Calligraphy Character Synthesis from Small Sample Set

Kai Yu and Zhenming Yuan

School of Information Science and Engineering, Hangzhou Normal University, China
{yk,zmyuan}@hznu.edu.cn

Abstract.A novel approach to synthesize calligraphy characters is presented in this paper. Only a small set of calligraphy characters written by the specific calligrapher is needed. A robust polygon based radical and stroke extraction method is introduced, which can generate strokes and radicals preciselywith a few manually marked pixels. A new radical and stroke selection method called component selection algorithm is described, which can decide whether to use radicals or strokes and can find out the most suitable ones from the candidate radicals and strokes.After putting the radicals and strokes together and form the calligraphy character, the style difference among them must be minimized. In order to do this better, a new way to adjust the stroke widths is presented. A sample set containing only 30 selected calligraphy characters are used to synthesize new calligraphy characters. The results show that our approach works effectively.

Keywords: calligraphy character synthesis, component selection algorithm, small sample set.

1 Introduction

Chinese calligraphy is an important and inseparable part of human cultural heritages. Its delicate aesthetic effects are generally considered to be unique among all calligraphic arts.

The historical Chinese calligraphic works arevaluable parts of Chinese cultural heritage. In the China Academic Digital Associative Library (CADAL), lots of famous ancientcalligraphic works are digitized. However, for some ancient calligraphers, only tens of calligraphic characters are survived till now. Many characters which are needed in the tablet generation are missing. Therefore, a method to synthesize calligraphic characters in a specified style with a small set of sample calligraphic characters is a necessity in order to generate calligraphy tablets.

A novel approach is proposed in this paper to synthesize calligraphy characters in a specific style with only tens of sample calligraphy characters. The strokes and radicals are extracted with a new polygon based extraction method. A new radical and stroke comparison and selection method called component selection algorithm is presented to find out the most proper components. Finally, these components are adjusted and put together to form the new calligraphy character.

K. Tuamsuk et al. (Eds.): ICADL 2014, LNCS 8839, pp. 379–384, 2014.

2 Related Works

Lots of researches on Chinese calligraphy appeared in the last few years, including calligraphy character retrieval [1] and recognition [2], calligraphy synthesis[3-6], etc.

Xu [3] used a hairy brush model to synthesize calligraphy characters. A synthesis approach was proposed with a six-level hierarchical model, which could produce Chinese calligraphic characters in a variety of styles.Some years later, an intelligent system for Chinese calligraphy[6] was proposed which could derive parameters in an interactive way and judge the beauty of calligraphy.With these methods, calligraphy characters that seemed to be written by human beings could be produced, but the style cannot be decided.

A style consistency calligraphy synthesis system [4] was presented in 2009. It could synthesize calligraphy characters in a specific style. Xia [5] introduced an ontology based model to synthesize calligraphy characters. For these methods, a large number of sample calligraphy characters were needed. However, for many ancient famous calligraphers, only a few calligraphy characters preserved up to now. Thus, the existed method may not work effectively.

3 Calligraphy Character Synthesis from Small Sample Set

The whole system contains two parts. One is the preparation part and the other is the synthesis part.The first part is the preparation part. The strokes and radicals will be extracted from the sample characters and the style features will be collected, too. The second part is the synthesis part. The character to be synthesized will be extracted and the most suitable components, including radicals and strokes, will be selected. The components are then adjusted and the new character is formed.

3.1 Synthesis Preparation

In calligraphy characters, the structures are very complex. Traditional minimum bounding box can only find out separate radicals, as shown in the left of Figure 1, but cannot find out the radicals (the red part) shown in the right of Figure 1.

In most cases, the radicals are crossed over or even connected together. It is very difficult to extract automatically with a high precision. A polygon-based character extraction is introduced.

Fig. 1. Different radicals in calligraphy characters

Polygon Extraction

Polygon is composed by a certain number of connected lines, which can form a cycle. Thus, polygon can be represented as an ordered list of pixels.

When the turning pixels are pointed out, we only need to connect them one by one, and a polygon can be found. However, the user may click a pixel which is slice difference from the accurate point. It may cause a mistake in character extraction.

To make the polygon extraction more robust, the most proper pixel in the small neighborhood area of the specified pixel is found out.

If a small part of a line crosses a solid part of a stroke, one of the terminal points is probably in a wrong place. Thus, the terminal points need to be corrected. In most cases, only one terminal point must be moved.

Stroke or Radical Representation

The solid area in the polygon found above is just a component which can be a radical or a stroke. It is part of a calligraphy character. The positions and relative sizes of the components are very important in Chinese calligraphy characters. The same stroke in different part in a character will be very different, and what's more, the same stroke in the same part with different relative sizes may still be different in shape.

To represent the component, the position and relative size must be record. A simple way is finding the minimum bounding box of the component, and recording the left-top point and right-bottom point of the bounding box.

3.2 Calligraphy Character Synthesis

To synthesize a calligraphy character, the character in print style is used as the template. It is extracted into radicals and strokes firstly, and then the most proper radicals and strokes are selected for each component. Finally, they are put together and form the new calligraphy character.

The polygon based radical and stroke extraction method can also be used in component extraction for printed characters.

Components Selection

To synthesize a calligraphy character, choosing the proper components including strokes and radicals is one of the most important tasks.

To make the synthesis easier, we try to find out the proper radicals firstly. If the proper one cannot be found, we will find the strokes which form the radicals.The component selection algorithm is designed to touch the target. It contains three steps. The first is component comparison, which can find out the most similar component. The second is radical decision, which is deciding where to use radicals or strokes. The last is components global selection.

The most proper component is the component whose positions and relative sizes fit the character best. It is determined by many factors.

The calligraphy styles for different calligraphers are different, and for the same calligrapher, the style may change with increasing age. However, sometimes, different calligraphers may have similar styles because of the disciple.

Thus, a two-stage comparison is needed. In the first stage, the positions, relative sizes, calligraphers and the disciple is considered, and a candidate radical set is found out for each radical. In the second stage, the difference between radicals in the same character is considered, including the source of the radicals, the difference in stroke width, and the age when the calligrapher wrote the works.

When the most proper radical is found out, we must decide whether to use the found radical or extract the radical into strokes and found out the most proper strokes. When the first stage different $Diff_1$ is larger than a threshold, it is dropped and the extracted strokes are used.

In component comparison, the style difference among different radicals and strokes is not considered. When synthesizing a calligraphy character with different radicals and strokes, the style difference must be very small. An average style feature is calculated and considered as the style feature of the combined calligraphy character. Thus, we want to minimize $D = \Sigma\, Diff(x_i,y_i)$.

$Diff(x_i,y_j)$ is the difference between the style feature of component x and the average style feature of the combined calligraphy character. It is defined as follows:

$$Diff(x,y) = p \cdot DAuthor(x,y) + q \cdot DDisciple(x,y) + r \cdot DAge(x,y) + s \cdot DFeature(x,y) \quad (1)$$

where $DAuthor(x,y)$ is 0 when $x$and y are written by the same author, and it is 1 otherwise. $DDisciple(x,y)$ is between 0 and 1 which represent the relationship between the two authors. $DAge(x,y)$ is the age difference when the x and y are written by the same author, and its value is 100 if $DAuthor(x,y)$ is 1.

$DFeature(x,y)$ is the style features difference of the two calligraphy characters. It is a combination of the difference in many features, such as the real size of the characters, the average stroke widths of the radicals, the stroke densities of the characters and so on. Here the stroke density is one of the important factors. Usually, the character part with large stroke density will have thinner strokes. Thus, there's a negative relationship between the stroke widths and the stroke densities.

Component selection algorithm

The component selection algorithm is proposed with the above key steps. The input of the algorithm is the hierarchical components needed to synthesize a calligraphy character.The output is the selected components which can used to compose the calligraphy character directly. The steps are as follows:

1. For each component, find out the candidate ones with the components comparison method.
2. For each candidate components, the first stage comparison is operated and the different ones are removed.
3. For each component, if there's no candidate, extract it and redo step 1 for the extracted components.
4. Select most proper ones with the global component selection method.

Strokes and Radicals Combination and Character Generation

When the strokes and radicals are prepared, they are to be combined together and the calligraphy characters are formed. Firstly, the strokes and radicals are put on the proper positions with the proper sizes. Then the stroke widths are adjusted. Thus, the new calligraphy character is generated.

The original positions and sizes of the strokes and radicals chosen are usually slightly different from the ones in the characters to be synthesized.The chosen strokes will be deformed to fit their areas in the characters to be synthesized.

The stroke widths of character components from different calligraphy characters may be very different. This may be worse when the components are scaled in different rates. It can be done with erosion and dilation operations in mathematical morphology. However, for some special strokes, such as long *Heng*(the horizontal curve) and *Shu*(the vertical curve), it may cause unscaled deformation on the ends of the strokes. To avoid this situation, an improved method will be used on such strokes.

Such stroke will be zoomed in or out to get a suitable stroke width. Then the zoomed stroke will be departed in to several parts, and the parts in the middle will be strengthened to get a new stroke. The whole process is shown in Figure 2. The first image shows the original stroke, whose stroke width need to be increased.The second one shows the segmentation result, and the last shows the stroke width adjust result.

Fig. 2. The whole process of the improved stroke width adjustment

4 Experiments and Discussion

A small set of 30 calligraphy characters from *Zhenqing Yan*, a famous calligrapher in *Tang* Dynasty (about 1300 years ago), were used as the learning sets. These calligraphy characters contained all the five basic strokes and some common used composed strokes. It is used to synthesize some calligraphy characters and compared them with the ones *Yan* really written. The result is shown in Figure 3.

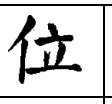
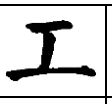
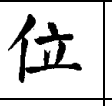
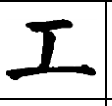
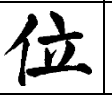
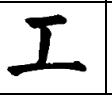
Synthesized characters with small sample sets				
Synthesized characterswith large sample sets				
Characters written by *Yan Zhenqing*				

Fig. 3. Comparison of the synthesized calligraphy character and the original ones

The synthesis result is a little different from the original ones because that the strokes and radicals in the calligraphy characters were not in the sample set. Thus, some other components are used to form the calligraphy character.However, these calligraphy characters are all in *Zhenqing Yan*'s style.

Figure 4 shows some more synthesis samples. The last three characters are in simplified Chinese which are impossible to be written by *Yan*.

Fig. 4. Some more synthesis results

5 Conclusion

In this paper, a novel approach is proposed to synthesize calligraphy characters in a specific style. Compared with the methods presented before, only a small collection of sample characters are needed. Experiments show that the proposed approach can synthesize style-consistent calligraphy characters.

References

1. Lu, W., Wu, J., Wei, B., Zhuang, Y.: Efficient Shape Matching for Chinese Calligraphic Character Retrieval. Journal of Zhejiang University - Science C 12(11), 873–884 (2011)
2. Yu, K., Wu, J., Zhuang, Y.: Skeleton-Based Recognition of Chinese Calligraphic Character Image. In: Huang, Y.-M.R., Xu, C., Cheng, K.-S., Yang, J.-F.K., Swamy, M.N.S., Li, S., Ding, J.-W. (eds.) PCM 2008. LNCS, vol. 5353, pp. 228–237. Springer, Heidelberg (2008)
3. Xu, S., Lau, F.C., Cheung, W.K., Pan, Y.: Automatic Generation of Artistic Chinese Calligraphy. IEEE Intelligent Systems 20(3), 32–39 (2005)
4. Yu, K., Wu, J., Zhuang, Y.: Style-Consistency Calligraphy Synthesis Systemin Digital Library. In: 9th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 145–152. ACM Press, New York (2009)
5. Xia, Y., Wu, J., Gao, P., et al.: Ontology-based model for Chinese Calligraphy Synthesis. Computer Graphics Forum 32(7), 11–20 (2013)
6. Xu, S., Jiang, H., Lau, F.C., Pan, Y.: An Intelligent system for Chinese Calligraphy. In: The 22nd Conference on Association for the Advancement of Artificial Intelligence, pp. 1578–1583. AAAI, New York (2007)
7. Lu, W., Zhuang, Y., Wu, J.: Discovering Calligraphy Style Relationships by Supervised Learning Weighted Random Walk Model. Multimedia Syst. 15(4), 221–242 (2009)