AUTOMATIC CARICATURE GENERATION BY ANALYZING FACIAL FEATURES

Pei-Ying Chiang   Wen-Hung Liao   Tsai-Yen Li

National Chengchi University, Taipei, Taiwan

{g9129, whliao, li }@cs.nccu.edu.tw

ABSTRACT

We developed a system for automatic generation of caricatures by analyzing unique facial features of the subject. The proposed face model contains 119 parameters derived from the MPEG-4 standard. These nodes are categorized into 8 groups with a pre-defined hierarchy to guide the placement of face components after shape exaggeration. Given an input image, the system determines which and how the face component should be altered. Using an artist's finished work as the source image, the system is capable of producing caricatures of a similar style effectively and efficiently.

1. INTRODUCTION

It is a common scene nowadays to see amateur painters at popular tourist spots drawing pictures for sightseers. Some of these pictures attempt to achieve high fidelity, while others are exaggerated, cartoon-like portraits that try to capture the essence of the subject with a bit flavor of humor or sarcasm. Experienced artists can often go beyond likeness, revealing the unique combination of character and appearance of the model through very simple, yet powerful sketches (e.g., refer to Fig. 1).

Figure 1. A caricaturist's rendering (left) of a model (right). From Hughes[1].

Cartoonists such as Hughes and Redman [2] published books on teaching the techniques of drawing caricatures. General guidelines do exist for beginners. However, caricature drawing is still considered a work of art. Existing image transformation algorithms [3] that convert a photo into oil painting or pencil touch can only change the overall drawing style, lacking the ability to selectively exaggerate facial features to generate interesting portrait caricatures.

A first attempt to develop computer-assisted caricature generation system is made by Brennan. In [4], an interactive system for producing sketches with exaggeration is reported. Murakami et al. [5] proposed a template-based approach to create facial caricatures with adjustable exaggeration rates. The sketches generated by PICASSO are line-based. The lines that comprise the sketch become fragmented when the deformation is too large. Yang *et al.* [6] took a somewhat different approach. A collection of face parts such as the eyes, the nose, and the mouth are first archived and stored in a database. The photo of the subject is analyzed and classified. A portrait caricature is then synthesized using corresponding parts from the database. Lin *et al.* [7] followed an example-based approach to learn the individual style of an artist. Satisfactory results have been obtained only with large amount of training data for each artist.

Experienced caricaturists are good at extracting and exaggerating facial features. However, with previous efforts as summarized above, the automatic process of locating unique facial features and determining a proper exaggeration rate remains a difficult problem. In this paper, we attempt to address some of the most important issues in automatic facial caricature generation from two respects. First, we define a face model based on MPEG-4 parameters. A feature extraction process is developed accordingly. Using a reference caricature as the basis, caricature generation is formulated as a metamorphosis process. Experimental results show that the proposed approach is both efficient and effective

The rest of this paper is organized as follows. In section 2, the proposed face model is defined. The computation and representation of average face data is also covered. Section 3 explains the automatic feature extraction process. Section 4 illustrates the key steps in determining the exaggeration rate and the placement of face components. Experimental results are depicted in Section 5. Section 6 provides a brief summary and comments on future developments.

2. FACE MODEL DEFINITION

This section defines the face mesh employed in our caricature generation system. The characteristics of the different types of constituent nodes are elucidated. We then proceed to compute the parameters of the average face and discuss the method to apply exaggeration to the face components.

2.1. Facial Feature Nodes

To control the shape and appearance of each facial feature, we define a set of 119 nodes based on the MPEG-4 face definition parameters (FDP) and face animation parameters (FAP). (Refer to Fig. 2a for the locations of the nodes.) These nodes are categorized into 8 groups, namely, face contour, left and right eyebrow, left and right eye, nose, upper and lower lip, as depicted in Fig. 2. The groups are related to each other by a pre-defined hierarchy shown in Table 1.

Table 1. Structure and hierarchy of the face components.

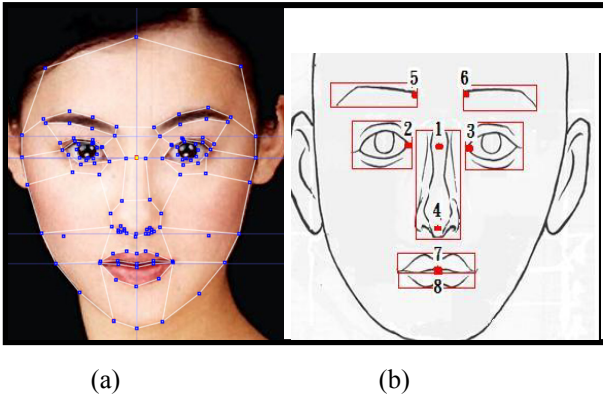| Group ID	Region	Rank	#of Nodes
G1	Face Contour	1	19
G2	Left eye	2	22
G3	Right eye	2	22
G4	Nose	2	22
G5	Left eyebrow	3	8
G6	Right eyebrow	3	8
G7	Upper lip	3	10
G8	Lower lip	3	8

| (a) | (b) |

Figure 2. (a) The proposed face mesh. (b) The 8 groups and the corresponding master nodes.

Each group contains three different types of nodes: master node, calibration node, and slave node, respectively. The master node is closely associated with the group. Fig. 2b depicts the 8 master nodes in our face model. Its main function is to define the position of the specific group. All other nodes in the same group move

accordingly with the master node. (Fig. 3a) The calibration node serves as the reference for shape measurement, normalization and deformation. (Fig. 3b) A group that contains complex shape can have more than one calibration node. For example, the nodes in family2 (left eye) can be further decomposed to four sub-groups: eye-lid, iris, eye contour and eye bag. These four sub-groups have their own calibration nodes, but share a common master node. All nodes that are neither master nodes nor calibration nodes are slave nodes.
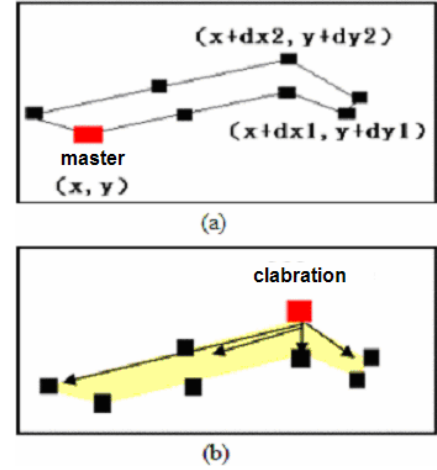

Figure 3. (a) The master node is used to position the group. (b) The calibration node serves as the reference for shape measurement, normalization and distortion.
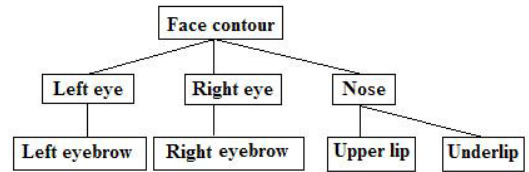

Figure 4. Spatial relationship among the different groups.

According to Table 1, the nodes that comprise the face shape possess the highest rank. The tree structure shown in Fig. 4 further defines the spatial relationship among the different groups of nodes. In placing the constituent parts of the face, the ranks of the groups must be taken into account. For example, G5 (left eyebrow) needs to be positioned accordingly with its father group G2 (left eye) to maintain a minimum distance to avoid collision/overlap when the nodes in the father group shift significantly.

2.2. Average Face Computation

It is necessary to perform quantitative analysis to obtain important statistics of the face components. At present, we have collected 100 photos of Asian female for the measurement. To ensure high accuracy, we manually

mark the nodes in each group and record their coordinates. These coordinates are then used to compute the size/ shape of each face component. To compensate for the possible difference in imaging distance, we set the unit length to be the distance between the two eyes. We also set the origin to be the midpoint between the two eyes. Aside from the normalized locations of the individual nodes, the coordinates of the bounding box for each face component are also recorded.

Table 2 lists some of the statistics of the resulting measurement. Note that these numbers were calculated using only middle 50% of the samples. This practice will limit the range of 'normal' samples and increase the chance of applying exaggeration effects, which is a key element in composing caricatures. We adopt a simple linear model, shown in Fig. 5, for assigning the exaggeration rate outside the normal range.

Table 2. Some statistics of the face components

Group	Normal Range	Average	Standard deviation
G3 x	0.5619~0.605	0.58504	0.0294839
y	0.0259~0.0636	0.041953	0.0262763
G2 x	-0.6017~ -0.5603	-0.57958	0.0318806
y	0.0276~0.0622	0.0465488	0.0252109
G6 x	0.477~0.5702	0.52158	0.0725435
y	-0.5479~-0.4465	-0.49647	0.0738605

```
if (value within normal range)
        value not changed
else if (value < range_min)
        value = value – (range_min -value)*scale
else
        value = value + (value - range_max)*scale
```

Figure 5. Linear model for computing the degree of exaggeration outside the normal range.

3. AUTOMATIC FACIAL FEATURE EXTRACTION

As discussed in Section 2, there are a total of 119 nodes in the proposed face model. Any exaggeration or modification of the face components must be performed on the face mesh defined by these nodes. Robust extraction of facial features is therefore indispensable if we wish to build an automatic system with minimal human intervention. This section discusses the methods we adopted to carry out the feature extraction task. Basically, the association between the master node in each group and the dominant features on the subject's face must be built accurately. However, some facial

features are more 'stable' than the others. In our approach, the positions of the eyes and the mouth are first estimated and the triangle formed by these parts is taken as the reference for locating other face components. The overall shape of the face is approximated with the boundary of the average face and subsequently refined using an active contour model.

3.1. Estimation of Eyes and Mouth Positions

We employ a measure known as *horizontal edge density* to estimate the locations of the eyes and the mouth. The input image (Fig. 6a) is first filtered with a high-pass kernel in the horizontal direction. A fixed percentage of edge points is preserved by applying a threshold (Fig. 6b). Finally, the horizontal edge density for each edge point is computed according to:

$$HED = \frac{\text{\# of edge points within distance D}}{\text{total \# of neighbors within distance D}}$$

Since the eyes and the mouth usually exhibit pronounced horizontal edges, the computed HED will be large near these regions. Thresholding the HED and applying connected component detection algorithm will result in a few clusters of edge points, whose bounding box can be calculated easily (Fig. 6c).
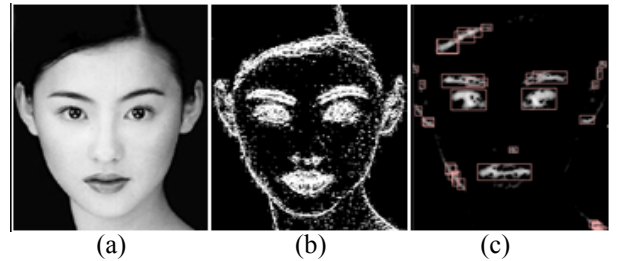

| (a) | (b) | (c) |

Figure 6. (a) Original image (b) Horizontal edges (c) Clusters of points with large horizontal edge density.

As the relative position of the face components is fixed, it is possible to use the following simple algorithm (Fig. 7) to obtain the possible sets of candidates that correspond to the eyes and the mouth regions (Fig. 8):

```
for(each bounding box B_i)
  for(other bounding box B_j)
   if (B_j is parallel to Bi and has a similar size with B_i)
      P_1 ← midpoint of B_iB_j
      for(the remaining bounding box B_k)
       if(Th1<distance between P_1 and B_k <Th2)
          then treat B_i、 B_j and B_k as a candidate set
```

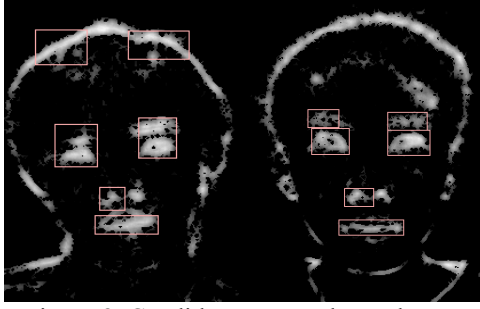Figure 7. Algorithm for selecting possible eyes and mouth regions.

Figure 8. Candidate eyes and mouth sets.

3.2. Iris Localization

The previous analysis might generate one or more sets of candidate regions. When it is the case, iris localization [8] can assist in defining the area that corresponds to the eyes. The process of finding the iris is summarized as follows:

1. Utilize canny edge detector to find connected edges.
2. Apply Hough transform to detect circles within each candidate region. Compute the weight for each detected circle by the following formula:

$$W(i) = CP(i) \bullet Contrast(i) \bullet Ratio(i)$$

where

$$CP(i) = \frac{\# \text{ of edge points on the circle}}{\text{perimeter}}$$

$$Constrast(i) = \frac{\text{average intensity outside the circle}}{\text{average intensity inside the circle}}$$

and

$$Ratio(i) = \frac{1}{1 + (10 - d/r)^2}$$ (d: the distance between two eyes, r: radius of the circle.)

The circle with the largest weighted, is chosen as the representative circle for this region.
3. For each possible pair of candidate regions, compute the probability of the representative circle being the iris according to:

$$P = \frac{1}{(1 - r_L / r_R)^2} + W_L + W_R$$

The pair with the highest probability is selected as the eye pair.

Figure 9. Iris detection helps locating the eyes.

3.3. Master Node Location and Shape Matching

The positions of the eyes and the mouth can be used to estimate the bounding boxes for the eyebrows. The next step is to identify the master node within each bounding box. Most master nodes (except that of group 1 and 4) are either boundary or high curvature points, which can be found by corner detection algorithm such as [9]. Regarding the nose and the face shape, we use the information obtained from the average face as the initial contour and apply an iterative approximation process to build the face mesh.
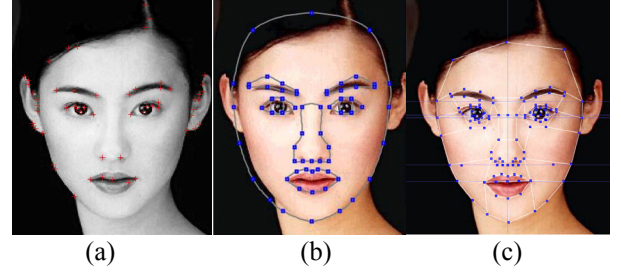

(a) (b) (c)

Figure 10. (a) Corner detection result. (b) Initial node positions derived from the average face. (c) Final face mesh obtained using active contour models.

4. CARICATURE GENERATION

The proposed caricature generation system is comprised of two stages: shape exaggeration and image metamorphosis. In the first stage, we attempt to discover the unique facial features of the subject by comparing the image with the average face. Quantitative analysis of these components reveals which and how they should be exaggerated. The second stage performs the actual shape deformation, together with texture alteration, using a pre-selected caricature as a reference.

4.1. Shape Exaggeration

We compare the face mesh of the input image with the average face to determine the exaggeration rate for each component. If the input component is classified as normal, i.e., falls within the middle 50% range, no shape exaggeration will be applied. Otherwise, a linear mapping is used to obtain the exaggeration rate. Given the amount of exaggeration for each component, a three-stage process is followed to adjust the positions of the nodes:

1. Initial node placement: At this stage, the positions of the nodes in each group are moved independently. Shift amount is directly proportional to the exaggeration rate.

2. Size adjustment: The size of the bounding box of the exaggerated shape is used to shrink or expand the component as a whole.
3. Final node placement: Spatial relationship among the face components has to be considered to ensure correct placement. This is achieved by placing the nodes with higher ranks first. The master point of each group is taken as the basis for detecting overlaps or collisions as a result of magnification or excessive shift. Other points in the group shift accordingly with the master node.

4.2. Image Metamorphosis

Caricature generation is formulated as a warping process in our approach. A feature-based image metamorphosis method is employed [10]. To begin with, an artist's work is selected as the source image, whose feature nodes have previously been labeled. For every input image, feature extraction and exaggeration rate estimation is performed as described in Section 3. We then use the exaggeration rate for each component to adjust the positions of the nodes, from high rank to low rank ones. A modified face mesh thus generated will contain the destination nodes for the warping process. If a different style of caricature is preferred, simply choose another caricaturist's work as the source image.

(a)   (c)   (e)
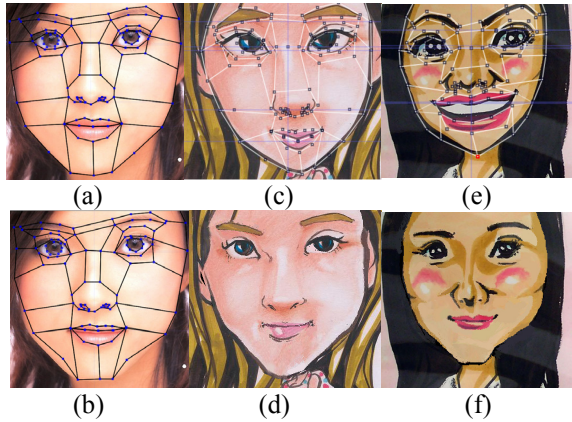
(b)   (d)   (f)

Figure 10. (a) A subject with the original face mesh overlaid. (b) Modified node positions after shape exaggeration. (c) An artist's work. (d) Caricature generated based on (c). (e) Another artist's work. (f) Caricature generated based on (e).

5. ILLUSTRATIONS

Figure 11 presents the caricature drawings of three different subjects using the same source image. Since the same artist's caricature is used, all drawings share a common style. Different parts of the face, however, are emphasized to various degrees. For example, the eyes of

subject in Fig. 11a are relatively small and have a unique orientation. These properties are amplified in the generated caricature. Similarly, the subject in Fig. 11c has large eyes and full lips, resulting in a caricature that further emphasizes these traits.
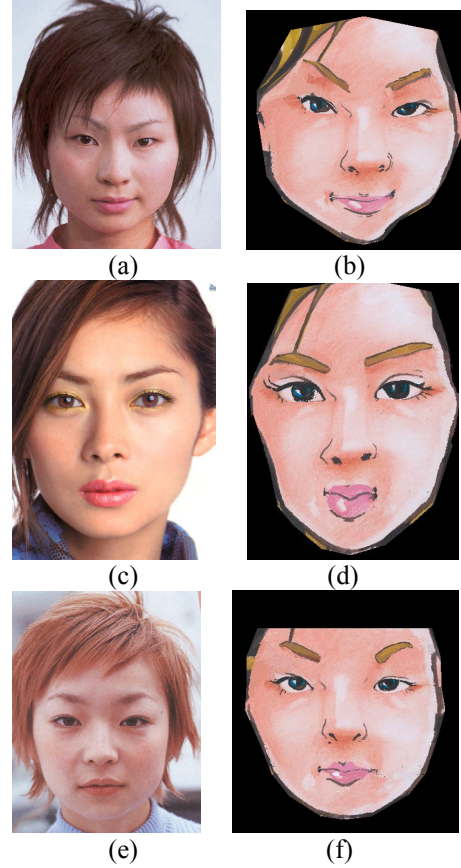

(a)   (b)

(c)   (d)

(e)   (f)

Figure 11. (a)(c)(e) Original input images. (b)(d)(f) Corresponding caricatures.

6. CONCLUSIONS

We have introduced an efficient approach for generating caricature drawings using only one existing caricature as the reference. An organized way for defining and estimating the face components and the associated parameters has been developed. By combining effective feature analysis with the proposed face model, caricature generation is converted into a simple image warping process.

The current prototype considers only the spatial relationship among the groups. Future system will also pay attention to the relative node position within each group to achieve more accurate shape definition. In addition, skin texture analysis and hair style classification will be included to enable the production of more appealing caricatures.

7. REFERENCES

.

[1] Hughes, *Caricatures (Learn to Draw)*, HarperCollins Publishers, London, 1999.

[2] L. Redman, *How to Draw Caricatures*, Contemporary Books, Chicago, 1984.

[3] A. Hertzmann, "Painterly Rendering with Curved Brush Strokes of Multiple Sizes," *Proceedings of the 25th annual conference on Computer Graphics and interactive Techniques*. 1998**.**

[4] S. Brennan, Caricature generator, Master's thesis, Cambridge, MIT, 1982.

[5] H. Koshimizu, M. Tominaga, T. Fujiwara, and K.Murakami. "On Kansei Facial Processing for Computerized Facial Caricaturing System Picasso," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 294-299, 1999.

[6] C.-Y. Yang, T-W Pai, Y-S Hsiao, C-H Chiang. "Face shape classification for 2D caricature generation," *Proceedings of Workshop on Consumer Electronics*, 2002.

[7] L. Liang, H. Chen, Y-Q Xu, H-Y Shum. *"*Example-Based Caricature Generation with Exaggeration," *Proceedings of 10th Pacific Conference on Computer Graphics and Applications***,** 2002.

[8] D. Toennies, F. Behrens, Melanie Aurnhammer ," Feasibility of hough-transform-based iris localisation for real-time-application." *ICPR 2002 , volume II, pp. 1053-1056,*2002. *"*

[9] C. J. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conf., Manchester*, pages 147-151, 1988.

[10] T. Beier and S. Neely, "Feature-Based Image Metamorphosis," *Proceedings of SIGGRAPH'92*, pp. 35-42, July 1992.