Caricature Synthesis Based on Mean Value Coordinates

Hongchuan YU and Jian J. Zhang
NCCA, Bournemouth University, UK
{hyu,jzhang}@bournemouth.ac.uk

Abstract

In this paper, a novel method for caricature synthesis is developed based on mean value coordinates (MVC). Our method can be applied to any single frontal face image to learn a specified caricature face exemplar pair for frontal and side view caricature synthesis. The technique only requires one or a small number of caricature face pairs and a natural frontal face 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 extended to facial expression transfer, interpolation and exaggeration, which are applications of expression editing. Moreover, the deformation equation of MVC is modified to handle the case of polygon intersections and applied to lateral view caricature synthesis from a single frontal view image. Using experiments we demonstrate that the transferred expressions are credible and the resulting caricatures can be characterized and recognized.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: animation

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 allowing 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.

Our proposed algorithm is based on the deformation property of Mean Value Coordinates (MVC) [4]. Our contributions can be summarized as follows:

- *Shape and relationship exaggeration.* We divide exaggeration into these two stages. The shape exaggeration of individual face components is computed by learning from one or a small number of examples' shapes, 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 maintains 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.* We further extend the proposed caricature exaggeration approach to the facial expression interpolation and exaggeration. 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" was 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;
- *Lateral View Caricature.* The MVC-based deformation is extended to lateral view caricature generation. We will modify the MVC deformation to deal with the cases of polygon intersection and show how to produce the lateral view caricatures based on a single frontal view image.

2. Exaggeration

Our approach consists of three parts, shape exaggeration, relationship exaggeration and optimization.

2.1. Mean Value Coordinates

For interpolation, if a value $F_i \in R^d$ is specified at each vertex $v_i \in R^2$, i.e. $F_i = F(v_i)$, then the mapping $F: R^2 \to R^d$ is defined by,

$$F(x) = \sum_i \lambda_i(\psi, x) F(v_i), \quad x \in R^2, \qquad (1)$$

in terms of the given polygon ψ and $v_i \in \psi$.

For deformation, consider two polygons $\psi, \hat{\psi}$ defined respectively on two planar domains $\Omega, \hat{\Omega}$. The bijection between ψ and $\hat{\psi}$ can be simply specified by $\begin{cases} f(v_i) = \hat{v}_i \\ f^{-1}(\hat{v}_i) = v_i \end{cases}$, where $v_i \in \psi$ and $\hat{v}_i \in \hat{\psi}$. Since $f(v_i) = \sum_i \lambda_i(\psi, v_i)\hat{v}_i = \hat{v}_i$, the deformation is therefore formulated as follows,

$$\begin{cases} f^{-1}(\hat{x}) = \sum_i \hat{\lambda}_i(\hat{\psi}, \hat{x}) v_i = x \\ f(x) = \sum_i \lambda_i(\psi, x)\hat{v}_i = \hat{x} \end{cases}, \qquad (2)$$

where $x \in \Omega, \hat{x} \in \hat{\Omega}$. 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 λ and $\hat{\lambda}$ are row vectors.

Another property of MVC is the Positivity, that is, sum of MVC is positive inside the kernel of polygons. We will utilize it to modify the scheme of Eq.(2) for handling the case of polygon intersection. (For more details, please refer to [4]).

2.2. Shape Exaggeration

Consider a training set of the neutral faces $\{X^{(i)}, i=1,...,n\}$. Each $X^{(i)}$ 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. Let $(X^{(0)}, \hat{X}^{(0)})$ be a given caricature face pair for learning. The basic idea is to first employ the training set $\{X^{(i)}\}$ to learning the shape style of the given neutral face $X^{(0)}$, and then apply them to learning the shape style of the given caricature $\hat{X}^{(0)}$.

To this end, we apply the scheme of Eq.(2) to the training set for learning the example $X_j^{(0)}$. This can be achieved by minimizing,

$$\min_T \sum_i^n \left\| X_j^{(0)} - \lambda\left(\psi^{(i)}, X_j^{(i)}\right) T\psi^{(0)} \right\|^2, \qquad (3)$$

where, control polygons $\psi^{(0)}, \psi^{(i)}$ might be the complement sets of $X_j^{(0)}, X_j^{(i)}$ respectively. The resulting T is a linear operator, which is in a matrix form of size NumControlPoint×NumControlPoint. Updating $X_j^{(i)}$ by using T, i.e. $X_j^{(i)\prime} = \lambda\left(\psi^{(i)}, X_j^{(i)}\right) T\psi^{(0)}$, we then apply the scheme of Eq.(1) to the updated $X_j^{(i)\prime}$ for learning the example $\hat{X}_j^{(0)}$, i.e.

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

Similarly, the solution $\hat{T}$ is also a linear operator. The operators $(T, \hat{T})$ indeed convey the caricature style of $(X^{(0)}, \hat{X}^{(0)})$ to others $\{X^{(i)}\}$. For the input X_j, one can firstly apply the scheme of Eq.(2) to it with the fixed control polygon $T\psi^{(0)}$, and then apply the scheme of Eq.(2) to the resulting X_j' with the fixed control polygon $\hat{T}\hat{\psi}^{(0)}$ for the caricature $\hat{X}_j$.

Moreover, the scheme of Eq.(3,4) can also be applied to the facial expression transfer by replacing the caricature examples with the facial expression ones. Figure 2 illustrates the procedure of transferring the facial expression from a given example pair to others. Our method can both transfer the facial expression between the same gender and between those of the opposite sex.

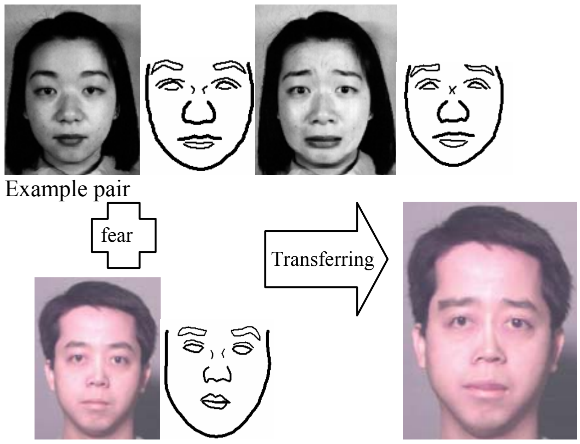

Fig.1. The illustration of facial expression transfer. The left column is the given facial expression pair, while the bottom right is the resulting face with the expression transferred from the left side. (The example pair is from the JAFFE facial expression dataset [5]).

2.3. Relationship Exaggeration

For a caricature, the relationship of facial components plays a dominant role. 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) [2]. The facial features are used to be described as a set of proportions respectively along the horizontal and vertical lines as shown Fig.2.

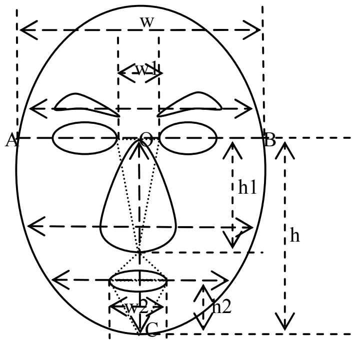

Fig.2. Illustration of facial features' contours.

In our algorithm, the Exaggeration Difference From Mean (EDFM) [3] is employed to the relationship exaggeration. For a given amount of exaggeration t and the specified gth feature, one can update the MVC of a face X by,

$$\begin{cases} \lambda\left(\psi, x^{(g)}, t\right) = \bar{\lambda}^{(g)} + t\Delta\lambda^{(g)}, \\ \Delta\lambda^{(g)} = \lambda\left(\psi, x^{(g)}\right) - \bar{\lambda}^{(g)} \end{cases}, \qquad (5)$$

where, $x \in X$, $\bar{\lambda}$ denotes MVC mean vector of training set and $t > 0$. Then, the exaggerated gth feature is updated by $\hat{x}^{(g)} \leftarrow \lambda\left(\psi, x^{(g)}, t\right)\hat{\psi}$.

For the purpose of likeness, MVC of subjects should remain unchanged (or little change) during the exaggeration. Thus, rigid transformation is applied individually to each facial component, e.g. the jth feature is transformed by $\hat{x}^{(j)} = G^{(j)}\left(x^{(j)}\right)$. The rigid transforms $G^{(j)}$ can be yielded by solving the following linear system,

$$G^{(j)}\left(x^{(j)}\right) = \lambda\left(\psi, x^{(j)}, t\right)\hat{\psi} . \qquad (6)$$

In terms of the resulting $G^{(j)}$, one can yield the exaggerated features and synthesize them into the whole caricatured $\hat{X}$ of X. Furthermore, to make the resulting $\hat{X}$ look like the original subject X, one can re-adjust the non-exaggerated features. MVC are re-computed based on the nested control polygons, including the facial contour and exaggerated feature contours, and then the procedure of Eq.(6) is employed to all non-exaggerated features as follows,

$$G^{(k)}\left(x^{(k)}\right) = \lambda\left(\psi, x^{(k)}\right)\hat{\psi}, k \neq j ,$$

where, the control polygons $\psi, \hat{\psi}$ contains both the facial contour and the jth feature contour. This allows the facial contour and exaggerated features fixed while adjusting the other non-exaggerated features in an optimal configuration.

Moreover, we address how to exaggerate the facial expression. In section 2.2, we can transfer the expression from the facial expression example pair $\left(X^{(0)}, \hat{X}^{(0)}\right)$ to some neutral face X for its $\hat{X}$ with expression. For further exaggeration, the scheme of Eq.(5) can be employed here. Note that the mean of MVC is replaced with MVC of X, i.e. $\lambda(\hat{X}, X)$ here. The amount of the expression exaggeration is first added into MVC as follows,

$$\begin{cases} \lambda(\hat{X}, X, t) = \lambda(\hat{X}, X) + t\Delta\lambda \\ \Delta\lambda = \lambda(\hat{X}, \hat{X}^{(0)}) - \lambda(\hat{X}, X^{(0)}) \end{cases} . \qquad (7)$$

Then, the exaggerated $\hat{X}$ is updated by $\hat{X} \leftarrow \lambda(\hat{X}, X, t)\hat{X}$. Note that the control polygons are the same here. This is because MVC of X is computed in the coordinate base $\hat{X}$, while the updating $\hat{X}(t)$ is indeed interpolated in the same coordinate basis $\hat{X}$.

2.4. Lateral View Caricature

In this paper, the given faces have been restricted to the frontal view. For the lateral view purpose, we prefer the estimation of the face horizontal pose to that of the vertical pose. Thus, in terms of facial symmetry, we can utilize a cylinder system to approximate a face, wherein the straight line traversing both eyes (i.e. $\overline{A_2B_2}$ in Fig.3) denotes the diameter w. The face horizontal pose can be estimated by the trigonometric computation as shown in Fig.3.

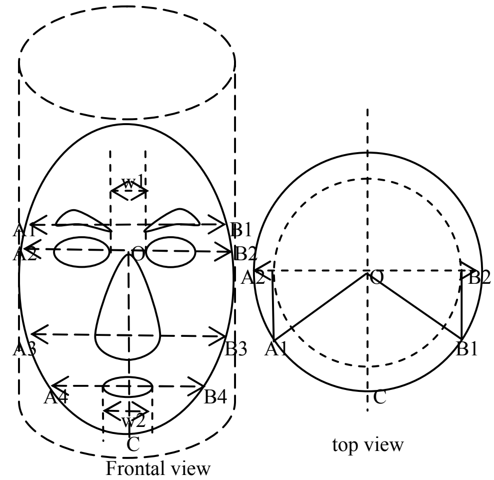

Fig.3. Illustration of horizontal pose computation

Usually, generating a lateral view will result in occlusion. For MVC computation, we need to deal with intersection or self-intersection of control polygons. This can be fulfilled by modifying the scheme of Eq.(1,2). Consider a set of simple polygons as shown in Fig.4. We first define the distance from x to the polygon ψ_1 (or ψ_2) as follows,

$$d(x, \psi_1) = \left\| \lambda\left(\{\psi_1, \psi_2\}, x\right) - \lambda\left(\{\psi_1, \psi_2\}, x, \{-\psi_2\}\right) \right\|_2 ,$$

and associated normalized weight,

$$w(x, \psi_i) = \frac{1}{d(x, \psi_i)} \Big/ \sum_j \frac{1}{d(x, \psi_j)} .$$

Then, the deformation formula of Eq.(2) is modified as,

$$\hat{x} = \begin{cases} \lambda\left(\psi_1, x\right)\hat{\psi}_1, & x \text{ lies inside } \psi_1 \\ \sum_i w(x, \psi_i)\lambda\left(\psi_i, x\right)\hat{\psi}_i, & \text{otherwise} \end{cases} . \quad (8a)$$

Note that $\lambda\left(\{\psi_1, \psi_2\}, x, \{-\psi_2\}\right)$ only retains MVC related to the polygon ψ_1 while setting the others ZERO. Then, we can approximate the inverse transform of Eq.(8a) based on the same idea of weighted sum as in Eq.(8a) as follows,

$$x = \begin{cases} \lambda\left(\hat{\psi}_1, \hat{x}\right)\psi_1, & \hat{x} \text{ lies inside } \hat{\psi}_1 \\ \sum_i w(\hat{x}, \hat{\psi}_i)\lambda\left(\hat{\psi}_i, \hat{x}\right)\psi_i, & \text{otherwise} \end{cases} . \quad (8b)$$

Applying the scheme of Eq.(8) to lateral view caricatures, one has to specify the source and target polygons manually. A single frontal face image cannot provide us 3D cues enough. But this also leaves users rooms free for creating art. Figure 5 shows some results of exaggerating on the lateral view caricatures, in which the rotation angle is $\pi/3$ and the noses are further exaggerated beyond the facial contour.

3. Experiments and Analysis

We only show two experiments in Fig.6 and Fig.7. due to space. The contours of the original facial features are extracted by using the AAM method [1].

4. Conclusions and Future Works

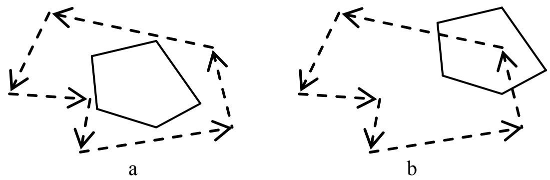

Fig.4. Illustration of polygon intersection. The arrows denote the positive orientation of polygon. a) before deformation; b) after deformation.

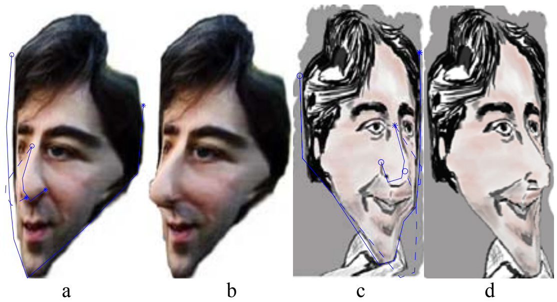

Fig.5. Illustration of lateral view caricature exaggeration. a) The original polygons (solid lines) and one exaggerated polygon (dashed nose contour); b) lateral view caricature with the exaggerated nose; c) original polygons (solid lines) and two exaggerated polygons (dashed facial contour and nose contour); d) lateral view caricature with the exaggerated nose and enlarged chin.

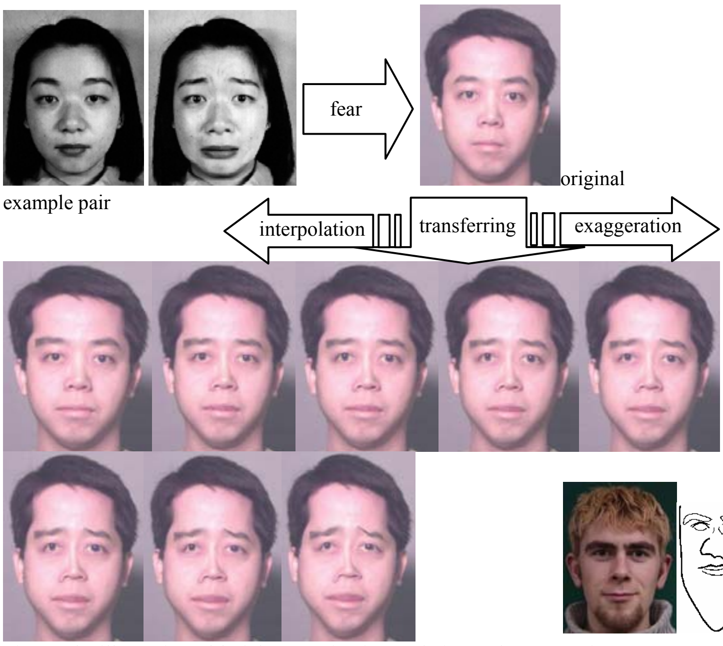

Fig.7. The illustration of facial expression interpolation and exaggeration.

Our future works will focus on the facial expression extracting and transferring on video under the MVC framework.

References

[1] 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.

[2] LENN R.: How to draw caricature. Contemporary Books, 1984.

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

[4] 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.

[5] 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.

Fig.6. Frontal view caricatures by using our approach. The 1st column shows the original face images. In the 2nd and 3rd columns, 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. Additionally, the columns from 6th to 7th show the normal cases of exaggeration.