2009 Ninth International Conference on Intelligent Systems Design and Applications

Study on mapping of basic elements in the Chinese character intelligent formation without character library system

Mingyou Liu, Wenzi Liao, Youguo Pi
College of Automation Science and Engineering, South China University of Technology
liumingyou1025@gmail.com

Abstract

The theory of Chinese character intelligent formation considers that Chinese characters are formed by components according to character structure; all the components are the topological mapping of basic elements in the character structure. The mapping method of basic elements in different character structures is one of the key technologies. This paper carried on a thorough analysis to the transformation of basic elements, proposed the topological mapping method based on affine transformation. 27533 Chinese characters in GB18030-2000 standard were taken as experiment subject, a platform for Chinese character intelligent formation system was developed and all the characters were formed in the platform.

1. Introduction

In the word information system, letters are taken as the smallest unit in the alphabetic writing, as well as characters as the smallest unit in the Chinese character. So far, Chinese character library has been adopted by China and international organizations to develop Chinese character information processing system, which basically meets the needs of Chinese character informationization. However, Chinese character library also has some drawbacks, such as unable to establishing a stable standard, not passing down the civilization of Chinese character well, not conforming to the mechanism of human cognition of Chinese characters, out of touching with education and unable to meeting the demand for application in various fields [1]. Paper [2] proposed the concept of Chinese character intelligent formation according to prototype authentication mechanism of cognitive psychology, reusing the basic elements which are hieroglyphic, self-explanatory and its symbols to form Chinese characters. Our research shows that Chinese characters are formed by some components according to character structure; all the components are the topological mapping of basic elements in the character structure.

Latin is piece together word, but with one-dimensional structure, which arrange in sequence in addition to complying with few rules. Basic elements mapping to character structure is very complex, the same basic element mapping to different structure even same structure that with different combination of basic elements, its size, location and shape are different. Therefore, it must study corresponding mapping method. In this paper, we study the mapping method of basic elements in the Chinese character intelligent formation system, the main idea is that use affine transformation as a basis for topology mapping, adopting scale invariant feature transformation (SIFT) and particle swarm optimization (PSO) to acquire affine coefficients of basic elements, and then building mapping knowledge library. According to the cognitive mechanism of human being's reading and writing of Chinese characters, likening the computer to the human brain, proposed the systemic structure of Chinese character intelligent formation without character library.

Structural arrangements for this paper: (1) introduce to Chinese character intelligent formation without character library system; (2) research on topology mapping of basic elements based on affine transformation; (3) experimental study.

2. Chinese character intelligent formation without character library system

Xu Shen summed up six kinds characters creation rules which are hieroglyphic, self-explanatory, meaning, echoism, mutually explanatory and borrowed in the book of 《Shuo Wen Jie Zi》. Chinese characters including single characters and synthesis characters, of which, the hieroglyphic and self-explanatory characters are single characters and the meaning and echoism characters that composed of hieroglyphic and self-explanatory are synthesis

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.31

character. The Chinese character is the ancestors of the Chinese nation based on the cognition and abstraction to the world, adopting the hieroglyphic method that abstract the symbol of objects, adopting the self-explanatory method that abstract the symbol of things. Because the numbers of hieroglyphic and self-explanatory symbols are limited, carries on the combination according to the meaning and echoism to express the complexity things. Sometimes in the composition of synthesis character, for the Chinese character overall construction, it needs to carry on some deformation to the hieroglyphic and self-explanatory characters, and becomes the hieroglyphic or self-explanatory symbols after the deformation, for example, "水" deforms to "氵". The hieroglyphic and self-explanatory symbols are the basic elements of Chinese characters, called basic element for short. The method of meaning and echoism of synthesis characters is the original structure, while the combination way of basic elements is the plane figure structure of Chinese characters, they determine the Chinese character of meaningful and structural;

Reading and writing of Chinese characters are a typical processing of pattern recognition. Therefore, combined the tradition theory of Chinese character formation with prototype theory of cognition psychology, Chinese character intelligent formation is formed based on cognition mechanism and can be expressed as follows. Comparing the computer with the human brain, we create database to store the basic elements of the Chinese character, the structure of the Chinese character as well as the mapping knowledge of the basic elements, to simulate the human long-term memory; to simulate our brain's control function while we read and write Chinese characters with inducing and explaining engine, the intelligent unit of Chinese characters formation. The Chinese character recognition corresponds to the process of reading Chinese characters, the intelligent Chinese characters formation corresponds to the process of writing Chinese characters. That extends the process of reading and writing Chinese characters to the computer. According to systemic structure of cognitive pattern recognition that proposed in paper [3], and combined the cognitive mechanism of Chinese characters, the systemic structure of Chinese character intelligent formation is shown in Fig.1.

The Chinese character intelligent formation system is composed of basic element library, knowledge library, inference engine, explain engine, Chinese character intelligent formation unit, user interface and storage unit. Knowledge library stores basic elements, character structure and mapping knowledge of basic elements. In the knowledge library, the knowledge acquired by analysis or calculation and organized by a certain form [4]. Inference engine are the procedure of rules and control strategies, it uses the knowledge of the knowledge library, so that the system can work harmoniously by the logical way. Explain engine analyzes the structures and basic elements of Chinese character. Chinese character intelligent formation unit extracts basic elements, structure knowledge and mapping knowledge, pieced together in sequence after mapping transformation to form a character. User interface has the function of input, display, save, open, print and so on. Storage unit is used to store data to facilitate the transmission of documents. Storage unit used to save user's data.

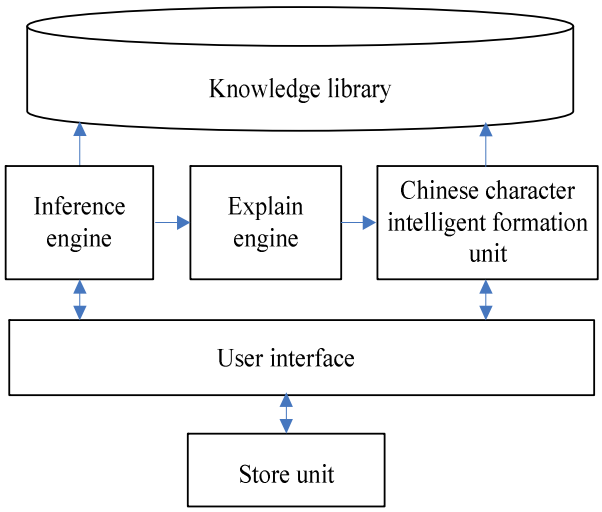

Figure 1. Systemic structure of Chinese character intelligent information

3. The mapping of basic elements based on affine transformation

3.1. The mapping of basic elements

The Chinese character is a combination of either single hieroglyphic or self-explanatory symbols, or several of them basing on meaning and echoism rules. The set of the character components corresponds to a certain basic element. The components of Chinese characters are the topological mapping of basic elements in the character structure, which is the position, size and shape of the basic elements may be different in different Chinese characters, but the basic elements are topological invariant. For example, basic elements of "氵" and "少" have different size, shape and position in the characters of "沙" and "莎". As another example of this, basic element of "寸" has different size, shape and position in the characters of "

讨”, “辱” and “褥”. Although “扑” and “把” are both left and right structure characters, the basic element “扌” in the two characters are different in size and shape. In the above-mentioned transformation, the same basic element corresponding to components of Chinese character are different in size, shape and position, but they have the same topology, namely topological invariance.

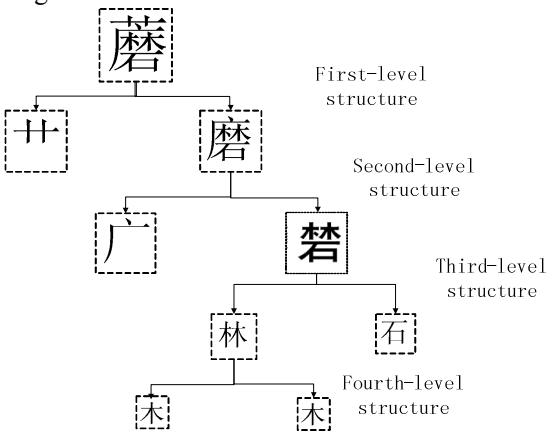

Figure 2. Example of the decomposition of structures and basic elements

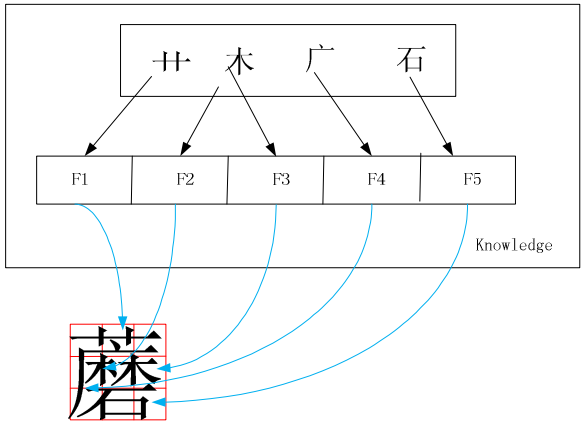

Figure 3. Example of the principle of basic element mapping

The decomposition of structures and basic elements of “蘑” is shown in Fig.2. This character contains four level structures and five components mapped by four basic elements. An example of the principle of mapping of basic elements is shown in Fig.3. The symbols “艹”、“木”、“广”、“石” are the basic elements of the Chinese character, $F_i(i=1,2,3,4,5)$ represents the knowledge of the topological mappings, such as the location, size and shape of the basic elements in a Chinese character. The final Chinese character “蘑” is created by mapping the basic elements into corresponding structure of the Chinese character.

3.2. The transformation model of basic elements

From the analysis of section 3.1, we can know that it needs a topological transformation method which can realize to transform basic elements to different position, size and shape in order to realize the mapping from basic elements to components of character, affine transformation meets this requirement.

Let W be a basic elements image and x be a point of the image. Let us define a geometric transformation of the basic element image by

$$AW+t=\begin{pmatrix} a_A & b_A \\ c_A & d_A \end{pmatrix}W+t=\begin{pmatrix} a_A & b_A \\ c_A & d_A \end{pmatrix}\begin{pmatrix} W_x \\ W_y \end{pmatrix}+\begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (1)$$

with A and t corresponding respectively to a linear transformation and a translation vector. a_A and d_A, b_A and c_A, t_x and t_y represent scaling factor, rotation factor and translation along the x, y axes respectively. Chinese characters are standard-type-face characters that square in a box, which determine the geometric transformation of basic elements involves only non-uniform scale, that is, $b_A = c_A = 0$. a_A and d_A reflect the changes in shape and size that map basic elements to characters, the basic element involves either only size change with the condition that $a_A \neq 1$ or $d_A \neq 1$, or size and shape change with the condition that $a_A \neq d_A \neq 1$.

3.3. Acquiring affine coefficients of basic elements

A approach was proposed that select corresponding points in the external connection rectangle of the basic element and mapping goal respectively to calculate the affine coefficients, but it need to remove other components manually when get the external connection rectangle of mapping goal [5]. A novel method for acquiring coefficients of basic elements automatically was proposed in this paper. Firstly, realized the key points match from basic element image to Chinese character image by SIFT [6, 7] algorithm, but in those matching points, a lot of are incorrect, then, three matching points pairs from all the matching points pairs were selected by PSO [8, 9] algorithm, adopting normalized correlation coefficient [10] as PSO’s fitness function. The process as shown in Fig.4, Firstly, determined the interested region by the structure of the Chinese character. Secondly, extracted scale and location invariant by SIFT algorithm, meanwhile removed rotation invariant in the interest region, and determined the matching points of

two images by nearest Euclidean distance of eigenvectors. Thirdly, choose three pairs non-collinear points randomly to constitute the initial solution space, each particle may obtains a group of affine transformation coefficients, then transform the basic element image to obtain transformation image, the fitness value of each particle determined by normalized correlation coefficient which measured the similarity between transformation image and Chinese character image. Finally, the optimal affine transformation coefficients were calculated by three pairs of optimal matching points.

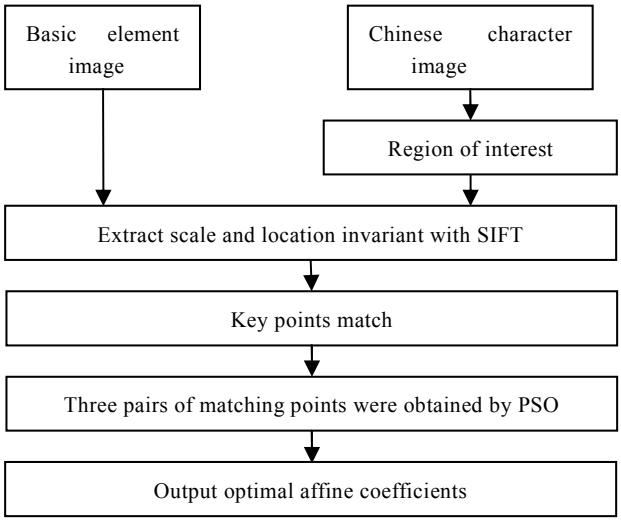

Figure 4. The flow of acquiring affine coefficients

4. Experimental and results analysis

4.1. Experimental process

27533 Chinese characters in GB18030-2000 standard were selected as experiment subject, we sum up the structures of the Chinese character and basic elements through the analysis of the structure of the Chinese character for all types, and then code for all characters. The knowledge database is constructed after obtaining all the affine coefficients for basic elements, the knowledge of basic elements, structures, character's coding and basic elements' mapping is organized and expressed by semantic network [4]. Experimental platform for the Chinese character intelligent formation is developed according to the systemic structure shown in Fig.1. All the characters are created on the platform. The user interface of Chinese character intelligent formation is shown in Fig.5.

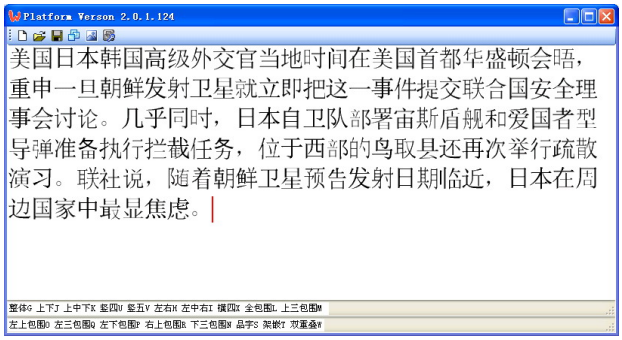

Figure 5.User interface of Chinese character intelligent formation

4.2. Results analysis

27533 Chinese characters in GB18030-2000 standard are formed correctly. The inputting rules are very simple, don't need any special input methods and remember the codes of basic elements and structures, just input Chinese characters according to its' corresponding structure.

The transforming process realizes the topological transformation of basic elements, but makes some strokes of a Chinese characters unaesthetic, for example, the downward vertical strokes of the components "申", "主" and "虫" corresponding in the characters "神", "注" and "融" are a little thinner. The reason is that basic elements are compressed too much in the horizontal direction, as well as the transforming method is linear.

5. References

[1] Youguo Pi, Jianping Lu, and Mingyou Liu, "Ponder for Chinese character library on Chinese character formation processing," *Contemporary academic research*, American Education Science and Technology Press, Hong Kong,2008,pp.68-70.

[2] Youguo Pi, The formation of Chinese character in computerization, China, CN1558314, 2004.

[3] Youguo Pi, Huailin Shu, and Tiancai Liang, "The Frame of Cognitive Pattern Recognition," *Proc of the 26th Chinese Control Conference*, Beijing, 2007, pp. 694-696.

[4] Youguo Pi, Mingyou Liu, and Jian Huang, "Knowledge Representation Based on Semantic Network in the Chinese Character Intelligent Formation System," *Proceeding of 2008 3rd International Conference on Intelligent System and Knowledge Engeneering*, 2008, (2), pp.785-790.

[5] Mingyou Liu, Jian Huang, and Youguo Pi, "Acquiring the Mapping Knowledge of Basic Element in the Chinese Character Intelligent Formation," *Proc of the 3rd international symposium on intellingece, WuHan*, 2008, pp.1025-1028.

[6] David G.Lowe, "Object recognition from local scale-invariant features," *Proc of international conference on*

computer vision .Corfu, Greece, 1999, 1150-1157.
[7] Lowe D G, "Distinctive image features from scale-invariant key-points," *International Journal of Computer Vision*, 2004, 60 (2): 91-110.
[8] Kennedy, J., and R. Eberhart, "Particle swarm optimization," *Proceedings of IEEE International Conference on Neural Networks*, Perth Australia, 1995, pp. 1942-1948.
[9] Shi, Y., and R. Eberhart, "A Modified Particle Swarm Optimizer," *IEEE World Congress on Computational Intelligence*, 1998, pp. 69-73.
[10] Lewis, J.P., Fast Normalized Cross-Correlation. 1995, www.idiom.com/~zilla/Work/nvisionInterface/nip.pdf.