Example-based Caricature Generation with Exaggeration

Lin Liang[1],Hong Chen[2*], Ying-Qing Xu[1], Heung-Yeung Shum[1]

[1] Microsoft Research, Asia
[2] Xi'an Jiaotong University, China

Abstract

In this paper, we present a system that automatically generates caricatures from input face images. From example caricatures drawn by an artist, our caricature system learns how an artist draws caricatures. In our approach, we decouple the process of caricature generation into two parts, i.e., shape exaggeration and texture style transferring. The exaggeration of a caricature is accomplished by a prototype-based method that captures the artist's understanding of what are distinctive features of a face and the exaggeration style. Such prototypes are learnt by analyzing the correlation between the image caricature pairs using partial least-squares (PLS). Experimental results demonstrate the effectiveness of our system.

1. Introduction

ex.ag.ger.ate v. 1. To describe or think about as greater than is actually the case. 2. To increase or enlarge beyond what is normal or expected. It is not the exaggeration of one's worst features. [2]

car.i.ca.ture n. A picture or description using gross exaggeration or distortion, as for humorous effect or in ridicule. [2]

A caricature can be defined as an exaggeration likeness of a person made by emphasizing all of the features that make the person different from everyone else. The key is that an exaggeration is not a distortion. "Exaggeration is the overemphasis of truth. Distortion is a complete denial of truth." [1].

Somehow artists have the amazing ability to draw a caricature of someone's face. An artist is capable of capturing distinguished facial features that make his/her subject different from others, and then exaggerating these features. In order to generate an exaggeration, one has to make the following observations. Which of the subject's features are significantly different from others'? How can one define and measure the difference? How does one know which of the subject's features are larger, smaller, sharper, or rounder than other people's? [1]

There has been some previous work on how a caricature can reveal characteristics of a face. Based on psychological hypotheses, for example, Rhodes et al. [12] reported on experiments where a caricature looks good.

There have been a few attempts to interactively synthesize facial caricatures. Akleman et al. [3] developed a procedure to make caricatures using an interactive morphing tool. Brennan [4] presented an interactive caricature generator. Tominaga et. al. [13] developed the template-based facial caricature system PICASSO and Web-PICASSO [14]. Many approaches have also been proposed to generate facial caricature [9, 11, 8]. But these approaches did not attempt to observe and learn from the artist's products, and thus produce stiff and inexpressive sketches.

On the other hand, only a few systems have been proposed to teach a computer to automatically generate a stylistic facial sketch by observing images drawn by artists. Librande [10] developed an example-based character drawing system called Xspace. Recently, Freeman et al. [7] presented an example-based system for translating a sketch into different styles. They focused on style transfer rather than generating a stylistic sketch from an image.

Based on a set of training images and their associated sketches drawn by an artist with a particular style, Chen et. al. [5] developed a system to automatically generate a stylistic sketch from an input image. However, their system does not generate exaggerated caricature.

In this paper, we present an example-based caricaturing system. We decouple the caricature generation into two parts, i.e., shape exaggeration and texture style transferring. From example caricatures drawn by an artist, a prototype-based method is proposed to capture the artist's understanding of what are distinctive features of a face and the exaggeration style. Combining this with sketches or images, we can automatically generate an effective caricature from an input image, with exaggeration.

*Visiting at Microsoft Research , Asia

Proceedings of the 10 th Pacific Conference on Computer Graphics and Applications (PG'02)
0-7695-1784-6/02 $17.00 © 2002 **IEEE**

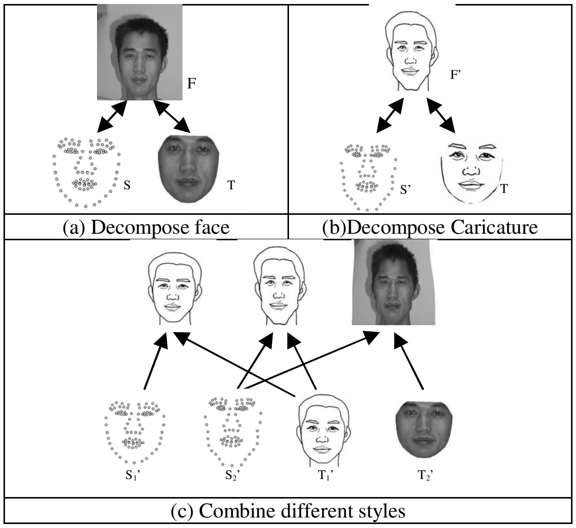

Figure 1. Decompose images and caricatures into shape and texture.

The rest of this paper is organized as follows. We present our caricaturing model and system framework in Section 2. After discussing ways of shape exaggeration in Section 3, we present the detailed algorithm of prototype-based shape exaggeration in Section 4. Results are shown in Section 5. In Section 6, we conclude this paper and discuss possible future research.

2. System Framework

2.1. A Caricaturing Model

An exaggerated caricature is very different from the original face image. It is rather complex to directly model the relationship between an image and its correspondence caricature in image space. One way to reduce such modeling complexity, as shown in statistical face modeling [6], is to separate a face image F into a shape S and a shape-free image (texture) T. Here, the shape S is defined by a set of feature points, and texture is the warped image using a standard shape (MeanShape), as shown in Figure 1 (a).

Naturally, we can also decompose a caricature image F' into shape S' and texture F' using the same set of feature points, as shown in Figure 1 (b). It is obvious that T' and T are highly correlated, so are S' and S. Therefore, we can decouple the caricaturing process into two parts: generation of S' and generation of T'. Compared to F and F', the relationships between S' and S and between T and T' are more clearly and easily modeled.

We call the process of synthesizing T' from T *texture style transferring*; and the process of generating S' from S *shape exaggeration*. These two processes can be combined to generate the final caricature.

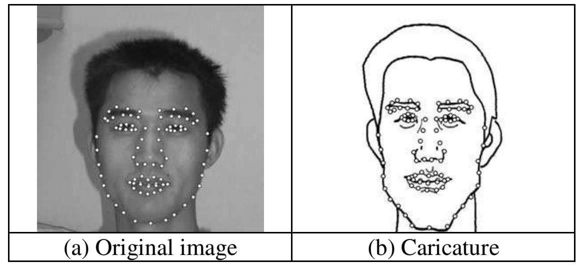

Figure 2. An example with labelled feature points.

Another benefit of this decoupled model is for style reuse. The texture style transferring models and shape exaggerating models, learned from different training data, can be combined to generate new styles of caricatures, as shown in Figure 1 (c). However, in this paper, we focus our work on generating shape exaggeration model. Most caricature examples we show here use sketches [13] for texture style transferring.

2.2. Training Data

Our examples (training data) include 92 pairs of original facial images and exaggerated caricatures drawn by an artist. All the original facial images have a frontal view without hat or glasses. We asked the artist to draw all the caricatures in the same stroke style. Also, we asked the artist to maintain the exaggeration style, e.g., the exaggeration rate. All of these assumptions simplify the learning process.

We manually label a set of feature points as a shape model on both the original image and the exaggerated caricature. These feature points are designed so carefully that they are easy to correspond in different faces and sufficient for obtaining reasonable results using Thin Plate Splines Warping between shapes. A pair of image and caricature with labeled points is shown in Figure 2.

2.3. The System Framework

Our system consists of a training phase and a runtime phase. In the training phase, we start with the set of exaggerated caricature and original image pairs, with manually labeled feature points both on the original images and the exaggerated caricatures, i.e. the original face shapes and exaggerated shapes. Then, we align the set of original face and the exaggerated shapes and train our prototype-based Shape Exaggeration Model.

Proceedings of the 10 th Pacific Conference on Computer Graphics and Applications (PG'02)
0-7695-1784-6/02 $17.00 © 2002 **IEEE**

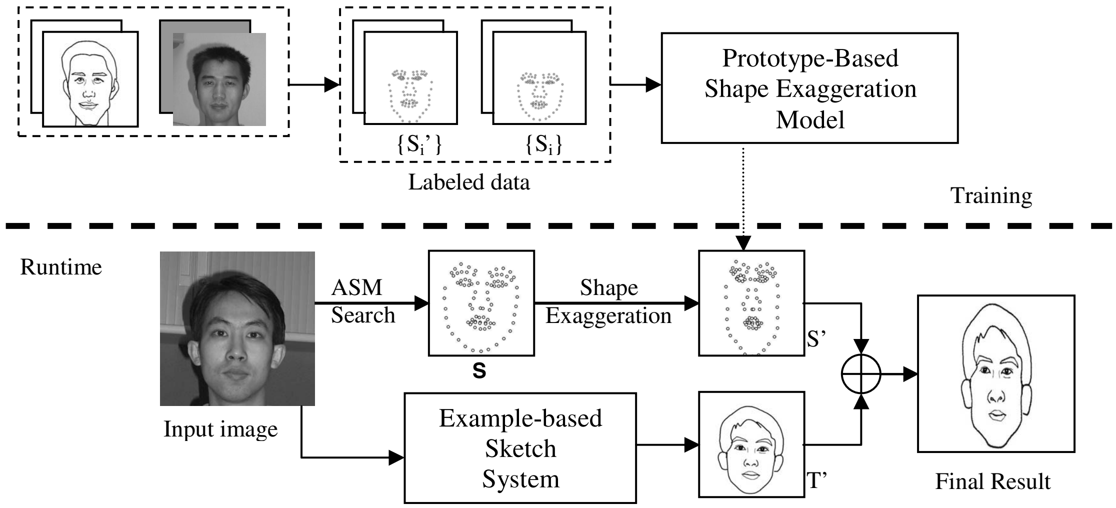

Figure 3. System framework for training phase and runtime phase.

At runtime, for a given image I, we extract the face shape using ASM [6]. Then, the Shape Exaggeration Model is employed to generate the exaggerated shape. Combined with an example-based sketch generation system, we can get the final sketch-style caricature with exaggeration.

Figure3 shows various steps in both the training phase and the runtime phase. Detailed algorithms of shape exaggeration are discussed in the next two sections.

3. Shape exaggeration

3.1. Scaling the difference

Training examples include the original shape S_i and exaggerated shape S_i', where both are aligned to the mean shape S_{mean}. Then S_i and S_i' can be reconstructed as: $S = S_{mean} + \Delta S$,$S' = S + \Delta S'$, where $\Delta S$is the difference between an individual face shape and the mean face shape, and $\Delta S'$ is the exaggerated part of the caricature shape with respect to the original shape. We aim to learn the relationship between ΔS and $\Delta S'$, i.e. $\Delta S' = f(\Delta S)$.

The simplest way to exaggerate is to scale, $\Delta S' = b\Delta S$, where b is the exaggeration weight. This is the approach used in previous work of the PICASSO system [13]. Obviously simple scaling is not enough to learn the artist's style of exaggeration; therefore the generated results are not satisfactory.

3.2. K-Nearest Neighbors (KNN)

Since the mapping relationship between ΔS and $\Delta S'$ cannot easily be described by a global parametrized model,

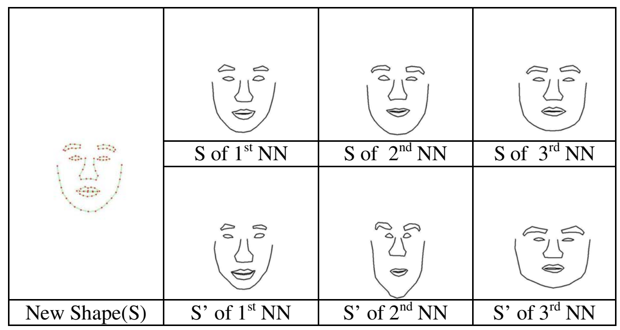

Figure 4. The k nearest neighbors(S) founded by Euclidean distance, exaggerated shapes(S') are exaggerated in different ways.

the kNN algorithm can be used to estimate $\Delta S'$ using a local linear model $\Delta S' = [\Delta S_1', \Delta S_2', ..., \Delta S_k']w$, where w is the weight which satisfies $\Delta S = [\Delta S_1, \Delta S_2, ..., \Delta S_k]w$ by means of least squares.

However, the k neighbor faces found using the Euclidean distance measurement may well be exaggerated by the artist in different ways, as in our training set shown in Figure 4. This means the weighed sum $\Delta S' = [\Delta S_1', \Delta S_2', ..., \Delta S_k']w$ will become an average of different exaggeration directions and degrees, thus the face cannot be exaggerated notably. Another problem of the KNN approach is that the distance measurement cannot capture the subtle facial features selected by the artist, such as the distance between the eyes. Note that we have tried other distance metrics such as Mahalanobis distance but there is no improvement.

3.3. Example-based Approach

In our work, an example-based method is used to learn how to identify facial features and exaggerate them using the artist's style, albeit implicitly. We ask our artist to select only several key facial features to exaggerate and to maintain consistent exaggeration styles throughout all examples.

In our system, a prototype is defined on a subset of training data that contains samples with similar exaggeration direction. Each prototype corresponds to the exaggeration style of some facial features, such as elongating a face. Based on the training data, we can select a set of such prototypes to represent different exaggeration directions of the artist. Then given a new input face, we judge to which prototype this face most likely belongs, then exaggerate it in the same direction as the samples that support this prototype.

Because the samples of one prototype are exaggerated in a similar exaggeration direction, the new input face will be exaggerated in the same direction. Thus the averaging effect caused by the previous methods is avoided. Prototype-based shape exaggeration method used in our system includes two phases.

At the training phase:

- From training examples, analyze the correlation between images and caricatures
- Construct a set of exaggeration prototypes

At the runtime phase:

- For an input shape, classify it into one of the exaggeration prototypes
- Exaggerate the input shape by the selected prototype

We explain the details in next section.

4. Prototype-based Exaggeration

We select exaggeration prototypes based on the following rules: each prototype represents an exaggeration trend corresponds to some facial features; a prototype should have enough supporting samples, and these samples are obviously exaggerated in the similar direction represented by the prototype; the complete prototype set is able to describe most exaggeration variance in the training data.

4.1. Exaggeration Analysis by PLS

To define such prototype set, we apply PLS (Partial Least Squares) [14] to the exaggerated shape space $\Omega_{\triangle S'}$ and normal shape space $\Omega_{\triangle S}$ to find different notable exaggeration trends correlate to facial features closely.

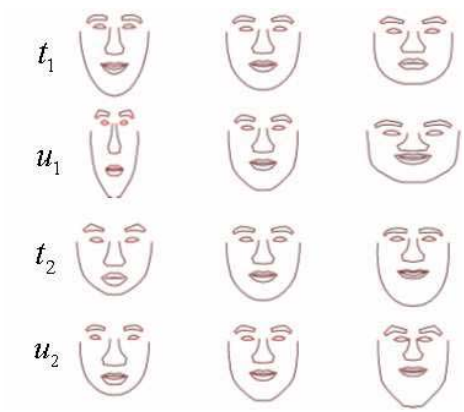

Figure 5. The first two pairs of modes of PLS.

PLS is similar to principle components analysis or PCA. But rather than treating input and output independently, PLS will consider the correlation between $\Omega_{\triangle S'}$ and $\Omega_{\triangle S}$ space when selecting principal components. During the iterative process of PLS, one pair of components t_i, u_i will be extracted in one step based on following rule:

$$\max Cov(t_i, u_i) \tag{1}$$

,where t_i and u_i is the ith component of $\Omega_{\triangle S'}$ and $\Omega_{\triangle S}$ respectively. Since

$$\max Cov(t_i, u_i) = \max \sqrt{Var(t_i)Varu_i}Corr(t_i, u_i) \tag{2}$$

, so t_i and u_i will not only represent the variation of each space, but also their correlation is very high.

Figure 5 shows the first two pairs of components found by PLS. Different exaggeration trends represented by components u_i are closely corresponding to some real facial features. It is very significant for our prototype selection. As u_2 shown in Figure 5, this mode describes how the artist exaggerates facial features: long and short jaws.

We also observe that the modes with large variance represent the exaggeration style of holistic facial features, such as the face contour, as shown in Figure 5. And modes with small variance correspond to some subtle face features captured by the artist, such as the size of the mouth.

4.2. Exaggeration prototypes selection

So we define exaggeration prototypes based on each component u_i, and use projection of samples on each principal components direction to select the samples supporting each prototype.

The whole process is summarized as follows:

- Select one pair of components t_i, u_i using PLS algorithm.

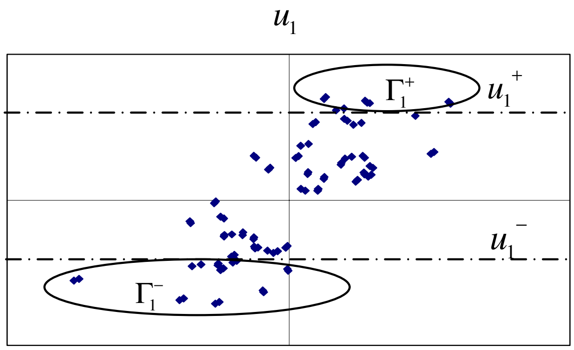

Figure 6. The projection of samples on the first pair of PLS components u_j and t_j.

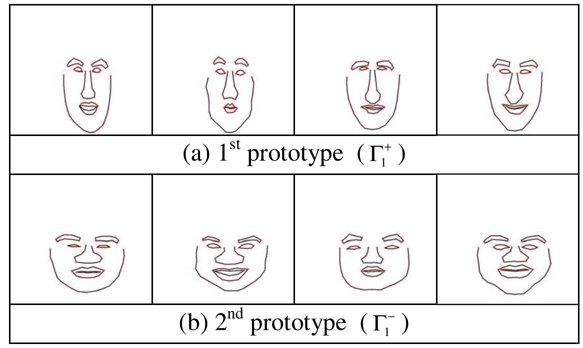

Figure 7. Some of the selected samples of the first two prototypes. Compared to Figure 4, they are exaggerated in more coherent way.

- Define two prototypes with opposite exaggeration directions based on component u_i. As shown in Figure 6, the corresponding sample sets are:

$$\Gamma_j^+ = \{(\Delta S_i, \Delta S_i')|u_j(\Delta S_i') > u_j^+\} \tag{3}$$

$$\Gamma_j^- = \{(\Delta S_i, \Delta S_i')|u_j(\Delta S_i') < u_j^-\} \tag{4}$$

  , where u_j^+ is the kth maximal value of projection coordinates on u_j , and u_j^- is the kth minimal value.

- Decompose $\Omega_{\Delta S'}^i$ and $\Omega_{\Delta S}^i$:

$$\Omega_{\Delta S'}^{i+1} = \Omega_{\Delta S'}^i - u_i q_i^T \tag{5}$$

$$\Omega_{\Delta S}^{i+1} = \Omega_{\Delta S}^i - t_i p_i^T \tag{6}$$

  , where q_i and p_i is the loading vector of u_i and t_i.

- Repeat the above steps until enough components are selected, i.e., the variance of $\Omega_{\Delta S}^{i+1}$ is less than a threshold λ.

Our experiment shows that $\lambda = 10\%$ is sufficient to guarantee that the prototype set will represent most exaggeration variance in the training data, and 28 prototypes are defined. Figure 7 shows the selected sample prototypes having a similar exaggeration direction.

4.3. Prototype classification for input shape

Given a new input shape S_{new}, we must judge which prototype will be used to exaggerate it. This step is to identify the main facial features of S_{new} that the artist is most likely to exaggerate.

Since t_j is most correlate to u_j, so we use the projection on t_j of the shape to judge whether it should be exaggerated at the direction of u_j. If we denote the projection of ΔS_{new} on t_j as $t_j(\Delta S_{new})$, for each u_j, we could define the evidence for the shape been exaggerated at the direction of u_j:

$$\alpha^+(t_j(\Delta S_{new})) = \frac{N_{\Gamma_j^+}(t_j(\Delta S_{new}))}{N(t_j(\Delta S_{new}))} \tag{7}$$

,where $N_{\Gamma_j^+}(t_j(\Delta S_{new}))$ is the number of samples in prototype Γ_j^+ that $t_j(\Delta S_i) > t_j(\Delta S_{new})$, and $N(t_j(\Delta S_{new}))$ is the number of training samples satisfying $t_j(\Delta S_i) > t_j(\Delta S_{new})$. The intuitive meaning of the evidence $\alpha^+(t_j(\Delta S_{new}))$ is the probability of the shape being exaggerated in the direction of u_j when its projection on t_j is bigger than $t_j(\Delta S_{new})$. The evidence for the shape being exaggerated at the opposite direction of u_j can be defined similarly:

$$\alpha^-(t_j(\Delta S_{new})) = \frac{N_{\Gamma_j^-}(t_j(\Delta S_{new}))}{N(t_j(\Delta S_{new}))} \tag{8}$$

Figure 8 shows the first three prototypes on which a new input shape has high evidences. As shown in the Figure 8, we exaggerate the new shape using these prototypes, and get different but reasonable caricatures. The result with higher evidence captures more notable facial feature, and looks more visually striking. Therefore, we select the one which has the maximum evidence in our system.

This process aims to identify the notable facial features of the input face, in a way very similar to what artists do in order to draw a caricature.

4.4. Exaggeration by prototype

Once the new input shape S_{new} has been classified into one prototype, a local linear model is used to estimate the exaggerated part $\Delta S_{new}'$ only using the samples of this prototype.

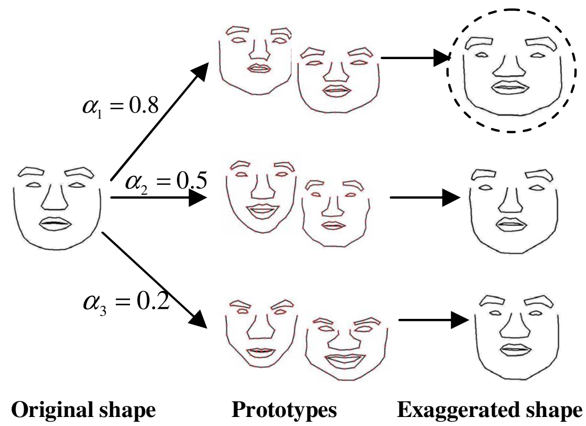

Figure 8. Selected prototypes for new shape with evidence measurement.

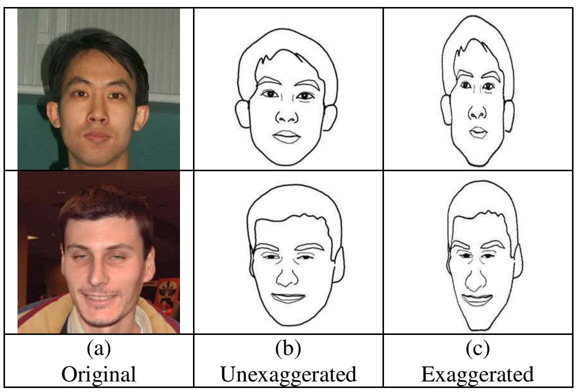

Figure 9. Two examples of exaggerated caricature.

Since the correlation between feature space T and U found by PLS is high, solving the local linear model in such feature spaces is more reasonable. So we first project ΔS_{new} into feature space T to get t_{new}, then in T find the least squares solution so that t_{new} can be best linearly approximated by the original shape features of prototype, such that $t_{new} = \sum_k w_k t_k$. In order to guarantee that ΔS_{new} is exaggerated in the same direction as the samples of the prototype, a least squares solution is found under constraints $w_k \geq 0$. The weights are then translated to the exaggerated shape feature space U to get $u_{new} = \sum_k w_k u_k$. Finally $\Delta S'_{new}$ is reconstructed from u_{new}.

5. Experimental Results

Finally we can generate caricatures for a given image of a frontal face by simulating the artist's style. We have applied "leave-one-out" approach to generate the caricatures,

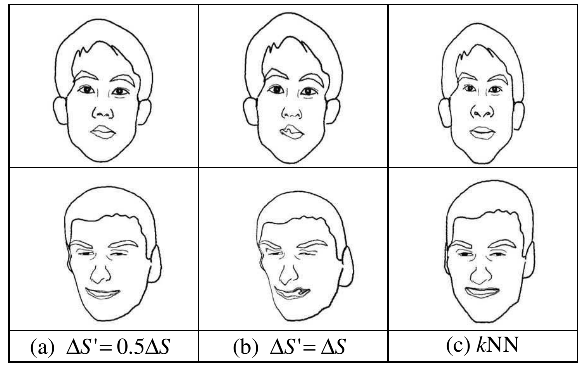

Figure 10. The results of compared methods. (a)$\Delta S' = b\Delta S$, $b = 0.5$ (b)$\Delta S' = b\Delta S$, b=1.0 ((a), (b) are PICASSO system's method) (c) kNN using Euclidean distance measurement.

without using its own caricature in the training process. Figure 9 shows some of the caricature sketches generated by our system. We can see that the exaggerated caricatures are more impressive than unexaggerated ones.

Figure 10 compares the prototype-based method with the kNN method and PICASSO system. As we can see, the prototype-based method can exaggerate facial features more markedly and impressively, without the averaging effect shown by other approaches.

More results are shown in Figure 11. We also apply the exaggeration to the original images and generate interesting results. Also, we compare them with the artist drawings. We can see that our system can capture the same notable features selected by the artist, and exaggerate these features in a similar style.

6. Conclusions

We have presented an example-based exaggerated caricature generation system. Our system can automatically identify facial features from input images, and exaggerate such features simulating the artist's style. By decoupling the caricature generation into two parts, shape exaggeration and texture style mapping, our caricature system can learn the artist styles of both parts. A new prototype-based exaggeration model is employed to learn how artists identify facial features and the way artists exaggerate them. Based on our model, some main facial features can be obviously exaggerated in a proper direction.

Since we define prototypes only based on principal components directions, the exaggeration directions determined by our model are limited. In fact the facial features selected by the artist to exaggerate may cover different prototypes, so a more effective way should be found to combine prototypes to exaggerate a new input face.

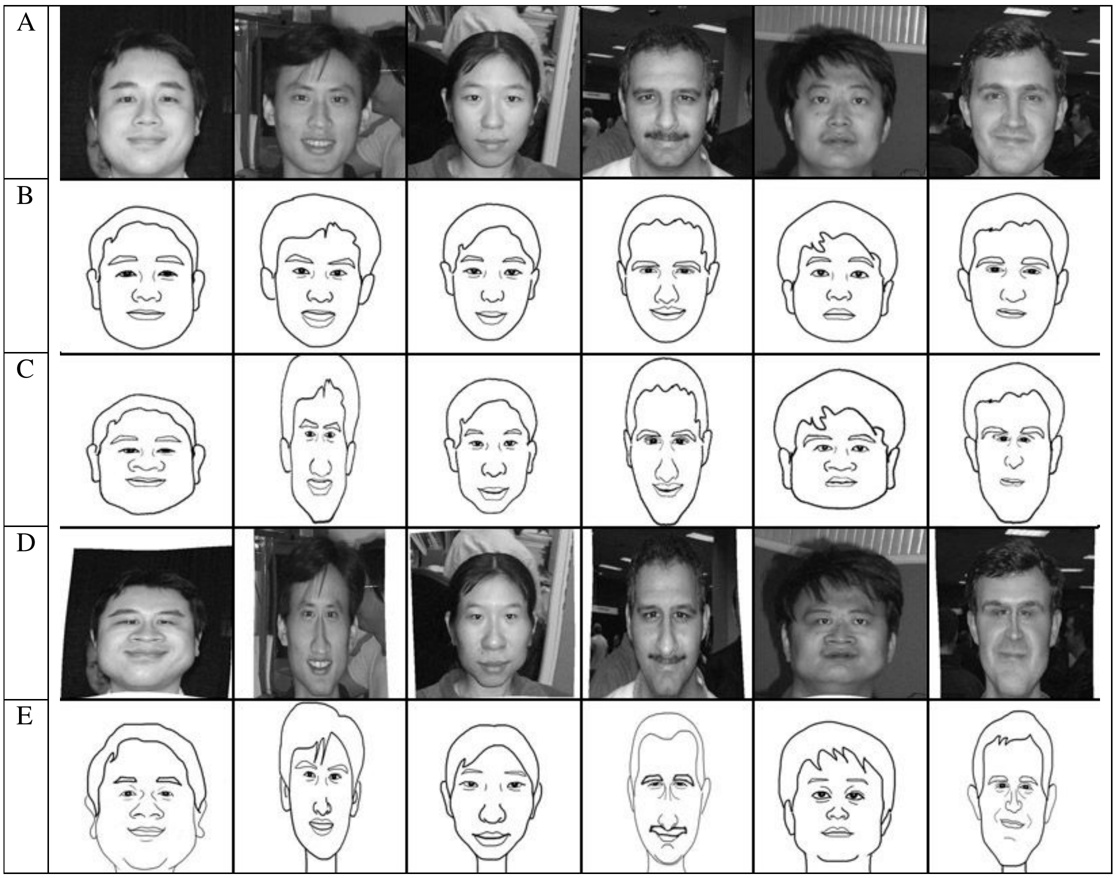

Figure 11. Result comparison. (A) Original image, (B) Unexaggerated sketch (C) Exaggerated caricature generated by our system, (D) The same exaggeration is applied to the image, (E) The caricature by the artist. Note that when generating the caricature (on rows C or D), the artist's caricature (on row E and the same column) was not used for training. This is so-called "leave-one-out" test in pattern recognition.

Proceedings of the 10 th Pacific Conference on Computer Graphics and Applications (PG'02)
0-7695-1784-6/02 $17.00 © 2002 IEEE

References

[1] *How to draw caricature*. Contemporary Books, 1984.

[2] *The new International Webster's Pocket Dictionary of the English Language*. Trident Press International, 1997.

[3] E. Akleman. Making caricature with morphing. In *Visual Proceedings of ACM SIGGRAPH'97*, page 145, 1997.

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

[5] H. Chen, Y. Xu, H. Shum, S. Zhu, and N. Zheng. Example-based facial sketch generation with non-parametric sampling. In *ICCV01*, pages II: 433–438, 2001.

[6] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer version. Technical report, University of Manchester, Manchester M13 9PT, U.K., 2000.

[7] W. T. Freeman, J. B. Tenenbaum, and E. Pasztor. An example-based approach to style translation for line drawings. Technical Report 11, MERL Technical Report, Cambridge, MA, Feb 1999.

[8] S. Iwashita, Y. Takeda, and T. Onisawa. Expressive facial caricature drawing. In *IEEE International conference on fuzzy systems*, volume 3, pages 1597 –1602, 1999.

[9] Y. Li and H. Kobatake. Extraction of facial sketch based on morphological processing. In *IEEE international conference on image processing*, volume 3, pages 316–319, 1997.

[10] S. E. Librande. Example-based character drawing. Master's thesis, Cambridge. MA. MIT, 1992.

[11] J. Nishino, T. Kamyama, H. Shira, T. Odaka, and H. Ogura. Linguistic knowledge acquisition system on facial caricature drawing system. In *IEEE international conference on fuzzy systems*, volume 3, pages 1591 –1596, 1999.

[12] G. Rhodes. Secrets of the face. *New Zealand journal of psychology*, pages 3–17, 1994.

[13] M. Tominaga, S. Fukuoka, K. Murakami, and H. Koshimizu. Facial caricaturing with motion caricaturing in picasso system. In *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, page 30, 1997.

[14] www.koshi-lab.sccs.chukyo-u.ac.jp/ fuji/pica2.

0-7695-1784-6/02 $17.00 © 2002 IEEE