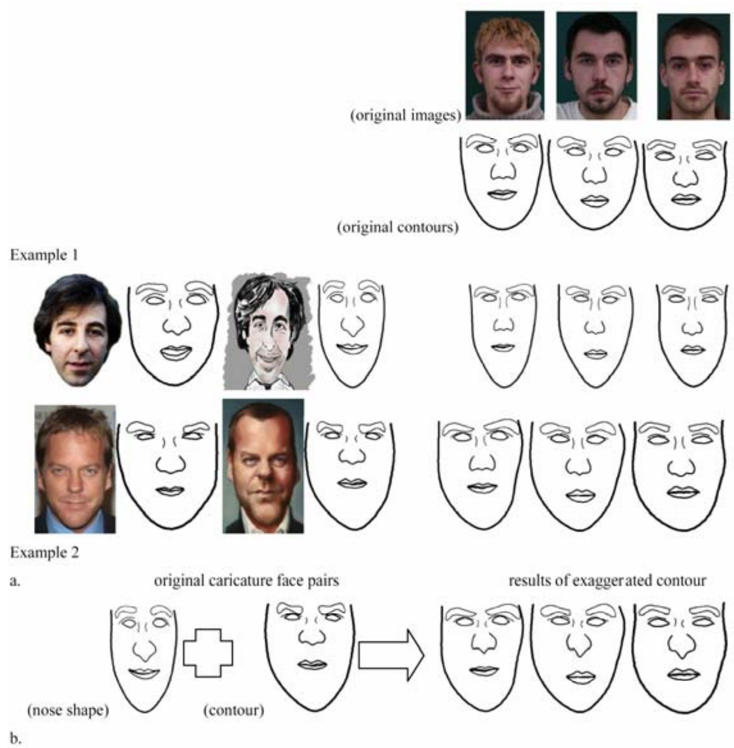

Figure 1. Illustration of shape exaggeration: (a) Shapes are exaggerated separately following the styles of two given caricature face pairs; (b) the nose shapes are exaggerated in agreement with the first example while the facial contours are exaggerated learning from the second example. All other facial elements are then merged into the caricatures, whilst keeping the original proportions.

(Example 1: http://www.quarehawk.com/category/cartoon;

Example 2: http://www.jasonseiler.com/illustrationa.html)

Method 1

Many existed approaches [1, 4, 5, 19] exaggerated all differences between the given face and an average one by the same amount. As a starting point, we emphasize all features along the horizontal and vertical axes. The relationships between the features can be intuitively defined by the distances between them. We normalize these distances by the length w of the horizontal axis and the length h of the vertical axis (which are called the *proportions*). Our basic idea is to emphasize the difference of the proportions between a given face and an average face in the direction of the horizontal and vertical axes respectively. Thus the facial features are represented by a set of proportions. Lenn [14] provided around 30 facial features. We employ only eight proportions, which are

$$\{w_1/w,\ w_2/w,\ w_3/w,\ w_4/w,\ w_5/w,\ h_1/h,\ h_2/h,\ h_3/h\}$$

as shown in Figure 2(a). Note that these proportions are defined along the horizontal and vertical axes except the proportions of $\{w_4/w,\ w_5/w\}$. In addition, we further add an area proportion of the triangle between the eyes and nose to establish the relationship of the width and height, i.e., w_3h_1/wh, as shown in Figure 2(a). Indeed, the triangle area is also viewed as one of facial features such as [17].

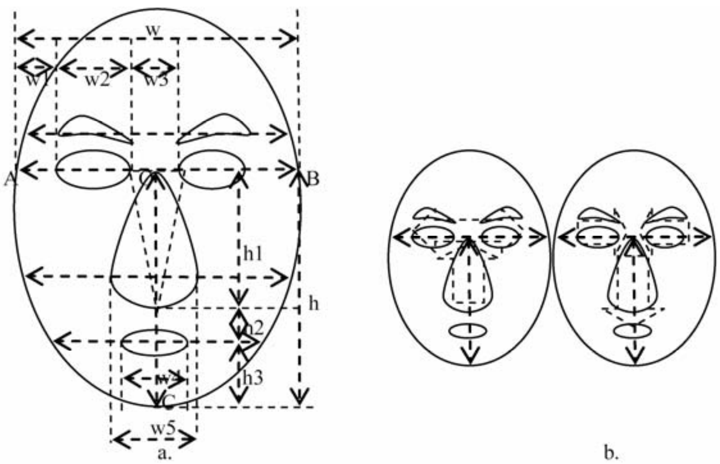

Figure 2. Illustration of the T-shape rule and facial elements' proportions accordingly: (a) Facial elements' proportions; (b) Motion of facial elements.

For a given natural face X_i, the horizontal and vertical axes are determined once the facial contour is fixed. We can calculate a set of proportions $P_i(j)$, $j = 1, \ldots, 8$ and further obtain the difference from the mean proportion $\overline{P}$, i.e., $\Delta P_i = P_i - \overline{P}$. (Note the area proportion $P_i(9)$ is not taken into account here. In Method 2, it will be included in the computation.) For a given amount of exaggeration *t*, we can update the proportion P_i by

$$P_i^{(t)} = \overline{P} + t\Delta P_i, \tag{5}$$

where $0 < t < \min\limits_{j=1,\ldots,8}\left(\dfrac{\overline{P}(j)}{\overline{P}(j) - P_i(j)} > 0,\ \dfrac{1 - \overline{P}(j)}{P_i(j) - \overline{P}(j)} > 0\right)$. We can understand from equation (5), if $t > 1$, then the proportions P_i are exaggerated. Otherwise, the differences from the mean are reduced as t decreases. This will then approach the mean face. Furthermore, it can be noted that equation (5) does not only exaggerate the positions of facial elements, but also their sizes and angles. Regardless of t (note $|\Delta P_i(j)|$ is always very small), the resulting coordinates of feature points can still stay within the range of the facial contour. All changes revolve around $\overline{P}$. On the caricature image plane determined by the horizontal and vertical axes, applying the exaggerated proportions to the facial elements, leads to the creation of a desired caricature. Additionally, since all of the facial features (not only a few) along the horizontal and vertical axes are exaggerated by the same t, the third item of the T-shape rule is not applicable here.

Figure 3 shows the results of exaggerating all features along the horizontal and vertical axes by using the same t. Because we only illustrate the relationship exaggeration here, shape exaggeration is turned off. It can be observed that even with large exaggerations, the proportion description can still merge all features within the range of the given facial contour, as shown in Figure 3(b).

Method 2

Caricaturists tend to emphasize only one or two salient features in a caricature [9]. The impression that one feature gives to the viewer also depends on the surrounding features. As a result, some caricaturists stress only one or a few distinctive features by enlarging its size and at the same time reduce the sizes of other features [14]. This makes the exaggerated features stand out from the others. This idea can be implemented easily using the mechanism of proportions.

Assume that t is given beforehand. The first problem we face is how to determine a small number of salient features of a face X_i to be exaggerated. Most present approaches exaggerate the difference from the mean, although the results of this approach have often been criticized. For example, in [15], Mo et al. emphasized that the distinctiveness of a displaced feature not only depends on its distance from the mean, but also on its variance. Normalizing the proportion differences of a feature by using its mean is viewed as an expression of the feature distinctiveness as follows:

$$distinct_i(j) = \left|\frac{\Delta P_i(j)}{\overline{P}(j)}\right|, \quad j = 1, \ldots, 9. \tag{6}$$

This is because equation (6) reveals the feature's variance in addition to its proportion difference. Note that the area proportion $P_i(9)$ is also used. The triangle that is between the eyes and nose determines if both eyes should be closer or further apart and also if the nose should be longer or shorter [17]. We therefore use $P_i(9)$ as a judgment condition of when and where to apply the third item of the T-shape rule. The distinctive features of X_i are determined by ranking the $distinct_i$ of X_i as follows:

$$k = \arg\max_{j=1}^{9} distinct_i(j). \tag{7}$$

The *k*th feature is regarded as the distinctive feature. If $k = 9,$ then the triangle of the eyes and nose should be enlarged or shrunk. In this case, the third item of the T-shape rule is not employed. Our algorithm does not operate directly on the triangle - instead, we choose the other feature that has the second highest rank *k* exaggeration. This is because enlarging or shrinking the triangle can be implemented by the other proportions. The third item of the T-shape rule is required only when $k < 9,$ for example, when a face has a small distance between two eyes and long nose. For exaggeration, the distance between the eyes is shortening, while the nose is also shortening in order to decrease the difference from the mean. Obviously, this violates the third item of the T-shape rule. We might keep the proportion of nose unchanged to avoid this difficulty.

The second problem is how to re-arrange all features when a salient feature $P_i(j)$ is exaggerated by

$$P_i^{(t)}(j) = \overline{P}(j) + t\Delta P_i(j), \tag{8}$$

where $j \neq 9.$ Due to the T-shape rule, all features vary only along the horizontal and vertical axes. When the *j*th feature is on the horizontal (or vertical) axis, the exaggerated proportion $P_i^{(t)}(j)$ will affect the other two proportions of the horizontal (or vertical) axis. The sum of these two proportions is changed in proportion with

$$\gamma = \frac{1 - P_i^{(t)}(j)}{1 - P_i(j)}.$$

Figure 3. Exaggerating all features by the same t at different levels. All facial elements can be merged within the given facial contour even when t is big. In our experiments, the eyes and eyebrows share the same proportions for convenience. Increasing t results in the eyebrows beyond the contour. In many caricatures, eyebrows may lie beyond the facial contour, so we do not modify them here.

Each of these two proportions (e.g., the *k*th proportion) can be easily updated:

$$P_i^{(\gamma)}(k) = \gamma P_i(k). \tag{9}$$

For other features, it is expected to reduce their difference from the mean. This can be implemented by using another t (which is set as $t < 1$) in equation (8). In our experiments, we let $t \leftarrow \max(1/t, 1 - 1/t)$ for other non-distinctive features. If the *j*th feature is the width of the nose (or mouth), i.e., $P_i(4)$ or $P_i(5)$, then we can simply set $t \leftarrow \max(1/t, 1 - 1/t)$, for all other features. Before updating P_i, we have to apply equations (6) and (7) to P_i for judging if the third item of the T-shape rule is applicable.

Similar to Method 1, applying the updated P_i to the facial elements makes the *j*th feature stand out in the resulting caricature.

The third problem is how to re-arrange all features when there are a few distinctive features exaggerated. For convenience, assume two proportions $P_i(j)$ and $P_i(k)$ are emphasized, respectively, by equation (8). Applying the scheme of equation (9) to $P_i(j)$, $P_i(k)$ yields the following four probable cases:

(1) The jth and kth features appear on the horizontal and vertical axes, respectively $(\text{e.g., } j \in [1, 3] \text{ and } k \in [6, 8])$. We can obtain two γ parameters for updating the other proportions on the horizontal and vertical axes with equation (9). For $P_i(4)$, $P_i(5)$, let $t \leftarrow \max(1/t, 1 - 1/t)$ for updating according to equation (8).

(2) Both features appear on the horizontal (or vertical) axis $(\text{e.g., } j, k \in [1, 3]$ or $j, k \in [6, 8])$. It can be observed that the third proportion is also exaggerated by the same t, i.e., all three features are exaggerated by t. Let the third proportion be $P_i(m)$. Due to $t\Delta P_i(j) + t\Delta P_i(k) + t_m \Delta P_i(m) = 0$, we have $t_m = -\dfrac{t\Delta P_i(j) + t\Delta P_i(k)}{\Delta P_i(m)}$. Because $\Delta P_i(j) + \Delta P_i(k) + \Delta P_i(m) = 0$, we have $t_m = t$. For other proportions, we let $t \leftarrow \max(1/t, 1 - 1/t)$ for updating.

(3) One feature appears on the horizontal (or vertical) axis while the other one does not (e.g., $j \in [1, 3]$ or $[6, 8]$ and $k = 4$ or 5). The other two proportions of the horizontal (or vertical) axis are updated by equation (9) with γ_j, while others by equation (8) with $t \leftarrow \max(1/t, 1 - 1/t)$ for their updating.

(4) The jth and kth features are of the width of mouth and nose $(\text{i.e., } j, k = 4, 5)$, which are not on either the horizontal or vertical axis. Applying $t \leftarrow \max(1/t, 1 - 1/t)$ to all the other proportions.

Before updating P_i, we have to judge whether to apply the third item of T-shape by using the scheme of equations (6) and (7). The same manner can also be applied to the exaggeration of more than two features.

By applying the proportions of facial features to exaggeration, the emphasized and non-emphasized features can be placed in a caricature in terms of their individual proportions. This makes the exaggeration both simple to achieve and intuitive. Although the distinctive features could be chosen in terms of equations (6) and (7), which features are to be exaggerated and the amount of exaggeration t could still, in practice, be chosen by the user. This gives the user a good deal of flexibility in deciding the relative influence of specific features. Our algorithm is capable of conveniently synthesizing all exaggerated and non-exaggerated features into a caricature.

4. Likeness

A good caricature is expected to look like its original subject. However measuring "likeness" remains very challenging. To our knowledge, likeness has not been well studied in literature on caricature synthesizing. When creating a synthesized caricature, the exaggerated features are highlighted while the non-exaggerated features should be adjusted to an optimal configuration, so that the resulting caricature looks like the original subject.

To implement likeness, we first need to establish a similarity metric to measure the "likeness" between or among face images. We believe the Modified Hausdorff Distance (MHD) [10] is a good candidate, which will be applied to measuring the likeness of a caricature to its original.

Thus we define the similarity of a caricature face pair (X_i, X_i^*), consisting of the exaggerated image X_i^* and the original one X_i, as,

$$MHD(X_i, X_i^*) = \max\left\{\frac{1}{|X_i|}\sum_{x\in X_i} d(x, X_i^*), \frac{1}{|X_i^*|}\sum_{x\in X_i^*} d(x, X_i)\right\}, \tag{10}$$

where $d(x, X_i) = \min_{y\in X_i} \| x - y \|^2$. Using this metric, in the following, we address the problem of how to adjust the facial elements (excluding the facial contour) of X_i^*, so that the created caricature still looks like the original X_i.

To adjust the facial elements is to change their individual positions, sizes and angles, i.e., rigid transforms. We follow the T-shape rule and restrict the movement of facial elements accordingly. For example, eyes (including eyebrows) move along the horizontal axis while nose and mouth move along the vertical axis. However, the variance of every element is small by optimizing the configuration of facial elements below. Hence, the third item of the T-shape rule is not required here.

The movements of facial elements include the lateral shift of eyes, scaling and rotation of eyes, lengthening or shortening of nose, broadening or narrowing of nose, and movement of mouth. The eyebrows have the similar movement with the eyes.

For convenience, let the eyebrows share the same movement with the eyes. Due to the translation of facial elements along the given horizontal and vertical axes (including lengthening or shortening of nose), the translation can be simply set as a scale *tr*. Taking account of the lateral shift of eyes, the left and right eyes share the same scaling and rotation but opposite translational scale *tr*, i.e., $+tr$ and $-tr$. Denote the scaling factor, rotation angle and translational scale of eyes as $(S^{(j)}, \theta^{(j)}, tr^{(j)})$, $j = 1, 2,$ the widening scale and lengthening scale of nose as $(S^{(3)}, tr^{(3)})$ and the scaling factor, rotation angle and translational scale of mouth as $(S^{(4)}, \theta^{(4)}, tr^{(4)})$. As $(S^{(1)}, \theta^{(1)}, tr^{(1)}) = (S^{(2)}, \pi - \theta^{(2)}, -tr^{(2)})$, there are eight undetermined parameters altogether. Note that the rigid transform is converted into three scalar parameters here.

To preserve the exaggerated features, the relative size or proportion must be fixed while adjusting the facial elements. For example, when the distance between both eyes is exaggerated, the eyes might be moved along the horizontal axis and their sizes and angles might also be changed. But, the distance between two eyes must be preserved. The constraints thus include the unchangeable exaggerated features (e.g., if the width of the nose is exaggerated, then set $S^{(3)} = 1$ during optimization), and facial elements are restricted from crossing the boundary of the facial contour.

It is straightforward to formulate the adjustment of the facial elements as the minimization of the similarity between the caricature face pair (X_i^*, X_i) with respect to a set of rigid transforms $(S^{(j)}, \theta^{(j)}, tr^{(j)})$, $j = 1, \ldots, 4,$ as follows:

$$\begin{cases} \min\limits_{(S^{(j)}, \theta^{(j)}, tr^{(j)})} MHD(X_i, X_i^*), \\ \text{subject to constraints.} \end{cases} \tag{11}$$

The cost function equation (11) is usually discontinuous, which renders many optimization techniques invalid. For this reason, we apply the simplex method to smooth it out, since the simplex method can effectively handle discontinuity, particularly, if the initial guess does not occur near the solution.

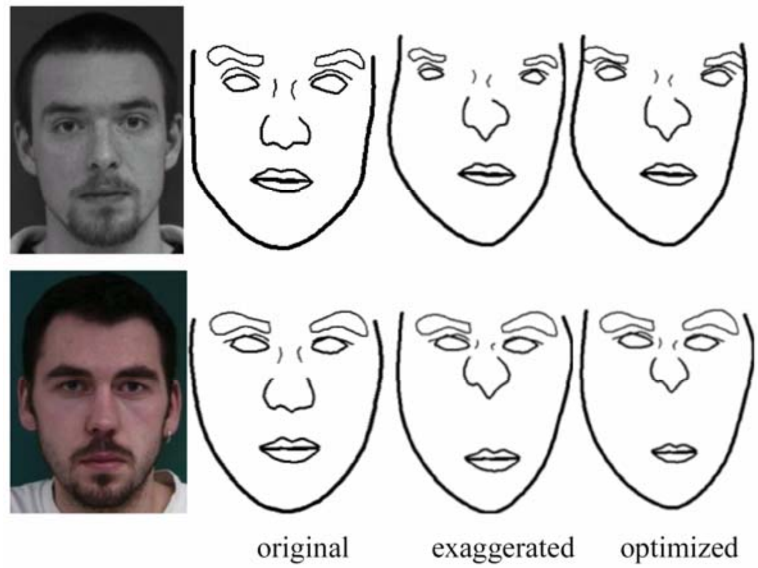

Figure 4. The results of the likeness optimization. The examples for shape learning are the same as in Figure 1.

Figure 4 shows the results of the likeness optimization. The facial contours and noses of the original faces are separately exaggerated first by equations (1)-(4), and then the distance between the eyes in the 1st row is exaggerated while the philtrum in the 2nd row is exaggerated at the relationship exaggeration step. After that, the likeness optimization of equation (11) is applied to the resulting caricatures. For the sake of demonstration, we set the amount of exaggeration t up to the extreme in relationship exaggeration. It can be observed that in the 1st row of Figure 4, the eyes are enlarged by the likeness optimization of equation (11) while the emphasized feature (i.e., the distance between two eyes) is maintained. In the 2nd row of Figure 4, the sizes of the mouths and noses are reduced while the exaggerated feature (i.e., philtrum) is unchanged.

5. Experiments and Analysis

Our experiments include two parts: one is to demonstrate the relationship exaggeration of all features, one feature and two features, respectively; the other is to illustrate the generated final caricatures, including shape and relationship exaggeration and likeness optimization. The contours of the original facial features are extracted by using the AAM method [3].

Experiment 1

In this experiment, we first fixed the original facial contour and applied Method 1 and Method 2 (with one and two exaggerated features, respectively). To illustrate relationship exaggeration, all shapes are unchanged here. When all features are emphasized (the eyes are enlarged, the nose is widened and shortened, and the mouth is widened, etc.) as shown in the 1st row of Figure 5, it is difficult to make the distinctive features stand out in a caricature. The 2nd row of Figure 5 shows the results of exaggerating the width of the mouth at different levels of exaggeration t. The 3rd row of Figure 5 shows the results of exaggerating the width of the mouth and length of the nose at different levels t. The emphasized features are chosen using equations (6) and (7). We can see from this figure that the selected features are exaggerated while the others are made less conspicuous. The emphasized features are prominent in the caricatures as shown in the 2nd and 3rd rows of Figure 5.

Experiment 2

In this experiment, the shape exaggeration includes the exaggeration of the facial contour and nose shape using equations (1)-(4). Two facial features are emphasized in the relationship exaggeration by Method 2. The resulting caricatures are then re-configured with the likeness optimization of equations (10) and (11). More examples are shown in Figure 6. The 2nd, 4th and 6th facial contours are exaggerated in the same style. The others are in a different style. The 3rd and 6th faces still include the exaggeration of the nose shapes besides facial contours in the shape exaggeration. For illustration, each face has two distinctive features to be exaggerated in the relationship exaggeration. In the final caricatures in the 2nd and 4th rows, the texture styles of given example caricatures are transferred respectively into the resulting face sketches.

In the 2nd face, it can be observed that the distance between the eyes shortens while the nose lengthens. Clearly, this obeys the third item of the T-shape rule. Compared with the 3rd face, it can be observed that the distance between both eyes shortens while the nose shortens too. This is because the rank of $distinct_i(9)$ of the triangle area proportion $P_i(9)$ is the biggest. In this situation, the third item of the T-shape rule is not applicable. For the 4th face, the width of eyes and the distance between the eyes are required to be exaggerated. Indeed, the three proportions of the horizontal axis are exaggerated by the same *t*. This is because both features (i.e., the width of eyes and the distance between two eyes) appear on the horizontal axis.

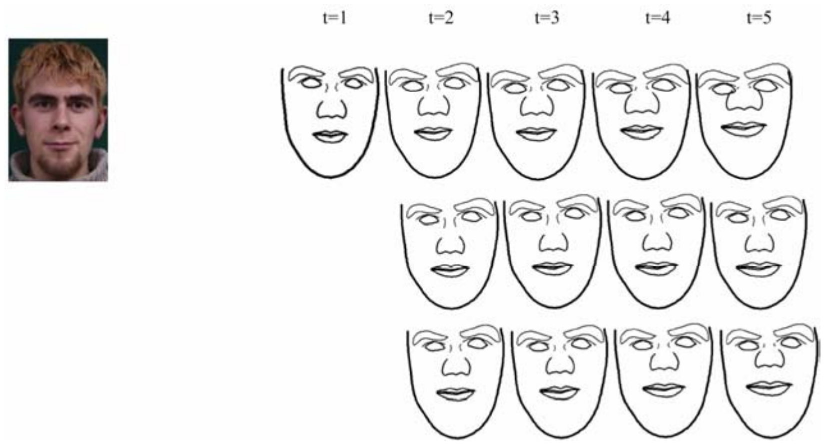

Figure 5. Comparison of exaggerating all features and few distinctive features. The 1st row shows the exaggeration results by using Method 1. The 2nd row shows the exaggeration results by using Method 2 with one exaggerated feature (mouth's width). The 3rd row shows the exaggeration results by using Method 2 with two exaggerated features (mouth's width and nose's length).

Additionally, it can also be noted that the final caricatures in the 2nd and 4th rows have hair contours and eyeballs. The eyeballs are drawn manually. The hair contours can be obtained by simple image processing as follows. After extracting the facial contours by AAM method, we first remove the background by the thresholding techniques, e.g., histogram method, and then apply the skin detection technique to the texture image. Incorporating the extracted facial contours, we can get a rough hair region. To obtain the hair contour, we view the two endpoints of the facial contour as two fixed vertices of the hair contour, and approximate the hair region by using spline curve.

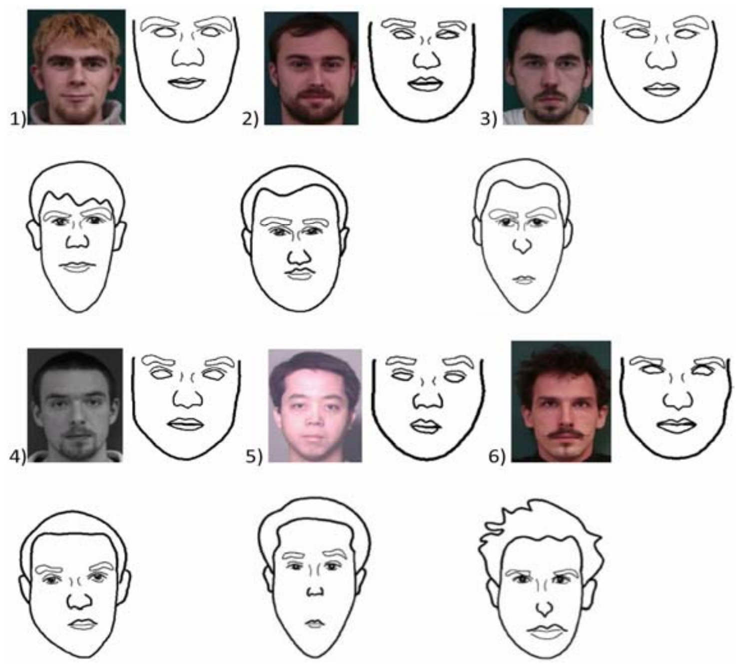

Figure 6. A number of generated caricatures using our approach. The 1st and 3rd rows show the original face images and contours, while the 2nd and 4th rows show the exaggerated results. The 1st face enlarges his mouth and shortens his nose. The 2nd face shortens the distance between two eyes and his philtrum. The 3rd face shortens the distance between two eyes and enlarges his philtrum. The 4th face exaggerates the width of eyes and the distance between two eyes. The 5th face enlarges his philtrum and shortens his mouth. The 6th face enlarges his mouth and lengthens his nose.

6. Conclusions and Future Work

In this paper, we present an example based caricature synthesis approach. It consists of three steps: shape exaggeration, relationship exaggeration and optimization for likeness. Unlike other published approaches, our new shape exaggeration method is based on only one or a few examples of facial components. We present a new relationship exaggeration technique based on the T-shape rule and introduce proportional descriptions of features, making global feature synthesis a simple task. We also introduce the Modified Hausdorff Distance (MHD) metric as a measure of the likeness of a caricature to the original image. Using this metric, we optimize the configuration of facial elements by maximizing the likeness while in the same time ensuring a number of constraints are maintained. Our experimental results demonstrate the effectiveness of this approach.

However, there remain a number of issues in our current development, which will be investigated in the future. The hair style and head shape have not been considered due to the problem of hair occlusion. This applies specifically to faces with long hair, which occludes a substantial part of the face, for example, the faces of female subjects. In addition, texture style transferring also needs to be considered. By modeling the style of drawing line, each artist will be able to produce more fantastic exaggeration effects.

Regarding the measurement of face likeness, there was little work done in the area of caricature synthesis. Our MHD based similarity metric definition attempts to tackle this issue. However, because likeness is both an objective measure and is also up to subjective interpretations, it is difficult to argue that our presented approach produces the best result. We believe that there is more work needed in the future to ascertain a more effective measurement of this property. What is promising here is that we have been able to establish a mechanism that can incorporate any likeness metric to optimization, hence, making it more relatively straightforward to add a new likeness metric. The lack of likeness measurements is also relevant to the objective comparison of different caricature synthesis techniques. Most caricature synthesis techniques are to 'learn' the styles of real caricatures. Without an objective measurement of likeness, it is impractical to compare which technique produces better or worse synthesized artefacts. This is why we have not attempted to compare our outcomes with those produced with different methods.

Acknowledgement

The authors are very grateful to Dr. Richard Southern for his careful proofreading of the manuscript.

References

[1] E. Akleman, Making caricature with morphing, Proc. of ACM SIGGRPH, Los Angeles, California, 1997, p. 145.

[2] E. Akleman, J. Palmer and R. Logan, Making extreme caricatures with a new interactive 2D deformation technique with simplicial complexes, Proc. of the 3rd International Conference on Visual Computing, Mexico, 2000, pp. 165-170.

[3] R. H. Davies, C. J. Twining, T. F. Cootes J. C. Waterton and C. J. Taylor, 3D statistical shape models using direct optimization of description length, Proc. of European Conference on Computer Vision 3 (2002), pp. 3-20.

[4] S. E. Brennan, Caricature generator: the dynamic exaggeration of faces by computer, Leonardo 18(3) (1985), 170-178.

[5] P. Y. Chiang, W. H. Liao and T. Y. Li, Automatic caricature generation by analyzing facial features, Proceedings of Asian Conference on Computer Vision, Jeju, Island, Korea, 2004.

[6] H. Chen, Z. Liu, C. Rose, Y. Xu, H. Y. Shum and D. Salesin, Example-based composite sketching of human portraits, Proceedings of 3rd International Symposium on Non-photorealistic Animation and Rendering, ACM, New York, 2004, pp. 95-102.

[7] H. Chen, Y. Q. Xu, H. Shum, S. C. Zhu and N. N. Zhen, Example-based facial sketch generation with non-parametric sampling, Proc. of the 8th International Conference on Computer Vision, Vancouver, Canada, 2001, pp. 433-438.

[8] H. Chen, N. N. Zheng, L. Liang, Y. Li, Y. Q. Xu and H. Y. Shum, PicToon: a personalized image-based cartoon system, Proc. of ACM Int'l Conf. on Multimedia, Juan-les-Pins, France, 2002, pp. 171-178.

[9] The American Heritage Dictionary of the English Language, 4th ed., Houghton Mifflin, 2000.

[10] M. P. Dubuisson and A. K. Jain, A modified Hausdorff distance for object matching, Proc. of Int'l Conf. on Pattern Recognition, Jerusalem, Israel, 1994, pp. 566-568.

[11] B. Gooch, E. Reinhard and A. Gooch, Human facial illustrations: creation and psychophysical evaluation, ACM Transactions on Graphics 23(1) (2004), 27-44.

[12] H. Koshimizu, M. Tominaga, T. Fujiwara and K. Murakami, On Kansei facial processing for computerized facial caricaturing system Picasso, Proc. of IEEE Int'l Conf. on Systems, Man and Cybernetics 6 (1999), 294-299.

[13] L. Liang, H. Chen, Y. Q. Xu and H. Y. Shum, Example-based caricature generation with exaggeration, Proc. of 10th Pacific Conf. on Computer Graphics and Applications, China, 2002, pp. 386-393.

[14] R. Lenn, How to draw caricature, Contemporary Books, 1984.

[15] Z. Y. Mo, J. P. Lewis and U. Neumann, Improved automatic caricature by feature normalization and exaggeration, Proc. of ACM SIGGRAPH 2004 Sketches, Los Angeles, California, 2004, p. 57.

[16] T. Richmond, How to draw caricature, 2008, http://www.tomrichmond.com.

[17] J. Seiler, The Art of Caricature with Jason Seiler (video tutorial), 2008, https://www.schoolism.com/.

[18] R. N. Shet, K. H. Lai, E. A. Edirisinghe and P. W. Chung, Use of neural networks in automatic caricature generation: an approach based on drawing style capture, Proc. of the Second Edition of the Iberian Conference on Pattern Recognition and Image Analysis, Portugal, 2005, pp. 343-351.

[19] C. C. Tseng and J. J. Lien, Synthesis of exaggerative caricature with inter and intra correlations, Proc. of Asian Conference on Computer Vision, Tokyo, Japan, 2007, pp. 314-323.

[20] G. Z. Xu, M. Kaneko and A. Kurematsu, Synthesis of facial caricature using eigenspaces, IEICE Trans. on Information and Systems (Japanese ed.) J84-D-II(7) (2001), 1279-1288.

[21] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless and D. Salesin, Image analogies, Proc. of SIGGRAPH, 2001.

[22] A. Efros and W. Freeman, Image quilting for texture synthesis and transfer, Proc. of SIGGRAPH, 2001, pp. 341-346.

[23] N. Andrew and A. Marc, Hybrid texture synthesis, Proc. of the Eurographics Symposium on Rendering, 2003, pp. 97-105.