

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(23), 2014 [14179-14185]

Online latin style anti-cursive handwriting input method of Chinese characters for non-Chinese speakers

Huang Tongcheng*, Fu Xin^{1,2}, Duan Xu^{1,2}, Chen Zhi^{1,2}, Xu Haizhong²

¹Provincial Key Laboratory of Informational Service for Rural Area of Southwestern Hunan, Shaoyang University, Shaoyang, 422004, (CHINA)

²Department of Information Engineering in Shaoyang University, Shaoyang, 422004, (CHINA)

E-mail: huangtc66@126.com

ABSTRACT

In order to facilitating the handwriting input of Chinese characters for non-Chinese speakers, this paper, based on profound information of cursive characters, deciphered the genetic code of ancient cursive symbols and made clear the rules for characters changing into its cursive style. It is directed toward to the development of the handwriting rules and vocabulary for Latin-style anti-cursive characters and the ways of their selection and classification, which aims to build a practical platform for a simple Latin-style online handwriting input and processing, with customary handwriting habits of non-Chinese speakers. As a result, it breaks the bottleneck which enables non-Chinese speakers to easily input information through handwriting Chinese characters.

KEYWORDS

Latin style online handwriting; Anti-cursive curve characters; Intelligent recognition; Non-Chinese speaker.



INTRODUCTION

With the development of the times, communications and exchanges between Chinese and overseas people become more active, which results in increasing needs and opportunities for non-Chinese speakers to process Chinese text information or communicate by using characters. Over the recent years, with the popularity of portable personal digital assistants with small keyboard such as palm computers and smart phones, various text input technologies came out and had been applied to text information exchanges and communications. Although a large number of scholars and researchers have made arduous efforts in the handwriting text recognition field (For example, a functional product, Hanwang, has been widely publicized and applied)^[1-10], there exists such a premise for using Chinese text input that users have to acquire the ability of recognizing and writing Chinese characters, which is an insurmountable barrier as Chinese characters, a strange, alien, complex and difficult one with complicated structures and strokes seems an incomprehensible text for non-Chinese speakers. Even if they learn to write, they have nothing to do but mimic characters on regular scripts with a limited speed. In the wake of this condition, it is quite difficult for potential consumers to input Chinese texts through handwriting with the use of commercially available products, which can't satisfy their need of handy handwriting characters on their personal computers and smart phones. They usually shrink as the sight of Chinese text input system, let alone learning to handwrite those picturesque characters, the cursive-style, which is totally opposite to their Latin-style writing habit. Therefore, it sounds rather contrived to help non-Chinese speakers to rapidly input Chinese text through handwriting Chinese characters on their handwriting terminals. The investigation has revealed that so far, no breakthroughs have been made in rapid input, processing and communication, based on handwriting terminals, of Chinese text for non-Chinese speakers. A simple handwriting input and intelligent recognition system in line with the handwriting habit for non-Chinese speakers have not come out yet^[11-15]. Thus, in order to genuinely enable them to fast input characters through handwriting, it is essential to develop a simple Chinese text input and intelligent recognition system mimicking the way of Latin handwriting, thus removing the barriers of inputting Chinese characters for the non-Chinese people and finally providing a more natural, humane online handwriting input system conducted by human and computer interactively.

On the other hand, we also know that the major goal of pattern recognition research is to create human perception capabilities in artificial systems. The task of automatically reading handwriting with close to human performance is still an open problem and the central issue of an active field of research^[16]. This may be due to the large degree of variability of human writing. Handwritten character recognition (HCR) systems have to address issues such as infinite variety of character shapes, similarity between characters, and distorted and broken characters^[17]. After painstaking researches, the genetic code of ancient cursive symbols have been deciphered, profound information of cursive characters has been acquired and the law of Chinese characters changing into its cursive style has been clarified. Finally, a simple Latin-style and anti-cursive handwriting input and a practical platform for Chinese text processing have been developed, which breaks the bottleneck of impossibility for non-Chinese people to fast and simply input characters through handwriting. As a result, non-Chinese speakers, familiar with Latin language and characters, can quickly and conveniently input Chinese characters, thus factually removing the "stumbling block" for them to input Chinese characters.

THE OVERALL RESEARCH ROADMAP

With a method integrating literature review, intensivism, comparative analysis, data mining, computer simulation and experimental methods and the highly representative methods, the feature extraction and the classification, this research is carried out profoundly. The road map of specific technology is shown in Figure(1).

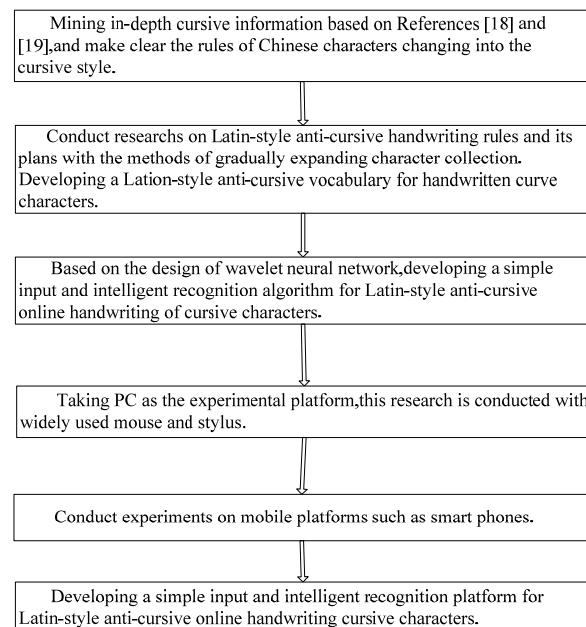


Figure 1 : The roadmap of specific technology

RESULT AND DISSCUSS

in-depth exploration and development of cursive chinese characters information

Guided by the core theory, Sinogram-based theory, with the help of precious literature, including References^[18-21], this research, based on exploration technology, studies the cultural codes of Chinese Cursive Script and seriously digs out profound information of cursive Chinese characters, finally deciphering the genetic codes and clearing the connective relationship between strokes, thus finding the rules for Chinese characters to turn into the cursive style. With the new point of view, ideas and methods, this research develops an innovative Latin-style anti-cursive handwriting rules and plans, based on ontology linguistics with Chinese character characteristics, which are applied to both Chinese-speaking and non-Chinese speaking peoples, and finally builds a Latin-style anti-cursive character collection for curve characters.

The reason of adoption of anti-cursive curve characters is that, after contrast and comparison of anti-cursive, cursive and regular scripts in Figure(2), the writing laws of anti-cursive Chinese symbols is identical to the Latin ones, which is in accordance with the writing habit of non-Chinese people. Both cursive and anti-cursive scripts maintain three Chinese elements, namely pronunciation, shape and meaning, which is an excellent example of a combination of Chinese and western elements. It makes the past serve the present and the Chinese serve the Foreign.

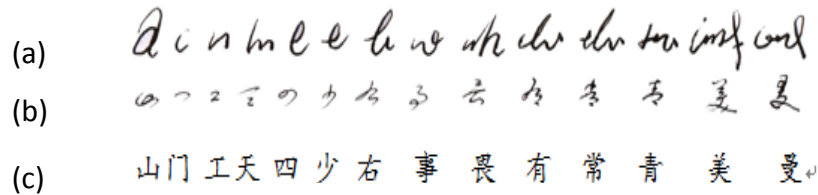


Figure 2 : Contracts of Parts of Anti-Cursive, Cursive and Regular Scripts : (a)Anti-cursive Script;(b)Cursive script;(c)Regular script

During the research process, it focuses on the simplified handwritten characters combining various calligraphies including Zhang and Jin cursive scripts, running and regular scripts, which forms a symmetrical seal-cutting mode where various scripts are contrasted, correspondent and complementary respectively, taking the character *Shi* as an example, seen in Figure (3).

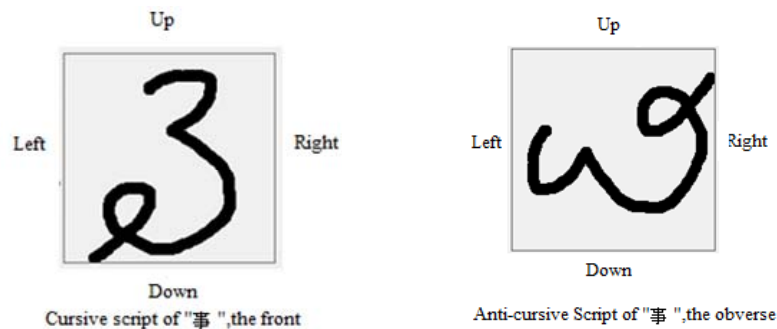


Figure 3 : A symmetrical seal-cutting mode of the character “事”

As Chinese characters have two different styles, namely, single and compound characters. It is relatively simple and easy to formulate the Latin-style anti-cursive handwriting rules for single characters, which can directly mimic standard cursive calligraphy, while in order to conform to the Latin-style writing habits of non-Chinese speakers after finishing anti-cursive process, compound ones, divided into up-down and left-right structures, have to be processed with structural transformation technology, in addition to exploring relative information of their standard cursive scripts.

THE ALGORITHM FOR PROCESSING ONLINE LATIN STYLE ANTI-CURSIVE HANDWRITING CHARACTERS

By referring to successful intelligent recognition technologies for handwritten Chinese characters and intelligent recognition technologies and algorithms for handwritten Latin characters, based on wavelet neural network where researches on profound feature extractions and the design of classifiers have been carried out^{[11][12][14]}, with the subject of anti-cursive handwritten curve characters, a simple input and intelligent recognition algorithm has been developed.

Wavelet neural network (WNN) is a network that is based on wavelet theory in which a discrete wavelet function is used as the node activation function. From an image application point of view, the proposed method, wavelet networks, is

divided into a series of stages for further concerning, each of which is described in the follows. The WNN has a three-layer structure with n_{in} nodes in the input layer, n_h nodes in the hidden layer, and n_{out} nodes in the output layer, described as in Figure(4)^[22].

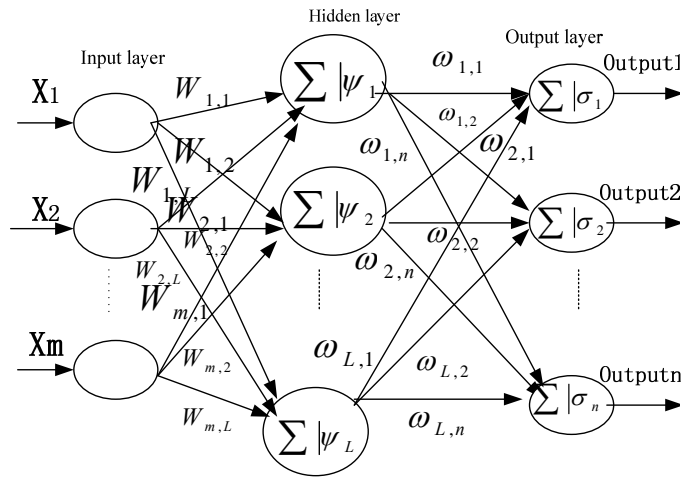


Figure 4 : WNN structure

It is choosing the mother wavelet. The Mexican hat wavelet^[11] is selected as the basis in this study. It is defined by:

$$\psi(x) = (1 - x^2)e^{-\frac{x^2}{2}} \tag{1}$$

The input to the k th neuron is

$$S_k = \sum_{j=1}^{n_{in}} W_{j,k} \times x_j \tag{2}$$

Where the x_j 's, $j=1,2, \dots, n_{in}$, are the input variables, and $W_{j,k}$ denotes the weight of the link between the i th input and the k th hidden nodes. In order to control the magnitude and the position of the wavelet, the multiscaled wavelet function is used as the hidden node transfer function. The dilation parameter a of the first hidden node ($j = 1$) is set as 1, i.e., $\psi_{1,b_1}(x) = \psi(x - b_1)$. For the second hidden node ($j = 2$), the dilation parameter a is set as 2, i.e., $\psi_{2,b_2}(x) = (1/\sqrt{2})\psi((x - b_2)/2)$, where the output of the wavelet is scaled down by $1/\sqrt{2}$. Similarly, for the j th hidden node, the dilation parameter a is set as j . Hence, the output of the hidden layer of the WNN is given by

$$\psi_{k,b_k} = \frac{1}{\sqrt{k}} \psi\left(\frac{S_k - b_k}{k}\right) \tag{3}$$

The output of the k th neuron is:

$$\psi_{k,b_k} = \frac{1}{\sqrt{k}} \left(1 - \left(\frac{S_k - b_k}{k}\right)^2\right) e^{-\frac{(S_k - b_k)^2}{2k^2}} \tag{4}$$

The output of the WNN is defined as:

$$O_l = \sum_{k=1}^{n_h} \psi_{k,b_k}(s(k)) \omega_{lk} = \sum_{k=1}^{n_h} \psi_{k,b_k} \left(\sum_{j=1}^{n_{in}} W_{j,k} \times x_j\right) \omega_{l,k} \tag{5}$$

Where $w_{l,k}$, $k=1,2,\dots,n_h$ and $l=1,2,\dots,n_{out}$, denotes the weight of the link between the k th hidden and l th output nodes. The weights, translating and scaling parameters are adjusted to minimize the error function during the training stage:

$$E = \frac{1}{2} \sum_{i=1}^I \sum_k^K (D_{ik} - O_{ik})^2 \tag{6}$$

Where $i=1,\dots,I$ is the number of training patterns, $k=1,\dots,K$ is the number of the objective, D_{ik} and O_{ik} represent the desired output value of Node $_{ik}$ and the actual net output value respectively.

THE IMPLEMENTATION OF ON-LINE LATIN STYLE ANTI-CURSIVE HANDWRITING INPUT PLATFORM

Members of the project group developed an algorithm and provide a platform for conveniently inputting, processing and communicating information with Chinese characters, which is consistent with the Latin-style handwriting habit for non-Chinese speakers. With the help of this platform, the characters which they input according to the Latin-style anti-curve online handwriting method can be intelligently recognized as Chinese characters in a rather short time. In addition, this platform is equipped with assistant contrast of Pinyin and English. The platform interface and its practical operating system can be seen in Figure(5).

The practical operating results of On-line Latin Style Anti-cursive Handwriting Input Platform of Chinese characters ‘美’ and ‘天’ are shown in Figure(6) a-b.

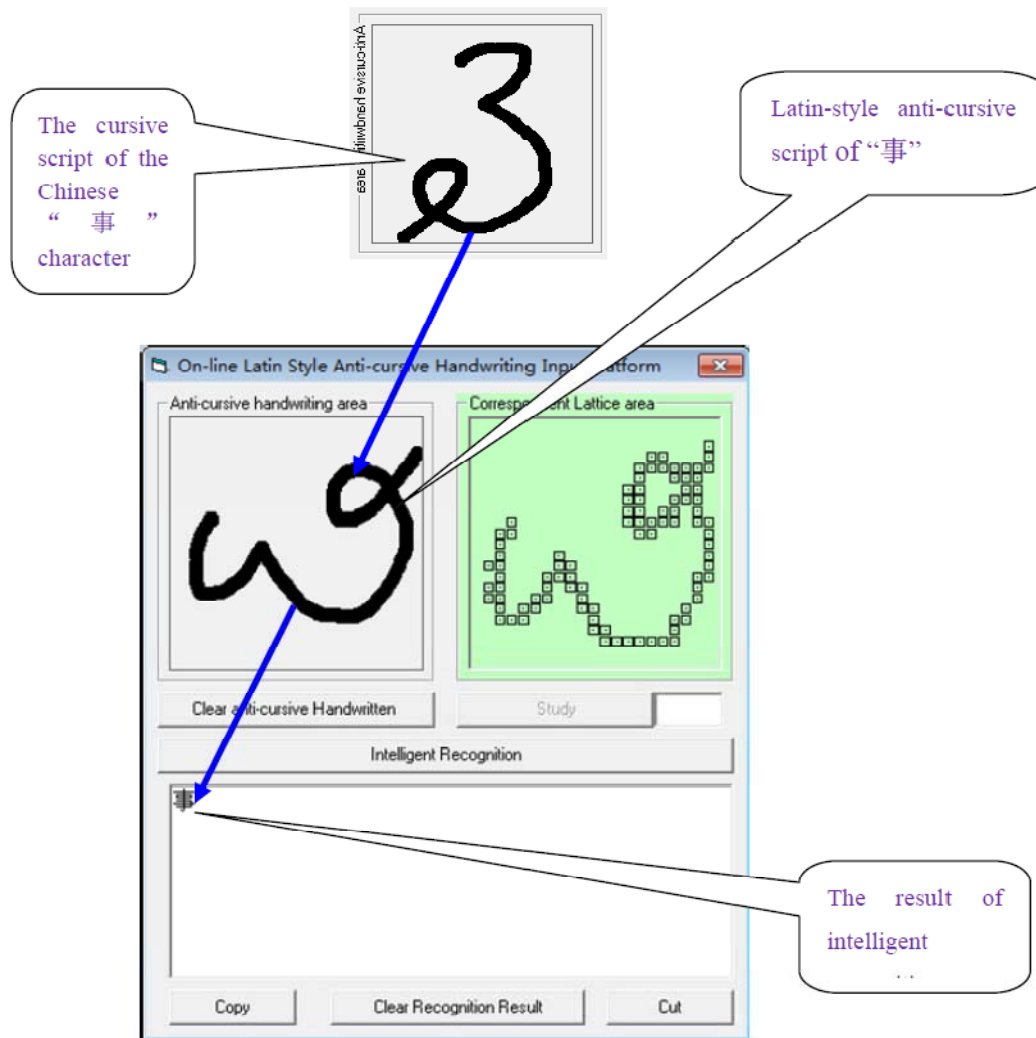


Figure 5 : Practical operating result of on-line latin style anti-cursive handwriting input platform of Chinese characters for non-Chinese speakers

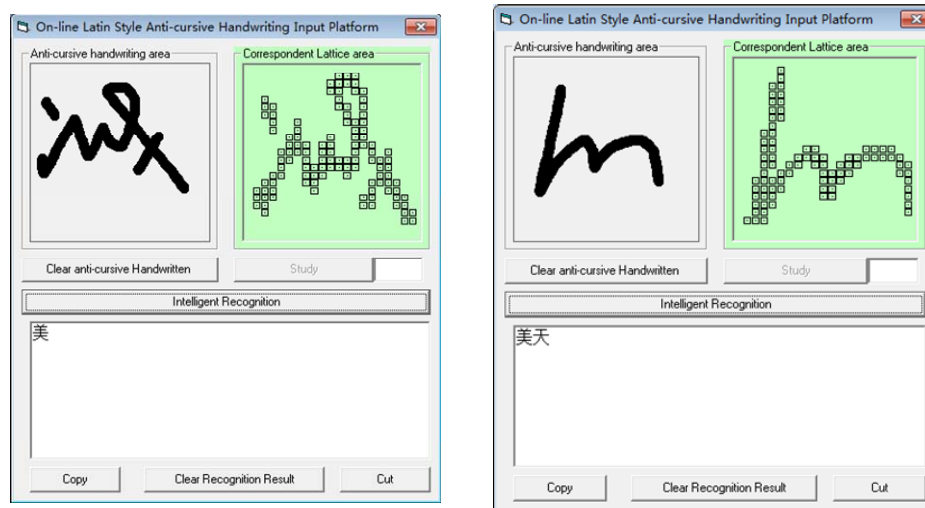


Figure 6 : Practical operating result of on-line latin style anti-cursive handwriting input platform of Chinese Characters ‘美’and ‘天’

Based on data mining to discover cultural genes of cursive Chinese characters, information and laws of cursive Chinese characters are sorted out in this paper. The project team invents laws and programs of Latin handwriting that encodes cursive handwritten Chinese characters, as well as a font of that Latin handwriting. Meanwhile, it develops Latin style handwritten cursive characters feature seculation algorithms research and classifiers research so as to implement their on-line portable input and intelligent recognition for non-native speakers of Chinese, and exploits intelligent recognition of Latin style handwriting for cursive Chinese characters.

In order to go through the bottle neck in handy inputting, dealing with and communicating Chinese characters information for non-native speakers of Chinese, the project team carries out the above studies independently and creates a practical platform for inputting and dealing with Chinese cursive characters by on-line Latin style handwriting that complies with habits of non-Chinese speakers. In this way, non-Chinese speakers can input Chinese characters simply in the way of inputting Latin alphabets.

ACKNOWLEDGEMENT

This paper is supported by Hunan Provincial Natural Science Foundation of China(14JJ7075), the key Scientific Research fund of Hunan Provincial Education Department(13A091), the key Scientific Research fund of Provincial Key Laboratory of Informational Service for Rural Area of Southwestern Hunan, and Aid program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province.

REFERENCES

- [1] Wei Deyang, Lei Wen; Dynamic evaluation on operational efficiency of publicly traded sports companies home and abroad, DEA-Malmquist TFP Index, Journal of Wuhan Institute of Physical Education, **46(2)**, 31-35, 42 (2012).
- [2] Xu Ying-Nan; The relationship between sporting goods companies and competitive sports in nearly 30 years, Taking NIKE and the NBA as an Example, Bulletin of Sport Science & Technology, **20(6)**, 97-99 (2012).
- [3] Li Gang, Wu Wei-Di, Wang Yong; Research of the performance of listed companies of port stocks, Journal of Tianji University of Technology and Education, **17(1)**, 49-53 (2007).
- [4] Kai Ding, Lianwen Jin, Xue Gao; A new method for rotation free online unconstrained handwritten Chinese word recognition: A holistic approach, In proceedings of the 10th international conference on document analysis and recognition, **7**, 1131-1135, (2009),
- [5] Z.L.Bai, Q.Huo, A study on the use of 8-directional features for online handwritten Chinese character recognition, ICDAR2005, 262 – 266 (2005).
- [6] F.Kimura, K.Takashina, S.Tsuruoka et al., Modified quadratic discriminate functions and the application to Chinese character recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, **9(1)**, 149-153 (1987).
- [7] H.Liu, X.Ding; Handwritten character recognition using gradient feature and quadratic classifier with multiple discrimination schemes, in Proceedings of the 8th ICDAR2005, 19-23 (2005).
- [8] T.Long, L.Jin; Building compact MQDF classifier for large character set recognition by subspace distribution sharing, Pattern Recognition, **41(9)**, 909-913 (2008).

- [9] Dzulkifli Mohamad et al.; On-line cursive handwriting recognition: A survey of methods and performances, In Proceedings of the 4th International Multi-conference on Computer Science and Information Technology, **4**, 70-84 (2006).
- [10] Alex Graves et al.; A novel connectionist system for unconstrained handwriting recognition, IEEE Transactions Pattern Analysis and Machine Intelligence, **31(5)**, 855-868 (2009).
- [11] Dapeng Tao, Lianwen Jin; Similar handwritten Chinese character recognition by kernel discriminative locality alignment, Pattern Recognition Letters, 1012-1016 (2012).
- [12] Nanxi Li, Lianwen Jin; A bayesian-based method of unconstrained handwritten offline Chinese text line recognition, International Journal on Document Analysis and Recognition, 1-15 (2012).
- [13] C.L.Liu, S.Jaeger, M.Nakagawa; Online recognition of Chinese characters: the state-of-the-art, IEEE Transactions on Pattern Analysis and Machine Intelligence, **26(2)**, 198-213 (2004).
- [14] Huang Tongcheng; Research on VOOCR and HOOCR Technology based on wavelet neural network theory, Ph.D.Thesis, Shanghai University, Shanghai, October, (2008).
- [15] K.P.Primekumar, Sumam Mary Idiculla; On-line malayalam handwritten character recognition using wavelet transform and SFAM, IEEE 3rd International Conference on Electronics Computer Technology (ICECT), 49 –53 (2011).
- [16] Binu P.Chacko et al.; Online sequential extreme recognition, In Proceeding of the 2011 IEEE Students' Technology Symposium, 14-16 (2011).
- [17] Binu P.Chacko et al.; Handwritten character recognition using wavelet energy and extreme learning machine, Int.J.Mach.Learn. & Cyber; **3**,149-161 (2012).
- [18] Albrecht Schmidt et al.; Interacting with 21st-century computers, PERVASIVE Computing, 22-31 (2012)
- [19] T.Plotz, G.A.Fink; Markov models for offline handwriting recognition: a survey, Int.J.Document Anal Recogn., **12**, 269-298 (2009).
- [20] G.Raju; Wavelet transform and projection profiles in handwritten character recognition-a performance analysis, in Proceedings of International Conference on ADCOM, 309-313 (2008).
- [21] Yu Yuan; Standard handwritten cursive chinese model by yu you-ren, Jindun Press,Beijing,China, (2009).
- [22] Sun Qingchang; New interpretation on cursive Chinese script, The World Publishing Corporation, Beijing, China, September, (2005).
- [23] Zhang Quan; Main points on standard cursive Chinese script, The World Publishing Corporation, Beijing, China, January, (2001).
- [24] Zhou Yongjin; Character collection of cursive Chinese characters, Rongbao Zhai Press, Beijing, China, January, (2009).
- [25] Tongcheng Huang, Liping Yin; A novel approach for text extraction from complex natural scene, Research Journal of Applied Sciences, Engineering and Technology, **3**, 415-430 (2012).