Article

Component-Based Cartoon Face Generation

Saman Sepehri Nejad [†] and Mohammad Ali Balafar *,[†]

Department of IT, Faculty of Electrical and Computer Engineering, University of Tabriz, 29 Bahman Boulevard, Tabriz 5166616471, Iran; samansepehri@gmail.com
* Correspondence: balafarila@tabrizu.ac.ir; Tel.: +98-914-409-4633
† These authors contributed equally to this work.

Abstract: In this paper, we present a cartoon face generation method that stands on a component-based facial feature extraction approach. Given a frontal face image as an input, our proposed system has the following stages. First, face features are extracted using an extended Active Shape Model. Outlines of the components are locally modified using edge detection, template matching and Hermit interpolation. This modification enhances the diversity of output and accuracy of the component matching required for cartoon generation. Second, to bring cartoon-specific features such as shadows, highlights and, especially, stylish drawing, an array of various face photographs and corresponding hand-drawn cartoon faces are collected. These cartoon templates are automatically decomposed into cartoon components using our proposed method for parameterizing cartoon samples, which is fast and simple. Then, using shape matching methods, the appropriate cartoon component is selected and deformed to fit the input face. Finally, a cartoon face is rendered in a vector format using the rendering rules of the selected template. Experimental results demonstrate effectiveness of our approach in generating life-like cartoon faces.

Keywords: cartoon face; facial features; Active Shape Model; cartoon templates

1. Introduction

The face is an invaluable characteristic for human identity, and has been studied in diverse areas of computer graphics and computer vision. Some researchers in the field of non-photorealistic rendering (NPR) have considered various types of artistic styles of human face. A cartoon face is an instance of these artistic representations, which is stylish, simple and usually humorous. Thanks to rapid developments in computer graphics, realistic human-like characters are becoming more widespread in digital world. The phenomenon of uncanny valley reveals that moving towards absolute human-like character surprisingly changes the viewer's feeling of it to a less life-like character [1]. Hanson [2] indicates the effect of cartoon-like style on reduction of uncanniness. On the other hand, rendering a high quality realistic character with complex texture map demands costly computation tasks. Over the recent years, cartoon face generation has become particularly interesting research topic due to the advantageous features of privacy, simplicity, funniness and their dominant digital applications such as in video games, virtual entertainment, social networks and mobile applications.

However, drawing cartoon faces is difficult for ordinary users. Hence, cartoon face producers attempt to imitate the artistic style using automatic or semi-automatic approaches. Generally, there are two prominent factors to consider in producing a cartoon face: likeness and aesthetic. Likeness can be addressed using facial feature extraction methods, and the artistic side is tackled by applying predesigned templates, learning the artistic style and NPR techniques. The system proposed by Koshimizu et al. [3], PICASSO, and the one developed by [4] generate stiff facial sketches according to the facial feature points using image processing techniques. Microsoft introduced a cartoon system,

MSN cartoon, which requires user interactions to choose the desired cartoon style, adjust the form and attach some add-ons. PicToon [5] is an automatic portrait generator based on a statistical model. Using training examples it captures an artistic style, and a non-parametric sampling approach is applied to render the vector-based facial sketch. In [6], an active shape model (ASM)-based facial feature extraction and an interactive template-based cartoon texture mapping is proposed.

Recent cartoon producers [7–9] are mostly based on fitting active models, commonly ASM [10] and active appearance model (AAM) [11], at the global level for face alignment and locating facial feature points. Although we make use of a similar model fitting approach, our system optimizes the contours at component level. Liu et al. [8] classified current approaches for cartoon depiction to contour and component representation. In their cartoon face producer, NatureFace, a component-based approach for modeling face and cartoon is integrated. A similar component-based method is used by Chen et al. [12] and Zhang et al. [13]. However, unlike our system, their cartoon matching is based on an imaged-based comparison approach such as k-nearest neighbor. In some recent works such as [9,14], to preserve the shape structure, the facial outline is drawn directly using feature points of the face component, and the remaining components are composed using template matching. Meng et al. [15] employs a dynamic thresholding scheme at component level to generate paper-cut portraits using predefined paper-cut templates. Ding and Martinez [16] train local classifiers for some facial components with large number of examples. In our proposed cartoon system, contour modification has been investigated for more components to retain the individualism of features and improve the template selection stage. Furthermore, cartoon results are evaluated in a subjective manner based on similarity of the cartoon face to the real face, which is a missing part in all previous research.

The rest of the paper is organized as follows. The description of our cartoon system is covered in Section 2. The proposed facial feature locating method is described in Section 2.1. Section 2.2 is dedicated to the cartoon sample collection and the consequent parameterization. In Section 2.3, processes required for cartoon rendering is presented. In Section 3, the final results are shown and our experiments are described in detail. Finally, in Section 4, we sum up our work, outline the benefits and limitations of the proposed system, and consider some future research topics.

2. Materials and Methods

The architecture of our cartoon face producer is shown in Figure 1. The proposed system can generate a cartoon face and a sketch representation from a given frontal face photograph using automatic techniques. The system incorporates three major components: cartoon template collection, feature extraction and cartoon rendering.

In the feature extraction stage, first, an extended ASM is employed to capture the initial locations and contours of the facial components. Contours of the components are optimized to better fit the input face. The hair region is segmented separately, as model fitting cannot address hair extraction properly [7].

Cartoon sample collection is an offline step, in which a number of different face images are painted by cartoonists and further decomposed into components. The same component-based feature extraction, except for the hair, has been applied to all the sample images in order to extract contours of their components. Using these original contours and those of the cartoons, we can match located components of the input face image to the fitting cartoon templates.

Finally, using the spatial information of extracted components from the input face, selected components are composited and deformed to match the input face. The output can either be rendered as a sketch or a cartoon face. Cartoon representation features such as shadows, highlights and supplementary curves for each component are captured from the corresponding template. Sketch output lacks cartoon features and is just composited with contours and supplementary curves of matched cartoon components. The matching system is based solely on the comparison of contours between extracted components and template components. Supplementary curves of facial components are ignored in template selection phase. However, in cartoon rendering, they are copied from the matched template to enrich the cartoon or sketch output.
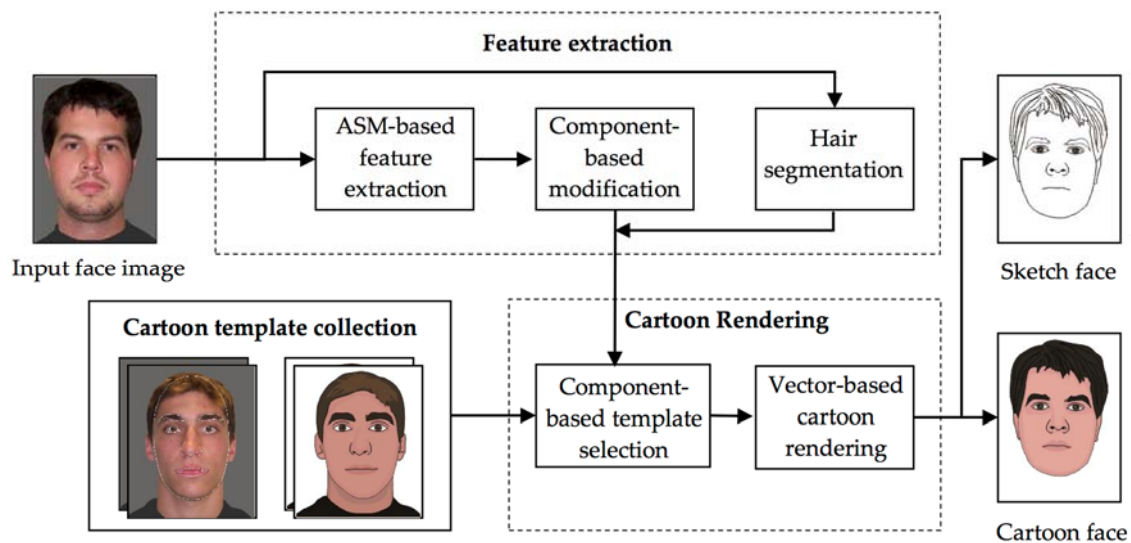
Figure 1. The framework of the proposed cartoon face system.

2.1. Feature Extraction

To address the shape-preserving necessity of cartoon face generation, most of the present systems employed either AAM [5,7,9,17] or a variation of ASM [6,15,18,19]. Generally, ASM shows better robustness to different conditions, as it searches through local regions oppose to the AAM, which uses models of holistic appearance.

An ASM-based facial feature extraction fails to capture the accurate contours required for artistic representation of all facial parts. It is well suited to be fitted at global level, as there is a trained constraint, which keeps the proportion of the facial components in ASM. Besides, the number of landmarks is not sufficient for some components. Thus, a local modification is proposed to enhance the quality of feature extraction by fine-tuning the contours of the captured components. Unlike current cartoon systems, which ask users to choose from hair templates [8] or determine the color distribution of the hair region using two brushes provided for the users [5], an automatic method for hair extraction is investigated to remove the requirement for user interaction and to enrich the appearance of the cartoon face by adding a cartoon-like hair. Since hair extraction is not a main focus in our proposed system, we consider a simple hair segmentation approach with some predefined conditions such as uniform background.

2.1.1. An Extended Active Shape Model

We utilize a modified ASM system named STASM, which is introduced by Milborrow et al. [20]. They proposed 2D profile areas around the landmarks. This modification is also considered by [14,21] to boost the accuracy of the ASM. For the early initialization of the ASM, a global face detector is employed by STASM to specify the initial search area. Starting from the mean shape adjusted by size and position to the detected face in the input image, STASM searches for the landmarks. As basic ASM can suffer from sensitivity to local extremum, good initial approximation improves the searching result. Gao et al. [14], locate irises and use them to calculate a rough starting approximation. However, STASM [20], which is employed in our system, uses the result of 1D profile ASM as a starting position for the searching with 2D profiles.

This global ASM-based feature locating, like STASM, cannot guarantee the accurate contour of all facial components at the same time. Hence, we extend the feature extraction stage with some optimization at component level using current shapes and spatial parameters of the facial elements that are discussed in the next section.

2.1.2. Component-Based Modification

As shown in Figure 2a,b, the facial contours are better fitted using our local modification technique. ASM, generally, captures contours for face, nose and lips parts more robustly than those for eyes and eyebrows. On the other hand, to keep the originality of appearance in cartoon level, we need to extract these features more precisely. Therefore, eyes and eyebrows are relocated specifically based on their starting regions. Additionally, Hermit interpolation and Bezier curve fitting are applied for the remaining components to satisfy the continuity and shape-preserving of the overall output.
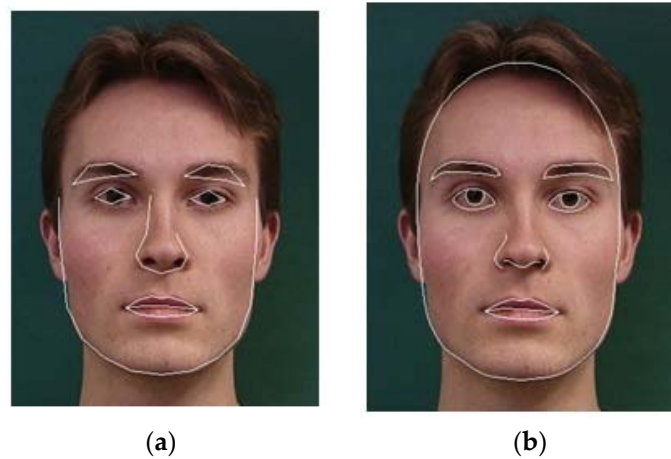

(**a**)　　　　　　　　　　　　　　　　　　　(**b**)

Figure 2. Active shape model (ASM) versus our contour modification result. (**a**) Extracted contours using the ASM; (**b**) Modified contours by our approach.

Eye Extraction

Many of the developed cartoon producers, caricature generators and sketching systems have considered the eye as a major factor for addressing likeness [22]. Generally, they do not investigate local approaches to optimize the extraction of eye. Although Gao et al. [14] indicate the high accuracy needed for these components, they do not modify the ASM-located eye contour. They directly represents the eyes by piecewise cubic Hermite interpolation.

Using a similar approach proposed by [23], and considering the feature points located by ASM, we can effectively extract irises and corners of the two eyes. First, the iris locating limits to fitting a circle using exhaustive search with a suitable window size around the initial region of the eyes. Then, the position of the corners are optimized by minimizing the energy function introduced in [24] (Equation (1)). For each candidate corner point, optimization is based on a triangle constructed by this point and the two ASM points on the top and the bottom of the eye. The triangle constructed by the best corner point is supposed to have the biggest white area. We extended the formula by considering a suitable weight for the initial ASM points. The energy function is as follows:

$$\frac{W_S}{A_{\text{triangle}}} \sum_{(u,v) \in A_{\text{triangle}}} [255 - S(x,y)] + W_C \times E(x,y) + W_D \times I(x,y) \tag{1}$$

where A_{triangle} is the area of the triangle, $S(x,y)$ is the color saturation, $E(x,y)$ is the edge intensity and $I(x,y)$ is the illumination value at pixel (x,y). Finally, two cubic Bezier curves are fitted to the upper and lower eyelids. The control points are calculated using the new corner points and former feature points on the top and bottom of the eye. Due to the lack of strong edges at the lower part of the eye, the position of the lower point is optimized first. This task is done by voting for a point that expresses stronger edge intensity and assigns higher saturation to the triangle constructed by this point and the two corners of the eye.

Eyebrow Extraction

Sadrô et al. [25] explored the role of eyebrow in the identification of faces. Their experimental results indicate the importance of the eyebrow as a facial characteristic. Despite the major impact of the eyebrow in face recognition, few studies have explored this [16,23,26]. Generally, current cartoon producers directly use contours extracted by ASM for the eyebrow. However, due to the wide range of eyebrows, in terms of color, thickness and compression of their hair, deformable models are not well suited for our purpose [23]. Some prior endeavors in this area are the works done by [23,26]. They used k-means clustering in a 5-patch profile of eyebrow. However, the final contour was not as smooth as required for our system.

With a big enough search area consisting the initial eyebrow, excluding the previously located eye region, and applying a Gaussian blur filter to the search region, we can effectively extract the strong edges at the upper part of the eyebrow. Binary morphological operations output a continuous edge, which is the longest one. Due to the presence of the shadow of the forehead in the lower side of eyebrow, we can extract the lower edge using the shadow removal method expressed in [23]. Using this method, a strong edge extracted pixel by pixel starting from an ASM point at the inner corner of the eyebrow. Finally, using the Hermit interpolation, the acceptable border of the eyebrow is obtained.

Other Improvements

Because of the impact of the hair, feature points of the upper face contour are omitted in facial feature locating methods. We calculate the approximate full contour of the face by adding a cubic Bezier curve at the upper part of the face component according to the starting and ending points of the ASM face landmarks. The completing part of face contour is proportional to the ratio of the face, which is calculated using ASM-based contour of the face. The complete contour of the face is beneficial for template matching process.

With overall analysis of the cartoon samples painted by our cartoonist, we considered three major curves for the representation of the nose. Therefore, using the ASM landmarks for the nose, three cubic Bezier curves are calculated to represent the shape of nose.

The shape of the lips, chiefly, is more sensitive to the emotion of a character, especially in cartoon representation. Predesigned cartoons of the lips can be used for diverse faces without major negative impact in likeness. Hence, we leave the contour of the lips untouched.

2.1.3. Hair Segmentation

Due to the wide range of variation in hair structure, model-based approaches are not successful in hair extraction. Moreover, presence of strong shadows and highlights in the hair region is a key problem for edge detection methods. In our proposed approach, the hair component is extracted using HSV color segmentation. HSV color space separates components for color, hue, and illumination (value) compared to RGB color space [27]. Hence, this color space is better for hair segmentation.

The initial region of the hair is extracted after skin detection and background segmentation. Using the extracted facial components and sampling color from the skin area, we can do a color segmentation to extract the skin region. With the same approach, we can separate the background. Considering Figure 3a, the initial region of the hair, Figure 3b is extracted after excluding skin and background. Now, an early color of hair is calculated from the region. With this early color of hair, regions extracted by HSV color segmentation are shown in Figure 3c. As can be seen in Figure 3d, binary morphology is employed to output a continuous hair region. The general location of the hair is taken into account to ignore other isolated areas. In the case of complex texture, we can utilize graph cut [28] technique for segmentation. Meng et al. [15] and Min et al. [7] used similar approaches for hair segmentation.

(a)   (b)   (c)   (d)

Figure 3. Hair segmentation. (**a**) Input face image; (**b**) Subtracting the detected skin and background region; (**c**) HSV segmentation with the early hair color calculated using binary image of (**b**); (**d**) Final hair region after binary morphological operations, considering the probable location of hair.

2.2. Cartoon Template Collection

The crucial issue of a templates-based approach in cartoon generation systems is cartoon collection, since it is very time-consuming and demands careful attention. As our template selection approach is based on feature points matching rather than pixel-based comparison, we used a well-known vector drawing software, Adobe Illustrator, to collect hand-drawn cartoon faces. In another part of our cartoon template collection, we extract facial components of the training set to be used further in the template matching section. The whole process of this phase is done offline.

2.2.1. Hand-Drawn Cartoon Faces

We asked a cartoonist to illustrate cartoon faces for different face images in predefined layers specified for each component. This components are hair, left eyebrow, right eyebrow, left eye, right eye, nose, lips, left and right ears, neck and shirt. For the matching process, we need a rendering contour for each facial component of the cartoons, so our artist illustrated a contour for each of the mentioned components. Moreover, the cartoon medium requires some specific features such as shadows, highlights and some sketches around each facial part. For example, unlike current cartoon systems, our cartoon faces have eyelashes for females. These features are drawn in sub-layers of the components. When a cartoon component from the template collection is selected, its cartoon properties such as shadows and additional curves are composited as well.

Shapes and curves are represented by different paths which contain anchor points with their two control points. This representation allows us to construct sequences of cubic Bezier curves for the contours. The whole properties of the components are collected automatically, i.e., spatial properties and rendering rules. The proposed approach is especially effective and convenient when we want to change the artistic style. Compared to the cartoon collection system proposed by [8], our approach is simple and also it can be done using many available software.

2.2.2. Component Extraction of the Sample Faces

As hand-drawn cartoon faces can differ from the corresponding real faces, direct comparison between cartoon components in the template set and the extracted components of the input face, cannot guarantee a proper selection of cartoon templates. Hence, we need to do one more comparison in the template selection procedure: a comparison between the extracted components of the input face and the extracted components of the template faces. More information about this matching procedure can be found in Section 2.3.1.

In this offline phase, facial components of the training images are extracted using our proposed feature extraction method. These components involve eyes, eyebrows, nose, mouth, and face contour. The matching procedure of the hair component is based on direct comparison between the extracted hair of the input face and the cartoon templates of the hair in the template collection.

Finally, in our cartoon template collection, there are some pairs of facial components in which we collected the original and its corresponding cartoon component, so each type of cartoon template can be compared and selected independently.

2.3. Cartoon Rendering

Although the shape of the extracted facial components is sufficient to preserve the likeness, they cannot bring the aesthetic aspect of the cartoon faces and we need to make use of suitable cartoon shapes for each component. For example, Figure 4 illustrates the general process for the cartoon rendering of eyes. Figure 4a illustrates the result of the component extraction phase. As shown in Figure 4b, in this phase corresponding cartoon components for the extracted eye are chosen from the pre-designed cartoon templates. Then, as shown in Figure 4c,d, the selected cartoon component is deformed according to the input face and further rendered based on the rendering rules of the component.
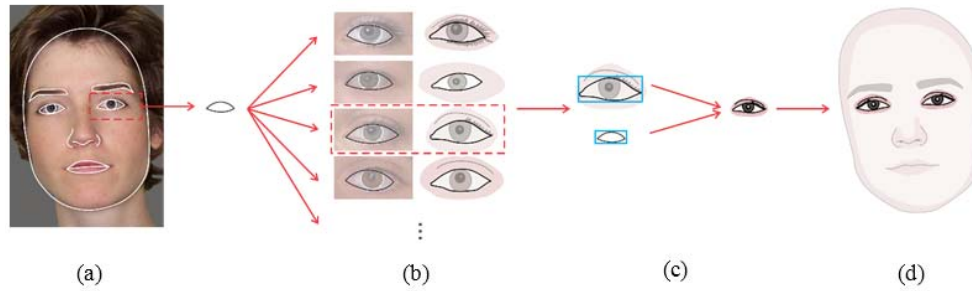

Figure 4. Template selection for eye and its cartoon representation. (**a**) Facial components of the input face image are extracted and the eye component is selected; (**b**) After normalization, the eye component is compared with both cartoon (**right**) and real contours (**left**) of all cartoon templates to select the closest template for the eye (**c**) The selected template is deformed according to the extracted component; (**d**) The cartoon template is adjusted in the final cartoon face.

2.3.1. Component-Based Template Selection

To match a proper cartoon template to each extracted component we need to measure a shape distance for each instance in the template collection set. Current cartoon systems use a fixed number of corresponding feature points in sample cartoon faces and the input face image. This means that specified number of corresponding points, which are extracted using ASM, should exist in both cartoon and input face.

However, as we applied a modification stage and extracted a more accurate contour for the components, there is no correspondence for the feature points and the number of feature points are not constant, so the components should be compared based on their contours. In the template collection phase, each component is represented as a sequence of Bezier curves. The same approach is applied to every component of the input face image after the extraction process. We employ Hausdorff distance as a shape distance measure since it requires neither one-to-one correspondence of feature points nor the same number of points in each point set [29]. The conventional Hausdorff distance is shown in Equation (2).

$$H_{a,b} = \max\left(h\left(P_a, P_b\right), h\left(P_b, P_a\right)\right), \text{where } h\left(P_a, P_b\right) = \max_{i \in P_a} \min_{j \in P_b} d\left(i, j\right) \qquad (2)$$

Point sets P_a and P_b belong to the components a and b. Points i and j are instances of point sets P_a and P_b, respectively. Consider $d\left(i, j\right)$ as a distance between two points, for example, Euclidian distance.

This distance is very sensitive to noise. Sim et al. [30] proposed a robust Hausdorff distance measure suitable for object matching in noisy images. However, our feature extraction uses ASM and also further local modification to extract continuous contours for each component (Figure 2b). Hence, the extracted components have less noises than images and the Hausdorff distance proposed by Sim et al. is not suitable for our purpose. We can make the Hausdorff distance more robust for our component matching using averaging the minimum distances for all points, despite the simple maximization over the minimum distances in the conventional one:

$$h\left(P_a, P_b\right) \; = \; \frac{1}{|P_a|} \sum_{i \in P_a} \min_{j \in P_b} d\left(i, \, j\right) \tag{3}$$

Moreover, Hausdorff distance is neither scale nor rotation and translation invariant. Before shape matching step, all templates should be normalized, i.e., rotated, scaled and translated if needed.

The amount of rotation is calculated by the angle of the line linking the two eyes that were extracted earlier.

When components are extracted, each component is compared with its corresponding templates in training set. As shown in Figure 4b, there are two comparisons for each instance in the template collection set: comparison with the cartoon component and comparison with the original component. The final distance of a component in input face from its correspondence in an instance of the template set is calculated using Equation (3). Let A be a facial component in input face image. Let S_F^i and S_C^i be the original component and the cartoon component in sample i of the template set, respectively.

$$D_{A,S^i} \; = \; w_1 H_{A,S_F^i} \; + \; w_2 H_{A,S_C^i} \tag{4}$$

where D_{A,S^i} is the final distance of component A from its corresponding component S^i in the template set. H_{A,S_F^i} is the Hausdorff distance of component A from component S_F^i and H_{A,S_C^i} is the Hausdorff distance of component A from component S_C^i. We assign weights for both cartoon shape and the extracted shape of samples. The weights w_1 and w_2 are determined experimentally with respect to each component.

Finally, the template giving the smallest error value is selected and further deformed and fitted to the input face according to result of feature extraction phase.

Due to the wide range of hair structure, we use shape context [31] for hair matching. After the contour of the hair component is extracted, we can select the closest hair component in our template collection using shape context distance. A thin plate spline model is used to warp the contours of the matched hair template to the contour of the extracted hair of input face image. A similar approach for hair matching has been used by Min et al. [7].

2.3.2. Vector-Based Cartoon Representation

Matched cartoon components should be adjusted to the input face and arranged together to have the final cartoon face rendered. Location, size and rotation of the original components are available using the information gained from the feature extraction stage. Additionally, rendering rules such as geometrical shape, supplementary curves and shadow regions of the final components are obtained by cartoon templates. For example, cartoon eye component is matched based on its contour, but the shadow region supplemented in the matched cartoon template is further added to the final cartoon face output (Figure 4c). Adjustment of color is done by searching through a color table which maps skin color ranges to some constant skin colors used in cartoon templates.

After normalization of the cartoon templates and the original components based on the rotation of the input face, cartoon templates are resized by their width and height considering the width and height of the corresponding original components (Figure 4c). Deformed cartoon templates are placed according to the spatial information of the components (Figure 4d). Components are accompanied by their add-ons such as neck and collar for face contour component and eyelash for eye component. These add-ons are deformed and rendered based on the basic shape of their components.

In our cartoon system, shapes are represented by Bezier curves. These shapes are rendered layer by layer with their color and stroke information. The final cartoon face can be generated in vector graphics, which is a suitable format for animation. Furthermore, cartoons are in vector format, the benefit of which is that they are pretty easy to modify in the shape, so exaggerations can be easily applied to create caricature like images that are out of scope of this paper.

3. Results

To prepare cartoon templates and evaluate the results of our system, our experiment is conducted based on frontal faces of different genders, ages and races. 30 images are collected from dataset of adult facial stimuli [32] and 30 images are obtained from an AR dataset [33] and IMM dataset [34]. Although we collect cartoon templates of all 60 face images, we consider 30 images as our training set and the rest for the test phase. Here are conditions considered in our cartoon collection set:

- Faces are frontal with no significant pose variance.
- Face images are with neutral expression.
- Training images are 640 × 480.
- Cartoon faces are drawn in a specific artistic style in Adobe Illustrator with open layers.
- Each pair of cartoon and original faces match perfectly.

Some of the cartoon faces generated using our approach are shown in Figure 5. Regarding the cartoon hair in Figure 5, it should be noted that cartoon hair rendering is a separate part of our system, and because of the wide range of hair styles, we incorporated all hand-drawn hair for hair cartoon selection. Since the matched hair templates already have the correct shape then the TPS warp for hair adjustment is essentially the identity mapping, which is why the artist-generated hair is identical to the automatically generated hair.

Figure 5. Examples of our cartoon face generator vs. hand-drawn cartoons. Images on the top row are the input face images and those on the bottom row are the cartoon faces generated by our cartoon system. The cartoons on the middle row are drawn by artist from the input face images.

To verify our system, we conducted two user studies and asked 94 participants of different genders and ages with various backgrounds to answer the questions. The parameters are considered constant among the evaluation process.

In one experiment, our aim is to verify the diversity of the cartoon faces generated by our system. To evaluate this, we consider 12 questions in which a real face image is depicted along with five cartoon faces. These cartoon faces are generated by five different faces using our proposed system and only one cartoon face belongs to the real face image in each question.

Participants choose the most similar cartoon face to the real face in each question based on their point of view. If the diversity of our system is high, users can properly select the right cartoon face. In order to make our verification in this experiment more robust, we omit the hair component for all cartoon faces to evaluate the production diversity even with the lack of hair as a prominent feature in identification. Moreover, we took five cartoon faces to increase the difficulty of the experiment, despite the two cartoon faces used in [19]. In Figure 6, the frequency of the number of correct answers among different users is illustrated. Around half of the participants correctly recognized the right cartoon face in more than eight questions (out of 12). On average, each user correctly answered seven questions. This shows that, considering the difficulty of this experiment, in more than half of the questions, the right cartoon face was recognizable and the diversity of our proposed cartoon generator is acceptable.
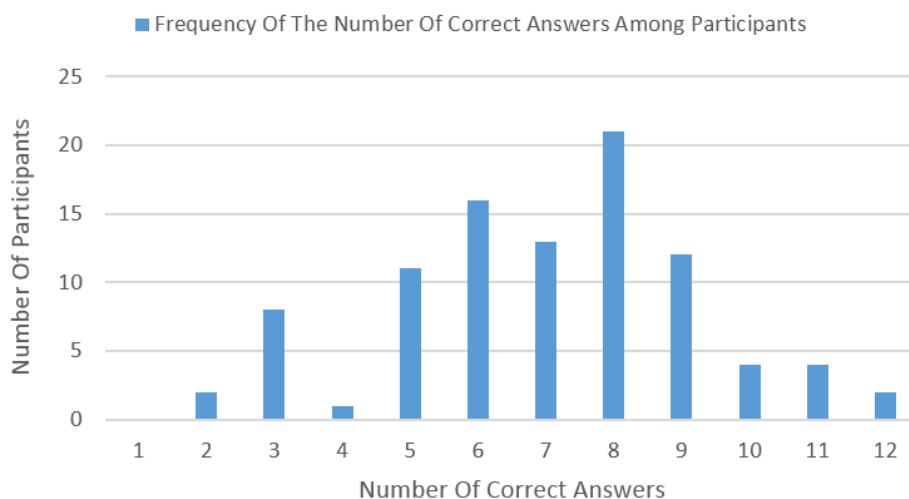

Figure 6. Frequency of the number of correct answers among 94 participants in Experiment 1. There are 12 questions in this experiment.

The other experiment aims to assess the likeness of the generated cartoon faces to their corresponding real faces. This test also consists of 12 questions. In each question, a real face image is shown along with its cartoon face produced using our system and involves five answers with values one to five. In each question, a rating of five means the generated cartoon face matches perfectly the real facial image. In this experiment, cartoon faces are depicted completely; all cartoon features, including the hair component, are illustrated. In this user study, the cartoon faces are evaluated generally from users' point of view. As each user has his or her opinion about the ideal level of likeness for a cartoon face, we put two hand-drawn cartoon faces as pivot cases among these 12 questions in order to balance the final ratings. In fact, these two questions are supposed to have the highest level of rating (close to five). In practice, experimental results show that the average ratings of the two standard questions are close to 4.5 (questions 2 and 8 in Figure 7). Excluding the two pivot cases in this experiment, the average rating of the users is greater than three. It means that the likeness of our generated cartoon faces is acceptable. In Figure 5, the second cartoon face from the right is assigned the highest rate (around four) among the remaining 10 questions (apart from the two standard cartoon faces).
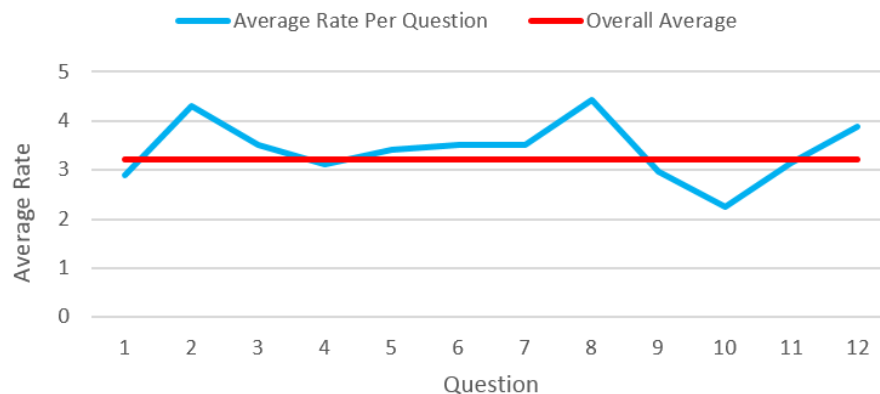
Figure 7. The average rating of each question in Experiment 2. Excluding the two pivot questions with the highest marks (question number 2 and 8), the overall average rating of 94 participants in this experiments is above 3.

Finally, using Hausdorff distance as a similarity measure, we objectively evaluate the error of extracted components as well as final cartoon components against hand-drawn cartoons. In Figure 8, the likeness error of lips, eyes, face and nose components are illustrated. We use hand-drawn cartoons to calculate the likeness error for the extracted components using our approach. Thirty hand-drawn cartoon faces are used in this evaluation. It shows the limitation of our system in extracting the nose component. However, other components have lower error, with eyebrows having the least error. Figure 9 shows the likeness error of the final generated cartoon components against hand-drawn cartoons. A reduction of the likeness error for eye component after template selection is justifiable. As discussed in Section 2.1.2, the final extracted eye comprises two cubic Bezier curves as illustrated in Figure 4a. Considering the differences around the eye corners between this contour for eyes and the corresponding hand-drawn ones (Figure 4b), it is clear that the matched eye templates will result in more similar eyes to hand-drawn eyes. As a result of this, we can see that in Figure 9, the eyes are improved. No similar situation happens for eyebrows because we do not consider predefined curves for them at the feature extraction stage. With further observation among cartoon faces whose ratings are low, we can admit the limitation of our proposed system in imitating the nose.
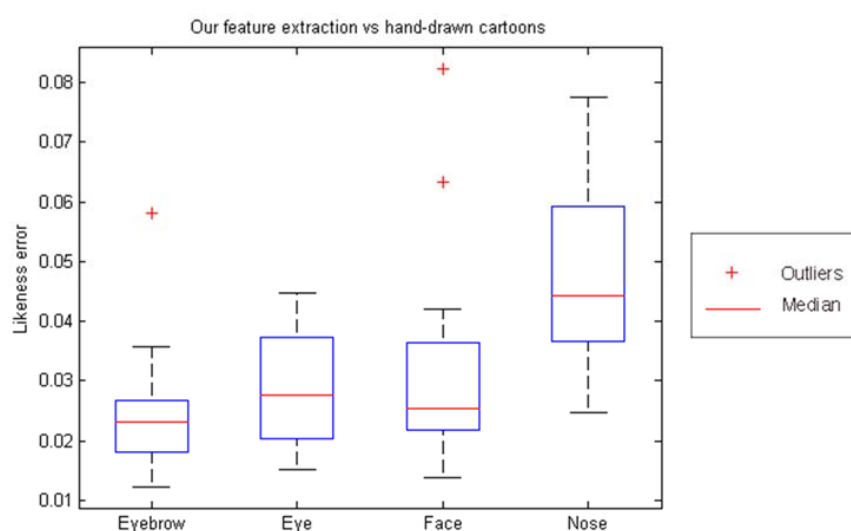

Figure 8. Likeness error of facial components based on the comparison of extracted contours and hand-drawn components. Notice the success of our system in extracting eyebrows using ASM and our local modifications.
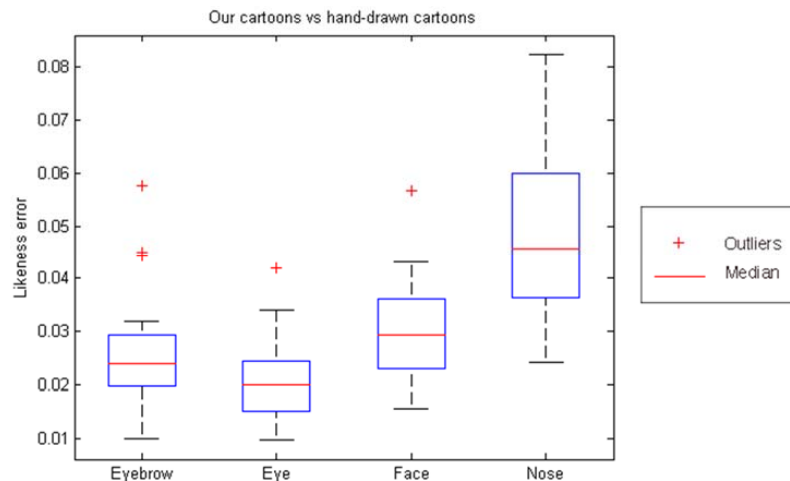
Figure 9. Likeness error of facial components based on the comparison of our generated cartoons and hand-drawn components. As can be seen, after cartoon template selection for the eye component, the likeness error has been reduced.

4. Discussion

Our proposed system is capable of automatically generating life-like cartoon faces vectorically from an input face image. The system consists of offline and online phases. In the offline stage, hand-drawn cartoon templates are collected and further decomposed into facial components. Components of the original faces in our collection set are extracted as well. We combine global and local approaches for feature extraction. In the online phase, facial components of the input face are extracted using a component-wise approach. The hair component is also extracted automatically. Then, the closest templates of each component are selected and deformed to be fit to the input face. Unlike current cartoon producers, for template matching we employed contour comparison to both cartoon contours and extracted shapes of sample faces.

Further research can be classified into two groups. Group one is related to the process of facial feature extraction and template matching. Figure 10 shows a failure case with respect to the feature extraction phase where a faulty extraction of face contour in ASM results in inappropriate cartoon face generation. Extraction of high level features such as the upper curve for the eye component and nostrils can be effective in matching procedure. These features lead to the selection of more suitable cartoon templates for facial components of the input face. In another group, future work can bring about better representations of final cartoons. Similar to the approach we conducted for cartoon template collection, we can offer the final cartoon face in Adobe Illustrator layer by layer, so users can modify the final result interactively using powerful graphical tools provided in this software, such as different brushes and filters.
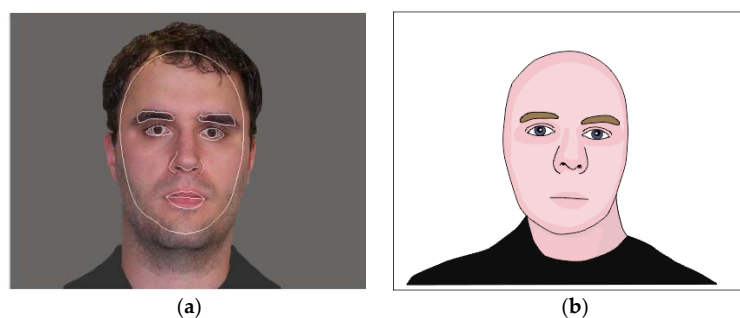

(**a**)   (**b**)

Figure 10. An example of the limitation of our system. (**a**) An input face image with the extracted contours using our approach overlaid on face. Notice the failure of ASM in fitting face component; (**b**) The synthesized cartoon face using our system which fails to meet the likeness requirement.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tinwell, A.; Grimshaw, M.; Nabi, D.A.; Williams, A. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Comput. Hum. Behav.* **2011**, *27*, 741–749. [CrossRef]
2. Hanson, D. Exploring the Aesthetic Range for Humanoid Robots. In Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, Vancouver, BC, Canada, 26 July 2006; pp. 39–42.
3. Koshimizu, H.; Tominaga, M.; Fujiwara, T.; Murakami, K. On KANSEI Facial Image Processing for Computerized Facial Caricaturing System PICASSO. In Proceedings of the 1999 IEEE International Conference on Systems, Man, and Cybernetics, EEE SMC'99 Conference, Tokyo, Japan, 12–15 October 1999; pp. 294–299.
4. Li, Y.; Kobatake, H. Extraction of Facial Sketch Image Based on Morphological Processing. In Proceedings of the International Conference on Image Processing, Santa Barbara, CA, USA, 26–29 October 1997; pp. 316–319.
5. Chen, H.; Zheng, N.-N.; Liang, L.; Li, Y.; Xu, Y.-Q.; Shum, H.-Y. PicToon: A Personalized Image-Based Cartoon System. In Proceedings of the tenth ACM international conference on Multimedia, Juan les Pins, France, 1–6 December 2002; pp. 171–178.
6. Che, J.; Tao, J.; Wang, X.; Mu, K.; Li, H. Feature-Based Multi-Style Cartoon System. In Proceedings of the 2010 International Conference on Audio Language and Image Processing (ICALIP), Shanghai, China, 23–25 November 2010; pp. 1012–1016.
7. Min, F.; Suo, J.-L.; Zhu, S.-C.; Sang, N. An Automatic Portrait System Based on and-or Graph Representation. In Proceedings of the Energy Minimization Methods in Computer Vision and Pattern Recognition, Ezhou, Hubei, China, 27–29 August 2007; pp. 184–197.
8. Liu, Y.; Su, Y.; Shao, Y.; Wu, Z.; Yang, Y. A Face Cartoon Producer for Digital Content Service. In *Mobile Multimedia Processing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 188–202.
9. Liu, S.; Li, H.; Xu, L. Face Cartoon Synthesis Based on the Active Appearance Model. In Proceedings of the 2012 IEEE 12th International Conference on Computer and Information Technology (CIT), Chengdu, Sichuan, China, 27–29 October 2012; pp. 793–797.
10. Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J. Active shape models-their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59. [CrossRef]
11. Cootes, T.F.; Edwards, G.J.; Taylor, C.J. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 681–685. [CrossRef]
12. Chen, H.; Liu, Z.; Rose, C.; Xu, Y.; Shum, H.-Y.; Salesin, D. Example-Based Composite Sketching of Human Portraits. In Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering, Annecy, France, 7–9 June 2004; pp. 95–153.
13. Zhang, Y.; Dong, W.; Deussen, O.; Huang, F.; Li, K.; Hu, B.-G. Data-Driven Face Cartoon Stylization. In *SIGGRAPH Asia Technical Briefs*; ACM: New york, NY, USA, 2014; pp. 14:1–14:4.
14. Gao, W.; Mo, R.; Wei, L.; Zhu, Y.; Peng, Z.; Zhang, Y. Template-Based Portrait Caricature Generation with Facial Components Analysis. In Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2009, Shanghai, China, 20–22 November 2009; pp. 219–223.
15. Meng, M.; Zhao, M.; Zhu, S.-C. Artistic Paper-Cut of Human Portraits. In Proceedings of the International Conference on Multimedia, Singapore, 19–23 July 2010; pp. 931–934.
16. Ding, L.; Martinez, A.M. Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2022–2038. [CrossRef] [PubMed]
17. Xu, Z.; Chen, H.; Zhu, S.-C.; Luo, J. A hierarchical compositional model for face representation and sketching. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 955–969. [PubMed]
18. Chen, H.; Liang, L.; Xu, Y.-Q.; Shum, H.-Y.; Zheng, N.-N. Example-based automatic portraiture. *Chin. J. Comput. Chin. Ed.* **2003**, *26*, 147–152.
19. Rhee, C.-H.; Lee, C.H. Cartoon-like avatar generation using facial component matching. *Int. J. Multimed. Ubiquitous Eng.* **2013**, *8*, 69–78.

20. Milborrow, S.; Nicolls, F. Locating Facial Features with an Extended Active Shape Model. In *Computer Vision–ECCV 2008*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 504–513.

21. Wang, K. Implementation of Face Cartoon Maker System Based on Android. In Proceedings of the 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP), Beijing, China, 9–11 June 2013; pp. 193–198.

22. Sadimon, B.; Sunar, M.S.; Haron, H. A review of facial caricature generator. *J. Comput.* **2011**, *3*, 6219–6234.

23. Kuo, P.; Hillman, P.; Hannah, J. Improved Facial Feature Extraction for Model-Based Multimedia. In Proceedings of the 2nd IEEE European Conference on Visual Media Production, London, UK, 30 November–1 December 2005; pp. 137–146.

24. Kuo, P.; Hannah, J. An Improved Eye Feature Extraction Algorithm Based on Deformable Templates. In Proceedings of the IEEE International Conference on Image Processing, ICIP 2005, Genova, Italy, 11–14 September 2005; p. II-1206-9.

25. Sadrô, J.; Jarudi, I.; Sinhaô, P. The role of eyebrows in face recognition. *Perception* **2003**, *32*, 285–293. [CrossRef]

26. Chen, Q.; Cham, W.-K.; Lee, K.-K. Extracting eyebrow contour and chin contour for face recognition. *Pattern Recognit.* **2007**, *40*, 2292–2300. [CrossRef]

27. Cheng, H.-D.; Jiang, X.; Sun, Y.; Wang, J. Color image segmentation: Advances and prospects. *Pattern Recognit.* **2001**, *34*, 2259–2281. [CrossRef]

28. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [CrossRef]

29. Veltkamp, R.C. Shape Matching: Similarity Measures and Algorithms. In Proceedings of the SMI 2001 International Conference on Shape Modeling and Applications, Genova, Italy, 7–11 May 2001; pp. 188–197.

30. Sim, D.-G.; Kwon, O.-K.; Park, R.-H. Object matching algorithms using robust Hausdorff distance measures. *IEEE Trans. Image Process.* **1999**, *8*, 425–429. [PubMed]

31. Belongie, S.; Malik, J.; Puzicha, J. Shape Context: A New Descriptor for Shape Matching and Object Recognition. In Proceedings of the Conference on Neural Information Processing Systems, Denver, CO, USA, 20 June 2000; p. 3.

32. Minear, M.; Park, D.C. A lifespan database of adult facial stimuli. *Behav. Res. Methods Instruments Comput.* **2004**, *36*, 630–633. [CrossRef]

33. Martinez, A.M. *The AR Face Database*; CVC Technical Report; Computer Vision Center: Arcelona Area, Spain, 1998; Volume 24.

34. Nordstrøm, M.M.; Larsen, M.; Sierakowski, J.; Stegmann, M.B. *The IMM Face Database—An Annotated Dataset of 240 Face Images*; Technical University of Denmark: Kongens Lyngby, Denmark, 2004.