Character-SIFT: a novel feature for offline handwritten Chinese character recognition*

Zhiyi Zhang, Lianwen Jin*, Kai Ding, Xue Gao
HCII Lab, College of Electronics &Information,
South China University of Technology, Guangzhou 510640, China
*Aleksis.zhang@gmail.com, *Lianwen.Jin@gmail.com*

Abstract

SIFT descriptor has been widely applied in computer vision and object recognition, but has not been explored in the field of handwritten Chinese character recognition. In this paper we proposed a novel SIFT based feature for offline handwritten Chinese character recognition. The presented feature is a modification of SIFT descriptor taking into account of the characteristics of handwritten Chinese samples. In our approach, global elastic meshing is first constructed and then the related gradient code of each sub-region is accumulated dynamically. Experiments using MQDF classifier show our feature's effectiveness with a recognition rate of 97.868%, which outperforms original SIFT feature and two traditional features, Gabor feature and gradient feature.

1. Introduction

Handwritten character recognition (HCR) has been one of the most active topics in pattern recognition field for several decades. Generally, a typical HCR system consists of four procedures: pre-processing, feature extraction, classification and post-processing. Among them, feature extraction is one of the most important factors in achieving high recognition performance. A great deal of effort has been made, and numerous methods of feature extraction were also proposed in the literature [1].

As for handwritten Chinese character recognition (HCCR), directional features are widely used and proven to be very effective [2], since they can catch the stroke directional pattern which is an important characteristic of Chinese Character. Previous researches [2] [4] demonstrated that among direction features the gradient feature [3] outperforms various other directional features.

A candidate competitor to the directional feature is Gabor feature, which was first proposed by Daugman [5]. A Gabor filter is a kind of local narrow band pass filter and selective to both orientation and spatial frequency. It is suitable for extracting the joint information in two-dimensional spatial and frequency domain. Researches have indicated that Gabor filters can fit the spatial and frequency profile of the simple cell in primary visual cortex very well [5], which justifies the wide applications of Gabor feature in many fields, such as face recognition and character recognition.

Recently, a successful feature extraction method, Scale Invariant Feature Transform (SIFT), has been proposed by D. G. Lowe [6]. SIFT comprises key-point localization and construction of key-point descriptor. SIFT, especially the SIFT descriptor, has been widely employed in computer vision and object recognition and proven to be very effective. Recently, SIFT was also applied to alphabetical character recognition [7] to achieve promising performance.

In this paper, inspired by SIFT descriptor, a novel offline handwritten Chinese character feature, character-SIFT Feature is proposed, which is based on local gradient histograms statistics. Our experimental results show that the proposed character-SIFT feature outperforms the original SIFT feature, and it can perform as good as (or even better than) the most popular gradient feature.

The rest of this paper is organized as follows. Section 2 reviews two state-of-the-art feature extraction methods for offline handwritten Chinese character recognition. This is followed by the introduction of the proposed feature in Section 3. In Section 4, experimental results are reported and discussed. Finally, conclusions are drawn in Section 5.

* This paper is partially sponsored by the following research projects: NSFC (no. U0735004, 60772116), GDNSF(no. 07118074)

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.27

2. Feature extraction in offline HCCR

Generally, linear normalization, nonlinear normalization or elastic meshing techniques are applied before feature extraction in offline HCCR. Linear normalization unifies all character images into the same size. Nonlinear normalization [8] or elastic meshing [9] is used to rectify the variation among the same Chinese character due to individual writing styles, and therefore improve the recognition accuracy. Although nonlinear normalizing is more commonly used, some experiments show that elastic meshing performs better in offline HCCR [10]. For this reason, all experiments in this paper are based on elastic meshing.

2.1. A feature extraction framework based on elastic meshing

In general, the diagram of feature extraction based on elastic meshing is shown in figure 1.

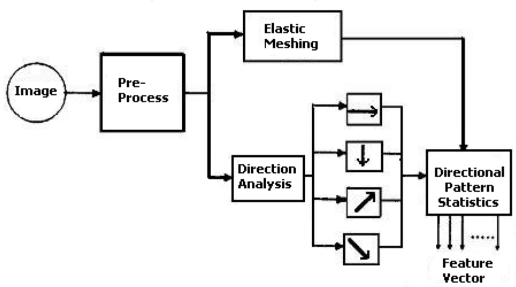

Figure 1. Process of feature extraction based on elastic meshing

Elastic meshing is the non-uniform region partition method for character images with imaginary grids. The principle of elastic meshing is that after partitioning, adjacent regions should have equal number of character pixel intensity (see reference [9] for details).

2.2. Gabor feature

A two-dimensional Gabor filter [5] is a band-pass spatial filter with selectivity to both orientation and spatial frequency, which can be expressed as follows:

$$G(x,y;l,\theta_k)=G_1(x,y)[cos(R)-exp(-\sigma^2/2)]+iG_1(x,y)*sin(R) \quad (3)$$

$$G_1(x,y)=\lambda^2 exp[-\lambda^2(x^2+y^2)/(2\sigma^2)]/\sigma^2,\ \sigma=\pi \quad (4)$$

$$R=2\pi[xcos(\theta_k)+ysin(\theta_k)],\ \lambda=2\pi/l \quad (5)$$

$$\theta_k=\pi k/D,\ k=0,1,2,\ldots,D-1 \quad (6)$$

Where l is the wave length, θ_k is the oscillation direction and D is the number of directions.

To extract Gabor feature, firstly elastic meshes is constructed on the character image, and let the center of each mesh be the sampling point. Then, the Gabor feature at the sampling point (x_m, y_n) is extracted as:

$$f_{gabor}(x_m,y_n)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)G(x-x_m,y-y_n;l,\vartheta_k) \quad (7)$$

Where M, N is the size of the filter region, $f(x, y)$ is the pixel value on point (x, y). In practice, the amplitude of $f_{gabor}(x_m, y_n)$ is usually used as the feature. Finally, the features from every sample points form the Gabor feature of the whole character image.

2.3. Gradient feature

Gradient feature is proven to be the most effective directional feature in the field of offline HCCR [2] [4]. To extract gradient feature, first we use 3×3 Sobel operators to obtain the horizontal and vertical gradient at each image pixel respectively.

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Figure 3. Sobel Operators

Then we define D directions with an equal interval $2\pi/D$, and decompose the gradient vector (g_x, g_y) into its two nearest directions in a parallelogram manner, as illustrated in figure 4 (in the figure, D is denoted as 8). In this way we obtain a D-dimensional gradient code at each image pixel. Then we divide the image into 8×8 sub-blocks through elastic meshing, and Gaussian blurring is applied to each sub-blocks with the center point of each sub-block as the sampling point to extract D Gradient features, resulting in a d-dimensional Gradient feature vector, where $d=8\times8\times D$. Finally a variable transformation $\boldsymbol{Y}=\boldsymbol{X}^{0.4}$ is applied on each element of the extracted feature vector to make its distribution more Gaussian-like.

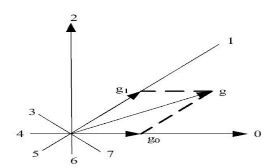

Figure 4. Decompsing of the gradient vector

3. Character-SIFT feature

In this section, we first give a brief introduction to SIFT descriptor, and then the particular details of the proposed feature extraction process will be discussed.

Research has found that complex neurons in primary visual cortex respond to a gradient at a particular orientation and spatial frequency, but the location of the gradient on the retina is allowed to shift over a small receptive field rather than being precisely localized. Scholars hypothesized that the function of these complex neurons was to allow for matching and recognition of 3D objects from a range of viewpoints [11].

Inspired by this idea, D.G. Lowe constructed a region description vector based on a local region gradient histogram, called SIFT descriptor [6]. Detailed experiment shows the robustness and distinctive characteristics of SIFT descriptor in rotation, scale changes and affine transformation.

However, from experiments we found that directly using SIFT for HCCR cannot obtain good enough performance. This is due to the fact that the key points detected by SIFT are not stable since great variations are commonly existed in handwritten samples with different writing styles.

Our goal is to modify the SIFT descriptor to suit for the description of the characteristics of handwritten Chinese character, and resulted feature is called character-SIFT. The process of extracting the character-SIFT feature consist of the following steps: elastic meshing, gradient code generation, dynamic gradient histogram computation, seed vector normalization and feature concatenation. Figure 5 gives a flow chart of the Character-SIFT Feature extraction procedure.

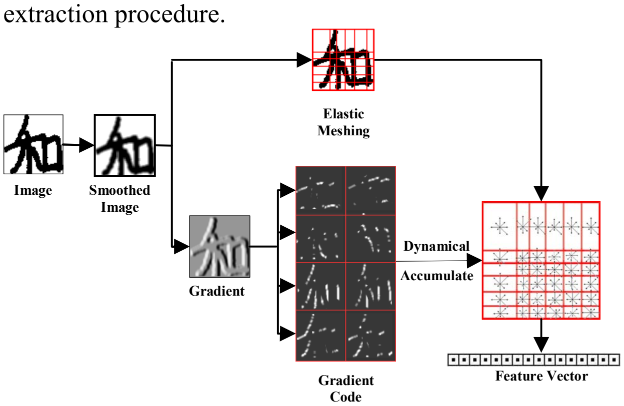

Figure 5. Character-SIFT Feature extraction procedure

3.1. Elastic meshing

Given a Chinese character image $I(x, y)$, a global elastic mesh with N lines in both horizontal and vertical direction is constructed through the method introduced in Section 2.1. Then the image is partitioned into $N \times N$ sub-regions and we set the center point of these regions as the seed point.

3.2. Gradient code generation

This stage is similar to the method of gradient feature illustrated in Section 2.3. Firstly, the image is mean-smoothed (with operator size of 3×3). After that, 3×3 Sobel operators are applied to every pixel in the image, and the horizontal and vertical gradient of each pixel are obtained. Then we define 8 directions with equal interval $\pi/4$ and the gradient vector of each pixel is decomposed into its two nearest directions in a parallelogram manner. Therefore every pixel corresponds to 8-dimensional gradient code.

3.3. Dynamic gradient histogram computation

First, we assign 8 dimensions vector referred to as seed vector to each seed point.

Then for each pixel in the image, its related sub-region is defined as below. The related sub-regions of a pixel p include the sub-region it belongs whose center denotes as m, and if p is at the down-right side of m, then the sub-regions at the down, right and down-right side of sub-region p belongs (if exists) are also the related sub-regions. Such is the cases with p is at the up-left, up-right or down-left side of m. Example of related sub-region is illustrated in Figure 6.

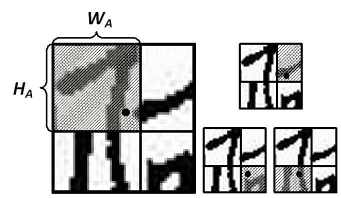

Figure 6. Example of related sub-region: black point denotes current pixel, the four blocks are its sub-regions and specially shadowed block indicates the region A current pixel belong to.

For every pixel, its gradient code is weighted by $w_x * w_y$ and accumulated to all its related sub-regions to avoid boundary effect. The value of w_x and w_y is determinate as follows. Let (x, y) denote current pixel, A be the sub-region current pixel belongs to. Let (x_A, y_A) denote the seed point of region A whose height and width is W_A, H_A respectively. Then the value of w_x and w_y is:

$$w_x = \begin{cases} 0.5 + \dfrac{\frac{1}{2}W_A - |x - x_A|}{W_A}, & \text{if the sub-region is A or on top of (or below) A;} \\ 0.5 - \dfrac{\frac{1}{2}W_A - |x - x_A|}{W_A}, & \text{other cases.} \end{cases} \quad (8)$$

$$w_y = \begin{cases} 0.5+\dfrac{\frac{1}{2}H_A-|y-y_A|}{H_A}, \\ \text{if the sub-region is A or on the right of (or on the left of) A;} \\ 0.5-\dfrac{\frac{1}{2}H_A-|y-y_A|}{H_A}, \\ \text{other cases.} \end{cases} \tag{9}$$

Equation (8), (9) indicate that the value of w_x and w_y are relevant to the sub-region sizes current pixel belong to, which varies with different writing style character due to elastic meshing. For example, w_x (or w_y) equal 1 when current pixel is at the center of sub-region A and goes down to 0.5 when moves toward the left or right border (up or down border for w_y) of A. This is the reason we call the stage as dynamic histogram computation.

3.4. Seed vector normalization

After dynamic gradient histogram computation, every seed vector contains the description of its related region. Yet, the size of these related regions are not equal. Therefore L2-normalization is applied to each seed vector.

3.5. Feature concatenation

To summarized, there are $N\times N$ sub-regions each described by an 8-dimensional seed vector and the character image is described by a d-dimensional feature, where d = $N\times N\times 8$. Finally a variable transformation $\boldsymbol{Y}=\boldsymbol{X}^{0.4}$ is applied on each element of the extracted feature vector $\boldsymbol{X}$ for the same reason stated in 2.3.

4. Experimental results and discussion

4.1. Experimental data

The experimental data is HCL2000 database, which is collected by Beijing University of Posts and Telecommunications for China 863 project. It contains 3,755 frequently used simplified Chinese characters, and all the character images are normalized to 64×64 pixels. A part of experimental samples is illustrated in figure 7.

Figure 7. Part of samples in HCL2000

4.2. Experiments setup

For determining the optimal meshing number, the experiments are carried out following the stages given below. First, character-SIFT features are extracted. Then Linear Discriminative Analysis (LDA) is used to compress the feature vector to dimension 256. And we use minimum Euclidean distance classifier to classify. Finally, both the raw feature recognition accuracy and LDA feature recognition accuracy are obtained and compared. To reduce the running time and improve efficiency, we use 100 sets for training and 30 sets for testing in each experiment.

To obtain more reliable experimental results when comparing between character-SIFT feature, Gabor feature and Gradient feature, we used the fast compact MQDF classifier [12]. The experiment procedure is given as below. First we extract character-SIFT feature, Gabor feature and gradient feature separately. The extracted feature vector is compressed to 256 dimensions by LDA, and the number of dominant eigenvectors in MQDF is selected as 12. A minimum Euclidean distance classifier is employed for coarse classification, and only the first 10 candidates are fed into the MQDF classifier. Moreover, 500 sets of samples are used for training and the rest 200 sets are for testing in the experiments.

Our proposed feature is called Char-SIFT for short in the tables appeared in the follow parts.

4.3. Determining the optimal meshing number

From the extraction method stated in Section 3, there is only one parameter for character-SIFT need to be determinate, that is the meshing number N. A set of experiments are carried out with different value of N. The Results are given in table 1.

Table 1. Comparison against different N

Sub-region number ($N\times N$)	Raw feature dimension	Recognition rate (raw)（%）	Recognition rate (LDA)（%）
6×6	288	90.531	94.411
7×7	392	**90.544**	**94.805**
8×8	512	89.857	94.668
9×9	648	89.386	94.642

From table 1 we can see that, although the recognition rates of N = 6,7,8,9 are very close, the best rate is of N = 7 and the corresponding raw feature dimension is 392 which means less computation time compared to N = 8 and 9.

4.4. Performance comparison between SIFT and character-SIFT

We give the recognition rate with the feature that extracted from character image using the original SIFT descriptor which divided the image into 8 × 8 sub-regions in table 2, comparing against character-SIFT with $N = 8$.

Table 2. Comparison between SIFT and Character-SIFT

Feat-type	Sub-region number (N×N)	Recognition rate (raw) (%)	Recognition rate (LDA) (%)
SIFT	8×8	77.252	89.054
Char-SIFT	8×8	**89.857**	**94.668**

From table 2 and table 1, we can see that character-SIFT of all value N significantly outperforms original SIFT. This shows that our modification do increase SIFT descriptor performance greatly for offline HCCR.

4.5. Performance comparison against Gabor feature and gradient feature

In this part, we compare the performance of character-SIFT with $N = 7$ against Gabor feature and gradient feature. Gabor feature and gradient feature are extracted through 8 × 8 elastic meshing and their original feature dimensions are both equal to 512. The results are given in table 3.

Table 3. Comparison of Character-SIFT feature against Gabor feature and gradient feature

Feat-type	Recognition rate (%)
Char-SIFT (N=7)	**97.868**
Gabor	96.580
Gradient	97.529

From table 3 it can be seen that both character-SIFT feature and gradient feature outperform Gabor feature and that character-SIFT slightly exceeds gradient feature in recognition performance.

Experimental results show the effectiveness of our proposed feature and the excellence of the idea behind the SIFT descriptor.

5. Conclusion and future work

SIFT descriptor is widely employed in computer vision and object recognition field. Yet, it has not been applied to the handwritten Chinese character recognition field.

In this paper we proposed a novel SIFT based feature for handwritten Chinese character recognition. The proposed feature is called character-SIFT, which is modification of SIFT descriptor according to the characteristics of handwritten Chinese character. Experiments have shown its effectiveness and merits while comparing with the original SIFT feature, Gabor feature and gradient feature.

It should be noted that although the experimental results of our feature are good, this is the first attempt to borrow the idea of SIFT descriptor for HCCR task. Further researches should be investigated to make this feature more suitable for HCCR, such as using more sophisticated meshing technologies, feature-fusing with gradient future and so on.

Reference

[1] Q. Due Trier, A.K. Jain, T. Taxt, "Feature Extraction Methods for Character Recognition: A Survey", *Pattern Recognition*, 1996, 29(4), pp. 641-662.

[2] H. Fujisawa, C.L. Liu, "Directional Pattern Matching For Character Recognition", *Proc.7th ICDAR*, Edinburgh, Scotland, 2003, pp.794-798.

[3] C.L. Liu, K. Nakashima, H. Sako, H. Fujisawa, "Handwritten Digit Recognition: Investigation of Normalization and Feature Extraction Techniques"， *Pattern Recognition*, 2004, 37(2), pp. 265-279.

[4] K. Ding, Z.B. Liu, L.W. Jin, X.H Zhu, "A Comparative Study Of Gabor Feature And Gradient Feature For Handwritten Chinese Character Recognition", *ICWAPR*, Beijing, China, 2007, pp.1182-1186

[5] J.G. Daugman, "Two Dimensional Spectral Analysis of Cortical Receptive Field Profiles", *Vision Research*, 1980, pp. 847-856.

[6] D. G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints", *Int.J.Comput.Vision*, 2004, 60(2), pp.91-110.

[7] J. Rodriguez and F. Perronnin, "Local Gradient Histogram Features For Word Spotting In Unconstrained Handwritten Documents", *ICFHR*, Montreal, Canada, 2008, pp.7-12.

[8] S.W. Lee, J.S. Park, "Nonlinear Shape Normalization Methods for the Recognition of Large Set Handwritten Character", *Pattern Recognition*, 1994, 27 (7), pp.895 - 902.

[9] T.L. Wu, S.P. Ma, "Feature Extraction By Hierarchical Overlapped Elastic Meshing For Handwritten Chinese Character Recognition", *Proc.7th ICDAR*, Edinburgh, Scotland, 2003, pp.529 - 533

[10] T.H. Su, T.W. Zhang, et al, "Off-Line Recognition Of Realistic Chinese Handwriting Using Segmentation-Free Strategy", *Pattern Recognition,* 2009, 42(1), pp.167-182.

[11] K. Mikolajczyk, C.Schmid ， A Performance Evaluation of Local Descriptors, CVPR, 2003, pp. 257-263.

[12] T. Long, L.W. Jin, Building Compact MQDF Classifier for Large Character Set Recognition by Subspace Distribution Sharing, *Pattern Recognition*, 2008, 41(9), pp.2916-292.