The Street View Character Recognition Based on Support Vector Machine

Min Peng, Jiawen Zhang, Ming Li, Yang Hu, Zhaohua He
School of Computer & the State Key Laboratory of Software Engineering Wuhan University
National Engineering Research Center for Multimedia Software, Wuhan University
Email: pengm@whu.edu.cn, zjw4891@gmail.com, liming751218@gmail.com

Abstract

The idea is intended for searching a wanted place in a foreign city with the business name, the scene text and the street address. There are two contributions in this paper. First the difficulty of locating the specified buildings exactly in an unfamiliar city is analyzed, and then a novel application associated with a mapping application to the real view of target place is presented in this paper. Another is to improve printed Chinese character classification with high accuracy. This paper conducts the problem to combine Euclidean distance and support vector machines. Euclidean distance is applied to rough classification stage, and then we apply support vector machines to classify characters that cannot be classified at rough classification stage.

I. Introduction

Looking for the target building in a unfamiliar place, some tools like Guidepost, Map, GPS and so on, are necessary. These tools can provide useful information but always is approximate situation and ambiguous expression without real view of the building. People must ask insider accord to give more visualized details about the appearance and the surrounding environment of the building, which waste lots of time to find it. So an automatic system directing the accurate position, precise description and real view of the target place with some key words are necessary and can effectively solve above mentioned problem. Extracting scene-text embedded in images from street views automatically is a crucial step in detection system. Associating the extracted text with the sign, location, appearance of the wanted buildings, people can easily find the target place by business names, store hours, and address in natural way. The outline of the application to locate the text is shown in Fig 1.

Fig 1. Marking text in virtual earth

Scene-text exists as part of the objects in a shooting scene image, when it is captured by a camera, which includes stereotypical forms such as street signs, hospital signs, and bus numbers, as well as variable forms such as shop signs, house numbers and billboards [1]. The text carries much definite and rich information, denoting the business name, street address, direction and other crucial messages. So the one of contributions in this paper is recognizing printed characters in natural images and locating the characters exactly axis in the street view. After extracting the embedded characters, the mapping application associating the scene-text and the locations with the captured street view images is proposed (as Fig 1 shown).

Microsoft is progressing the research of handwritten Chinese character recognition (HCCR) and collaborates with some distinguished Universities in China to develop the advanced techniques, one of which member is us, Wuhan University. Google as a search engine is also committing them to capture text from street view images and associate the locations with these images, so the extracted text from street scene images can be indexed and associated with a mapping application [2] in Google earth. So we receive enlightenment and propose a state-of-the-art method to address the problem of character recognition.

The next of paper is organized as follows. In section 2, preprocessings and feature extraction are described. Section 3 described the rough classification. Section 4 mainly describes the classification with SVM. Section 5 discusses the system implementation and presents the experimental results and performance, explanations. Section 6 provides the conclusion and the future work.

II. Preprocessing and feature extraction

In this section, preprocessings and feature extraction are described.

A. Preprocessing

There are four preprocessings in our method, noise reduction, size normalization, smoothing, and thinning. Noise reduction removes small size of connected components from original images. Then, we smooth surface of the image. After that, the image is normalized into the size of 64×64 pixels. Lastly, Hilditch's thinning process is applied to find a skeleton of character. The thinning process is necessary to extract directional element features, shown below.

B. Feature Extraction

Fig. 2-(a) shows an example of an image after applying all preprocessings. Directional element features (DEFs) [3] represent characteristic features of Japanese characters very well. Thus, we use DEFs as feature vectors of Chinese characters. The following explains what DEFs are. Consider Fig 2-(a). The image is divided into the 64 squares, each of them is in size 8 × 8 pixels, and then four adjacent squares group one sub-region. As the result, we have 49 sub-regions, each of them is a 16×16 pixel square. Then, every sub-region is divided into four parts 4, 3, 2 and 1, as shown in Fig 2-(b). Part 4 is a 4 × 4 pixel square in the center of the sub-region. Part 3 is a 8×8 pixel square, but exclusive part 4. Parts 2 and 1 are obtained in a similar manner to part 3.

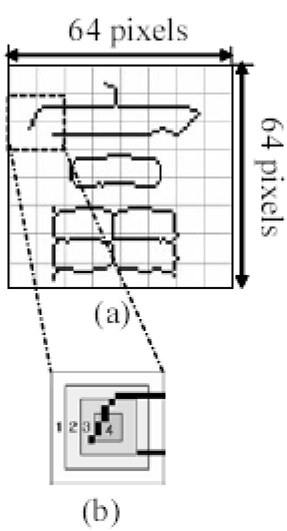

Fig 2. Example of DEF

Weighting factors are defined in each part of sub-regions. The weighting factors are graded according to the position of parts. Practically, we define the weighting factors as 4, 3, 2, and 1 from the center to the edge. A four-dimensional vector (x_1, x_2, x_3, x_4) is defined for every sub-region, where x_1 , x_2 , x_3 , x_4 represent element quantities of four orientations, vertical, horizontal and two oblique lines slanted at ±45°.Each four-dimensional vector (x_1, x_2, x_3, x_4) is calculated below.

$$x_i = 4x_i^{(4)} + 3x_i^{(3)} + 2x_i^{(2)} + x_i^{(1)} \quad (i = 1, 2, 3, 4)$$

Where $x_i^{(4)}, x_i^{(3)}, x_i^{(2)}, x_i^{(1)}$ denote element quantities of each vector in parts 4, 3, 2 and 1. Since each sub-region has four dimensions, DEF of a character is a 196 (=49×4) dimensional vector.

III. Rough classification

In application of support vector machines for large categories of handwritten Chinese character recognition, whether it is training or identification stage, to reduce storage and computation is an important issue. Therefore, we added a rough classifier before SVM classifier [4][5]. The classification flow of method we apply is shown in Fig 3

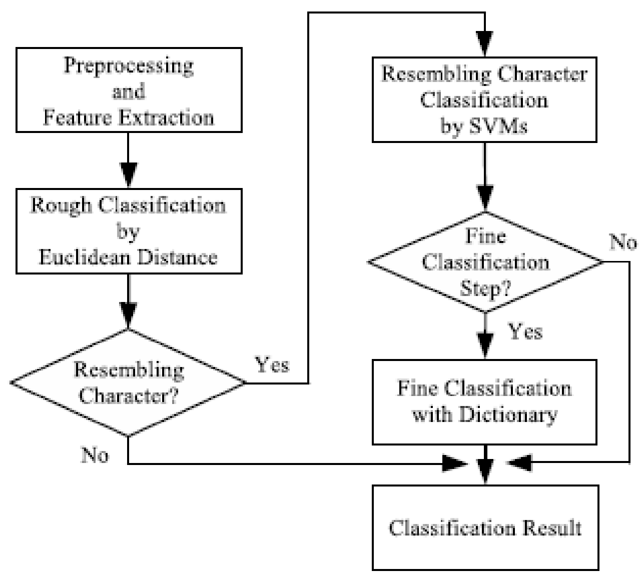

Fig 3. The method of two-classifier

In rough classification, we use the following Euclidean distance as discriminant function.

$$D_i(x) = \sum_{j=1}^{196} \left(y_{ij} - x_j \right)^2, \quad (1 = 1, \ldots, N) \qquad (1)$$

where y_i i is a reference vector of the i-th character, $\boldsymbol{x}$ is a DEF vector of an input pattern, and N is the number of characters. A reference vector is a mean DEF vector of training samples, in this paper. Then, in rough classification, an input pattern x is classified into the following class c.

$$c = \arg \min_{i=1,\ldots,N} D_i(x) \quad (2)$$

IV. Classification with SVM

Since there are many resembling Chinese characters, it is difficult to accomplish to classify Chinese characters correctly when we use only Euclidean distance as the discriminant function. Support vector machine is a remarkable machine learning algorithm. This is because many reports show that support vector machines perform well on various pattern classification problems. In this paper, we introduce Chinese character classification method with SVMs.

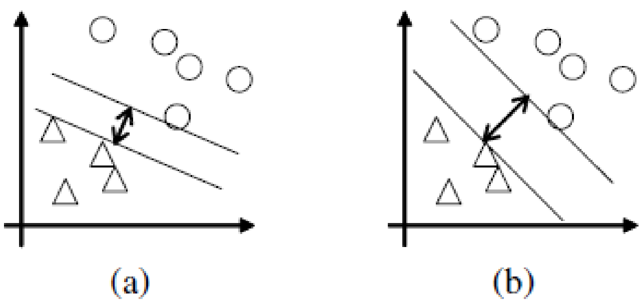

Fig 4. An example of two-class classification in SVM

In the 1990s, a new type of learning algorithm was developed, based on results from statistical learning theory. It is called support vector machines (SVMs). One of the advantages is that once we found support vectors of two classes, the SVM gives us the maximal margin between the two classes (see Fig 4-(a)).

The tutorial by Burges [6] gives a good overview of two category classification SVMs. Given a set of linear separable training samples $S = \{(x_1, y_1), \ldots, (x_s, y_s)\}$, where $x_i \in R^n$ is a training vector and $y_i \in \{-1, +1\}$ is its corresponding target label. The main task in training a SVM is to solve the following quadratic optimization problem.

$$\text{Maximize:} \quad W(\alpha) = \sum_{i=1}^{8} \alpha_i - \frac{1}{2} \sum_{j,j=1}^{8} y_i y_j \alpha_i \alpha_j \langle x_i \cdot x_j \rangle$$

$$\text{Subject:} \quad \sum_{i=1}^{8} y_i \alpha_i = 0, 0 \le \alpha_i \le \ C(i = 1, \ldots, s)$$

Where $\alpha_i \ (i = 1, \ldots, s)$ a Lagrangian multipliers and C is is an upper bound of all Lagrangian multipliers, controlling the penalty to the training errors.

This quadratic optimization problem can only solve linearly separable problems. If the inner product $\langle x_i, x_j \rangle$ of the objective function is exchanged with a special function $K(x_i, x_j)$, called a kernel function, then SVMs can solve nonlinear separable problems. The discriminant function f of two-class SVM is given below.

$$f(x) = \sum_{i=1}^{8} y_i \alpha_i K(x_i, x) + b \quad (3)$$

Where b is a bias. Then, an input pattern $\boldsymbol{x}$ is classified in the following way.

$$belongs \quad to \quad class \begin{cases} +1 & if \quad D(x) > 0 \\ -1 & if \quad D(x) < 0 \end{cases}$$

Recently, many methods for solving the optimization problem are proposed [7]. In this paper, we apply LIBSVM [8], which is a powerful tool of training SVMs.

For multi-class problems, 1-against-1 or 1-against-all is common strategies. Hsu *et al.* [9] gives a detailed report on performance of the two methods, and their experimental results suggest that 1-against-1 is superior to 1-against-all. Therefore, we apply 1-against-1 for our Chinese character classification problem.

Consider an n-class problem. For a 1-against-1 SVM, we convert the n-class problem into $n(n-1)/2$ two-class problems, which cover all pairs of classes. Let f_{ij} be a decision function between classes i and j, with the maximum margin that separates class i from class j, below.

$$f_{ij}(x) = \sum_{i=1}^{8} y_k \alpha_k^{ij} K^{ij}(x_k, x) + b^{ij} \quad (4)$$

Let N_i be the number of classes j which satisfy $f_{ij} > 0$.

An input pattern $\boldsymbol{x}$ is classified into the largest vote, that is, it is classified into the following class c.

$$c = \arg \max_{i=1,\ldots,N} N_i$$

In case where two classes have identical votes, thought it may not be a good strategy, we simply select the one with the smaller index. We selected the following radial basis function (RBF) $K(x_i, x_j)$ as kernel function.

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (5)$$

Where $\|\cdot\|$ is the Euclidean norm, and γ is a parameter.

V. Experiment

To evaluate the performance of the proposed method, we did experiments with randomly chosen street view images.

A. The datasets

The proposed method has been evaluated through experiments with two large and diverse datasets, which one is used for training two classifiers and the other is used for testing them.

The training dataset is supplied by Recognition Research Platform; the platform is sponsored by Microsoft Research Asia (MSRA).

Fig 5. 3S integration car

The testing dataset of 128 images is got by the 3S car (Fig 5) with multi-sensors integration. The moving data-capture system used GPS/INS, Lidar and four cameras mounted in front of a car. The system processes GPS, image and Lidar data at the same time. The images of street view can be got (as Fig 6 shown) by this car. Some preprocess can be done in the same time.

B. The proceedings of experiment

After preparing the training dataset and testing dataset, the following steps demonstrate the flow of our experiment.

1) **Creating samples** Before the recognition, characters must be detected from images. The detection was responsible for my partner. The correct results of detection are selected as samples of testing; some correct results are shown in Fig.6

Fig 6. The correct results of detection

2) **Training and testing classifier** In each training stage, more than 3,000 typical Chinese characters were used at one time to train two-classifier. During the testing stage, 100 samples are used to examine the performance of two classifiers. The can classify almost all Chinese characters correctly, but some resembling characters are misclassified. However, these resembling characters eventually can be classified correctly by SVM.

3) **The results** Our final classifier performed well with 100 samples, 431characters. Tab 1 shows classification of all classifiers, the examples of recognition result are shown in Fig 7. Most characters are classified by rough classifier. Fig 7(a) (b) shows the correct recognition. Even if after fine classification, some characters are misclassified because of their small size, excessive distortion or low contrast, as shown in(Fig 7(c)(d)).

Fig 7. Text detection results

Tab 1.
The classification rate of all classifiers

Classifier	Rough classifier	Fine classifier	Tow-classifier
Classification rate	83.29% (359/431)	45.83% (33/72)	90.95% (392/431)

VI. Conclusions

A novel application of Tow-Classifier to recognize Chinese character in images from street views to the complex natural scene, and map the classifying results to the real surroundings of target place is presented in this paper. Because of the variability of the street view images, the accuracy of recognition is limited. The further research is how to recover the character from distortion during detection; more effective preprocessings which can improve the accuracy of recognition are going to research in the future.

Acknowledgement

This paper was supported by

(1)The Research on Self-adaptive Service Strategies Mec hanism in Ubiquitous Computing Environment, Open Fund o f the State Key Laboratory of Software Engineering, Wuhan , China. SKLSE20080714, 2009-2010.

(2)The Research and Application of Auto-abstracts for W eb Page Documents Based on Graph Models, Hubei Natural Fund Project, 2008CDB343, 2009-2010.

(3) Feature Extraction Based on Multi-Feature Integration for Handwritten Chinese Character Recognition, Asia Research Centre of Microsoft Co.

References

[1] X. R. Chen, Alan L. Yuille. “Detecting and reading text in natural scenes”, *in IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol.2, pp. 366-373, June 2004

[2] Bill Slawski. “Google on reading text in images from street view, store shelves, and museum interiors”, Available: http://www.seobythesea.com/?p=952

[3] N. Sun, M. Abe and Y. Nemoto, “A Handwritten Character Recognition System by Using Improved Directional Element Feature and Subspace Method”, *IEICE Transactions on Information and Systems*

[4] GAO Xue1,JINLian-wen1,YIN Jun-xun1,HUANG Jian-cheng2,” A New SVM-Based Handwritten Chinese Character Recognition Method”, Motorola China Research Center,Shanghai200002,China

[5] Platt J C. Sequential minimal optimization: a fast algorithm for training support vector machines. advances in kernel methods-support vector learning [A].Cambridge, MA: MITPress [C].1999.185-208.

[6] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, Data Mining and Knowledge Discovery, Vol.2, No.2, pp.121-167, 1998.

[7] J. X. Dong, A. Krzy˙zak and C. Y. Suen, “Fast SVM Training Algorithm with Decomposition on Very Large Data Sets”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27,

[8] C. C. Chang and C. J. Lin, “Libsvm: A Library for Support Vector Machines”, *Technical Report of Dept. of Computer Science and Information Eng.*, National Taiwan University, 2007.No.4, pp.603-618, 2005.

[9] C. W. Hsu and C. J. Lin, “A Comparison of Methods for Multi-class Support Vector Machines”, *IEEE Transactions on Neural Networks*,Vol.13, pp.415-425, 2002.