RECOGNITION OF CALLIGRAPHY STYLE BASED ON GLOBAL FEATURE DESCRIPTOR

*Yi Zhang, Yanbin Liu, Jianing He, Jiawan Zhang**

Graphics and Visual Computing Laboratory
Tianjin University, Tianjin, China
yizhang@tju.edu.cn, csyanbin@gmail.com, jianinghe111@gmail.com, jwzhang@tju.edu.cn

ABSTRACT

The study of digital Chinese calligraphy has become more valuable nowadays. While existing researches on Chinese calligraphy analysis are primarily focused on stroke-based character recognition and simulation, we propose a global feature descriptor to deal with style recognition problem in this paper. The proposed method extracts three categories of character features: position features, proportion features and projection features. These features are then used to train an SVM classifier of calligraphy style. We test the global feature based classifier on five-style Chinese calligraphy character set. The experimental results show that the proposed method can achieve good classification accuracy, proving the effectiveness of the global feature descriptor in calligraphy style recognition.

***Index Terms*—**recognition, calligraphy style, global feature

1. INTRODUCTION

Chinese calligraphy plays a special role in the long river of Chinese history and culture. Due to the significant value of aesthetics and culture, outstanding calligraphers and their works were praised highly all over the world.

With the development of digital library and image processing techniques, calligraphy creation and calligraphy appreciation have extended from manual mode to digital mode. Lots of calligraphy works were digitized for the purpose of preservation; and the calligraphy characters were unified and integrated into special font libraries for usage. That great amount of digital calligraphy resource facilitates calligraphy researches in the field of art, in the meanwhile, poses some scientific problems in analysis, recognition, and retrieval of digital calligraphy characters and images.

There has been some image processing work reported in the field of multinational handwriting texts such as Hebrew characters, Western characters, numeral characters and Chinese calligraphy. Therein, Chinese calligraphy, especially the stylized writing, is complicated because the characters are composed of multiple strokes which are written by soft hairy brush and ink. Existing researches on digital Chinese calligraphy processing could be primarily divided into three categories: character recognition and retrieval [1-2], calligraphy simulation and generation [3-5] and calligraphy analysis. Character recognition and retrieval are useful in the applications such as reprinting documents and optical character recognition (OCR). Calligraphy simulation and generation focus on creating artistic writing using digital tools. Other works such as writer identification [6], style recognition [7], and calligraphy evaluation [8] are classified into the third category.

As the style plays a crucial role in calligraphy training, appreciation and retrieval, we work on style recognition of traditional Chinese calligraphy in this paper. A global feature descriptor is introduced to model the calligraphy style. Particularly, the feature elements are not extracted from individual strokes and their textures, but structural characteristics such as position, proportional relationships between parts and so on. Our idea is inspired by traditional calligraphy training method. In the training point of view, there are two most important skills in good calligraphy writing. One is the stroke shapes and the other is the structure manipulations which include the stroke balance, stroke arrangement and combination. In other words, a good calligraphy character should make its strokes elegant and stay in the proper positions. The calligraphy exercises usually start from copying the character samples and then spend years or longer exercising in order to write the characters in similar shapes with the samples. Intuitively, the character looks weird with good stroke shapes but terrible structures. To the contrary, a character with good structures is much easier to be accepted even with childish writing. Therefore, we suppose that the calligraphy style could be represented by structural features.

Our method defines calligraphy style by three categories of statistical features: position of the character within grid, proportional relationships between parts and projection centers in X/Y direction. For a segmented calligraphy character image, these structural features are extracted and normalized to train a multi-class SVM

---

*Corresponding author: Jiawan Zhang

This work is supported by the National Science and Technology Support Project (2013BAK01B01, 2013BAK01B05).

classifier. We experimentally test 5-style classifier for both printing characters and handwriting characters, and prove the effectiveness of the calligraphy style representation.

The rest of the paper is organized as follows. In Section 2, related work about digital calligraphy processing and feature extraction is discussed. Section 3 describes the global features which represent calligraphy styles and how these visual descriptors work in our method. In Section 4, we show calligraphy style classification results to evaluate the performance of global feature descriptor. Finally, Section 5 concludes the paper.

2. RELATED WORK

2.1. Digital calligraphy processing

Digital calligraphy processing has been an important research topic and a body of work exits.

To solve the problem of Chinese character recognition, a wide variety of methods have been proposed. Liu et al. [9] adopt a model-based structural matching method for Chinese character recognition where the strokes and inter-stroke relations are used to achieve consistent matching by heuristic search. Yu et al. [2] build a skeleton database for multi-style Chinese characters. The recognition is achieved by comparing the skeleton extracted by MFITS method with the skeleton of the character samples in the database. They report that their method has a 95% recognition rate for dataset including more than 300 calligraphy characters.

In terms of Chinese calligraphy creation [10], the existing methods so far can be further classified into two groups: stroke syntheses methods and brush-writing simulation methods.

Xu et al. [5] generate artist Chinese calligraphy by segmenting the character strokes and combining the strokes to build the reference models on the basis of constructive ellipses. More recently, Xu et al. [4] explore a calligraphy generation and evaluation system by means of supervised learning on the extracted strokes. Stroke syntheses methods are efficient in speed but are not sufficient in free manipulation. Relatively, brush-writing simulation methods are more flexible as the users are able to write what they want using a virtual brush. For instance, Wong et al. [11] simulate the physical process of the stroke creation using Virtual Brush which is a parameterized model capturing the 3D geometric parameters, the brush hair properties and the variations of ink deposition along a stroke trajectory of the writing brush. Mi et al. [12] propose a 2D droplet model for improving the performance of brush simulation while maintaining expressive painting effects.

In the field of calligraphy analysis, we will briefly review Zhuang's method [7] and Han's method [8] because their work is most closely related to our work. Zhuang et al. introduce a calligraphy style representation which is the multinomial probability distribution over visual words defined on texture features. Then Latent Style Model is adopted to discover the style of the calligraphy works. They also develop a calligraphy browser to organize and exhibit the calligraphy resources according to the styles. They similarly deal with the problem of style classification as our work, but employing the different feature descriptors. Han et al. propose an interactive calligraphic guiding system to grade the written characters by measuring three quantized features: the center, the size and the projections of the character. Comparing to our work, their work also pays attention to the structural features of the character but serve the different purpose.

2.2 Feature extraction

Feature extraction is the key problem in the research of calligraphy style analysis. The feature extraction techniques of Chinese characters generally consist of two categories: stroke extraction approaches and global feature extraction approaches. The stroke approaches stress on extracting separated strokes and modeling their relationships. On the other side, the global feature approaches concern on statistical patterns in character images.

Extracting strokes from images is a difficult task. The traditional stroke segmentation is achieved by thinning process [13]. To better solve the distortion case, Liu et al. [9] propose a stroke extraction method by matching feature points and line segments with reference strokes. Su et al. [14] extract stroke segments based on a directional filtering technique and then refine the strokes' shapes by iteratively minimizing the reconstruction error. Basically, the strokes play the decisive role in character recognition because they are able to describe particular details and connections within the character. However in style recognition, too many stroke details may cause misclassification and loss of performance.

Among the commonly used global features, color feature is not applicable in calligraphy analysis because calligraphy works are grayscale images or binary images consisting very limited level color. Texture has been used to differentiate the calligraphy styles by modeling strokes in an implicit form in [7] and achieve good classification accuracy. According to our observation, the layouts of the strokes carry the structural measurements of the calligraphy style; thus we employ the distributions of the strokes and their strengths to identify the calligraphy style in our paper.

3. GLOBAL FEATURE DESCRIPTOR

As stated before, we think that the statistical features are effective in describing the structure information of Chinese calligraphy, which shows good potential in automatic calligraphy style classification. In our work, we build the descriptor by extracting three categories of statistical features including position features, proportion features and projection features. Therein, the position features stress the balance of the whole character; the proportion features focus on the distributions of the character strokes and the relationships between the character subdivisions; and the projection features pay attention to the balance of the character strokes in horizontal direction and vertical direction.

3.1 Position Features

Chinese calligraphy writing devotes particular care to the harmony and the balance within a limited square space. To keep the script nest, the calligrapher is required to write the character in proper size and position regardless of whether the strokes of a character are complex or simple. Thus one essential factor in calligraphy exercise is the character position within the grid. The beginners start exercising writing in the formatted grid to remind them keeping good structures of the characters. There are two common-used grids: tianzi-grid and mizi-grid. In particular, we use tianzi-grid in our method as Figure 3 shows. Tianzi-grid divides the character into four same-size sub-blocks from the midpoints of the horizontal direction and the vertical direction.

Normally, the barycenter of the character and the center of the grid are not exactly the same point. The difference between these two centers changes when the stroke compositions of a character differ in strength and distribution. As each calligraphy style has its stroke layout characteristics, we take the difference between the center of the grid and the barycenter of the character as a measure to describe the calligraphy style.

For a N*N character image, assuming that b(x,y) is the function for the pixel located at (x, y) that segments and binaries the character foreground from the background of paper, the barycenter of the character is extracted as Eq.1. With the barycenter g, we calculate the distance between the barycenter and the geometrical center via Eq.2. Mostly, the barycenter distributes randomly within a narrow range around the geometrical center of the grid.

$$g = \sqrt{g_x^2 + g_y^2}$$
$$g_x = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} b(x,y)*x}{\sum_{x=1}^{N}\sum_{y=1}^{N} b(x,y)}, g_y = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} b(x,y)*y}{\sum_{x=1}^{N}\sum_{y=1}^{N} b(x,y)} \quad (1)$$

$$Dist_g = g - N/\sqrt{2} \quad (2)$$

3.2 Proportion Features

The proportion features are defined as the ratios between each pair of the character sub-blocks. The proportional relationships have the advantages of resolution independen-ce and can be calculated in multiple scales by repeatedly block partitioning. Basically, smaller blocks reduce the style-indicative ability of the proportion features while enhancing the character-discriminating ability. Therefore, we employ coarse scale in our method. According to the tianzi-grid template, the character can be naturally divided into four parts: top-left block, top-right block, bottom-left block and bottom-right block. The proportion features of each pair of blocks are obtained from Eq.3.

$$R_{b1-b2} = \frac{\sum_{x=1}^{M_1}\sum_{y=1}^{N_1} b(x,y)}{\sum_{x=1}^{M_2}\sum_{y=1}^{N_2} b(x,y)} \quad (3)$$

Here, M_i, N_i are the resolution of the block i along the horizontal direction and the vertical direction respectively. For four sub-blocks in two scale levels, six ratios including $R_{top\text{-}bottom}$, $R_{topLeft\text{-}bottomLeft}$, $R_{topRight\text{-}bottomRight}$, $R_{left\text{-}right}$, $R_{topLeft\text{-}topRight}$, $R_{bottomLeft\text{-}bottomRight}$ are calculated to measure the balance of the character. We randomly select 40 regular script characters and illustrate their proportion curves in Figure 1. The top-bottom curves and left-right curves are separately displayed for the purpose of clearance. From Figure 1, several facts can be found:

(i) Most ratio values range from 0.5 to 2. In other words, the inking of two symmetrical regions, such as left part and right part, top part and bottom part, don't vary too much. Thus for a character with simple left component and complicated right component, the barycenter should shift to left from the center point and the left component may use stronger strokes.

(ii) The proportion relationships of the left-right region and the top-bottom region respectively appear the similar tendency in different scales, but have no obvious correlation with each other.

3.3 Projection Features

The coordination of the strokes is important in Chinese calligraphy, especially in horizontal direction (X axis) and vertical direction (Y axis). For instance, the character with a long horizontal stroke seems wired if other horizontals are very short; even it requires the non-horizontal strokes to extend forward horizontal direction to ensure the coherence of the character. To assess the properties in X and Y directions, we devise three projection features in each direction: average projection value, middle projection value and projection center.

First, the character pixels are projected into X and Y axes, respectively. The projections of the character in X/Y axis are represented as $Pv_x(r)$ and $Pv_y(c)$, computing by accumulating the foreground pixels in corresponding row/column. Then the projection values are defined by the number of the rows/columns whose projections are more than the corresponding threshold. And projection center is defined as the middle projection value. Formally, for a character located in a $n \times m$ grid, its ordered projection sequences are denoted by Pv_x and Pv_y as Eq.4, then the average projection values $Pf_{ave\text{-}x}$, $Pf_{ave\text{-}y}$ are computed via Eq.5, and middle projection values $Pf_{mid\text{-}x}$, $Pf_{mid\text{-}y}$, projection centers $Pf_{c\text{-}x}$, $Pf_{c\text{-}y}$ are computed as Eq.6-Eq.7 respectively .

$$Pv_x:\{Pv_x(r_1),Pv_x(r_2),\cdots,Pv_x(r_n)|Pv_x(r_1)>Pv_x(r_2)>\cdots>Pv_x(r_n)\}$$
$$Pv_y:\{Pv_y(c_1),Pv_y(c_2),\cdots,Pv_y(c_m)|Pv_y(c_1)>Pv_y(c_2)>\cdots>Pv_y(c_m)\} \quad (4)$$

$$Pf_{ave-x}=\sum_{i=1}^{n}B_{pf-x}\left(i,\frac{\sum_{j=1}^{n}Pv_x(j)}{n}\right) \quad (5)$$
$$Pf_{ave-y}=\sum_{i=1}^{m}B_{pf-y}\left(i,\frac{\sum_{j=1}^{m}Pv_y(j)}{m}\right)$$

$$Pf_{mid-x}=\sum_{i=1}^{n}B_{pf-x}\left(i,\frac{Pv_x(r_1)+Pv_x(r_n)}{2}\right) \quad (6)$$
$$Pf_{mid-y}=\sum_{i=1}^{m}B_{pf-y}\left(i,\frac{Pv_y(c_1)+Pv_y(c_m)}{2}\right)$$

$$Pf_{c-x}=Pv_x(r_{\lfloor n/2\rfloor}) \quad (7)$$
$$Pf_{c-y}=Pv_y(c_{\lfloor m/2\rfloor})$$

where

$$B_{pf-x}(r,th)=\begin{cases}1, Pv_x(r)\ge th\\ 0, Pv_x(r)<th\end{cases}$$
$$B_{pf-y}(c,th)=\begin{cases}1, Pv_y(c)\ge th\\ 0, Pv_y(c)<th\end{cases}$$

The projection features of 40 regular script characters are shown in Figure 2. Figure 2 illustrates the comparison of two projection values under different thresholds in X and Y direction. From the observation, most two projection values appear similar changing trends in both directions, and the differences of two projection values in X direction are smaller than those in Y direction. A smaller difference suggests that the projection distributes more balanced between maximum and minimum. The results may support the following facts:

(i) If a character has a special long stroke such as horizontal or vertical, others strokes have the trend of stretching in the same direction.

(ii) The degree of stretching is larger in vertical direction than in horizontal direction.

3.4 Global feature descriptor used in training

All three categories of features, namely position features, proportion features and projection features make up an 18-dimenison descriptor for each character. These features will be used as the support vectors in training to build an SVM classifier. When more than two styles are considered, we adopt one-versus-one and max-wins voting strategies to deal with the multi-class classification. The performance of the classifier will be discussed in more details in the experiment section.

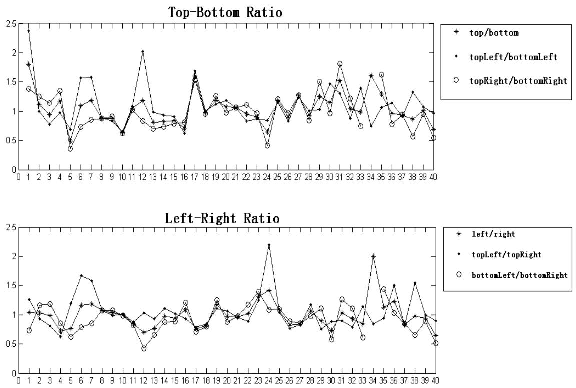

Fig.1. Proportion curves of 40 regular script characters

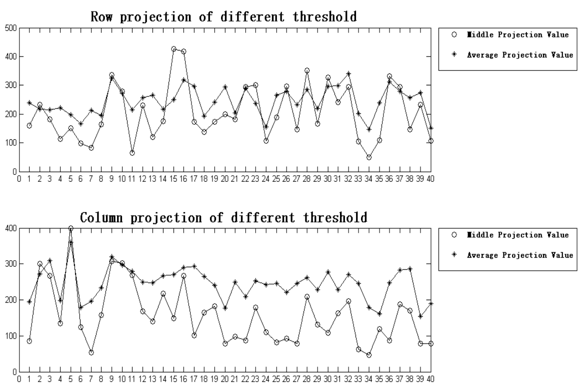

Fig.2. Projection curves of 40 regular script characters

4. EXPERIMENTS AND RESULTS

In order to verify the effectiveness of the global feature descriptor in calligraphy style recognition, we organize the experiments in two steps. First, all three categories of the features are extracted for the experimental character set including five calligraphy styles. Then we perform the training process and obtain the multi-class style classifier. Then the classifier is tested and evaluated on the same character set with no overlap with the training data. Second, the trained style classifier is tested by a general calligraphy dataset collected from Internet. Usually, the characters written by different calligraphers are more diversified than the characters from the standard font library because of personal writing habits. Thus the robustness of the classifier can be measured in that way.

In the experiments, we prepare a handwriting character dataset including five calligraphy styles: regular script, clerical script, running script, seal script and cursive script. Figure 3 gives the example of these five calligraphy styles. The dataset is collected from two widely used Chinese font library: Fangzheng font library and Wending font library.

We perform the training on totally 500 characters with 100 characters for each style; then the classifier is tested on a calligraphy dataset including another 500 characters with 100 characters for each style. The training and testing are text-dependent, that is to say that the characters from the different styles are the same. We think the features from these characters are efficient for other characters because all commonly used character structures have been included in our training and testing data. There are four common structures in Chinese characters, which are left-right structure, top-bottom structure, surrounded by structure and single structure; their ratios in training data are 52%, 10%, 13% and 25% respectively; and the ratios in testing data are 58%, 12%, 13% and 17% respectively.

In the results, the recognition performance is evaluated by precision and recall. Assuming that a given testing set has N characters in total for classification, within there are N_1 style-1 characters; the trained classifier correctly recognizes M_1 style1 characters and wrongly recognizes M_2 style1 characters, the precision ρ and recall γ for style1 are defined as Eq.8.

$$\rho=\frac{M_1}{M_1+M_2}, \gamma=\frac{M_1}{N_1} \tag{8}$$

The training and testing are performed for every two styles at first. Therefore 10 2-class classifiers are generated for five styles in all. Then we extend them and develop a 5-class classifier. The classification results are listed in Table 1 and Table 2. According to the experimental results, we conclude the following facts:

(i) Almost all the 2-class classifiers can achieve high recognition accuracy except the classifier of running script and cursive script. By analyzing the training characters, we find that the accuracy degradation is caused by high similarity between the characters of these two styles used in experiments. Figure 4(a)-(b) give the training character examples of running script and cursive script. Cursive script is special because it has been given a lot of freedom. Figure 4(c)-(e) show three other ways to write the same cursive script character. Thus the performance of cursive script classifier relies heavily on the training data selection.

(ii) Based on success of 2-class classifiers, the 5-class classifier is proved promising in recognizing calligraphy styles. Similarly, the classifier suffers from a performance reduction in classifying running script and cursive script because of approximate training data.

To further verify the effectiveness and robustness of the descriptor in describing the calligraphy style, we conduct an experiment by expanding the testing dataset. In the extended experiments, the testing characters are collected from Internet, including 500 characters with 100 characters for each style. The extended testing experiment is text-independent, so these 500 characters are basically different regardless of the styles. These characters are written by multiple calligraphers from various dynasties, thus the shapes of the characters are not as uniform as the shapes of the training data. Table 3 and Table 4 are respectively the 2-class recognition results and 5-class recognition results for

Fig.3. Examples of five Chinese calligraphy styles

Fig.4. Examples of character "Love". (a) Running script character used in experiment; (b) Cursive script character used in experiment; (c)-(e) other Cursive script characters

Table1. Recognition performance of 2-class classifiers
1- Regular Script, 2- Clerical Script, 3- Running Script, 4- Cursive Script, 5- Seal Script

Classifier Former style & Latter style	Former style		Latter style	
	Precision	Recall	Precision	Recall
1 & 2	0.9604	0.9700	0.9697	0.9600
1 & 3	0.9803	0.9900	0.9899	0.9800
1 & 4	0.9519	0.9900	0.9896	0.9500
1 & 5	1.0000	0.9900	0.9901	1.0000
2 & 3	0.9515	0.9800	0.9794	0.9500
2 & 4	0.9252	0.9900	0.9892	0.9200
2 & 5	1.0000	1.0000	1.0000	1.0000
3 & 4	0.8470	0.8500	0.8333	0.8500
3 & 5	0.9794	0.9500	0.9515	0.9800
4 & 5	0.9785	0.9100	0.9159	0.9800

Table2. Recognition performance of 5-class classifier
1- Regular Script, 2- Clerical Script, 3- Running Script, 4- Cursive Script, 5- Seal Script

	1-type	2-type	3-type	4-type	5-type
Precision	0.923	0.904	0.843	0.822	0.867
Recall	0.96	0.940	0.750	0.740	0.980

extended experiments. The results in Table 3 indicate that our method can achieve relative high recognition accuracy for most testing data regardless of the characters vary in size and resolution, even slight deformation. Most 2-class classifiers are proved robust. However, multi-class classifier doesn't perform well especially in recognizing running script and cursive script. The results in Table 4 reveal that the recognition accuracy heavily depends on the similarity between training data and testing data. Our method has the limitation in classifying calligraphy styles with internal variety. On the other hand, the results suggest that our method may be used in the applications of calligrapher identification.

Table3. Recognition performance of 2-class classifiers for multi-source characters
1- Regular Script, 2- Clerical Script, 3- Running Script, 4- Cursive Script, 5- Seal Script

Classifier Former style & Latter style	Former style		Latter style	
	Precision	Recall	Precision	Recall
1 & 2	0.8966	0.7800	0.8053	0.9100
1 & 3	0.5474	0.7500	0.6032	0.3800
1 & 4	0.8816	0.6700	0.7339	0.9100
1 & 5	0.9688	0.9300	0.9327	0.9700
2 & 3	0.8349	0.9100	0.9011	0.8200
2 & 4	0.9327	0.9700	0.9688	0.9300
2 & 5	0.9901	1.0000	1.0000	0.9900
3 & 4	0.1167	0.700	0.3357	0.4700
3 & 5	0.8870	0.5500	0.6739	0.9300
4 & 5	0.8924	0.8300	0.8411	0.9000

Table4. Recognition performance of 5-class classifier for multi-source characters
1- Regular Script, 2- Clerical Script, 3- Running Script, 4- Cursive Script, 5- Seal Script

	1-type	2-type	3-type	4-type	5-type
Precision	0.575	0.716	0.11	0.257	0.765
Recall	0.420	0.830	0.100	0.270	0.880

5. CONCLUSION

In this paper, we presented an approach to capture the style representation of calligraphy characters. Different with stroke-based calligraphy analysis methods, we extract three categories of global features, named position features, proportion features and projection features. Based on the proposed feature descriptor, we trained a multi-class SVM classifier to identify the calligraphy style. Our experimental results support that our style classifier can achieve good recognition performance for the standard and extended testing data. In the future, non-stroke local features will be investigated and combined to refine the style descriptor. In addition, further researches may include calligrapher recognition and variety identification within one style.

6. REFERENCES

[1] Lu W, Wu J, Wei B, Zhuang Y, “Efficient Shape Matching for Chinese Calligraphic Character Retrieval,” *Zhejiang University Press*, pp. 873-884, 2011.

[2] Yu K, Wu J, Zhuang Y, “Skeleton-Based Recognition of Chinese Calligraphic Character Image,” *Advances in Multimedia Information Processing-PCM 2008*, pp. 228-237, 2008.

[3] Leung H, Wong S T S, Ip H H S, “In the Name of Art,” *IEEE Signal Processing Magazine*, 25(4): 49-54, 2008.

[4] Xu S, Jiang H, Lau F C M, Pan Y, “An Intelligent System for Chinese Calligraphy,” *AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence*, pp. 1578- 1583, 2007.

[5] Xu S, Lau F C M, Cheung W K, Pan Y, “Automatic Generation of Artistic Chinese Calligraphy,” *IEEE Intelligent Systems,* 20(3): 32-39, 2005.

[6] Li X, Ding X, “Writer Identification of Chinese Handwriting Using Grid Microstructure Feature,” *Advances in Biometrics*, pp. 1230-1239, 2009.

[7] Zhuang Y, Lu W, Wu J, “Latent Style Model: Discovering Writing Styles for Calligraphy Works,” J*ournal Of Visual Communication And Image Representation*, 20(2):84-96, 2009.

[8] Han C C, Chou C H, Wu C S, “An Interactive Grading and Learning System for Chinese Calligraphy,” *Machine Vision And Applications*, 19(1):43-55, 2008.

[9] Liu C L, Kim I J, Kim J H, “Model-Based Stroke Extraction and Matching for Handwritten Chinese Character Recognition,” *Pattern Recognition*, 34(12):2339-2352, 2001.

[10] Dong J, Xu M, Zhang X, Gao Y, Pan Y, “The Creation Process of Chinese Calligraphy and Emulation of Imagery Thinking,” *Intelligent Systems, IEEE*, 23(6):56-62, 2008.

[11] Wong H T F, Ip H H S, “Virtual Brush: A Model-Based Synthesis of Chinese Calligraphy,” *Computers & Graphics*, 24(1): 99-113, 2001.

[12] Mi X F, Tang M, Dong J X, “Droplet: A Virtual Brush Model to Simulate Chinese Calligraphy and Painting,” *Journal Of Computer Science And Technology*, 19(3):393-404, 2004.

[13] Liu K, Huang Y S, Suen C Y, “Robust Stroke Segmentation Method for Handwritten Chinese Character Recognition,” *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, pp. 211-215, 1997.

[14] Su Y M, Wang J F, “A Novel Stroke Extraction Method for Chinese Characters Using Gabor Filters,” *Pattern Recognition*, 36(3):635-647, 2003.