Complete Font Generation of Chinese Characters in Personal Handwriting Style

Jeng-Wei Lin
Dept. Information Management
Tunghai University
Taichung, Taiwan
jwlin@thu.edu.tw

Chian-Ya Hong, Ray-I Chang,
Dept. Engineering Science and Ocean Engineering
National Taiwan University
Taipei, Taiwan

Yu-Chun Wang, Shu-Yu Lin, Jan-Ming Ho
Inst. Information Science
Academia Sinica
Taipei, Taiwan

***Abstract*—Since a complete Chinese font has typically several thousand or more Chinese characters and symbols, and most of them are much more complicated than English alphabets, it takes a lot of time and efforts for even professional font engineers to create a Chinese font. Although several attempts had been made to synthesize Chinese characters from strokes and components, it is still not easy to synthesize so many Chinese characters at one time. In this paper, we present an easy and fast solution for an ordinary user to create a Chinese font of his or her handwriting style. We adopt the approach: to synthesize Chinese characters using components extracted from the user's handwritings. In the preprocessing phase, we built a Web interface for crowds to label the positions and sizes of components of every Chinese character in the target character set. The standard Kai font was selected as a reference. We also devised an algorithm to find a small subset of Chinese characters having all required components to synthesize other Chinese characters. To create a personal handwriting font, with commonly-used 3,914 traditional Chinese characters, a user only has to handwrite 400 or so Chinese characters on a pad. One character by one character, our system can track every stroke, recognize and extract components from the user's handwritings. Then, every target Chinese character is synthesized from the extracted components, by placing them properly according to their position and size information. The experiment results show that although manually fine-tune is still required for few synthesized Chinese characters, users can create a Chinese font of their personal handwriting styles more easily and quickly.**

Keywords—Chinese font; personal font; handwrite font; glyph synthesis.

I. Introduction

It is believed that Chinese calligraphy can reflect one's emotion, education level, and personality. Many Chinese have been long looking for personal Chinese fonts, so that they can not only present their ideas in their digital works, but also show their personality and feeling to readers. However, there are few choices available in Chinese font libraries or markets currently. Chinese character set standards [1-8] used in computer systems typically have several thousand or more Chinese characters, as enumerated in Table I, and most Chinese characters are more complicated than English alphabets. Hence, it is not easy for an ordinary user to make a font in his or her personal handwriting style. Nowadays, Chinese fonts are still made by professional font designers or companies. They usually took very long time, probably from several months to years, to draw and beautify the outlines (or glyphs) of all Chinese characters.

As computer technology improves, several tools had been developed for font engineers. For example, FontForge [9], an open source font editor, allows users to create or modify the glyph of every character, one by one, in a font file. These tools usually require a certain level of knowledge and skills of computer systems and charset (character set) encodings [1-8], font specifications [10], computer graphics and image processing, and so on. These tools usually do not support Chinese characters well.

In the literature, many researches related to glyph synthesis of Chinese characters have been reported. In [20][21], analysis of Chinese character structure was studied. Some researchers proposed to synthesize the glyph of a target character from its strokes [11][12], and some proposed from its components (or radicals) [13][14]. In [15][16], via several image processing procedures, the scanned image of ancient Chinese calligraphy was digitized into vectored glyphs. The methods presented in [17-19] were proposed to beautify the glyphs for handwritten Chinese characters. However, most of them aimed at creating the glyph of a single Chinese character at a time. There is still a lack of a systematic solution to create a complete Chinese font in one's personal handwriting style.

In [22], we had presented an on-demand approach for digital art designers, who at a time usually need a very small subset of Chinese characters. In FontCloud, they only have to

TABLE I. Number of Chinese Characters in Standards

	Publish Year	*No. of Chinese characters*	*Note*
BIG5 [1-2]	1984	13,051	Traditional Chinese
CNS 11643 [3-4]	1992	48,027	Traditional Chinese
GB 2312 [5]	1980	6,763	Simplified Chinese
GB 18030 [6]	2000	27,533	Simplified Chinese
Unicode 1.0 [7-8]	1993	20,902	Chinese, Japanese, Korean, etc.
Unicode 3.1 [7-8]	2001	70,195	

This study was supported in part by Ministry of Science and Technology of Taiwan (R.O.C.), under grants MOST 104-2221-E-029-025.

978-1-4673-8590-9/15/$31.00 ©2015 IEEE

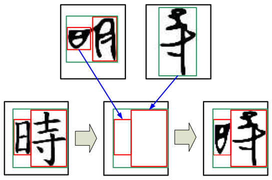

Fig. 1. Concept of glyph synthesis of Chinese characters from components extracted from a user's handwritings.

create needed glyphs to quickly make an incomplete-but-useful Chinese font. Incrementally, the glyphs are accumulated and the users can make a complete Chinese font. In [23], we tested preliminarily the approach: to synthesize Chinese characters using components extracted from the user's handwritings, as shown in Fig. 1. The experiment results showed that many synthesized glyphs of Chinese characters were satisfied.

In this paper, we present an easy and fast approach for an ordinary user to create a complete Chinese font of his or her personal writing style. To make a complete Chinese font from components, we need to answer the following questions.

1) What are the required components?

2) How many Chinese characters a user has to handwrite?

3) How to extract the required components from a user's handwritings?

4) Given a target Chinese character, where to place these components? How large these components are?

There are two phases in the proposed solution. (1) In the preprocessing phase, we built a user-friendly Web interface for crowds to manually label the positions and sizes of components of every Chinese character in the target character set. The Kai font was selected as a reference. We also devised an algorithm to find a small subset of Chinese characters having all required components to synthesize other Chinese characters. (2) In production phase, to create a personal handwriting font, a user has to handwrite the small subset of Chinese characters on an smartphone or electronic pad. One character by one character, our system tracks every stroke and extracts components from the handwritings. Then, every target Chinese character is synthesized from the extracted components, by placing them properly according to their position and size information.

To create a personal handwriting font, with commonly-used 3,914 traditional Chinese characters, a user has to handwrite only 400 or so Chinese characters. The experiment results show that although manually fine-tune is still required for few synthesized Chinese characters, users can create a Chinese font of their personal handwriting styles more easily and quickly.

The paper is organized as follows. After this introduction, we present our solution in section II, i.e., the proposed methods in the preprocessing and production phases to conquer each of the four questions. In section III, we describe the experiments and discuss the results. Finally, we conclude the paper and give some intended future work.

TABLE II. GLYPH CONSTRUCTION EXPRESSIONS

	Operators	Explanation	Example
Two components	⿰	Left and Right	順＝川⿰頁
	⿱	Top and Bottom	含＝今⿱口
	⿴	Outer and Inner	圍＝口⿴韋
More than two components	[illegible]	Components enclosed between Begin ([illegible]) and End ([illegible])	牖＝[illegible]片户甫[illegible]
	[illegible]		
Duplications	[illegible]	Two same components in vertical	炎＝[illegible]火
	[illegible]	Three same components in vertical	[illegible]=[illegible]户
	[illegible]	Two same components in horizontal	朋＝[illegible]月
	[illegible]	Three same components in horizontal	[illegible]=[illegible]去
	[illegible]	Three same components in triangular	焱＝[illegible]火
	[illegible]	Four same components in horizontal	
	[illegible]	Four same components in vertical	
	[illegible]	Four same components in 2x2 array	燚＝[illegible]火

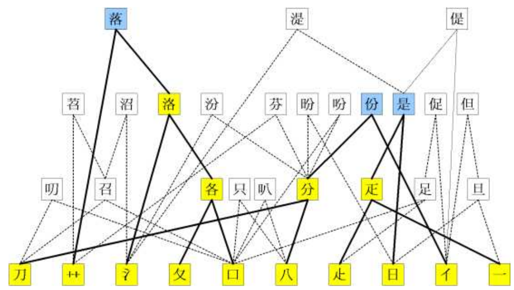

Fig. 2. Synthesis of 30 Chinese characters from 3 Chinese characters. Blue ones are handwritted, yellow ones are extracted from blue ones, and white ones can be synthesized from others.

II. PROPOSED SOLUTION

A. Component Analysis

In the literature, there are many studies about the structure of Chinese characters. In this paper, we leverage the result of [21], in which all Chinese characters recorded in Hanyu Da Zidian (漢語大字典) were decomposed into about 1,000 root radicals, as shown in TABLE II. We had extended [21] and decomposed other Chinese characters encoded in Unicode 3.1 into radicals [24-25].

In the preprocessing phase, we analyze the structures reported in [24-25] to find a small subset of Chinese characters having all components required to synthesize other Chinese characters in a target Chinese charset encoding. The concept is shown in Fig. 2. In this example, when a user handwrites three Chinese characters, we can synthesize other 30 Chinese characters. In this study, a Chinese character is synthesizable when all of its components are available, i.e., which have been extracted from some other Chinese characters.

We adopted a simple greedy algorithm, as shown in Fig. 3. Initially, all target Chinese characters were put in the candidate pool. A loop followed. In each iteration, a target Chinese character was selected into the output subset and removed from the candidate pool. We then checked other Chinese characters left in the candidate pool whether they could be synthesized. If yes, they were removed from the candidate pool. The loop stopped when the candidate pool was empty, that means all target Chinese characters were either selected into the output subset or synthesizable.

1. Decompose all target Chinese characters into Root Radicals
2. Sort all target Chinese characters according to the specified criterion
3. Put all target Chinese characters into the candidate pool
4. While the candidate pool is not empty
5. Remove a Chinese characters from the pool in the sorted order
6. Insert the character into the ***output subset***
7. Remove all other synthesizable Chinese characters from the pool
8. End of while
9. Return the ***output subset***

Fig. 3. Greedy algorithm to find a small subset of Chinese characters

We must note that the sorting criterion in step 2 is variable. There are many criteria, such as daily-used frequency, number of components, and so on. Therefore, there exist many small subsets of Chinese characters having all required components.

Thus, we answer the aforementioned questions *1)* and *2)*.

B. Positons and Sizes of Components in a Chinese Character

In the preprocessing phase, we also created a Web interface to label the positions and sizes of components of every Chinese character in the target character set, as shown in Fig. 4. The standard Kai font was used as a reference. The user first clicks a component of the assigned Chinese character and then draws a minimal rectangle covering the component. It is possible that two rectangles have an overlay.

We must note that similarly to the previous subsection, we had to do this preprocessing only once. Therefore, we adopted crowd-sourcing. We hired 7 students to manually label the information. At the same time, a simple evaluation was made. We required the crowds to evaluate the glyph synthesized from the components extracted from the standard Kai font. These feedback were somehow subjective. However, we still got a lot of information related to glyph synthesis of Chinese characters. For example, we realize that it is not a good idea to decompose the Chinese character 升 into two components 丿 and 廾, as shown in the fourth example in Fig. 4.

We must also note that handwriting variation of users is usually very large, even for a same user. The position and size information of components is for reference, just giving a start point to the aforementioned question *4)*.

C. Component Extraction

To conquer the aforementioned question *3)*, i.e., to extract components from a user's handwritings, there are several possible solutions. For example, we can require the user to handwrite selected Chinese characters on a customized gridded paper, scan the paper, crop the images of characters, and then use image processing techniques to extract components from the handwritings [22]. However, the solution has some flaws. If the user is unsatisfied with only a few characters just written on the paper, the user has to rewrite all characters again on a new empty paper from the scratch. In this study, we have developed an APP on smartphones or electronic pads. Each time, the APP shows a selected Chinese character to the user and requires the user to handwrite the character on the touch screen. Since the APP knows all components of the character, it can easily track and classify every stroke into components according to the stroke's starting point, ending point, and

Fig. 4. Web interface to label the positions and sizes of components of Chinese characters

ordering, and shows the result to the user with components in different colors. If the APP fails to extract the component correctly, the user can simply redo the character again.

The solution is incremental. The user can use any short free time to handwrite a few Chinese characters. As well, when new components are extracted, synthesizable Chinese characters are synthesized without having to wait for the finish of all selected Chinese characters.

Currently, the APP is still sensitive to the sensing capacity of touch screen and user's finger actions. As a result, we only consider normally handwriting in this study. I.e., stroke orders must be obeyed. Starting and ending of each stroke must be clear and space between consecutive strokes is required.

D. Glyph Synthesis

Given a Chinese character, when all of its components are available, we can synthesize its glyph, vectorize the glyph [26], and use tools such as FontForge [9] to merge the glyph into the user specific font file.

A component may appear more than once in the select subset. For example, Chinese characters 的 and 時 were both selected when daily-usage frequency was considered in the greedy algorithm. They both have a same component 日, but in different positions. Chinese characters 妻 and 姊 were another group for component 女, so were Chinese characters 做 and 變 for the component 攵. As discussed in [23], the following principals are adopted.

(a) A component at a similar relative position is better than the one in other positions. For example, given a target Chinese character 姑, the left component 女 is chosen from 姊 or 媽 rather than from 妻.

(b) A handwritten component is better than synthesized one. For example, given 侍, a handwritten component 寺 extracted from 時 is better than the one synthesized from 士 and 寸.

(c) When more than one components are satisfied with (a) and (b), we randomly choose one.

This gives freedom and variety in the glyph synthesis of Chinese characters, but still keeps the synthesized glyphs in the user's handwriting style.

Then, we place the components according to size and position information extracted from the standard Kai font in the preprocessing phase, as shown in Fig. 1.

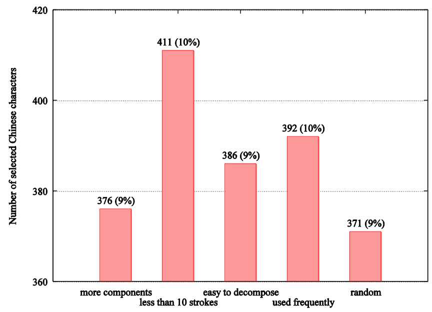

Fig. 5. Number of selected Chinese characters using different criteria

Fig. 6. Synthesized Chinese characters from components extracted from volunteer testers' handwritings. In a cell, the left is in standard Kai style, and the right is synthesized.

Fig. 7. Synthesized Chinese characters that have flaws. In a cell, the left is in standard Kai style, and the right is synthesized.

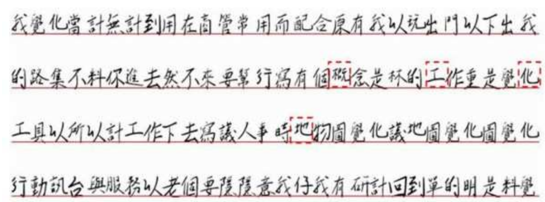

Fig. 8. A demonstation of another user's personal handwriting Chinese font

III. Experiments

In this section, we present the experiments and results. We initially aimed at the 3,914 commonly-used traditional Chinese characters encoded in both BIG5 [1-2] and T1 plane of CNS 11643 [3-4].

We first analyzed the component structure based on our results reported in [24-25]. The greedy algorithm shown in Fig. 3 was used with 5 different criteria to find 5 subsets of Chinese characters having all required components to synthesize the 3,914 Chinese characters. We then recruited several volunteer testers using our APP to generate their personal handwriting fonts.

A. Number of Chinese characters to be written

We tested the following 5 different criteria to find 5 small subsets of Chinese characters having all required components to synthesize the 3,914 commonly-used Chinese characters.

- Characters having more components first: we expected fewer Chinese characters would be selected. Indeed, the number was small. However, it unnecessarily selected many complicated Chinese characters. For example, the first five selected characters were 讓, 廳, 曬, 灑, and 釀. Furthermore, it is probably not a good idea to require users to write such complicated Chinese characters on a small screen of a smartphone.
- Characters having less than 10 strokes first: in the beginning the algorithm selected many simple Chinese characters. This favors the user to write simple Chinese characters, which however have less components. In the ending, the algorithm selected some very complicated Chinese characters such as 獵 and 黎. The number of selected Chinese characters was largest.
- Characters easy to decompose first: in this study, if the structure of a Chinese character is left-and-right or top-and-bottom combination of two components, it was considered easy to decompose. Thus, the proposed APP can perform better to extract components from it. The first five selected characters were 的, 是, 了, 有, and 他.
- Characters more frequently-used first: as described, we prefer handwritten components rather than synthesized components. The first five selected characters were 的, 一, 是, 不, and 了.
- Random selection: using this criteria, the algorithm in each iteration randomly selected a Chinese characters from the candidate pool. We ran the algorithm with this criteria with 100 times. The average number was the smallest.

Fig. 5 shows the numbers of selected Chinese characters with different criteria. In general, the numbers were 400 or so. We finally adopted the daily-usage frequency of Chinese characters so that most frequently-used Chinese characters are handwritten, but not synthesized. However, other subsets can be used.

B. Synthesized Personal Handwriting Chinese Fonts

Fig. 6 shows some synthesized Chinese characters from components extracted from users' handwritings. In general, the synthesized glyphs were satisfied by the volunteer testers.

However, some synthesized glyphs had flaws. As shown in Fig. 7, some components attached to or crossed with each other improperly. Some tiny flaws are acceptable, such as 絮 shown in Fig. 6. Some are not acceptable. The component position and size information extracted the standard Kai font is a good reference. However, an avoidance algorithm is needed to avoid attachment of components.

In Fig. 6 and 7, we also can see that the thickness of strokes is different, even within a same characters. When the font was used, as shown in Fig. 8, it is easy to see that these Chinese characters were not well aligned. Some shifted far away from the baseline. Some were too small or big. Since the variation of users' handwritings is typically large, furthermore fine-tuning should be done, such as component attachment avoidance, stroke thickness tuning, and so on.

IV. Conclusion

In this paper, we present an easy and fast solution for an ordinary user to create a complete Chinese font of his or her personal handwriting style. The user has to handwrite only 400 or so Chinese characters on a smartphone or pad. The proposed APP will track every stroke and extract components from the user's handwriting. One character by one character, the APP will collect more and more useful components, and synthesize glyphs of other Chinese characters from these components.

To create a complete Chinese font of the 3,914 commonly-used Chinese characters in one's personal handwriting style, the user has to handwrite 400 or so Chinese characters. In the experiment, at first glance, the synthesized Chinese characters were satisfied by the volunteer testers, although some of them still require fine-tuning. A further evaluation is undergoing, including the cost and performance analysis of the proposed solution and satisfication of the APP and quality of synthesized Chinese characters.

Currently, we are conducting a larger experiment to create personal handwriting Chinese font of Unicode 1.1, which encodes 20,901 CJKV Hanzis, including traditional, simplified and other Chinese characters encoded in BIG5 and GB 2312 respectively. Based on the result of reported in [25-26], the proposed component structure analysis for glyph synthesis of Chinese characters is undergoing.

As well, we are going to study the component attachment avoidance problem, stroke thickness tuning problem, and related typography problems of handwriting Chinese font.

References

[1] BIG5-1984, or simply BIG5, a de facto standard for Traditional Chinese character.

[2] BIG5-2003, in appendix of CNS-11643 expansion, 2003.

[3] CNS 11643-1992, "Chinese Standard Interchange Code," Obsoletes CNS 11643-1986. Chinese National Standard, 1992.

[4] CNS 11643-1986, "Standard Interchange Code for Generally-Used Chinese Characters," Chinese National Standard, 1986.

[5] GB 2312-80, "Code of Chinese Graphic Character Set for Information Interchange Primary Set." Technical Standards Press, People's Republic of China, 1981.

[6] GB 18030-2000, "Information technology - Chinese Ideograms Coded Character Set for Information Interchange - Extension for the Basic Set." Technical Standards Press, People's Republic of China, 2000.

[7] ISO/IEC 10646-1:2000(E). International Standard - Information technology -- Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane, 2000.

[8] The Unicode Consortium, "The Unicode Standard." Available at , http://www.unicode.org/unicode/standard/standard.html.

[9] G. Williams, "Font creation with FontForge," Proceedings of EuroTeX 2003, TUGboat, vol. 24, pp. 531–544, 2003.

[10] T. W. Phinney, "TrueType, PostScript Type 1, & OpenType: What's the difference," version 2.36, 2004.

[11] S. B. Lim and M. S. Kim, "Oriental character font design by a structured composition of stroke elements," Computer Aided Design, vol. 27, no. 3, 193–207, 1995.

[12] H. T. F. Wong and H. H. S Ip, "Virtual brush: a model-based synthesis of Chinese calligraphy," Computers & Graphics, vol. 24, pp. 99–113, 2000.

[13] P. Y. C Wong and S. C. Hsu, "Design Chinese typeface using components," Computer Software and Applications Conference, 1995.

[14] S. Xu, F. C. M Lau, W. K. Cheung, Y. Pan, "Automatic generation of artistic chinese calligraphy," IEEE Intelligent Systems, vol. 20, no. 3, pp. 32–39, 2005.

[15] J. Zhang, G. Mao, H. Lin, J. Yu, C. Zhou, "Outline font generating from images of ancient Chinese calligraphy", Transactions on Edutainment V, Lecture Notes in Computer Science, vol. 6530, pp. 122–131, 2001.

[16] J. Zhang, J. Yu, and H. Lin, "Capturing character contours from images of ancient Chinese calligraphy," the 2nd Workshop on Digital Media and its Application in Museum & Heritage, 2007.

[17] W. Xia and L. Jin, "A Kai style calligraphic beautification method for handwriting chinese character," International Conference on Document Analysis and Recognition, 2010.

[18] C. Zitnick, "Handwriting beautification using token means," SIGGRAPH, 2013.

[19] X. Zhu, L. Jin, "Calligraphic beautification of handwritten Chinese characters: a patternized approach to handwriting transfiguration", Proceedings of ICFHR 2008, pp. 135–140, 2008.

[20] H. Shu and R. C. Anderson, "Role of radical awareness in the character and word acquisition of Chinese children," Reading Research Quarterly, vol. 32, pp. 78–89, 1997.

[21] D.-M. Juang, J.-H. Wang, C.-Y. Lai, C.-C. Hsieh, L.-F. Chien, and J.-M. Ho, "Resolving the unencoded Character problem for Chinese digital libraries," 5th ACM/IEEE Joint Conference on Digital Libraries, 2005.

[22] J.-W. Lin, F.-S. Lin, Y.-C. Wang, J.-M. Ho, R.-I Chang, "FontCloud: Web Font Service for Personal Handwritten, Ancient, and Unencoded Characters," Future Information Technology - II, Lecture Notes in Electrical Engineering, vol. 329, pp. 113–119, 2015.

[23] J.-W. Lin, C.-Y. Wang, C.-L. Ting and R.-I Chang, "Font Generation of Personal Handwritten Chinese Characters," 5th International Conference on Graphic and Image Processing, Hong Kong, China, 2013.

[24] International Encoded Han Character and Variants Database. Available at http://chardb.iis.sinica.edu.tw/.

[25] J.-W. Lin and F.-S. Lin, "Unicode Han character lookup service based on similar radicals," International Journal of Smart Home, vol. 6, no.3, pp. 99–106, 2012.

[26] P. Selinger, "Potrace: a polygon-based tracing algorithm," (2009). Available at http://potrace.sourceforge.net/.