World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:2, No:8, 2008

Interactive Chinese Character Learning System though Pictograph Evolution

J.H. Low, C.O. Wong, E.J. Han, K.R Kim K.C. Jung, and H.K. Yang

Abstract—This paper proposes an Interactive Chinese Character Learning System (ICCLS) based on pictorial evolution as an edutainment concept in computer-based learning of language. The advantage of the language origination itself is taken as a learning platform due to the complexity in Chinese language as compared to other types of languages. Users especially children enjoy more by utilize this learning system because they are able to memories the Chinese Character easily and understand more of the origin of the Chinese character under pleasurable learning environment, compares to traditional approach which children need to rote learning Chinese Character under un-pleasurable environment. Skeletonization is used as the representation of Chinese character and object with an animated pictograph evolution to facilitate the learning of the language. Shortest skeleton path matching technique is employed for fast and accurate matching in our implementation. User is required to either write a word or draw a simple 2D object in the input panel and the matched word and object will be displayed as well as the pictograph evolution to instill learning. The target of computer-based learning system is for pre-school children between 4 to 6 years old to learn Chinese characters in a flexible and entertaining manner besides utilizing visual and mind mapping strategy as learning methodology.

Keywords—Computer-based learning, Chinese character, pictograph evolution, skeletonization.

I. Introduction

MASTERING Chinese character is a difficult task because Chinese character is categorized as sinograph and alphabet-less type of language. On average, a student with well language aptitude requires 1320 hours of instruction in order to achieve level 2 language proficiency compared to the Latin-based languages which the latter requires 450 hours [1]. As for young children to learn and recognize a language, a special mechanism towards effective learning is desired to advance in learning the Chinese language.

Realizing the complexity in learning and memorizing the shape of Chinese characters, our motivation to research on an assistive platform based on the grounding base of how the Chinese character is evolved through pictograph evolution initiates our idea in this paper. Therefore, a learning mechanism involves interactive Chinese character learning through pictograph evolution is proposed.

Although not all Chinese characters are created from pictograph evolution, but our target is towards pre-school children aged between 4 to 6 years old. Therefore it is sufficient to use the limited characters that are able to be represented through pictograph evolution as pre-school children starts learning things mostly through a more visualized method using pictures or objects around them.

Taking the advantage of learning through visual and mind mapping [2], our proposed system is deployed on the similar concept but directly focusing on the learning process of Chinese characters. Learning the Chinese character involves visual recognition and continuous practice and this creates a problem of frustrating the young learners and creating an unpleasant learning experience. Therefore, our approach is relevant to ease the learning and to certain extend providing an exciting and fun in learning the language.

The proposed interactive Chinese character learning system is designed based on the original creation of the character itself. Our mechanism involves matching procedure whereby children are required to either sketch an image or write a word as an input and the matching will obtain the related image, word or pictograph evolutionary animation to re-enforce the learning process. In other words, our proposed system uses simple matching techniques to match the character or object drawn by freehand and relate it with an animated pictorial evolution of the Chinese character as a visual and mind mapping strategy.

Skeletonization is applied as a representation of the sketched image and word. Skeletonization is chosen as it is a flexible and efficient shape representation using a compact shape representation that is able to maintain the original structure, topology and physical appearance of the original shape. Shortest skeletonization path similarity Matching technique is used based on Xiang et. al [3].

The notable contribution in this paper includes presenting an interactive way to learn Chinese character based on their origin of creation, using one algorithm for both sketched image and word matching instead of two different matching mechanism based on shortest path similarity concept. There are several advantages in our proposed method. It is suitable to match both singularity appearance object and non-singularity appearance

J.H. Low, C.O. Wong and K.R. Kim are postgraduate student of the HCI Lab, Soongsil University, Seoul, Korea (e-mail: jinghong2007@gmail.com, cheeonn@ssu.ac.kr, raks@ssu.ac.kr).

E.J. Han is a Post-Doctorial at the School of Media and Graphic Interdisciplinary Centre , University of British Columbia, Vancouver, Canada (e-mail: ythghbn@nate.com).

K.C. Jung is a professor at the School of Digital Media, Soongsil University, Seoul, Korea (e-mail: kcjung@ssu.ac.kr).

H.K Yang is a professor at the Department of Computer Information Engineering, Dongseo University, Busan, Korea (e-mail: hykyang@dongseo.ac.kr).

object. In addition, simple statistical method is used for image and word matching utilizing Hungarian algorithm [16] as the matching strategy between query and model.

The remaining of this paper is structured as follow. In section II, we discuss the related work on learning and matching Chinese character. Section III outlines the main concept and proposed system. Section 4 shows the proposed method used in our system and system prototype. A conclusion is presented in the next section.

II. Related Work

Our work is related to interactive Chinese characters learning. Much of the application had been done base on this direction with different approaches. Some of the related work on research and system application is described in this section.

There are several applications that have been developed to help in the Chinese language learning. Tang et al. have proposed a web-based Chinese handwriting system that allows student to learn Chinese handwriting at anytime and anyway [4]. This system combines several important technologies that are able to provide intelligent and effective learning environment to users. This system also able to reduce the time used to learn Chinese character compare with traditional approach.

Chen at al. have developed augmented interface for Children Chinese learning [5]. Basically, this study tries to create a more efficient and interesting learning environment through Augmented Reality. In addition, this research describes the exertion of AR to construct a leaning system, where rich and effective interactions in leaning were considered. The prototype of this AR system enables school children to have a better understanding in pronunciation and memorization of Chinese.

Several applications have been developed specifically for memorizing the Chinese character. Flash card is one of the learning tools that aid in memorizing the Chinese characters. In general, flashcards are widely used as a learning drill to aid memorization by way of space repetition. [6] and [7] have created an online Chinese character flashcards application. Children might able to remember Chinese characters by using it for several time but this blindly memorize approach may cause boringness to the users.

Another approach that has been adapted by the researcher to aid learning and mastering the Chinese characters is "grab and match it" as in [8]. In general, children are required to recognize and match the relevant character that enables children to remember the Chinese character. This approach is an interactive approach by allowing the user to grab those randomly moving Chinese character and match it base on understanding. Nevertheless, this idea generally is similar to the flash card whereby the user still needs to memorize blindly on the Chinese character before using the application. No prior knowledge on the character will result much errors or leading to frustration and terminate to use the system further.

Word and pictures is another example of application that had been developed for Chinese characters learning as in [7]. A diagram comes along with a Character and enable user to learn the Chinese character visually. Nevertheless, this application just allows the user to remember the Character by image rather than understand the unique structure of the Chinese character.

In terms of technology wise, skeletonization approach is implemented to match the Chinese character. Bai et al. have used skeletonization approach to represent the shape [3]. This line like presentation is taking advantage of preserving the originality by minimum deformation representation. For the matching procedure, beside just consider the shortest skeleton path between end node for matching, a modification is added by considering the shortest skeleton path between the junction node.

III. Main Concept

The concept behind this paper deals with interactive and effective learning of Chinese characters for pre-school children. Utilizing the concept of pictograph evolution and simple matching techniques for both words and pictures, a new approach in learning Chinese characters is introduced. This approach is consistent with the concept of pictograph representation that is based on the oldest form of the Chinese character classification [9][10].

Our proposal is intended to solve the problem of difficulty to learn Chinese characters for pre-school children and utilize a simple and fast matching technique that does not require heavy computation. The system is targeted to ease learning and create fun in learning as learning Chinese character is proven to be a difficult task especially involves memorization of each character [11][12][13]. Learning the character has no other way except heavy memorization as each of the character differs with their meaning and pronunciation.

A. Chinese Character Classification Based on Pictograph Evolution

Chinese is a language with the largest number of users in the world, and the Chinese character is created with the longest in history of language. It is very hard to determine the time when has the Chinese character emerged and the character that we see today are the character on the tortoise shells and animal bones. Therefore, Chinese characters are very unique in shape and each character has a story behind the stone.

There are several principles that are used to define and explicate the Chinese character and one of the principles is pictographs. Pictographic Chinese character is composed of different symbols that often resembled the objects that it represents and sometimes are combined in some way to represent a whole new word or idea. After an evolution of time, those symbol or diagram are simplified and adjusted as Chinese writing to be further developed and expanded. So, each pictographic Chinese character has a story behind the shape as its appearance. Today, roughly 600 Chinese characters are created based on the pictograms principles and among those are claimed to be the oldest form of character. Figure 1 shows the example of the Chinese character falling in the pictogram category and its evolution. In this system, the advantage of the pictogram Chinese characters is taken whereby each Chinese character is an image vision of character. This will drastically facilitate the learning of the character besides creating fun in

learning.

B. Interactive Chinese character learning system (ICCLS)

In the recent years, computer-based learning has emerged to be a new genre of instructional media which utilized the capabilities of multimedia personal computer to animate software designed to both educate and entertain young children. Learning Chinese character is more relevant to have a fun and proper learning method due to the nature of its complexity. Therefore, our proposed system consists of the elements of visual, mind mapping, learning and entertaining.

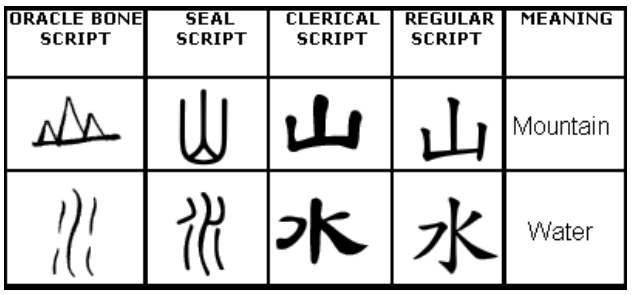

Fig. 1 Chinese character in evolution

Figure 2 shows the interactive Chinese character learning system overview. The user in our domain is pre-school children aged between 4 to 6 years who are starting to learn languages, specifically learning the Chinese characters. The child sits on a computer terminal and able to input Chinese words or sketch a simple picture on the input panel and the matching takes place to display the object or words together with an animation of how the Chinese character evolved from the pictograph evolution.

The input from the sketch panel representing either a word or a simple 2D picture will first represented using a skeletonization approach. Based on the skeletonization representation of the input, the shortest skeleton path pair between end node and junction node is generated. Therefore, each input and model is represented by a sequence of shortest skeleton path between the end node pair and junction node pair. The dissimilarity distance matrix between the model and query is generated. Next, based on the dissimilarity matrix, the matching procedure is taken place by using the Hungarian algorithm and the possible matched result is generated. If a word is drawn on the input panel, matching of the word will identify the correct word and at the same time retrieve the image and animation about the word out from the database and vice versa.

Our proposed system has an advantage of flexibility to input as a word or as a sketch image. Both inputs use the same algorithm although they are from two types of representation. The animated pictograph evolution is a learning strategy that we propose for children to learn Chinese characters. It also serves as an edutainment element in the computer-based learning system thus creating a sense of entertaining while learning a new language that is closely related to visual representation.

Figure 3 shows an example of input panel for our prototype system. The input panel accepts both forms of input; word and drawing. The same matching technique is used to match the words or pictures and in return display the output as in Figure 4.

In Figure 4, the example shows an input of a word "山" as an input by the user and the picture of a "mountain" retrieved together with the animated pictorial representation displayed to enhance the learning in the children. In this way, the user will be able to visualize and relate the word with the object and be able to remember the word in future.

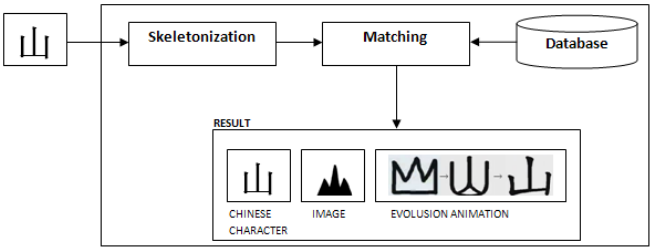

Fig. 2 System overview

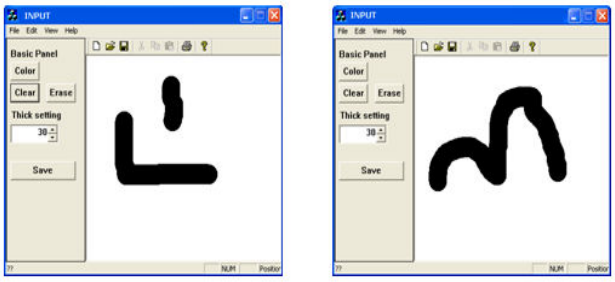

Fig. 3 Input panel for writing and drawing

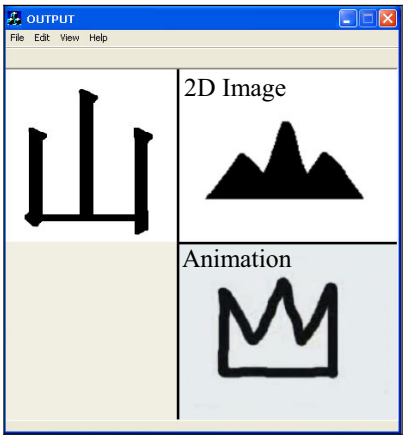

Fig. 4 Output panel

IV. OVERALL MACTHING PROCEDURE AND SYSTEM REPRESENTATION

Figure 5 shows the overall structure of matching technique that is implemented in our system. This system starts with using better skeletonization method to generate the skeleton [14]. Based on the skeleton, our system computes the quantity of the

end node and junction node, and the matching procedure is conducted within the chosen category whereby the category is define by the quantity of the end node and junction node. On the other hand, Chinese character representation for matching is computed base on the generated skeleton. We use the shortest path similarity for matching. Lastly, the matched data will be displayed to the user in the viewing panel.

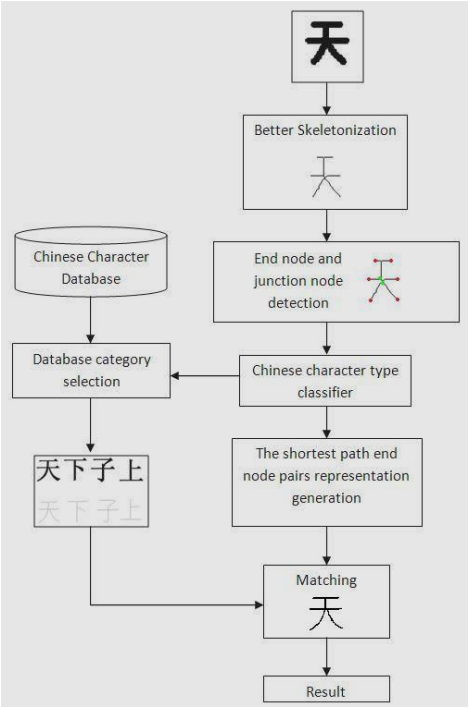

Fig. 5 Overview of the matching procedure

A. Input

Since we are targeting on low cost and common system, we use the mouse as an input device in CLLS. We have set the brush size to 30 points as a default input. This is due to the robustness issue especially for pre-school children to handle mouse input and write a character or draw a simple picture it is much more easier that the font appears to be thicker representation. In general, more noise is generated by using a mouse to write. Therefore, we use better skeleton method to generate the skeleton because this skeletonization method is able to reduce the noise which is generated by strokes written with a mouse as input.

B. Skeletonization

Skeletonization is a shape deformation technique that is introduced by H. Blum as the result of Medial transform [15]. A very explanatory definition of the skeleton is given by prairie-fire analogy which the boundary of an object is set on fire and the skeleton is the loci where the fire fronts meet the quench each other. The advantage of skeletonization is ensuring the topology and structure of the original shape under the deformation situation. Therefore, skeletonization is an approach that uses minimum information to represent a shape and we utilize it for Chinese character representation for matching.

C. Chinese Character Representation

We had modified Bai et al object representation for our Chinese character representation [3]. This representation uses the shortest skeleton path of each end node pair to represent an object in 2D. Since Chinese character can be recognized as objects in 2D, we have included junction node shortest skeleton pair to represent the Chinese character. End node is a node that only has 1 neighbor node and junction node is a node that has more than 2 neighbor nodes. Figure 6 and 7 shows a representation of the Chinese character "画" (draw) and "羊" (goat), and the representation set is listed as Table 1.

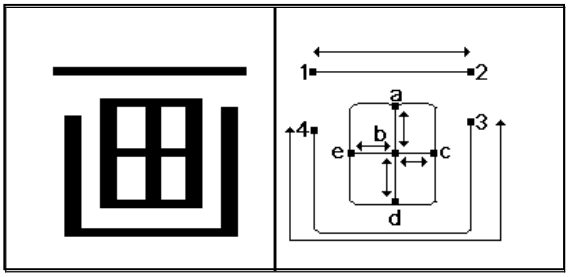

Fig. 6 Representation of Chinese character "画"

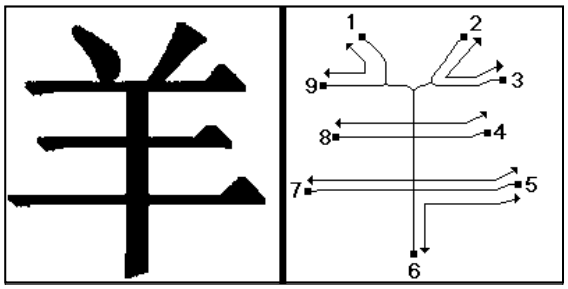

Fig. 7 Representation of Chinese character "羊"

TABLE I
SET REPRESENTATION OF CHINESE CHARACTER

Chinese Character	Shortest skeleton path Set in pair.
画	{{(1,2),(3,4)},{(a,b),(b,c),(d,b),(e,b)}}
羊	{(1,9),(2,3),(4,8),(5,7),(6,7)}

This representation is particularly good to integrate into our learning system because our system is a small scale system. Hungarian algorithm is used as a simple statistical matching approach to implement this representation. Our argument for using a simple statistical Chinese character matching instead of complex algorithm is due to a more cost effective and time saving approach [3].

In our representation, each shortest path distance is calculated based on the geodesic distance. It is the actual distance along the path. Each pair of shortest skeleton path is represented by a series of pixels that connects between the end node pair or junction node pair whereby each pixel is represented by the distance transform, *dt*, of the original Chinese character shape. The representation is shown in mathematic equation as follows:

$$E_{n_s,n_e} = \{dt_1, dt_2, ..., dt_n\} \quad (1)$$

$$J_{n_s,n_e} = \{dt_1, dt_2, ..., dt_n\} \quad (2)$$

Equation (1) denotes the shortest skeleton path between each end node, *E* and equation (2) denotes as the shortest skeleton path between each junction node, J.

Next, sampling took place by sampling K nodes along the path for each shortest skeleton end node pair and junction node pair. Let $E_{n_s,n_e}(g)$ denotes the sampling distance transform pixels with index *g* and $J_{n_s,n_e}(g)$ denotes the sampling distance transform pixels with index *g* as shown in the equation below:

$$E_{n_s,n_e} = (E_{n_s,n_e}(g))_{g=1,2,...,K} = \{dt_{(p_1=1)}(1), dt_{(p_2=p_1+\frac{n}{K})}(2), ..., dt_{(p_g=p_{g-1}+\frac{n}{K})}(K) \quad (3)$$

$$J_{n_s,n_e} = (J_{n_s,n_e}(g))_{g=1,2,...,K} = \{dt_{(p_1=1)}(1), dt_{(p_2=p_1+\frac{n}{K})}(2), ..., dt_{(p_g=p_{g-1}+\frac{n}{K})}(K) \quad (4)$$

Figure 8 is a visualized explanation of equation (3) and (4). Let the highlighted pixels as the shortest skeleton path of end node pair mark with distance transform value and *K = 5*. Since the shortest skeleton path is 25, therefore:

$$E_{n_s,n_e} = \{i, ii, iii, iv, v\}$$

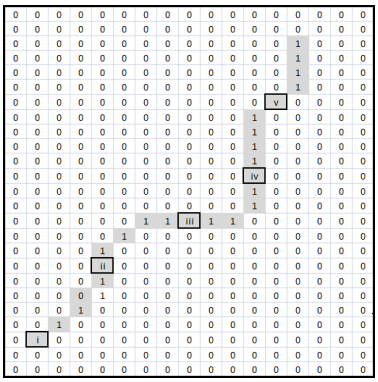

Fig. 8 Shortest skeleton path between the nodes

Next, we normalize the sampling nodes based on the average distance transform of the shape that is able to solve the scale problem. The average of distance transform as in equation *(5)* is equal to the sum of all distance transform in Chinese character divide by the size of the Chinese character in pixels, N_o. Each shortest skeleton path between end node and junction node will be normalized in respect to the average distance transform of the original Chinese character as in equation (6) and (7). Therefore, each Chinese character is represented by a normalized set of the sampling shortest skeleton path of end node pair and junction node pair.

$$\overline{dt} = \frac{\sum_{o}^{n=1} dt_n}{N_o} \quad (5)$$

$$(NE_{n_s,n_e}(g))_{g=1,2,...,K} = \frac{(E_{n_s,n_e}(g))}{\overline{dt}} = \frac{\{dt_1, dt_2, ..., dt_3\}}{\overline{dt}} \quad (6)$$

$$(NJ_{n_s,n_e}(g))_{g=1,2,...,K} = \frac{(J_{n_s,n_e}(g))}{\overline{dt}} = \frac{\{dt_1, dt_2, ..., dt_3\}}{\overline{dt}} \quad (7)$$

D. Database

Our database contains 50 pictogram Chinese characters which are suitable for pre-school children of aged between 4-6 years old. We had classified those Chinese characters by the need of our matching system based on the node type of Chinese character. Table 2 shows the category of Chinese character and some example data of each category in our database. The purpose of classification is to increase the matching speed by minimize the brute force search

TABLE II
DIFFERENT CATEGORIES OF CHINESE CHARACTER TYPE

Chinese Character Type	Example
Singularity with Looping	“目”
Non-singularity with Looping	“晶”
Singularity without Looping	“天”
No-singularity without Looping	“林”
Singularity combine Looping	“电”
Non-singularity combine Looping	“画”

E. System Prototype

A prototype system is developed to test the efficacy of the proposed method for Chinese character learning. Figure 9 shows the interface of our prototype with a breakdown example of input, output and animated pictograph evolution. User can input a word or a picture of their preference.

The figure shows user input a word “山” with the mouse. It literally means “mountain”. Therefore, in the Output section, the accurate word of “山” is displayed along with an image of “mountain” is retrieved. The animated pictograph evolution of how the word “山” is derived from the object “mountain” based on pictograph evolution theory is also displayed to the user. Fig. 10 shows animation sequence of the evolution of Chinese Character from the object itself to the word for the word “山” and “水”.

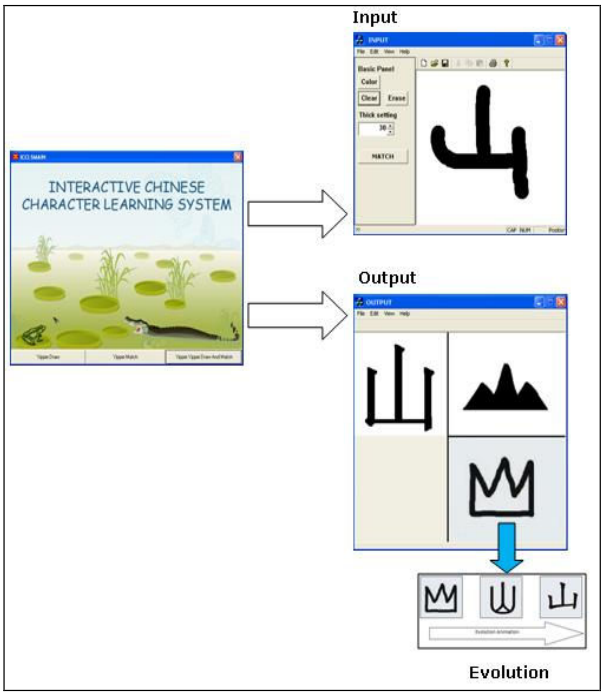

Fig. 9 Prototype system overview

V. LEARNING PEDAGOGY

The key pedagogy learning of Chinese language is listening, reading, writing, memorizing and speaking [13]. Our system is emphasis on memorizing learning pedagogy because the existing computer-based learning of memorizing is no very helpful enough as discussed in III Main Concept.

Our interactive learning system provides a frame work for student centered learning focus. Our system acts as an assistant for children to memories the Chinese Character. Besides that, our system further extern by adding the animation display that allow children to understand the origin and evolution of the Chinese Character, from the shape itself to the word itself as shows in Fig. 10. This allows creating an interesting, fancy and funny element in the process of Chinese Character learning.

Diagram and animation is utilized in our learning system because we taking on the advantage of image are ease to allow the children to remember [2] the appearance of the Chinese Character since diagram is able to create a profundity memory in their mind. Therefore, the evolution of the Chinese Character and profundity memory is use as the pedagogy in our interactive Chinese Character learning system and this idea is able to reduce the burden of the children in memories of complex languages learning especially in Chinese Character learning.

VI. CONCLUSION

We proposed a new approach in edutainment for pre-school children learning Chinese language. Chinese character is a complex character in linguistic. Learning the character require memorization and practice. Therefore, learning could be frustrating especially to pre-school children.

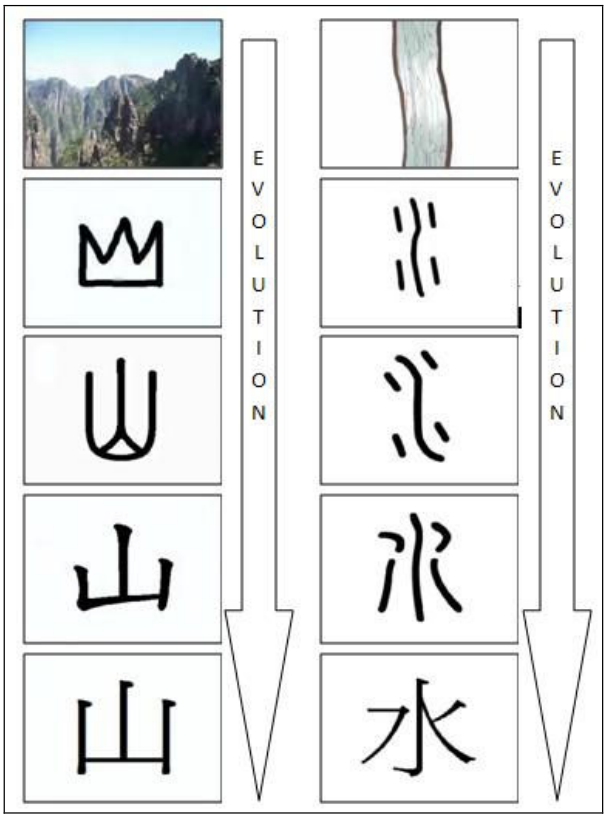

Fig. 10 Evolution sequence of Chinese Character "山" and "水" that displayed as animation

Our approach uses character and object matching with skeletonization and shortest path matching algorithm to identify the written input both in words and in sketched diagram in order to relate it with pictograph evolution of the Chinese character. Our strategy is to use the early evolution of Chinese character as a learning mechanism involving visual and mind mapping concept to instill the learning of the language.

Statistical matching technique using Hungarian algorithm is able to solve the problem of single paired end node as well as junction nodes. This special characteristic in Chinese character is able to match using the mentioned algorithm. We also take the advantage of using a pictograph evolution animation as a tool for learning the character in a more effective manner. The animated pictograph evolution also serves as an entertaining component in learning the language. This

ACKNOWLEDGMENT

This research was supported by the Ministry of Culture, Sports and Tourism, Korea, under the CTRC (Culture Technology Research Center) support program supervised by the ITRC (Information Technology Research Center) support program.

REFERENCES

[1] W. Baxter, "How hard is Chinese?," The Defense Language Institution, Monterey, California. Available: http://www-personal.umich.edu/~wbaxter/howhard.html
[2] A. Hollingworth, C.C. Wiliam and J.M. Henderson, "To see and remember: Visually specific information is retained in memory from

previously attend objects in natural scenes," Psychonomic Society Publication, Phychomic Bulletin & Review.8,4 (Dec 2001), 761-768(8).
[3] X. Bai and L.J. Latecki, "Path Similarity Skeleton Graph Matching," PAMI, IEEE Trans, Jul 2008, Vol.30, No.7, pp 1282-1292.
[4] K.T. Tang, K.K. Li and H.A. Leung, "Web-based Chinese handwriting education system with automatic feedback and analysis," LNCS Springer, Oct 2006, Vol 4181/2006, session 3, pp 176-188.
[5] C.H. Chen, C.C. Su, P.Y. Lee and F.G. Wu, "Augmented Interface for Children Chinese Leaning," ICALT 2007, IEEE, July 2007, pp 268-270.
[6] E. Peterson, "Chinese charater flashcards," 1996-2005 Online Chinese Tool. Available: http://www.mandarintools.com/flashcard.html
[7] J. Lau, "Yellow Bridge flash cards," Yellow Bridge-Chinese Language Center, 2003-2008. Available : http://www.yellowbridge.com/chinese/flashcards.php
[8] J. Winters, "Language Lab Project: Activity Builder." Available : http://www.csulb.edu/labs/langlabs/builder/
[9] L. Wieger, "Chinese characters: Their Origin, Etymology, history, classification and signification a thorough study from Chinese documents," Dover Publication, 1965. Available : http://books.google.com/books?hl=en&lr=&id=odrkZvbqJQoC&oi=fnd&pg=PA1&dq=history+of+chinese+character+&ots=qjOysgvFIL&sig=JcWQKE2zG0YmZdFAfg6OKL2Hy-o#PPA2,M1
[10] Y.X. Jiang, "Carvings may rewrite history of Chinese characters," Xinhua net, May 2007. Available : http://news.xinhuanet.com/english/2007-05/18/content_6118003.htm
[11] "How hard is it to learn Chinese?," Jan 2006. Available : http://news.bbc.co.uk/2/hi/uk_news/magazine/4617646.stm
[12] Y.B. Liu, "Teacher- student talk in Singapore Chinese language classrooms: a case study of initiation/response/follow-up (IRF)," Asia Pacific Journal of Education, Mar 2008, Vol. 28, Issue 1, pp. 87-102.
[13] C. Tan, "Change and continuity: Chinese Language Policy in Singapore," Language Policy, Springer, Mar 2006 Vol. 5, No. 1, pp. 41-62(22).
[14] D.Reniers and A. Telea, "Tolerance-based Feature Transforms," VISAPP, Springer, Feb 2006, Vol. 4, pp 187-200.
[15] A.K. Robert and M.P. Stephen, "Untangling the Blum Medial Axis Transform," IJCV, 2003, Vol. 55, Issue 2-3, pp. 139-153.
[16] J.G. Wang, P.K. He and W. Coo, "Study on the Hungarian algorithm for the maximum likelihood data associated problem," Journal of System Engineering and Electronics, Mar 2007, Vol. 18. Issues 1, pp 27-32.