# Mean Value Coordinates Based Caricature and Expression Synthesis

**Hongchuan YU and Jian J. Zhang**

**NCCA, Bournemouth University, Poole, UK**

**{hyu,jzhang}@bournemouth.ac.uk**

**Abstract**

We present a novel method for caricature synthesis based on mean value coordinates (MVC). Our method can be applied to any single frontal face image to learn a specified caricature face pair for frontal and 3D caricature synthesis. This technique only requires one or a small number of exemplar pairs and a natural frontal face image training set, while the system can transfer the style of the exemplar pair across individuals. Further exaggeration can be fulfilled in a controllable way. Our method is further applied to facial expression transfer, interpolation and exaggeration, which are applications of expression editing. Additionally, we have extended our approach to 3D caricature synthesis based on the 3D version of MVC. With experiments we demonstrate that the transferred expressions are credible and the resulting caricatures can be characterized and recognized.

**Keywords:** Caricature and expression synthesis, Mean value coordinates.

## 1. Introduction

In this paper we present a new technique for the synthesis of novel human face caricatures, learning from existing examples. The purpose is twofold. The first is to facilitate caricaturists to produce caricatures efficiently by providing them with initial templates which they can change. This will save them time and allow them to concentrate on their creative work. The second is to enable a novice to learn and produce caricatures for entertainment purposes by mimicking one or more existing caricature styles.

Caricature is a form of 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.

Current approaches of caricature synthesis are typically based on one or a set of frontal face images. The resulting caricatures might be produced by learning a specified artistic style associated with a training set or a set of semi-regular rules. From a practical perspective, a general user can only get one or a small number of caricature samples from a caricaturist or with the same artistic style. The first problem we encounter is which features to use and how to exaggerate them in terms of one or a small number of given caricature face pairs, each consisting of a natural face image and its corresponding caricatured face image. The second problem is how to allow users to modify the result to add personality to the subject. The third is propagating these exaggerations to a 3D model. Indeed, an interactive 3D editing tool is very useful in practice. From a frontal view face, the change of facial expression is prone to be perceived by people. The fourth is therefore how to transfer, interpolate and exaggerate the facial expression. Moreover, users can further expect the visual similarity of the resulting caricature to the subject.

Our work tackles these above-mentioned problems. In this paper we present a new synthesis algorithm based on the deformation property of Mean Value Coordinates (MVC) [16,21]. Our contributions can be summarized as follows:

- Training set of caricature face pairs. We divide exaggeration into two stages, shape and relationship exaggerations. The shape exaggeration of individual face components is computed by learning from one or a small number of caricature face pairs rather than a large training set of caricature face pairs, while the exaggeration of relationship among facial components depends on the user preferences. In this paper we apply MVC to shape learning and exaggeration, since MVC stores the features of the original subjects and deforms them in terms of the specified control polygons (or polyhedrons). It proves both simple and intuitive;

- Facial expression interpolation and exaggeration. Facial expression can usually be regarded as a special case of facial caricature. We will show how to transfer the facial expression to a neutral frontal face, and how to interpolate and exaggerate facial expressions;
- Optimization for Likeness. In existing methods, “likeness” is seldom considered for caricature synthesis due to lack of a “likeness” metric. We incorporate a likeness metric in our caricature model. By optimizing the configuration of the facial components we ensure the resulting caricature resembles the original subject;
- 3D Caricature. The 3D version of MVC is introduced to 3D face caricature generation. We will show how to model a specified human face and interactively produce its 3D face caricatures based on a single frontal face image.

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

### 1.1. Related Work

The relevant approaches to caricature generation can be categorized into three groups. The first is template based morphing where the user manually deforms a template to produce a new caricature, such as [1,2,11,8]. This kind of 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. This idea has been employed in many caricature systems, such as PICASSO [12]. However different opinions exist regarding the effectiveness of EDFM. The central question is whether one 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 its variance”. Many researchers focused on distinguishing the distinctiveness of the facial features, such as [5,19,18]. These approaches essentially formulate some semi-regular rules to exaggerate the difference. Indeed, distinguishing the distinctiveness of the facial features depends on users. There are no objective standards. Thus, our proposed approach provides an interactive way to exaggerate the given features. We concentrate on simplicity, intuitiveness and likeness in this paper.

The third group includes the example based learning methods. These approaches usually need a training database containing a large number of caricature face pairs from a particular artistic style, such as [13,6,7,18]. In practice, however, it is difficult to get a large training set of caricatures that have the same style or are 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.

Synthesis of facial expressions has been studied in both real face images [22] and in computer animation [20]. One of the challenges is to generate expression details, such as the wrinkles caused by skin deformation. This usually requires example data of the facial component movements for the expression synthesis. To this end, Blanz et al. in [10] proposed the Morphable Model of 3D face to generate the reasonable facial component movements. Zhang, et al. in [9] proposed a technique to infer the feature point motion from a given training set. However, these methods are difficult to set up the required feature tracking. In this paper, we present a simple method to transfer facial expressions, and show that by interpolation and exaggeration the expressions can be exaggerated without tracking the feature points.

3D caricature is becoming an active research topic in recent years [23-25]. The challenging problem is how to model a 3D face from images. In addition, it is difficult to apply learning methods to 3D caricature synthesis because of the unavailability of the training set of 3D caricatures. Practically the way around this issue is to provide the caricaturists tools for 3D face modeling and interactive editing. Our aim in this paper is therefore to develop a simple and effective approach to 3D caricature synthesis.

## 2. System Overview

Our goal is to synthesize a caricature by example using one or a small number of given caricature face pairs. The presented caricature synthesis method is developed based on Mean Value Coordinate framework. Basically, it consists of three main steps as shown in Figure 1. They are (1) example based shape learning for learning the style of given caricature face pairs; (2) relationship exaggeration for further exaggerating a specified facial component; and (3) the optimization for maximizing the likeness.

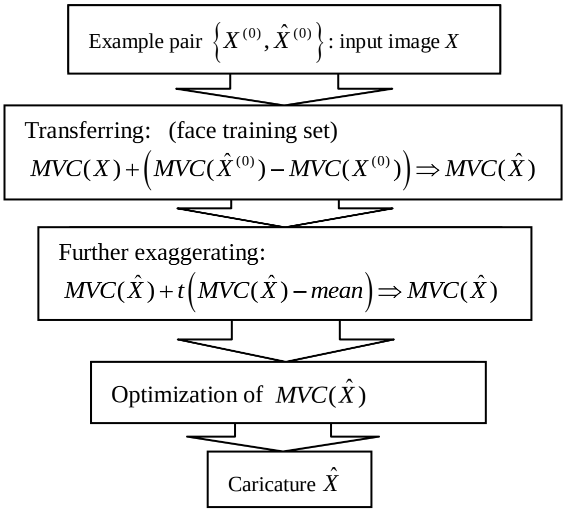



Figure 1. Flowchart of caricature synthesis system.

A human face can be decomposed semantically into seven facial components, which are facial contour, left and right eyebrows, left and right eyes, a nose and a mouth. Each facial component may be further divided into several prototypes (e.g. the eyebrow component has two prototypes, thick and thin) based on its appearance in the individual training datasets. A caricature is usually represented by shape exaggeration of individual facial components, and the exaggeration of the relationship between these facial components. The latter includes position, size and angle of the facial components [14]. For example, eyebrows are exaggerated in the shape of a thin curve (shape exaggeration) while they may be moved apart from each other (relationship exaggeration). Shape exaggeration can be implemented by learning the style of a given shape, while relationship exaggeration usually depends on a global model which handles the overall arrangements. But, capturing the global model implicitly needs a very large training set, such as in [13], since the seven facial components and their parameters (including scaling, position and orientation) lead to a huge number of combinations. Usually, any given subject might have several different interpretations with respect to the exaggeration of the relationships of its facial components and each may be as successful [14]. Clearly, relationship exaggeration usually depends on users. We thus think these two kinds of exaggerations might be handled independently in a drawing. In shape exaggeration, our approach attempts to learn the shape style of some specified facial component by example using one or a small number of given caricature face pairs. In relationship exaggeration, our approach attempts to exaggerate one or a few features, which are specified and controlled by the user. This will assist caricaturists to produce caricatures semi-automatically while allowing them to adjust exaggeration effects at anytime.

A good caricature is expected to look like its original subject. But measuring “likeness” remains very challenging. To our knowledge, likeness has not been well studied in existing literature on caricature synthesis. When creating a synthesized caricature, the exaggerated features are highlighted while the non-exaggerated features should be adjusted to an optimal configuration as well, so that the resulting caricature looks like the original subject. Thus, our approach attempts to handle the “likeness” under the MVC framework.

3D caricature synthesis is becoming a new research issue in recent years. For completeness of caricature synthesis, we extend our presented approach to 3D caricature generation and develop an interactive 3D editing tool under the 3D MVC framework.

The rest of this paper is organized as follows. Section 3 addresses our approach, including shape exaggeration, relationship exaggeration and how to maximize the “likeness” of a caricature. Section 4 addresses a series of applications of our approach in the 2D caricature synthesis, facial expression exaggeration, and 3D caricature generation based on a single frontal view image. Section 5 concludes the paper by looking into the areas of improvement for our future work.

## 3. Exaggeration

### 3.1. Mean Value Coordinates

Mean Value Coordinates (MVC) presented in [16,21] can provide a simple means to linearly interpolate the interior and exterior of any polytope without self-intersection. For arbitrary planar polygons without self-intersection, any point $x$ can be expressed by using the vertices of the polygons $v_1,...,v_n$ in an affine combination form [16], i.e. $\sum_{i=1}^{n}\lambda_i(x)v_i = x$, and the MVC of $x$ satisfy $\sum_{i=1}^{n}\lambda_i(x)=1$. Our basic idea is to make use of the deformation formula of MVC based on arbitrary planar polygons, i.e. $f: R^2 \to R^2$,

$$\begin{cases} f^{-1}(\hat{x}) = \sum_i \hat{\lambda}_i(\hat{\psi},\hat{x})v_i = x \\ f(x) = \sum_i \lambda_i(\psi,x)\hat{v}_i = \hat{x} \end{cases}, \qquad (1)$$

where, $\psi,\hat{\psi}$ are the corresponding control polygons separately on two planes, $v_i \in \psi$ and $\hat{v}_i \in \hat{\psi}$ consist of a set of the corresponding vertex pairs $\{(v_i,\hat{v}_i)\}$, points $x,\hat{x}$ are defined respectively on two planes and $\lambda_i,\hat{\lambda}_i$ are their corresponding MVC. Note that MVC $\lambda_i$ of $x$ (or $\hat{\lambda}_i$ of $\hat{x}$) depend on the control polygon $\psi$ (or $\hat{\psi}$). Thus, we denote MVC as $\lambda_i(\psi,x)$ (or $\hat{\lambda}_i(\hat{\psi},\hat{x})$) here. It can also be formulated in a matrix form as $\begin{cases} \hat{x} = \lambda(\psi,x)\hat{\psi} \\ x = \hat{\lambda}(\hat{\psi},\hat{x})\psi \end{cases}$, where, $\lambda$ and $\hat{\lambda}$ are row vectors, and $\psi,\hat{\psi}$ are composed of the point list in a matrix form. Furthermore, for a set of control polygons $\{\psi_i\}$, it is required that the nested polygons should preserve opposite orientations, i.e. clockwise vs. counterclockwise, while the nearby polygons should keep the same orientation. MVC based texture mapping can achieve a continuous texture mapping without any pre-triangulation. The mapping $M$ from the source image $I(x)$ to the target $\hat{I}(\hat{x})$ is expressed as,

$$\begin{cases} M: I(x) \to \hat{I}(\hat{x}) \\ x = \sum_i \hat{\lambda}_i(\hat{\psi},\hat{x})v_i, v_i \in \psi \end{cases}, \qquad (2)$$

where points $x,\hat{x}$ are defined respectively on the 2D image planes, and $\psi,\hat{\psi}$ might be a set of the nested control polygons. It is noted that the MVC framework first establishes a one to one mapping of the point coordinates between $I(x)$ and $\hat{I}(\hat{x})$, and then the pixel at $x$ in $I(x)$ is mapped into that of $\hat{x}$ in $\hat{I}(\hat{x})$ accordingly. In practice, we prefer to an inverse mapping $M^{-1}: \hat{I}(\hat{x}) = I(x(\hat{x}))$.

**Remark**

The distinct features of MVC are the smoothness and affine precision, i.e. $\lambda_i$ is $C^\infty$ everywhere expect at the vertices $v_i$ of the control polygons where it is only $C^0$; and a linear function can be reproduced exactly from the interpolation values. Most of the classical 2D deformation techniques used to apply affine transformations to a triangulation of the deformation region. From a perspective of numerical computation, they don't have as good smoothness and affine precision compared with MVC deformation. Other warping techniques [17] with B-splines and Radial Basis Functions have similar numerical properties. However, compared to other approaches, another important property of MVC deformation is that it is linear along the edges of the control polygons, which is useful to preserve some basic geometrical structures during deformation. Moreover, the areas of surrounding the edges of polygons still remain smooth.

### 3.2. Shape Exaggeration

For shape exaggeration, our goal is to create a caricature by learning the shape style of some specified facial component. Without loss of generality, we firstly consider the case of the training set containing a set of caricature face pairs with the same style, i.e. $\{(X^{(i)}, \hat{X}^{(i)}), i = 1,...,n\}$, where $X^{(i)}$ denotes the neutral face, while $\hat{X}^{(i)}$ denotes the corresponding caricature. Each $X^{(i)}$ (or $\hat{X}^{(i)}$) contains a set of the given polygons of the 7 facial components, and moreover, each polygon describes the shape of a specific facial component as shown Fig.2. For convenience, these polygons are usually stored in a point list form. To remove the influence from position and scaling of face images, the vertex's coordinates are usually normalized by shifting the origin to the midpoint O between two eyes and quantifying by using the width $\|\overline{AB}\|$ of facial contour as shown in Fig.2.

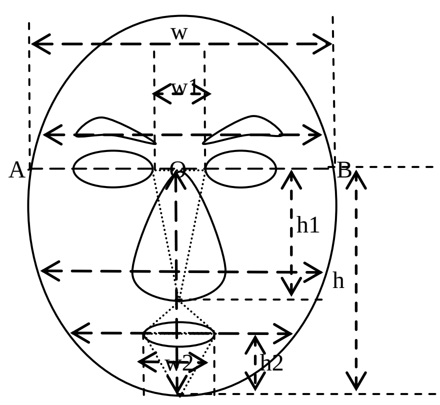


Figure 2. Illustration of facial features' contours.

We specify some facial component for shape learning here, i.e. the *j*th specified component $X_j^{(i)}$ is a subset of $X^{(i)}$, and assume that the shape style of the specified facial component is the same as all the exemplar pairs in the training set. The basic idea is to first build the two eigenspaces respectively based on the training sets $\{X_j^{(i)}\}$ and $\{\hat{X}_j^{(i)}\}$, and then apply them to a new input neutral face for its caricature synthesis. To this end, we apply PCA technique to the training set. The exemplar pair $\left(X_j^{(i)}, \hat{X}_j^{(i)}\right)$ can be linearly represented as follows,

$$\begin{cases} X_j^{(i)} = \bar{X}_j + \sum_k \alpha_k \mathbf{x}_k \\ \hat{X}_j^{(i)} = \bar{\hat{X}}_j + \sum_k \hat{\alpha}_k \hat{\mathbf{x}}_k \end{cases},$$

where $\mathbf{x}_k, \hat{\mathbf{x}}_k$ denote the eigenvectors of $\{X_j^{(i)}\}$ and $\{\hat{X}_j^{(i)}\}$, and $\bar{X}_j, \bar{\hat{X}}_j$ denote their means respectively. Applying the scheme of Eq.(1) to $\left(X_j^{(i)}, \hat{X}_j^{(i)}\right)$ yields a mapping of $\hat{X}_j^{(i)} = \lambda\left(\psi^{(i)}, X_j^{(i)}\right)\hat{\psi}^{(i)}$, where control polygons $\psi^{(i)}, \hat{\psi}^{(i)}$ might be the complement sets of $X_j^{(i)}, \hat{X}_j^{(i)}$ respectively or the polygons of other facial components, e.g. learning the shape of nose, one can utilize facial contour as control polygon.

For a new input $X_j$, we hope to seek a most probable exemplar pair $\left(X_j^{(*)}, \hat{X}_j^{(*)}\right)$ in the training sets $\{X_j^{(i)}\}$ and $\{\hat{X}_j^{(i)}\}$, and then transfer the shape style of $\left(X_j^{(*)}, \hat{X}_j^{(*)}\right)$ to $X_j$ for its caricature synthesis. This can be implemented as follows. One can represent the input as $X_j'$ by a linear combination of principal components of $\{X_j^{(i)}\}$, and further identify the closest example $X_j^{(*)}$ to $X_j'$ in the training set $\{X_j^{(i)}\}$. Please note here we denote this linear representation as $X_j'$, since $X_j'$ is only approximate to the original $X_j$ but not identical to it. Our goal is to transfer the shape style of $\left(X_j^{(*)}, \hat{X}_j^{(*)}\right)$ to $X_j'$. It can be achieved by minimizing,

$$\min_{\hat{\alpha}} \left\| \hat{X}_j^{(*)} - \lambda\left(\psi', X_j'\right)\hat{\psi}(\hat{\alpha}) \right\|^2, \qquad (3)$$

where, $\hat{\psi}(\hat{\alpha})$ is expressed by a linear combination of the eigenvectors $\hat{\mathbf{x}}_k$ and $\hat{\alpha}$ denotes the weight vector. This usually yields a linear system. Solving $\hat{\alpha}$ yields the caricature $\hat{X}_j$ of $X_j$ with the shape style of $\left(X_j^{(*)}, \hat{X}_j^{(*)}\right)$.

Then, consider the case of the training set only containing the neutral faces, i.e. $\left\{X^{(i)}, i=1,...,n\right\}$. Let $\left(X^{(0)}, \hat{X}^{(0)}\right)$ be a given caricature face pair for learning. For learning purposes, it is usually not sufficient to use only one or a small number of training samples. Due to the practical difficulties mentioned in Section 1.1, we model this problem with some restraints, e.g. only one facial component is exaggerated in shape exaggeration every time.

Due to lack of the caricature training set $\left\{\hat{X}^{(i)}\right\}$ here, we need to firstly generate the counterpart of $\left\{X^{(i)}\right\}$ by minimizing the following functional,

$$\min_{\hat{T}} \sum_{i}^{n} \left\| \hat{X}_j^{(0)} - \lambda\left(\psi^{(i)}, X_j^{(i)}\right) \hat{T} \hat{\psi}^{(0)} \right\|^2 . \qquad (4)$$

Note that $\hat{\psi}^{(0)}$ is the control polygon of the exemplar $\hat{X}_j^{(0)}$ instead of others here. The resulting $\hat{T}$ is a linear operator, which is in a matrix form of size N×N, where N denotes the number of the vertices on Control Polygons. Producing $\left\{\hat{X}_j^{(i)}\right\}$ can be achieved by using $\hat{X}_j^{(i)} = \lambda\left(\psi^{(i)}, X_j^{(i)}\right)\hat{T}\hat{\psi}^{(0)}$. One can construct the eigenspace of the resulting counterpart $\left\{\hat{X}^{(i)}\right\}$ just like $\left\{X^{(i)}\right\}$. For an input *X*, we firstly project it onto the eigenspace of $\left\{X^{(i)}\right\}$ for its linear representation $X'$ and then follow the scheme of Eq.(3) as,

$$\min_{\hat{\alpha}} \left\| \hat{X}_j^{(0)} - \lambda\left(\psi', X_j'\right) \hat{\psi}(\hat{\alpha}) \right\|^2 , \qquad (5)$$

where the vector $\hat{\alpha}$ contains the weights of the linear combination of $\left\{\hat{X}^{(i)}\right\}$. The operator $\hat{T}$ obtained from Eq.(4) is employed to construct the eigenspace of $\left\{\hat{X}^{(i)}\right\}$, while the vector $\hat{\alpha}$ further convey the caricature style of $\left(X_j^{(0)}, \hat{X}_j^{(0)}\right)$ to the input shape *X* for its caricatured $\hat{X}_j = \lambda\left(\psi', X_j'\right)\hat{\psi}(\hat{\alpha})$.

### 3.3. Relationship Exaggeration

The relationship exaggeration is to exaggerate the relationship of the 7 facial components as shown in Fig.2. The relationships include position (e.g. the relative distances between facial components), size (each component 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) [14]. Unlike shape exaggeration, the relationship is exaggerated in terms of the Exaggeration Difference From Mean (EDFM) [4] rather than learning an example. In our algorithm, the facial contour is assumed fixed while the other six components are placed into it. We expect to emphasize a small number of facial features rather than all features. The facial features are used to be described as a set of proportions respectively along the horizontal and vertical lines. We herein determine the horizontal line by connecting points A and B of the facial contour, and the vertical line by drawing a line between the midpoint O of the two eyes and the bottom point C of the chin as shown Fig.2. Instead of facial proportions, we employ a set of polygons to describe facial components, such as solid polygons to denote eyebrows, and eyes etc, respectively. The facial proportions hide in the coordinates of the polygons' vertices. Moreover, we add three triangles respectively to describe the distance between the eyes, philtrum and chin, as shown by the dotted triangles in Fig.2.

For a given amount of exaggeration *t* and the specified *j*th facial component $X_j$, one can update the MVC of $X_j$ by,

$$\begin{cases} \lambda\left(\psi, X_j, t\right) = \bar{\lambda} + t\Delta\lambda, \\ \Delta\lambda = \lambda\left(\psi, X_j\right) - \bar{\lambda} \end{cases}, \qquad (6)$$

where, $t > 0$, $\bar{\lambda}$ denotes MVC mean vector of training set and the control polygon $\psi$ is the facial contour. Then, the exaggerated *j*th facial component is updated by $\hat{X}_j \leftarrow \lambda\left(\psi, X_j, t\right)\psi$ . It can be noted that changing the MVC $\lambda$ by varying $t$ results in $X_j$ changed under the same control polygon $\psi$, such as facial proportions. But there is no guarantee to preserve the $X_j$ 's shape here.

For the purpose of likeness, the $X_j$ 's shape should remain unchanged (or only a little change) during the relationship exaggeration. It can be achieved by affine transformation, i.e. the *j*th component is rigidly transformed by $\hat{X}_j = G^{(j)} X_j$, where $G^{(j)}$ denotes affine transformation. $G^{(j)}$ can be yielded by solving the following linear system,

$$G^{(j)}X_j = \lambda\left(\psi, X_j, t\right)\psi . \qquad (7)$$

After that, the exaggerated and non-exaggerated features are merged into the input $X$ for its caricature $\hat{X}$. Furthermore, to make the resulting $\hat{X}$ look like the original subject $X$, one can re-adjust the non-exaggerated facial components, that is, MVC are re-computed based on the nested control polygons $\psi, \hat{\psi}$, i.e. $\psi = \left\{X_1, X_j\right\}, \hat{\psi} = \left\{X_1, \hat{X}_j\right\}$ where $X_1$ denotes the facial contour. And then the procedure of Eq.(7) is applied to all non-exaggerated components, i.e. $G^{(k)}X_k = \lambda\left(\psi, X_k\right)\hat{\psi}, k \neq 1, j$, to update all the non-exaggerated facial components. This allows the facial contour and the exaggerated component contour fixed while adjusting the other non-exaggerated features in an optimal configuration.

### 3.4. Likeness Evaluation

Caricatures as a type of artifacts are conventionally evaluated by 'subject experts' due to its artistic nature. Quantitatively it is so far impractical to compare which technique produces better or worse synthesized caricatures. This presents a difficulty for software developers due to lack of guide. In order to get around this difficulty and to inspire other research efforts among the research community, in this paper we propose to use the Modified Hausdorff Distance (MHD) [27] to measure the likeness of an exaggerated face to its original. This metric is defined as,

$$\begin{cases} MHD = \max\left\{d(A,B), d(B,A)\right\} \\ d(A,B) = \dfrac{1}{N_A}\sum_{a \in A} d(a,B) \\ d(a,B) = \min_{b \in B}\left\|a-b\right\| \end{cases},$$

where $N_A$ denotes the number of the elements in the set $A$. The Hausdorff distance (HD), involving its various modified versions, is a non-linear operator, which measures the mismatch of two sets. Because it takes into account various features of data sets, we believe the MHD captures the key shape indicators of a caricature. Although it is not the only way to assess similarity of spatial point sets, it has proven to be effective in image registration applications. In the following applications, we employ the MHD as a likeness metric to evaluate the likeness of exaggerated face images to the corresponding real face images.

# 4. Applications

In our implementation, we first constructed the AAM models [3] based on a training set and then applied the AAM algorithm to the input face images for extracting the contours of the original facial features. The MVC computation in our experiments is based on the pseudo-code available in [16,21].

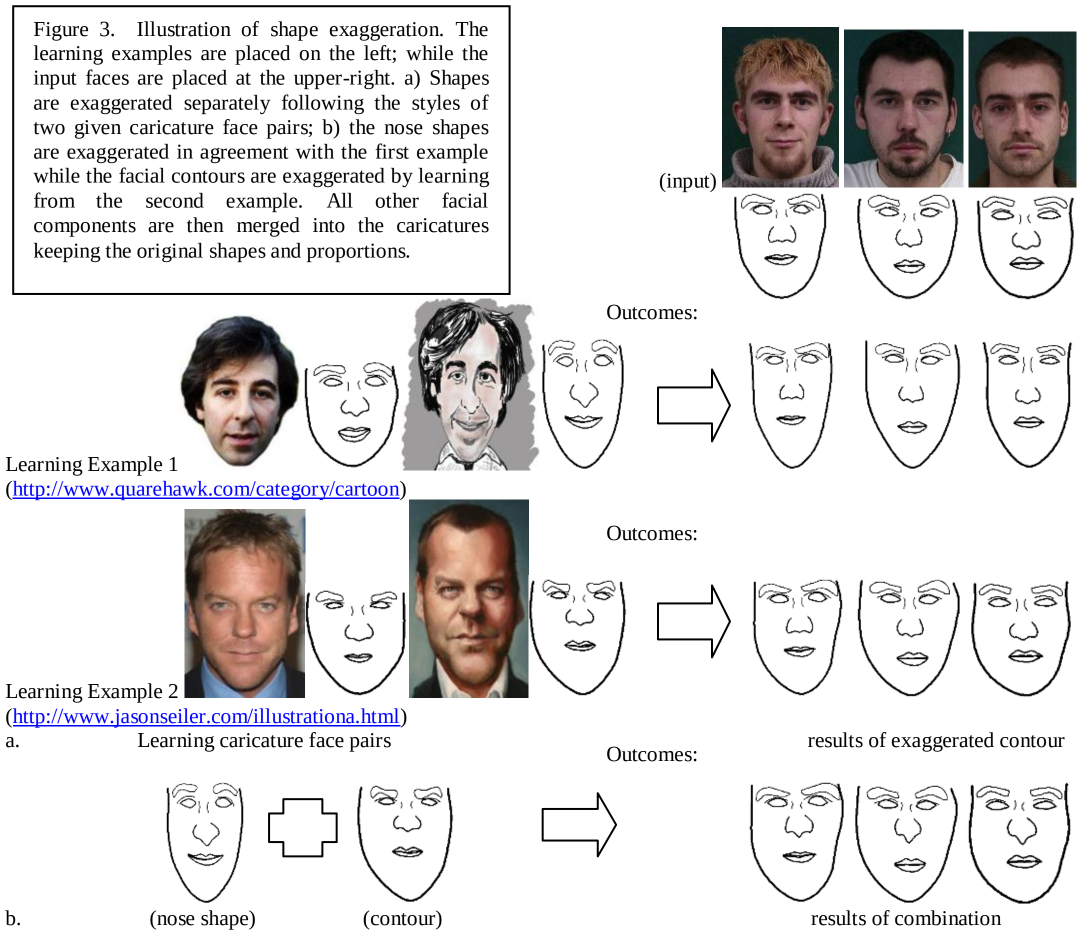



Figure 3. Illustration of shape exaggeration. The learning examples are placed on the left; while the input faces are placed at the upper-right. a) Shapes are exaggerated separately following the styles of two given caricature face pairs; b) the nose shapes are exaggerated in agreement with the first example while the facial contours are exaggerated by learning from the second example. All other facial components are then merged into the caricatures keeping the original shapes and proportions.

## 2D Caricature synthesis

In this application, we apply Eq.(4-7) to the usual 2D caricature synthesis. Firstly, we show that our approach can achieve shape learning without a large training set of caricature face pairs in shape exaggeration, and for the different facial components, our approach can mimic these individual components from several caricature examples respectively. In general, the different caricaturists and artistic traditions draw the facial components differently which give caricatures a distinct style. Therefore a new caricature is expected to be created by taking these individual components from several caricature examples. For instance, one may want to exaggerate a face with a narrow facial contour and short nose. If both features are present in different examples, the solution is to pick up the necessary features from the respective example caricatures.

Figure 3a only shows the results of facial contour exaggeration with Eq.(4,5). The other facial components (e.g. the eyes and mouth) are merged into it by keeping their individual original shapes and proportions unchanged. To illustrate the shape learning, we only show the original and exaggerated contours here. Figure 3b shows the results of two exaggerated facial components, facial contour and nose shape. The exemplar pairs, which are used respectively for exaggerating the facial contour $\left(X_1^{(0)}, \hat{X}_1^{(0)}\right)$ and nose shape $\left(X_j^{(0)}, \hat{X}_j^{(0)}\right)$, are from two different caricature face pairs. The other non-exaggerated components keep their original shapes and proportions. It can be observed that apart from the facial contour, little shape change arises within the noses, i.e. the exaggerated nose does not obviously stand out in the caricature. Relationship exaggeration might help us to further highlight the noses.

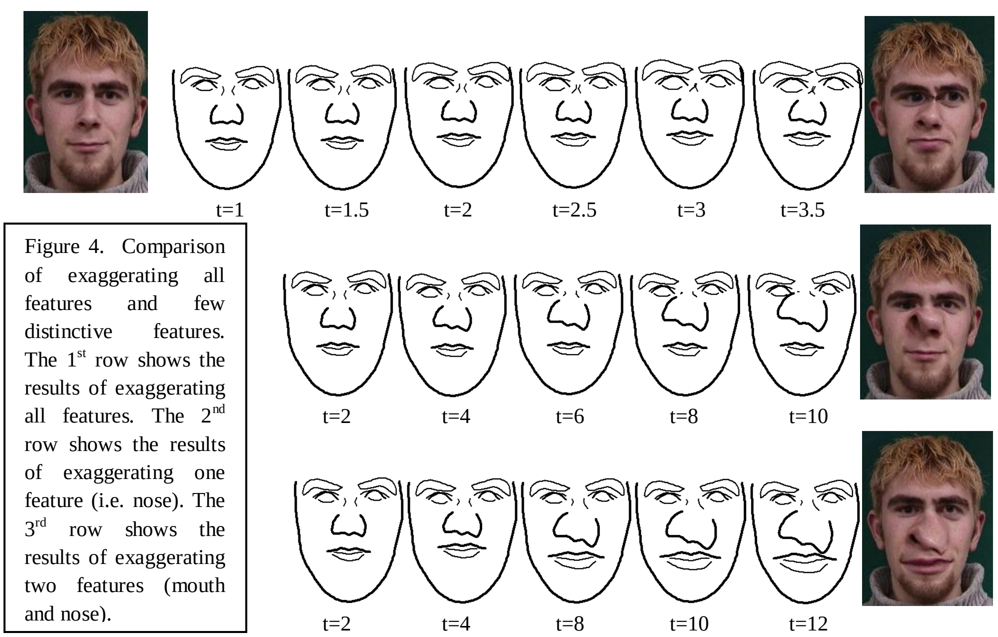



Figure 4. Comparison of exaggerating all features and few distinctive features. The 1st row shows the results of exaggerating all features. The 2nd row shows the results of exaggerating one feature (i.e. nose). The 3rd row shows the results of exaggerating two features (mouth and nose).

Then, we show that our method can not only perform relationship exaggeration in terms of the specified features, but also make the resulting caricature look like the original subject. Figure 4 shows the results of the relationship exaggeration with one or a few features emphasized and all features emphasized. We first fixed the facial contour and applied the original scheme of Eq.(6,7) to all other features (i.e. exaggerating all feature at different levels of exaggeration t). To illustrate relationship exaggeration, all shapes are unchanged here. When all features were emphasized (e.g. the eyes are enlarged, the nose is widened and tilted, and the distance between the eyes is shortened etc.) as shown in the first row of Fig.4, it is difficult to make the distinctive features stand out in a caricature. The second row of Fig.4 shows the results of only exaggerating the nose at different levels of exaggeration t. The third row of Fig.4 shows the results of exaggerating the mouth and nose at different levels $t$. One 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 row of Fig.4. For comparison, we also placed the textured caricatures of the extreme cases of exaggeration on the right side.

To illustrate the efficiency of the optimization scheme of Eq.(7) for likeness, we show the results of the relationship exaggeration with and without the re-optimization procedure of Eq.(7) in Fig.5. Only the size and position of nose are emphasized here. Obviously, re-optimization can change and further modify the relationship of facial components. Note that the affine transformations $G^{(j)}$ depend on facial components. For example, to Fig.5a, we expected to further emphasize the nose tilting. This could be fulfilled by simply adding rotation into $G^{(j)}$ of the nose as shown in Fig.5b.

Consequently, we also show the results of shape exaggeration plus relationship exaggeration in Fig.6. The shape exaggeration includes the exaggeration of the facial contour and nose shape by using Eq.(4,5). Then, these two facial features are emphasized in the relationship exaggeration by using Eq.(6), and the resulting caricatures are further re-configured in an optimal configuration by Eq.(7). For comparison, we gave out both the contours and the textured caricatures here. In the 2nd and 3rd columns of Fig.6, the facial contours are first exaggerated in the same style. And then, the nose and mouth are further emphasized in the relationship exaggeration. The 4th and 5th columns still include the exaggeration of the nose shape beside facial contours in the shape exaggeration. However, the columns from 2nd to 5th show some extremes of exaggeration, while the columns from 6th to 7th show the normal cases of exaggeration. Additionally, one can also note that in the bottom row of Fig.6, the mouth shapes of the caricatured sketches are almost unchanged while those of the textured caricatures appear to have bigger changes. This is because the caricatured sketches are fairly imperceptible compared to the textured caricatures in general. Painting texture may improve the visual impact.

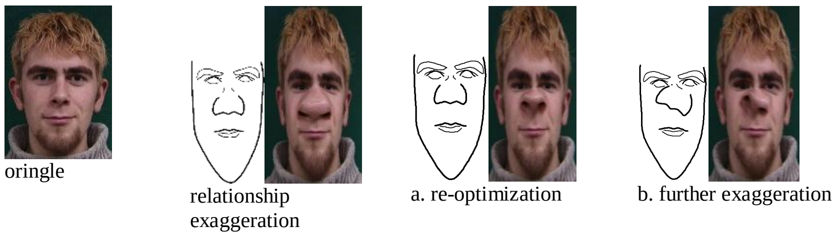



Figure 5. a) Comparison of relationship exaggeration and re-optimization; b) further rotating the nose by relationship exaggeration.

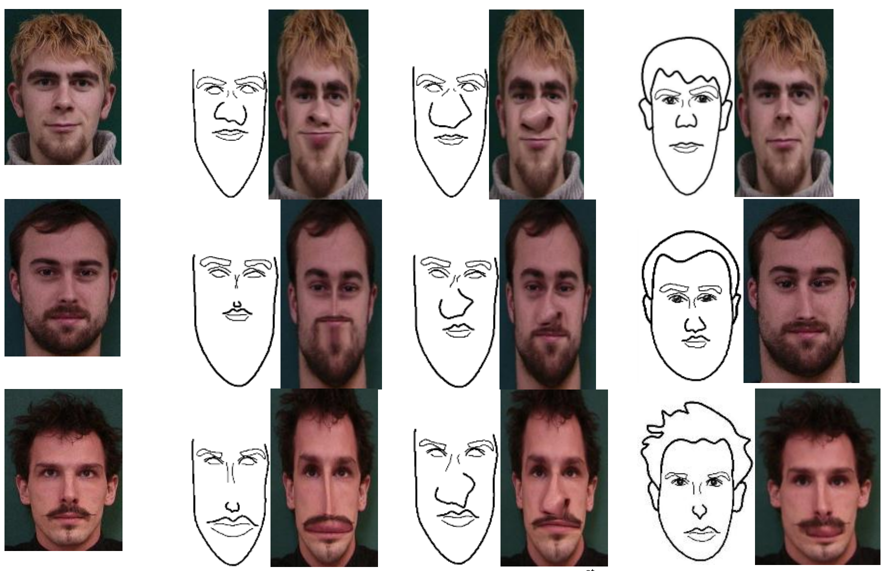
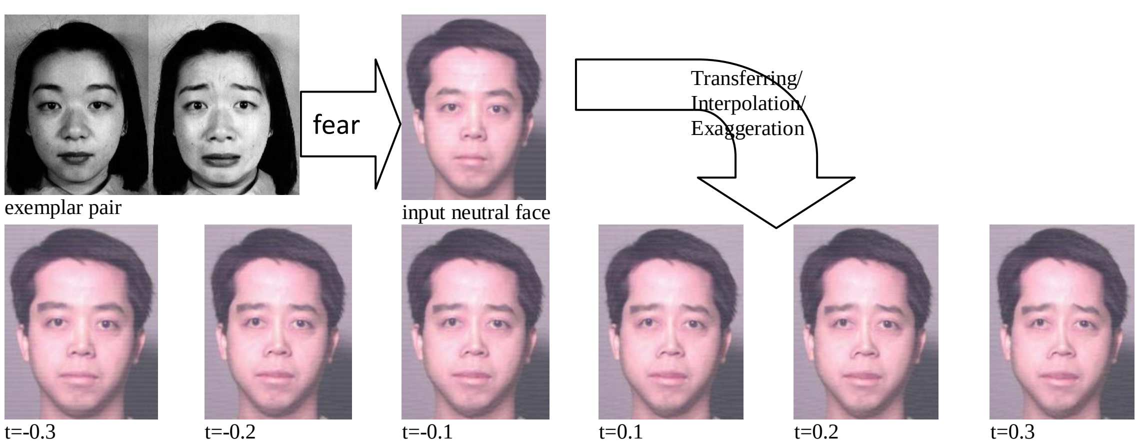


Figure 6. Frontal view caricatures by using our approach. The 1st column shows the original face images, while the other columns show the exaggerated results. The final column shows the texture transferring results. The learning samples for the shape exaggeration are the Example 1 in Fig.3.



t=0.4 t=0.5

Figure 7. The illustration of facial expression interpolation and exaggeration.

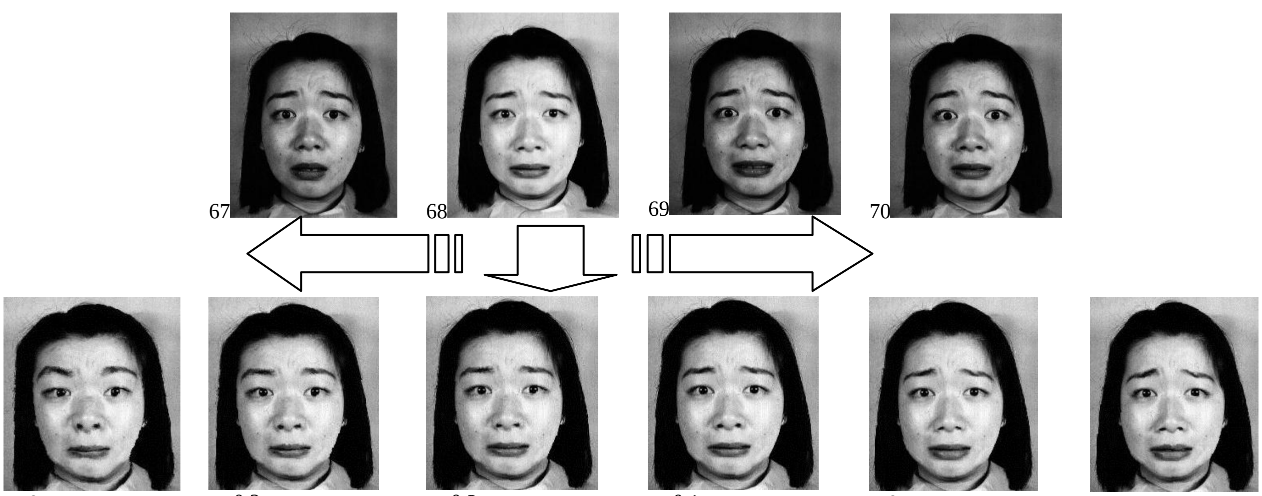
## Facial Expression Transfer

Unlike the preceding caricature synthesis, facial expression transfer usually needs to copy all the changes between a given exemplar pair to a new face. For a given facial expression pair $\left(X^{(0)}, \hat{X}^{(0)}\right)$, it is expected to transfer the expression of $\left(X^{(0)}, \hat{X}^{(0)}\right)$ to an input face X. Eq.(6,7) can be employed here. We summarise the procedure of facial expression computation as follows,

(1) Difference of expression: $\Delta\lambda = \lambda\left(\bar{X}, \hat{X}^{(0)}\right) - \lambda\left(\bar{X}, X^{(0)}\right)$;

(2) Transfer, interpolation and exaggeration: $\lambda(\bar{X}, \hat{X}, t) = \lambda(\bar{X}, X) + t\Delta\lambda$.

Herein, we utilize the mean $\bar{X}$ of the training set $\left\{X^{(i)}\right\}$ as a common control polygon, so that $X, \hat{X}, X^{(0)}, \hat{X}^{(0)}$ share a common base of mean value coordinates. This is in favor of the expression difference transfer. The new expression $\hat{X}$ of $X$ can be obtained by $\hat{X} = \lambda\left(\bar{X}, \hat{X}, t\right)\bar{X}$.




t=-0.4 t=-0.3 t=-0.2 t=-0.1 t=0.1 t=0.2


t=0.3


t=0.4


t=0.5

Figure 8. The comparison of the synthesized faces of expression exaggeration with the real facial expression images. The real face images are in the first row, while the synthesized faces in the next rows. The 2nd image (first row, no.68) is the input face.

Interestingly, the procedure of expression computation allows the amount of exaggeration t < 0. Usually, caricature synthesis requests t > 0 to avoid distortion. Herein, the expression exaggeration with t < 0 leads to expression interpolation. Figure 7 shows the facial expression interpolation and exaggeration by using the procedure of expression computation. The facial expression of the given exemplar pair is first transferred into the input neutral face, and then the expression interpolation and exaggeration are fulfilled by decreasing and increasing the exaggeration amount t. It can be noted that our method can generate a sequence of continuously various facial expressions. Our method can both transfer the facial expressions between the same gender and between those of the opposite sex.

Moreover, we have also compared the resulting interpolated and exaggerated faces based on the given exemplar pair with the real facial expression images of the same example. The JAFFE facial expression dataset [22] only provides 4 successive frames with the fear expression (No.67-70). We exaggerated from the No.68 face image here, and showed the results in Fig.8 for comparison. It can be observed that although the resulting faces do not match the real faces precisely, the facial expressions can be plausibly conveyed to the synthesized faces.

## Likeness Evaluation

Table 1. Likeness metrics to the input face in Fig.7. Images 1-8 correspond to the 8 images in the 2nd and 3rd rows of Fig.7.

| | Image1 | Image2 | Image3 | Image4 | Image5 | Image6 | Image7 | Image8 |
|---|---|---|---|---|---|---|---|---|
| MHD | 0.121 | 0.215 | 0.355 | 0.495 | 0.591 | 0.762 | 0.831 | 0.957 |

Table 2. The 1st row shows the likeness metrics to the face No.68 in the 1st row of Fig.8. The 2nd row shows the likeness metrics respectively to the 4 input faces in the 1st row of Fig.8. Images 1-9 correspond to the 9 images in the 2nd and 3rd rows of Fig.8.

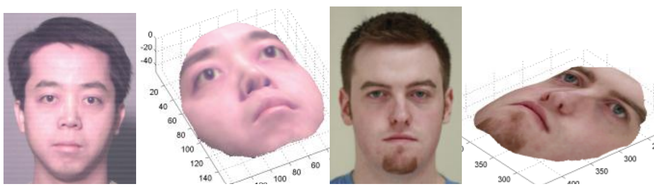
| | Image1 | Image2 | Image3 | Image4 | Image5 | Image6 | Image7 | Image8 | Image9 |
|---|---|---|---|---|---|---|---|---|---|
| MHD | 0.502 | 0.386 | 0.269 | 0.190 | 0.182 | 0.282 | 0.404 | 0.580 | 0.727 |
| MHD | | | 0.411 (No.67) | | | 0.378 (No.69) | 0.493 (No.70) | | |

To test the effectiveness of our approach, we apply the MHD to measure the likeness between the exaggerated facial expressions and their real facial images in Fig.7 and Fig.8. Table 1 shows the MHD values of the likeness between the exaggerated faces and the original input image in Fig.7. Table 2 shows the MHD values of the likeness between the exaggerated faces and the real face images with expression in Fig.8. Moreover, the exaggerated faces in Fig.8 are similar to the successive face images (No.67, 69 and 70). The second row of Table 2 shows the MHD values of the 3 exaggerated faces to the 3 successive face images. The likeness values of the exaggerated faces to the real face images with expression fall in an acceptable range (i.e. MHDs are less than 0.5). This means our method can transfer a specified facial expression to other faces effectively, and further conveys the style of a given caricature face pair to others as well.

**3D Caricature Synthesis**

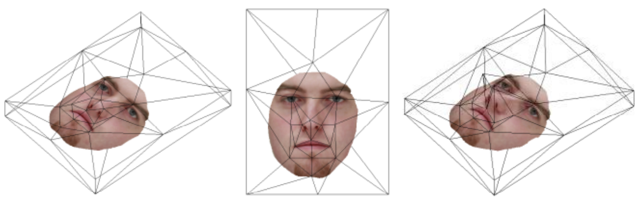
Moving on from the 2D cases, 3D caricature has received increasing attention recently. In this application, we describe an interactive method to generate 3D caricatures based on the 3D version of Mean Value Coordinates [21]. Our work aims at interactive 3D caricature generation.

Our method consists of two steps, 3D face modeling and interactive 3D face exaggeration. Usually, modeling a face from 2D photographs requires multiple 2D images, including frontal and side view ones, such as FaceGen. But there is only one frontal face image available in our application. For the purpose of 3D face modeling, our basic idea is to deform a reference 3D face model based on the morphed feature points on 2D face image. In our experiment, the reference 3D textured face model is generated through a 3D frontal face range data associated with the texture image as shown in Fig.(8a,8b). Note that the reference texture image has been registered with the reference 3D face surface. For a given target face image, we firstly apply the MVC deformation formula of Eq.(1) to morphing the feature points of the reference 2D image to the target image on the 2D image plane. Then, applying harmonic map [26] to the reference 3D surface with the morphed feature points as the fixed boundary, one can get the deformed 3D face model corresponding to the given target face image. Note that feature point morphing is carried on the 2D image plane. To utilize the harmonic map to 3D surface deformation, we take the depth information (i.e. the z-coordinates) of the feature points on the reference 3D face surface as that of the morphed feature points on the target image plane. Figure 9d shows the result of 3D face modeling based on the given target face image shown in Fig.9c. It is straightforward to get a 3D caricatured face model based on the feature point set of a given 2D caricature in the same manner.



a b c d

Figure 9. Illustration of 3D face modelling



a b c

Fig.10. Original 3D face model with the control mesh in (a). The top view in (b). The exaggerated 3D face model with the deformed control mesh in (c).

In order to exaggerate the resulting 3D face, we introduce the 3D version of MVC here. Similar to the 2D version of MVC, the 3D version also needs a control mesh, which is called a "cage" in [21]. The deformation formula of 3D MVC is expressed as, $\hat{x} = \frac{\sum_i \lambda_i(x)\hat{v}_i}{\sum_i \lambda_i}$, where, $\lambda_i$ are the 3D MVC about the vertex x in the surface with respect to each vertex $v_i$ in the unchanged control mesh, and $\hat{v}_i$ are the vertices in the deformed control mesh (refer to [21] for details). One can specify an arbitrary closed triangle mesh as the control mesh. This allows the artist to deform the control mesh by specifying the positions of the vertices in the control mesh for generating a desired 3D caricatured face. In our experiments, we utilize a cubic box as the control mesh and select some feature points as the vertices in the control mesh, such as eye balls, nose tip and mouth corners as shown in Fig.10b. Figure 10a depicts a 3D face before exaggeration and the surrounding control mesh with black line. Changing the positions of the vertices in the control mesh yields the exaggerated 3D face shown in Fig. 9c. Moreover, the above procedure of expression computation can also be applied to a 3D face model as follows,

$$\begin{cases} x(t) = \dfrac{\sum_i \left(\lambda_i(x) + t\Delta\lambda_i\right) v_i}{\sum_i \lambda_i(x)} \\ \Delta\lambda_i = \lambda_i(\hat{x}) - \lambda_i(x) \end{cases},$$

where, $x, \hat{x}$ denote the feature points respectively in the neutral and exaggerated facial surfaces, $v_i$ denotes the vertex of the control mesh, $\lambda_i$ are the 3D MVC of the vertices $x, \hat{x}$ respectively with respect to the same control mesh. It can be noted that only the feature points are further exaggerated here. Keeping the updated feature points as the fixed boundary, one can utilize the harmonic map to update the 3D face model with a new exaggerated expression.

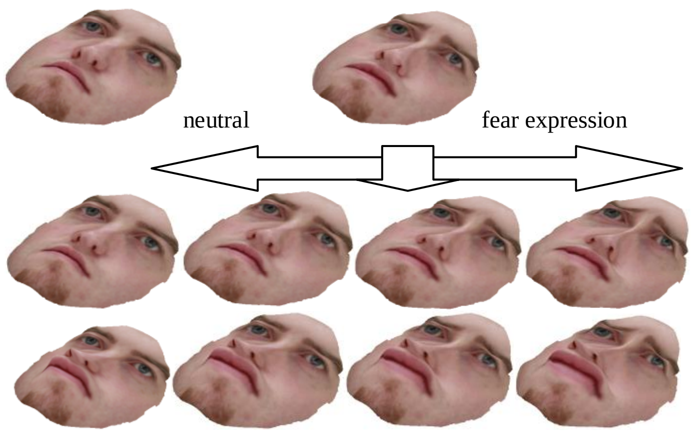



Figure 11. Illustration of 3D caricatures.

Figure 11 shows some 3D facial caricatures. The face models with neutral and fear expression shown in the 1st row of Fig.11 are based on the frontal face images. Then, the interpolation and exaggeration of facial expression were carried out on the resulting 3D face models as shown in the 2nd row of Fig.11. For comparison, we further exaggerated the mouth and nose of 3D models by manually adjusting the control mesh as shown in the 3rd row of Fig.11.

**Remark**

The distinctive difference of 2D and 3D versions of MVC is that the control polygons of 2D MVC might be a set of nested simple polygons, while the control mesh of 3D MVC must be a closed mesh. Thus, the control mesh is usually defined as a closed cage as shown in Fig.10. This is similar to the previous freeform deformation (FFD) techniques [28]. FFD usually requires specifying volumetric cells on the interior of the control mesh. The resulting deformations depend on how to decompose the control mesh into volumetric cells. In contrast, 3D MVC accepts an arbitrary closed triangular surface as the control mesh and does not require volumetric cells to span the interior. Moreover, 3D MVC can generate smooth, realistic looking deformations even with a small number of control vertices and is quite fast. (For more comparison, please refer to [21]).

## 5. Conclusion and Future Works

In this paper we have presented a caricature synthesis technique based on mean value coordinates. It consists of three steps: shape exaggeration, relationship exaggeration and optimization for likeness. Unlike other methods, our shape exaggeration is implemented by learning some specific facial components from one or a small number of exemplar pairs. Using MVC the relationship exaggeration can be conveniently implemented to maximize likeness. Moreover, our approach can be applied to facial expression interpolation and exaggeration. We further extend our approach to 3D caricature synthesis whereby 3D caricature synthesis is produced based on a single frontal face image. One novelty of our approach is to transfer facial expressions across individuals based on one or a small number of exemplar pairs and to interpolate expressions by a controllable factor. For the 3D case, we have presented an interactive 3D editing tool for 3D caricature synthesis. The experimental results demonstrate the effectiveness of our technique.

There remain a number of issues requiring further investigation. The main challenges include texture style transfer and 3D manipulation. Each artist has his or her own painting style, which could be learned from the examples in a similar manner to our shape learning. Although we have used existing techniques to transfer styles such as image analogies and image quilting, the effect on images is unimpressive. Texture style transfer needs to be developed in future. Our method of 3D caricature synthesis focuses on the face. A natural extension of our technique is to define a head model, allowing us to generate a caricature for the whole head. Due to hair occlusion and styles, it raises the difficult question of how to model hair. This challenging problem is an area of our future investigation.

Regarding the likeness metrics, there was little works done in the areas of caricature or facial expression synthesis. We attempted to use the MHD to tackle this issue in this paper. However, because likeness is both an objective measure and also up to subjective interpretations, it is difficult to argue which approach produces the best result. Our contribution here is that we have given out a quantitative comparison of the synthesized face images with the real face image with expression for evaluating our approach. We believe there is more work needed in the future to ascertain a more effective measurement of this property.

**References**

[1] AKLEMAN, E.: Making Caricature with Morphing. In Proc. of ACM SIGGRPH 1997, Los Angeles, California, pp.145-145.

[2] AKLEMAN, E., PALMER, J. and LOGAN, R.: Making Extreme Caricatures with a New Interactive 2D Deformation Technique with Simplicial Complexes. In Proc. of the Third International Conference on Visual Computing (2000), Mexico, pp.165-170.

[3] DAVIES, R. H., TWINING, C. J. and COOTES, T. F. et al.: 3D Statistical Shape Models Using Direct Optimisation of Description Length. In Proc. of European Conference on Computer Vision (2002), Vol.3, pp.3-20.

[4] BRENNAN, S. E.: Caricature Generator: The Dynamic Exaggeration of Faces by Computer. Leonardo (1985), Vol.18, No.3, pp.170-178.

[5] CHIANG, P. Y., LIAO, W. H. and Li, T. Y.: Automatic Caricature Generation by Analyzing Facial Features. In Proc. of Asia Conf. on Computer Vision (2004), Korea.

[6] CHEN, H., LIU, Z. and ROSE, C. et al.: Example-based composite sketching of human portraits. In Proc. of International Symposium on Non-photorealistic Animation and Rendering (2004), New York, pp.95-102.

[7] CHEN, H., XU, Y. Q. and SHUM, H. et al.: Example-based facial sketch generation with non-parametric sampling. In Proc. of the 8$^{th}$ International Conference on Computer Vision (2001), Vancouver, Canada, pp.433-438.

[8] CHEN, H., ZHENG, N. N. and LIANG, L. et al.: PicToon: A Personalized Image-based Cartoon System. In Proc. of ACM Int'l Conf. on Multimedia (2002), Juan-les-Pins, France, pp.171-178.

[9] ZHANG, Q., LIU, Z. and GUO, B. et al.: Geometry-driven photorealistic facial expression synthesis. IEEE Trans. on Visualization and Computer Graphics (2006), Vol.12, No.1, pp.48-60.

[10] BLANZ, V. and BASSO, C. et al.: Reanimating Faces in Images and Video. in Proc. of Eurographics 2003, (Eds. P. Brunet and D. Fellner), Granada, Spain, Vol.22, No.3, pp.641-650.

[11] GOOCH, B., REINHARD, E. and GOOCH, A.: Human Facial Illustrations: Creation and Psychophysical Evaluation. ACM Transactions on Graphics (2004), Vol.23, No.1, pp.27-44.

[12] KOSHIMIZU, H., TOMINAGA, M., FUJIWARA, T. and MURAKAMI, K.: On KANSEI Facial Processing for Computerized Facial Caricaturing System Picasso. In Proc. of IEEE Int'l Conf. on Systems, Man and Cybernetics (1999), pp.294–299.

[13] LIANG, L., CHEN, H., XU, Y.Q. and SHUM, H.Y.: Example-based Caricature Generation with Exaggeration. In Proc. of 10th Pacific Conf. on Computer Graphics and Applications (2002), China, pp. 386–393.

[14] LENN, R.: How to draw caricature. Contemporary Books (1984), pp.61-90.

[15] MO, Z. Y., LEWIS, J. P. and NEUMANN, U.: Improved Automatic Caricature by Feature Normalization and Exaggeration. In Proc. of ACM SIGGRAPH 2004 Sketches (2004), Los Angeles, California, pp.57-57.

[16] HORMANN, K. and FLOATER, M. S.: Mean value coordinates for arbitrary planar polygons. ACM Trans. on Graphics (2006), Vol.25, No.4, pp.1424-1441.

[17] ARAD, N. and DYN, N. et al.: Image warping by radial basis functions: applications to facial expressions. CVGIP: Graphical Models and Image Processing (1994), Vol.56, No.2, pp.161-172.

[18] SHET, R.N. and LAI, K.H. et al.: Use of Neural Networks in Automatic Caricature Generation: An Approach Based on Drawing Style Capture. In Proc. of The second edition of the Iberian Conference on Pattern Recognition and Image Analysis (2005), Portugal, pp.343-351.

[19] TSENG, C.C. and LIEN, J.J.: Synthesis of Exaggerative Caricature with Inter and Intra Correlations, In Proc. of Asian Conference on Computer Vision 2007, Tokyo, Japan, pp.314-323.

[20] XU, G.Z., KANEKO, M. and KUREMATSU, A.: Synthesis of Facial Caricature Using Eigenspaces, IEICE Trans on Information and Systems (Japanese Edition) (2001), Vol.J84-D-II, No.7, pp.1279-1288.

[21] JU, T., SCHAEFER, S. and WARREN, J.: Mean value coordinates for closed triangular meshes, in Proc. of ACM SIGGRAPH 2005, pp.561-566.

[22] LYONS, M., BUDYNEK, J. and AKAMATSU, S.: Automatic Classification of Single Facial Images, IEEE Trans. on Pattern Analysis and Machine Intelligence (1999), Vol.21, No.12, pp.1357-1362.

[23] FUJIWARA, T., KOSHIMIZU, H. and FUJIMURA, K. et al.: 3D modeling system of human face and full 3D facial caricaturing, in Proc. of IEEE 7th Intl. Conf. on Virtual Systems and Multimedia, Oct. 25-27, 2001.

[24] XINGYU, G., YIQIANG, C. and JUNFA, L. et al.: Personalized 3D Caricature Generation Based on PCA Subspace, Proc. of 9th Pacific Rim Conference on Multimedia, Taiwan, Dec. 2008.

[25] AKLEMAN, E.: Modeling Expressive 3D Faces, in Proceedings of ACM SIGGRAPH'2004 (Siggraph Sketch), Los Angeles, California, July 2004, pp.61.

[26] FLOATER, M. S. and HORMANN, K.: Surface Parameterization: a Tutorial and Survey, in Advances in Multiresolution for Geometric Modelling, N. A. Dodgson, M. S. Floater, and M. A. Sabin (eds.), Springer-Verlag, Heidelberg, 2005, pp.157-186.

[27] DUBUISSON, M. P. and JAIN, A. K.: A modified Hausdorff Distance for Object Matching. In Proc. of Int'l Conf. on Pattern Recognition (1994), Jerusalem, Israel, pp.566-568.

[28] MACCRACKEN, R. and JOY, K. I.: Free-form deformations with lattices of arbitrary topology. In Proc. of ACM SIGGRAPH 1996, pp.181-188.