

Component-stroke Corresponding System for Chinese Characters of Different Styles

Yunlong Che, Xiaohui Liang, Pengfei Yue

State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University
Beijing, China

Author*: blizzak@163.com

Abstract—Stroke and shape of the Chinese character with different script has changed hugely among different style. To present the correspondence between Chinese characters component-strokes, this paper designs a component-stroke corresponding system for Chinese characters based on semantic knowledge. Firstly, we acquire the correspondence between components by the way of semi-automatic. Then we get the feature points from the intersection of skeleton and contour by skeleton algorithm, finally we finish the Chinese character component-stroke correspondence process. Our system can present a robust and fast component-strokes correspondence results.

Keywords—Chinese character; component-stroke; Feature points; correspondence; semantic information

I. INTRODUCTION

Character is the basic unit of Chinese writing, which uses the latest before the Shang Dynasty began. Development of the Ancient Chinese Characters, there is a long process of evolution, modern Chinese characters evolved from the ancient Chinese character based on evolved. Since the same characters in different times have different strokes form, coupled with extremely complex topology Chinese Characters, Chinese evolution rule base is large, clear description of the correspondence between the different strokes of Chinese characters become a serious problem.

This paper describes a correspondence algorithm of components and contour based on semantic. This component correspondence algorithm consists of two parts:

Correspondence between components and corresponding to the contour and stroke. Components corresponding algorithm has two parts, the first part in section III is based on automatic gravity position corresponding algorithm, then with the expert interactively adjust based on Semantics generate the correspondence between components of characters.

The stroke contour corresponds algorithm in section IV is the corresponding Chinese character stroke between contour points. Stroke corresponds algorithm consists of three parts: Component-stroke skeleton extraction algorithm, automatically corresponding algorithm and Chinese characters stroke feature points corresponding algorithm.

With the results of Chinese character component correspondence shown in section V, we get the

correspondence between different component-strokes. It will play important role in future studies.

II. CONTOUR OF CHINESE CHARACTER

A. Component-stroke Detection

The main object of this paper is contour component of Chinese characters, in order to measure properties of the character's contour, it must be described. According to the shape description based on the whole or partial elements. Contour description method is subdivided into global methodology and structured approach.

1) *The global method* take advantage of overall shape information, the smaller the amount of calculation, but only distinguish quite different shapes.

2) *Structured approach* needs to be provided more features, usually through a combination of primitive and therefore a relatively large amount of calculation, to distinguish more capacity.

The global method based on the contour is usually calculated based on the shape of a multi-dimensional feature vector boundary information, Converted the matching problem into pure numerical problems. Simpler methods are described in the perimeter area ratio, eccentricity (shape minor axis ratio), spindle direction, bending energy. These method are simple and robust, but less ability to distinguish, and can't be used for precise matching. The more ability of description methods are Fourier method, wavelet description method, measuring distance [1], shape context [2] [3], integral invariants [4], Heat Kernel [5].

The basic idea of structured approach based on contour is coding S as type of $S = s_1, s_2 \dots s_n$. And the s_i is the metadata of the contour. The difference between various methods lies in the different decomposition way, like polygonal approximation, curve decomposition, spline approximation method. Polygonal approximation is currently very active method. Lateck^[6] proposed an adaptive algorithm based on the gradient profile side length and angle relationship between the shapes of the contour of a polygon can be simplified description.

Felzenszwalb [7] proposed the concept of the shape-tree, the tree will use contour shape as primitive data, layer by layer construct contour tree by contour segmentation method. The

shape of the tree on the global information and local information are described, it can be used to form an exact match, but higher computational complexity.

B. Correspondence of Contour

Correspondence of contour can measure the similarity between different contours by some certain criteria [8]. In recent years, some study proposed the corresponding algorithm with dynamic programming [9] or graph cuts [10]. However, the current corresponding method usually cost much time and space, It can't meet the requirements of practical application, particularly in the case of mixed noise contour, the effect of existing methods is not good enough. The currently consideration of the conventional method only take use of geometric features.



Fig. 1. Chinese characters of different styles

Fig. 1 shows the evolution process of the character “采”. They are the same in Semantics, but have different topologies.

III. KNOWLEDGE-BASED COMPONENT-STROKE CORRESPONDENCE

Combined with Chinese characters formation details about the evolution of Chinese characters, Chinese Strokes follow the semantics of The corresponding Chinese characters, As used herein, Expert Character interaction was used with the corresponding Chinese character semantics. This method consists of two steps: The first is the automatic algorithm based on the corresponding part of the stroke center, The main purpose of this step of the process is to reduce the workload of human interaction, For the most part, the stroke of which corresponds to the relative positional relationship has not changed much. For this reason, first uses the initial coordinates corresponding the center of gravity-based, so it can find the corresponding original parts of Chinese characters. The second step is to adjust the semantics, Due to the automatic mapping algorithm can't guaranteed out of the correspondence between the semantically, it also needs to adjust the characters interact with experts, The basic procedure is shown in Fig 2:

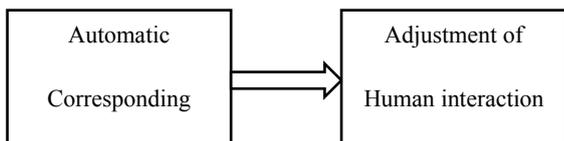


Fig. 2. Flow chart of Structure correspondence

A. Automatic mapping algorithm

Automatic mapping algorithm is based on the idea of greedy, Initial conditions algorithm is to find the center of gravity of the beginning and end of the two Chinese characters

components collection.as $S \{s_i, i = 1, 2, \dots, m\}$ and $T \{t_j, j = 1, 2, \dots, n\}$, The basic idea is compute the distance between the center of gravity of each calculation of stroke, as $Dist. \{dist. i, j\}$, ($dist. i, j$ means the distance from i center of gravity to j center of gravity), The minimum distance value is the set of matching matches, Then find the nearest match in turn, The matched result is the desired result.

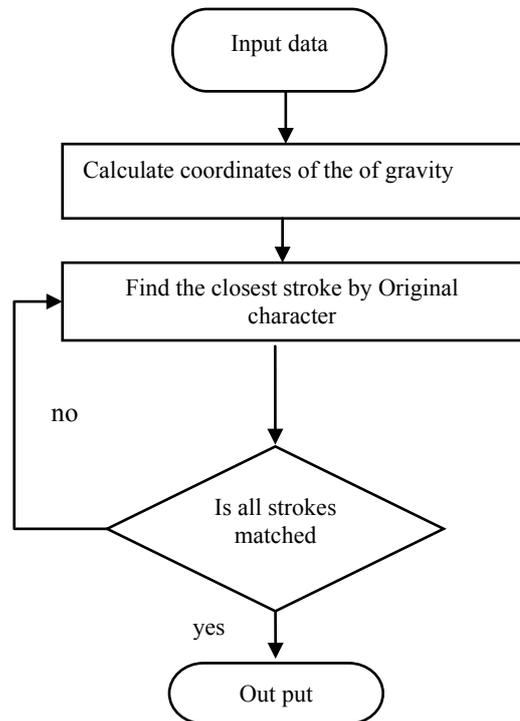


Fig. 3. Flow chart of Automatic mapping algorithm



Fig. 4. Small Seal script “幸” and Clerical script “幸”

B. Adjustment of human interaction

Human interaction is remedy to stroke corresponding, because automatic stroke extraction algorithm can't fully determine the corresponding components. The reason is the corresponding Chinese character strokes of Chinese characters is based on semantics rather than geometric features. However, we need characters expert interactive to change corresponding relationship, then obtain corresponding components.



Fig. 5. “心” before adjustment and “心”after adjustment

IV. STROKES CONTOUR CORRESPONDING ALGORITHM

After Determined the correspondence between strokes, it also need to determine the corresponding relationship between stroke contour points. Contour points corresponding is to determine the corresponding relationship between the strokes corresponding of each points. As we all know, Features characters can be represented by the skeleton information characters. Character contour information may be skeleton contour strokes of Chinese characters will be divided into two characters, so by finding the intersection of the strokes of Chinese characters skeleton and contour to these two points as the split characters contour feature points. Determining the correspondence between the characters of these two feature points can get the whole upper contour feature point correspondence between strokes.

Fig 6 shows the Strokes contour corresponding algorithm workflow.

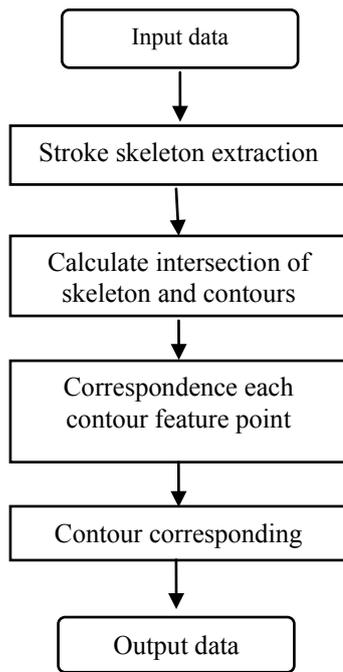


Fig. 6. Flow chart of Strokes contour corresponding algorithm

A. Stroke feature extraction algorithm

Stroke feature points is key component of corresponding contour. In this section, we describes the concepts of stroke point, and then introduce how to extract the feature point from stroke.

Correspondences between characters stroke contour also need to comply with the rules of writing Chinese characters stroke. One stroke have a beginning and an end, and the skeleton between the start and end points constitute the basic characteristics of stroke. In order to generate the corresponding relationship between the strokes of characters also require the corresponding character strokes of each contour points. The method here can find the intersection of characters skeleton and characters stroke contour firstly, then

the point obtained will recorded as feature of Chinese characters.

Skeleton information is foundation of extracted feature point from stroke. As the graphic character stroke topology is simple, easy to strike a skeleton strokes of Chinese characters. Feature point here refers to the characteristic point stroke starts and ends. The traditional extraction algorithm based on the time frame refined high complexity. Feature extraction algorithm used in this paper is based on the stroke contour of triangulation.

In this paper, stroke feature extraction algorithm step comprises the following steps:

- a) Sampling Some Polygons to simulate strokes approximate polygon contour.
- b) Polygon will be sampling as the Constrained Delaunay Triangulation triangle.
- c) For each subsequent Constrained Delaunay Triangulation triangle, find the adjacent triangle, connecting the center of gravity, to get a series of short line segments.
- d) Chinese characters can be connected segments by these short segments set. Find the longest line segment, it will approximate the skeleton as characters.
- e) Obtained the stroke contour intersection points, these points are the feature points of Chinese characters.

B. Feature points stroke correspondence

After obtaining feature points corresponding to the stroke, and the need for further strike correspondence between feature points. According to the contour and the feature point position, this article uses the minimum distance algorithm to solve the corresponding equation of feature point.

Fig.7 shown the two different strokes, the algorithm find the correspondence between four feature points A, B, C, D ,Here with a method based on shortest distance, put the two strokes of Chinese characters together, because the relationship between the four match points only two, so can calculate the distance between two points. Use the minimum distance between the two. Set it as the "closer" match.as shown in Fig 7. The point A should match C, and The point B match D.

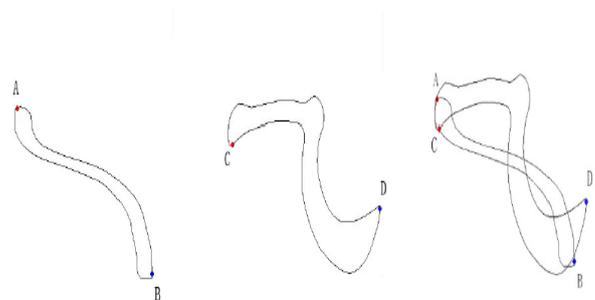


Fig. 7. Key point of char characters and stroke normalization

V. RESULTS

A. Component correspondence

Fig. 8 shows the results of component correspondence for Chinese character "心". The results will be used for Stroke skeleton extraction.

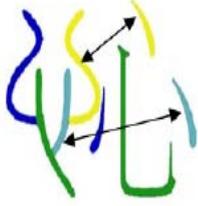


Fig. 8. Results of component correspondence

B. correspondence of strokes by feature points

As shown in Fig. 8, we give the component segmentation results for some Chinese characters. Then we get extraction skeleton from the stroke. The results shown in Fig. 9. And the left one is small seal script and the right one is clerical script.

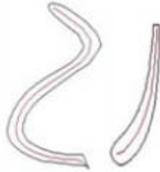


Fig. 9. Stroke skeleton extraction

Fig. 10 shows the results of Stroke skeleton extraction. The red feature point means the start point of the each strokes and the blue points are end points of them;

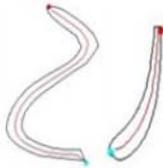


Fig. 10. Results of correspondence by feature points

VI. CONCLUSION

This article describes the steps to establish the correspondence between the different forms of the same font in the character strokes, Relationship between these is built on two levels, First of all corresponding between the components of characters, followed by stroke between components corresponding to the contour of the establishment. Stroke corresponding relationship building process is an interactive semi-automatic process, while the stroke corresponding the process is completely automatic process based on the skeleton information established.

Acknowledgment

This paper is supported by National Key Technology Research & Development Program of China (2014BAK18B01).

References

- [1] Asi Elad Elbaz and Ron Kimmel. On bending invariant signatures for surfaces[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2003, 25(10):1285–1295.
- [2] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2002, 24(4):509–522.
- [3] Natasha Gelfand, Niloy J. Mitra, Leonidas J. Guibas, and Helmut Pottmann. Robust global registration.[A]. In Symposium on Geometry Processing[C]. Eurographics Association,2005:197-206.
- [4] Siddharth Manay, Daniel Cremers, Byung-Woo Hong, Anthony J. Yezzi, and Stefano Soatto. Integral invariants for shape matching[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2006, 28(10):1602–1618.
- [5] Jian Sun, Maks Ovsjanikov, and Leonidas J. Guibas. A concise and provably informative multi-scale signature based on heat diffusion[J]. Comput. Graph. Forum, 2009, 28(5):1383–1392.
- [7] Longin Jan Latecki and Rolf Lakemper. Convexity rule for shape decomposition based on discrete contour evolution[J]. Computer Vision and Image Understanding, 1999,73(3):441–454.
- [8] Pedro F. Felzenszwalb and Joshua D. Schwartz. Hierarchical matching of deformable shapes.[A]. In CVPR[C]. 2007.
- [9] Michela Mortara and Spagnuolo. Similarity measures for blending polygonal shapes[J]. Computers & Graphics, 2001, 25:13–27.
- [10] Ligang Liu, Guopu Wang, Bo Zhang, Baining Guo, and Heung-Yeung Shum. Perceptually based approach for planar shape morphing[A]. In Pacific Conference on Computer Graphics and Applications[C]. 2004:111-120.
- [11] Clayton Scott and Robert D. Nowak. Robust contour matching via the order-preserving assignment problem[J]. IEEE Transactions on Image Processing, 2006, 15(7):1831–1838.