Automatic Generation of Caricatures with Multiple Expressions Using Transformative Approach

Wen-Hung Liao and Chien-An Lai

Department of Computer Science, National Chengchi University,
Taipei, Taiwan
{whliao,g9202}@cs.nccu.edu.tw

Abstract. The proliferation of digital cameras has changed the way we create and share photos. Novel forms of photo composition and reproduction have surfaced in recent years. In this paper, we present an automatic caricature generation system using transformative approaches. By combing facial feature detection, image segmentation and image warping/morphing techniques, the system is able to generate stylized caricature using only one reference image. When more than one reference sample are available, the system can either choose the best fit based on shape matching, or synthesize a composite style using polymorph technique. The system can also produce multiple expressions by controlling a subset of MPEG-4 facial animation parameters (FAP). Finally, to enable flexible manipulation of the synthetic caricature, we also investigate issues such as color quantization and raster-to-vector conversion. A major strength of our method is that the synthesized caricature bears a higher degree of resemblance to the real person than traditional component-based approaches.

Keywords: Caricature generation, facial animation parameters, image morphing, facial feature localization.

1 Introduction

Caricatures are exaggerated, cartoon-like portraits that attempt to capture the essence of the subject with a bit of humor or sarcasm. These portraits are usually drawn by professional cartoonists/painters based on their personal judgment or representative style. Recently, interests in creating systems that can automatically generate 2D caricatures using photo as the input have surged due to the proliferation of image capturing devices as well as advances in facial feature analysis. Some system (e.g., Mii editor in Wii) employs a *component-based* approach, in which a collection of facial components are pre-defined and then selected to assemble a cartoon face either manually or by similarity analysis of individual facial parts. A major issue with component-based approach is that the composite cartoons exhibit limited varieties in their styles due to the nature of template selection. Even worse, the resulting caricature sometimes bear little resemble to the person even when the individual parts are matched flawlessly. In *example-based* approaches [1][2], a shape exaggeration model is learned from a labeled example set. At runtime, the facial shape is extracted using

F. Huang and R.-C. Wang (Eds.): ArtsIT 2009, LNICST 30, pp. 263–271, 2010.
© Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering 2010

active shape model (ASM) and then classified into one of the exaggeration prototypes for further deformation. It appears that shape plays a primary role (compared to texture) in this type of approach. As a result, the computer-generated caricatures are usually monochrome, lacking details in textural elements.

To address the issues described above, we have developed and implemented an automatic caricature generation system using a *transformative framework* [3]. In this type of approach, a new caricature is synthesized by transferring the style from a reference cartoon drawing via matching of the two sets of pre-defined control points in the input photo and the reference image, as depicted in Fig. 1. In [4], we further improve the precision of the feature localization result using active appearance model (AAM). We also propose to create a 3D caricature model by merging two planar caricatures using principles of binocular vision.

Caricature by cartoonist

Input photo

Transformed image

Fig. 1. Transformative approach for generating caricature from a single reference drawing [3]

The research described in this paper marks an extension to our previous efforts. Specifically, we incorporate the MPEG-4 facial animation parameters into the shape model to enable of automatic generation of caricature with multiple expressions using a single reference drawing. We also exploit the idea of creating new drawing styles by blending features from multiple sources using polymorph technique [5]. As the collection of reference drawings becomes larger, it will be convenient to select/recommend a proper template based on shape similarity measure. Finally, to facilitate efficient storage and flexible manipulation of the synthesized caricature, we look into issues regarding color reduction and raster-to-vector conversion.

The rest of this paper is organized as follows. In Section 2, we first review the facial mesh model designed for our application. We then present two general approaches for creating caricatures from an input photo and a reference cartoon. The second part of Section 2 is concerned with template selection using shape information, as well as style blending using polymorph technique. Section 3 defines the 2D FAP set employed in our system and the associated facial component actions, which are further combined to derive archetypal expression profiles, i.e., different emotion types. Results of synthesized images with multiple expressions are presented as Emoticons. Section 4 concludes this paper with a brief summary.

2 Caricature Generation

The overall procedure for generating caricature from input photos is illustrated in Fig. 2. In the first step, a frontal-view photo is presented as the source image. Using active appearance model (AAM), we can locate important facial components and the corresponding mesh, as shown in Fig. 3. The mesh is comprised of 8 components, with a total of 66 control points and 114 constrained conforming Delaunay triangles (CCDT) for defining the facial configuration.

The caricature can then be generated using two different approaches: image morphing or rotoscoping. In image morphing, a collection of reference cartoons with different styles needs to be gathered. Control points as well as the corresponding mesh of these cartoons will be labeled manually. The user will choose a template to transfer the style from. We then proceed to identify the corresponding triangular meshes between the reference cartoon and the input image. After the correspondence has been established, it is a simple matter to apply image morphing technique to transfer the texture from the reference drawing to the input photo. In our current implementation, barycentric coordinate is employed to speed up the morphing process. Details regarding the coordinate transformation on CCDT can be found in [6]. Typical morphing results are shown in Fig. 4.

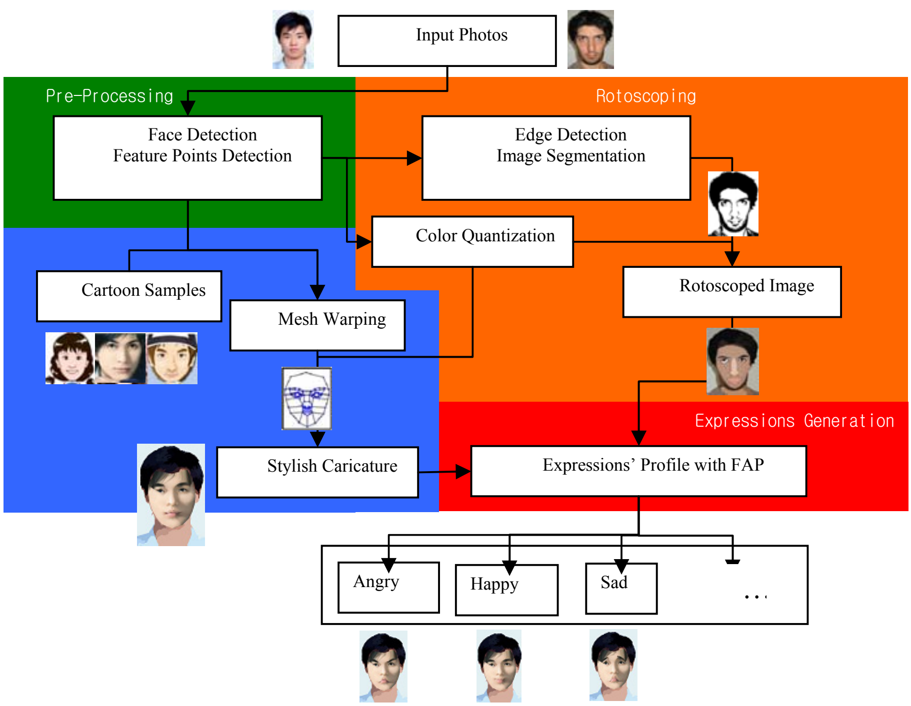


Fig. 2. Proposed framework for generating caricatures with multiple expressions

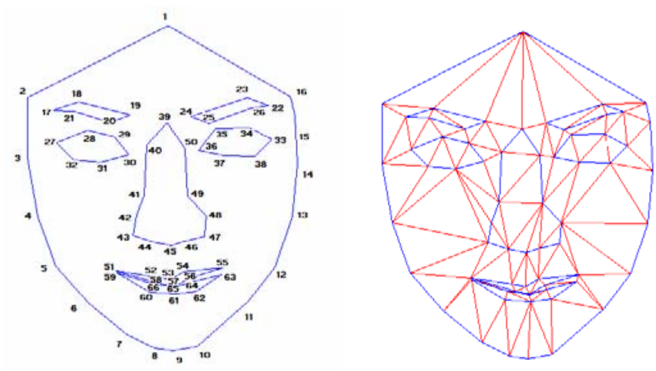


Fig. 3. (a) Control points for the 2D face model, (b) triangular facial mesh

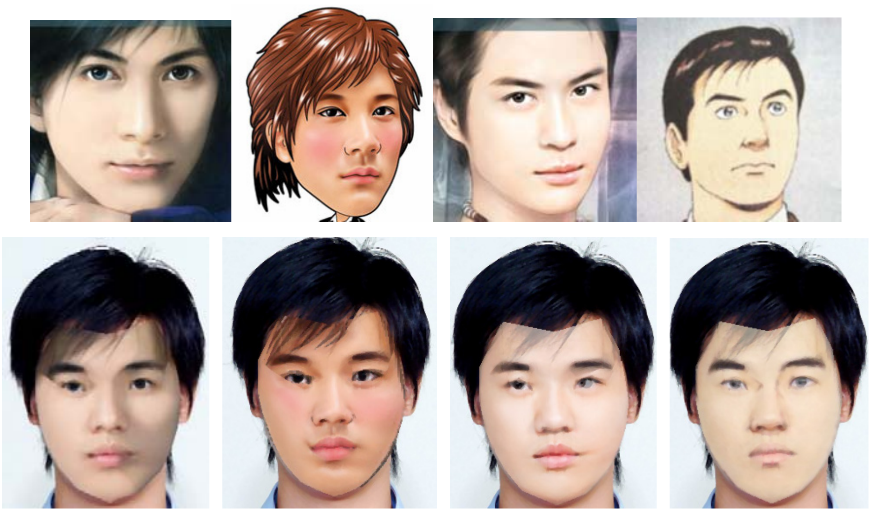


Fig. 4. Caricatures generated from different reference drawings

The results depicted in Fig. 4 (bottom row) are generated using different reference cartoons (top row). Generally speaking, each synthesized image bears certain degree of likeness to the original photo, yet at the same time exhibits the drawing style originated from the reference cartoon.

As the number of available templates increases, it becomes convenient for the system to select or recommend best references based on shape similarity. Since each component in our face model forms a closed contour, Fourier descriptor can be utilized to represent the individual parts. Let $(x_i(n),y_i(n))$ denote the n-th contour point of the i-th facial component, and let $u_i(n)=x_i(n)+jy_i(n)$, the facial shape descriptors (FSD) can then be defined according to:

$$a_i(k) = F\{u_i(n)\} = \sum_{n=0}^{N_i-1} u_i(n) \cdot e^{-j2\pi kn/N_i} \tag{1}$$

where N_i indicates the total number of control points in the i-th component.

Using Eq. (1), the difference between two facial shapes P (extracted from photo) and Q_s (template in the database) can be expressed according to:

$$D(P, Q_s) = \sum_{i=1}^{M} Dist(FSD_i^P, FSD_i^{Q_s}) \tag{2}$$

where

$$Dist(FSD_i^P, FSD_i^{Q_s}) = \left[\frac{1}{L_i} \sum_{k=0}^{L_i} \left(\left| a_i^P(k) \right| - \left| a_i^{Q_s}(k) \right| \right)^2 \right]^{0.5} \tag{3}$$

M denotes the number of facial components, and L_i denotes the number of Fourier coefficients retained to represent the i-th component.

Fig. 5 presents the top 5 choices for photos of different gender suing FSD distance. An additional benefit of defining FSD arises from the blending process. It is known that blending of styles can be easily accomplished using polymorph technique. With FSD, the weight for each template can be set to be inversely proportional to the distance between the input shapes and the respective reference caricatures, as shown in Fig. 6.

The second method to generate caricature is rotoscoping. In rotoscoping, we combined contour detection, mesh generation, histogram specification and mean-shift algorithm to obtain a simplified region-based representation of the face with user-defined

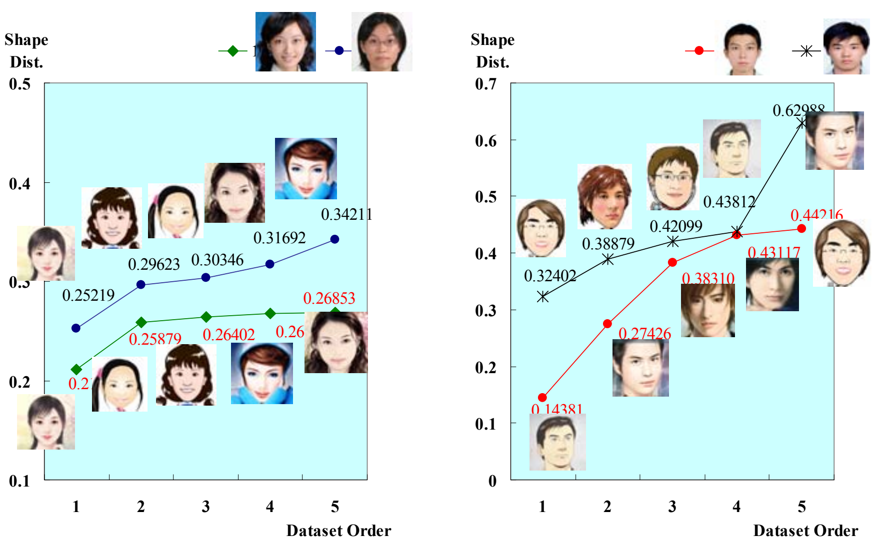


Fig. 5. Top template recommendations using shape similarity analysis

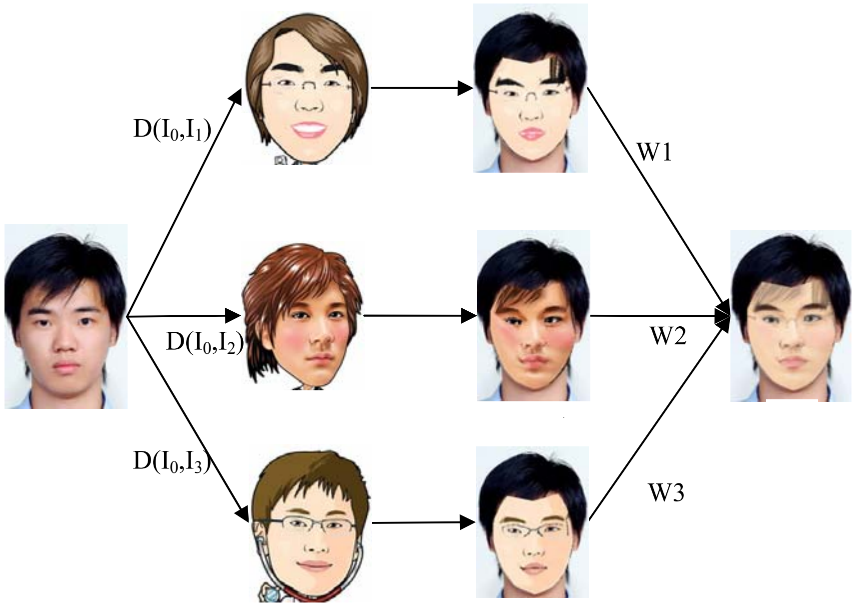


Fig. 6. Blending of styles using polymorph with FSD-controlled weights

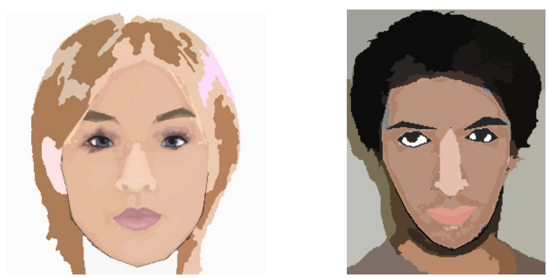


Fig. 7. Rotoscoped images

color palette. Details regarding the rotoscoping process can be found in [6]. Fig. 7 illustrates some rotoscoped images. Due to the nature of mean-shift segmentation, some triangular meshes are merged and contain the same color. These simplified regions are readily subject to raster-to-vector conversion to obtain vector-based representation such as Scalar Vector Graphics (SVG).

It should be noted that the color quantization stage will effectively reduce the number of colors used in the resulting picture, where as the histogram specification stage will attempt to preserve color information from the original photo. These steps can also be incorporated into the image morphing approach to retain the color tone from the input photo.

3 Synthesis of Multiple Expressions

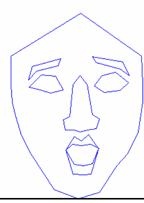
The FAP in MPEG-4 defines the set of parameters for 3D facial animation. Here we are mainly concerned with the synthesis on a 2D model. Consequently, we decompose the archetypal expression profile defined in [7] to arrive at a reduced set of facial animation parameters (FAP) for 2D caricatures, as shown in Table 1.

Using facial component action definition, the FAP can be set to manipulate the 66 facial definition points (FDP) in our model directly to synthesize different expressions. We have also defined 12 profiles according to popular emoticons. An example is illustrated in Table 2, where the emotion type (surprise) and its corresponding facial component action definition (FCAD), applied FDP and adjusted FAP are tabulated. Fig. 8 depicts the cartoons generated using the emoticon settings.

Table 1. Extracted 2D FAP Set

Group	Description	FAP No.	No. of Points	No. of Points Included in Reduced 2D Set
1	Visemes, Expressions	1~2	2	0
2	Jaw, chin, inner-lowerlip, corner-lips, midlip	3~18	16	11 (69%)
3	Eyeballs, pupils, eyelids	19~30	12	4 (33%)
4	Eyebrow	31~38	8	8 (100%)
5	Cheeks	39~42	4	0
6	Tongue	43~47	5	0
7	Head rotation	48~50	3	0
8	Outer-lip positions	51~60	10	8 (80%)
9	Nose	61~64	4	0
10	Ears	65~68	4	0
	Total		68	31 (45.6%)

Table 2. An example of emoticon profile setting

Emoticon	FCAD	Applied FDP	Related FAP
 Surprise	Brow: up Eye: wide Jaw: oh		$F_{31,33,35,\ 32,34,36}$ (1), $F_{37,38}$ (1.2) $F_{19,21}$(1.2), $F_{20,22}$(1.2) F_{3}(0.6), $F_{4,8,9}$ (0), $F_{5,10,11}$ (-0.6)$F_{12,59}$ (-0.8), $F_{13,60}$ (-0.8)$F_{6,53}$(0.5), $F_{7,54}$(0.5)$F_{55,56,57,58}$(1.3)

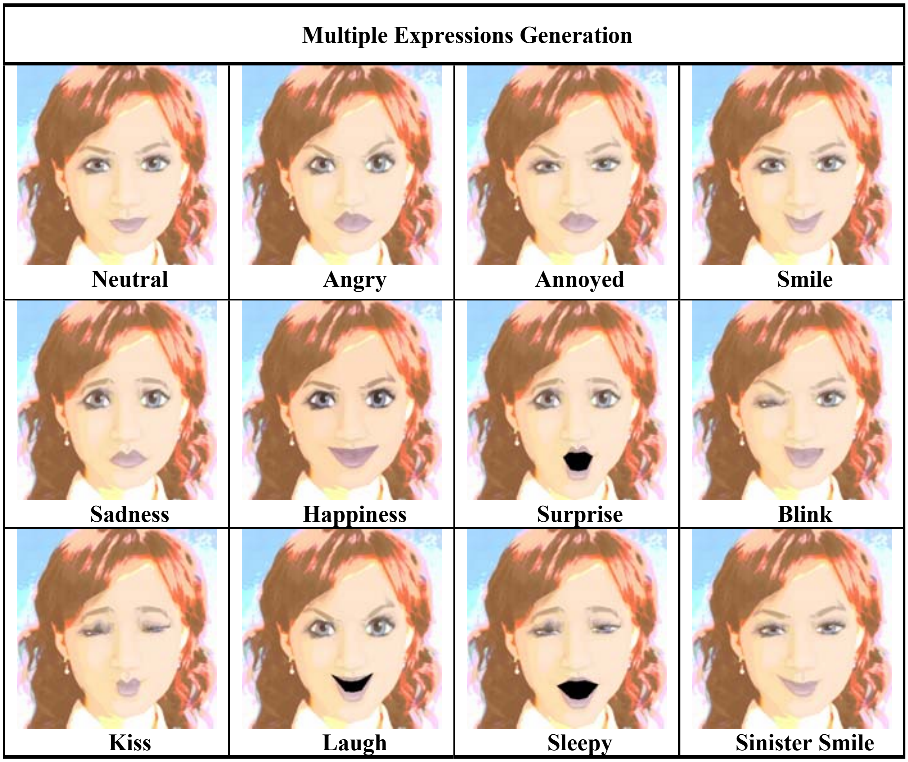


Fig. 8. Synthesized cartoons with multiple expressions

4 Conclusions

We have demonstrated an automatic caricature generator that is capable of synthesizing personalized cartoons with multiple expressions using only one reference image. New styles can be easily created by blending existing samples. The incorporation of FSD facilitates selection of templates in a cartoon database, as well as weight calculation in polymorph process.

Compact representation of the synthesized caricatures remains an issue for further investigation. Current model cannot strike a balance between fidelity and color quantization/region segmentation factors. Raster-to-vector conversion usually generates over-segmented patches. Novel types of vector representation such as diffusion curves may be applied to tackle this problem.

References

1. Liang, L., Chen, H., Xu, Y.Q., Shum, H.Y.: Example-Based Caricature Generation with Exaggeration. In: 10th Pacific Conference on Computer Graphics and Applications (PG 2002), pp. 386–393 (2002)

2. Mo, Z., Lewis, J.P., Neumann, U.: Improved Automatic Caricature by Feature. Normalization and Exaggeration. In: Proceedings of ACM SIGGRAPH Conference on Computer Graphics and Interactive Technique (2004)
3. Chiang, P.Y., Liao, W.H., Li, T.Y.: Automatic Caricature Generation by Analyzing Facial Features. In: Proc. of 2004 Asian Conference on Computer Vision (2004)
4. Chen, Y.L., Liao, W.H., Chiang, P.Y.: Generation of 3D Caricature by Fusing Caricature Images. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 1, pp. 866–871. IEEE Press, Los Alamitos (2006)
5. Lee, S., Wolberg, G., Shin, S.Y.: Polymorph: Morphing Among Multiple Images. IEEE Computer Graphics and Applications 18, 58–71 (1998)
6. Lai, C.-A.: Automatic Generation of Caricatures with Multiple Expressions Using. Transformative Approach. Master's Thesis, National Chengchi University (2008)
7. Perlin, K.: Layered Compositing of Facial Expression. In: ACM SIGGRAPH Technical Sketch (1997)