IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-5, Issue-8 E-ISSN: 2347-2693

Comparison of Structure Based Models for Handwritten English Character Recognition

Bhavana Shastry M.[1*], Pradeep N.[2]

[1*]Dept. of CSE, Bapuji Institute of Engineering and Technology, Davanagere, India
[2]Dept. of CSE, Bapuji Institute of Engineering and Technology, Davanagere, India

**Corresponding Author: bhavanashastry13@gmail.com*

Available online at: www.ijcseonline.org

Received: 12/Jul/2017, Revised: 23/Jul/2017, Accepted: 16/Aug/2017, Published: 30/Aug/2017

***Abstract*:** Characters are the symbols made by man that are composed of different structure and strokes for easy communication. The intrinsic characteristics of the characters can be utilized to design the stroke and structure based models for handwritten character recognition. This paper focus to learn the part based and the stroke detector based models to recognize the characters by detecting the elastic strokes. The Tree Structured Model (TSM) and the Mixture of parts Tree Structured Model (MTSM) are the part based models that uses the trained part models on the images to recognize the characters. These models require manually labelled key points. In order to learn the discriminative stroke detectors automatically, the discriminative spatiality embedded dictionary learning-based representation (DSEDR) is used for character recognition. A comparative study is made on all the three models on the chars74k dataset to determine the model that shows the best performance.

***Keywords*:** Character recognition, stroke detector, codewords, spatially embedded dictionary, part based model.

I. INTRODUCTION

Text provides the semantic information and is the most important way to express and communicate. Thus, over the past several decades, text recognition has been a hot topic. Text recognition finds its application in several fields such as image understanding, automatic sign recognition and translation, navigation, automatic geocoding. Many algorithms are proposed to deal with text detection, recognition and the end-to-end recognition [1]. However, in recent years, more and more researchers pay their attention on the recognition of hand written characters, as even if the character is perfectly detected, the recognition accuracy is still far from satisfactory for practical requirement, due to the challenges in the characters because of various styles of hand writing [2].

Although conventional Optical Character Recognition (OCR) has been considered as a solved problem and has been successfully applied in many fields, handwritten character recognition still has many obstacles. Most recently published methods consider the characters as a special category of objects. They take advantages of feature extraction, representation or classification methods that perform well in object detection or recognition tasks and directly apply these methods on character recognition [19]. These methods have shown promising performance compared to conventional OCR ones. However, there is still much room for improvement, since most of the existing object-recognition based methods are not specially designed to make full use of the unique characteristic of these man-made characters [8].

This paper explores the feasibility of using the stroke and structure information for the handwritten character recognition and gives an analysis of the performance and suitability of stroke detector based methods for character recognition. Considering the success of part-based models on object detection as well as the elastic structure information used by the model, it is required to first explore the suitability of applying part-based models on character recognition. The part-based models such as the part based tree structure model (TSM) and mixtures of parts tree structured model (MTSM) [1] can be used to recognize characters by detecting the elastic stroke like parts.

The TSM and MTSM [2]rely on the manually designed tree structure and the size of the parts which is designed by placing the points on the characters. In order to learn the discriminative stroke detectors automatically, the discriminative spatiality embedded dictionary learning based representation (DSEDR) [4] can be used for character recognition using the BOW framework. A comparative study on performance is performed on these three models by considering the chars74k dataset. The experimental results shows a better performance on the dataset.

II. RELATED WORK

Optical character recognition (OCR) has been studied for several decades and many effective approaches have been proposed for conventional OCR. This section reviews recently published work on recognition of hand written characters based on object recognition techniques, which includes the techniques to recognize the challenging hand written characters.

Cun-Zhao *et al*. [1] proposed the use of the structure based models for the character recognition. The work presents the four structure based models for the recognition of the characters collected from different datasets.

Shi *et al.* [2] described the usage of the part-based TSM in order to find out the characters present in scenic images. As this model can utilize both the local structure knowledge and the global appearance information, it is possible to cope up with characters that are challenging in the aspects of resolution and writing styles. Yao *et al.* [3] presented a stroke based model that used multiple scale presentation for recognition of characters. The presentation includes a set of findable primaries, that captured the needed sub strokes of the characters at unique positions.

Song GAO *et al.* [4] proposed a higher way of presentation called stroke bank for the recognition in order to train the detectors and use their maximally obtained output as features. This model learns the strokes for the character detection with the use of SED.

De Campos *et al.* [5] benchmarked the performance of various features to assess the feasibility of posing the problem as an object recognition task and showed that Geometric Blur and Shape Context in conjunction with Nearest Neighbor (NN) classifier, performed better than other methods.

Smith *et al.* [9] also proposed to incorporate character similarity information to improve recognition performance. Tian *et al.* [6] proposed to use co-occurrence of histogram of oriented gradients to recognize scene characters and reported better results than HOG.

Netzer *et al.* [11] proposed to recognize digits in natural scenes using unsupervised feature learning methods and experimental results demonstrated the major advantages of learned representations over hand crafted ones. Neither the learned features or the hand crafted ones could deal with characters with large deformations, low resolution or distortions, especially when the training samples could not include all the factors mentioned above.

Wang *et al.* [15] used Convolutional neural networks (CNN) to recognize English and digits characters in natural scene images and achieved satisfactory performance when using the original training set as well as those synthetic ones.

Yang and Ramanan [12] proposed to detect articulated pose of human using flexible mixtures-of-parts. Tree structure is used to model co-occurrence and spatial relations, and the model could be efficiently optimized with dynamic programming. Experimental results on standard benchmarks for pose estimation demonstrate that their approach outperforms past work by 50% while being orders of magnitude faster.

Xiangxin and Ramanan [8] proposed to jointly address the tasks of face detection, pose estimation, and landmark estimation using mixtures of trees with a shared pool of parts. Although their model is only trained with hundreds of faces, it compares favourably to commercial systems trained with billions of examples.

III. METHODOLOGY

The character recognition model takes the image as the input and these images are recognized by different character recognition models. The proposed system for the recognition of the characters are as shown in the figure 1.

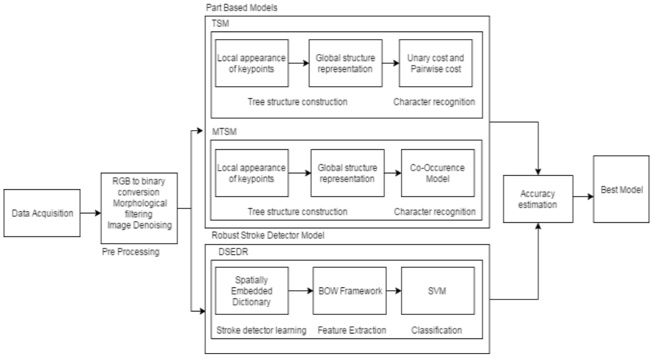
Part Based Models
TSM
Local appearance of keypoints
Global structure representation
Unary cost and Pairwise cost
Tree structure construction
Character recognition
MTSM
Local appearance of keypoints
Global structure representation
Co-Occurence Model
Tree structure construction
Character recognition
Data Acquisition
RGB to binary conversion Morphological filtering Image Denoising
Pre Processing
Robust Stroke Detector Model
DSEDR
Spatially Embedded Dictionary
BOW Framework
SVM
Stroke detector learning
Feature Extraction
Classification
Accuracy estimation
Best Model

Figure 1: Proposed system for the character recognition accuracy for different structure based models

3.1 Data Acquisition

The data acquisition step consists of collection of related data from the dataset. The dataset used in this project is **chars74k** [5]. The chars74k dataset consists of handwritten character images. There are 62 character classes [0-9] [A-Z] [a-z] into which images can be categorized. Each class consists of 50 instances of the image. All of these images are written in black and on a white background. The images consist of 900*1200 pixels resolution [20] and the images have been placed at different positions. Figure 2 an example of the dataset '0'.

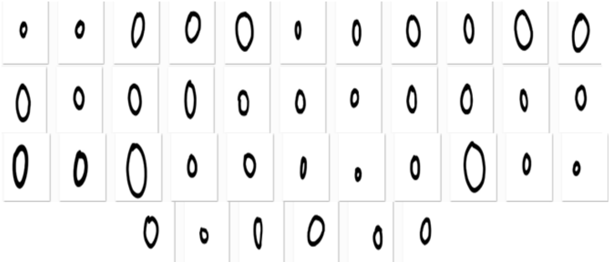

Figure 2: Image samples of chars74k dataset

3.2 Tree Structured Model (TSM)

TSM is a part based model [1]. Each category of characters is represented by a tree $T_k = (V_k , E_k)$, where k is the index of the model for different structures, V_k represents the nodes and E_k specifies the topological relations of nodes. Each node represents a part of the character. Let I represents the input image and $l_i = (x_i , y_i)$ denotes the location of part i . The score of the configuration could be defined as

$S(L, I, k) = S_{App}(L, I, k) + S_{Str} (L, k) + \alpha k$

As it can be seen, the total score of a configuration L for model k consists of the local appearance score, the structure or shape score, and the bias αk [2].

Local Appearance Model*:* The equation shown below is the local appearance model which reflects the suitability of putting the part based models on the corresponding positions [10].

$$S_{App}(L, I, k) = \sum_{i \in V_k} w_i^k \cdot \phi(I, l_i)$$

w^k_i represents the filter for part i , structure k, and $\varphi(I, li)$ denotes the feature vector extracted from location li . HOG is chosen as the local appearance descriptor due to its good performance on many computer vision tasks. The part models reflect the shapes of certain parts of characters, which could also be regarded as a stroke or sub-stroke of characters [17].

Global Structure Model*:* The equation below shows the structure or shape model which scores the character-specific global structure arrangement of configuration L.

$$S_{Str}(L, k) = \sum_{ij \in E_k} w_{ij}^k \cdot \psi(l_i - l_j)$$

Here we set $\psi(l_i - l_j) = [d_x\ d_{x2}\ d_y\ d_{y2}]$, where $d_x = x_i - x_j$ and $d_y = y_i - y_j$ are the relative distance from part i to part j . Each term in the sum acts as a spring that constrains the relative spatial positions between a pair of parts. The parameters w^k_{ij}, which are learned in the training process, could control the location of each part relative to its parent and the rigidity of each spring [1]. The part models could be located within a certain range and each location would be linked with a cost according to the learned $w^k_{ij.}$

3.3 Mixture of parts Tree Structured Model (MTSM)

MTSM is also a part based model. The part-based tree-structured model (TSM) only uses one component to represent each part. It might fail to detect characters with large deformation, rotation or distortion. Mixtures-of-parts tree-structured model (MTSM) [2]can be used to overcome this problem. Let $l_i = (x_i , y_i)$ be the pixel location of part i and t_i be the mixture component of part i . Here t_i is the type of part i. $G = (V, E)$ represents a K-node relational graph whose edges specify which pairs of parts are constrained to have consistent relations.

Co-occurrence model: This model uses a specific type of assignment for a part and the pairwise parameter finds the co-occurrence of the parts [5]. The formula below shows the co-occurrence model

$$S(t) = \sum_{i \in V} b_i^{t_i} + \sum_{ij \in E} b_{ij}^{t_i,t_j}$$

MTSM also considers the local appearance model to find the correctness of placing the part model in right position. The characters with different writing styles are made to recognise by considering the Tree structured model and a large number of combinations of trees can be obtained considering the different parts together [10].

3.4 Discriminative Spatially Embedded Dictionary learning based Representation (DSEDR)

Unlike TSM, MTSM and DMSDR which need the part or key points labels, DSEDR learns the spatiality embedded codewords automatically. Different from conventional codeword learning methods, which use K-means to cluster the descriptors regardless of their positions, the DSEDR generates position related codewords by associating each codeword with a response region. Given the spatiality embedded dictionary, coding could be performed locally and also spatially according to the spatial relationship between codeword and descriptor [1].

Learning Spatiality Embedded Dictionary**:** Each training image is normalized to W = H = 64 and partitioned into $n_h \times n_w$ blocks. Suppose the HOG features within each block have n_{hog} dimensions, a feature vector of $n_h \times n_w \times n_{hog}$ dimensions is used to represent each image. Based on the overall representations, for each category of characters, K-means clustering is used to get centers. For each of these clustering centers, the 1D overall representation is reshaped to 3D matrices with three dimension sizes of n_h , n_w and $n_{hog.}$ For each codeword d_j , a response region $r j$, is recorded which should include the codeword sampling window. All the codewords sampled from the clustering centers of all the categories constitute the learned spatiality embedded dictionary $D_{SED} = \{(d_1, r_1), (d_2, r_2), (d_3, r_3), . . . , (d_n, r_n)\}$.[4]

Feature Extraction and Classification: Given the learned codewords, features are extracted for each character image using the BOW framework. Since each codeword has its own response region, coding could be performed locally and spatially according to codewords' response regions. For each local descriptor extracted from position *(x_i, y_i)* of a certain image, it is coded with those codewords whose response region contain point *(x_i, y_i)*. Linear SVM is used for classification.[3]

IV. RESULTS AND DISCUSSION

This section gives the thorough evaluation of the performance of TSM, MTSM, and DSEDR on the chars74k dataset. Comparative experiments and analysis of the proposed three methods are given using different numbers of training samples. Testing is performed by 70-30 rule on all the three models. That is, out of 3100 images 2170 images are considered for the purpose of training and the rest 930 images are considered for testing. The results obtained for testing are tabulated as shown.

Table 1: Comparison results of different structure based models on chars74k dataset

Model	Training images	Test Images	Classified images	Misclassified images	Recognition rate
TSM	2170	930	722	208	77.63
MTSM	2170	930	779	151	83.76
DSEDR	2170	930	810	120	87.09

The table shows the characters correctly classified characters and the misclassified characters. The recognition rate for TSM is 77.63 and for MTSM 83.76 and for DSEDR it is 87.09. The recognition rate for DSEDR is highest and hence DSEDR is considered to be the best model.

V. CONCLUSION

This paper uses the part based models, TSM and MTSM and the stroke detector model such as DSEDR for character recognition. Experimental results on the chars74k dataset demonstrate the effectiveness of the part based methods and stroke detector based methods for recognizing characters. The results indicate that TSM and MTSM could be effectively trained with labelled samples. The results also show that MTSM and DSEDR are even more effective when compared with TSM. Furthermore, the results reveal that, the DSEDR model which does not need the part labels, can learn the position related codewords (strokes) automatically. In the future, these methods can be applied the stroke based methods on recognition of other languages.

REFERENCES

[1] Cun-Zhao Shi, Song Gao, Meng-Tao Liu, Cheng-Zuo Qi, Chun-Heng Wang, "Stroke detector and structure based models for Character Recognition: A Comparative Study" *in IEEE transactions on image processing, volume 24*, no. 12, Dec 2015.

[2] C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multi-scale representation for scene text recognition," in the Proceedings CVPR, Jun. 2014, pp. 4042–4049.

[3] S. Gao, C. Wang, B. Xiao, C. Shi, and Z. Zhang, "Stroke bank: A high level representation for scene character recognition," *in the Proceedings on 22nd International Conerence Pattern Recognition* (ICPR), Aug. 2014, pp. 2909–2913.

[4] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, "Scene text recognition using part-based tree-structured character detection," *In the Proceedings of IEEE Conference Computer Vision and Pattern Recognition* (CVPR), Jun. 2013, pp. 2961–2968.

[5] T. E. de Campos, B. R. Babu, and M. Varma, "Character recognition in natural images," *In the Proceedings of VISAPP, 2013, pp. 273–280.*

[6] S. Tian, S. Lu, B. Su, and C. L. Tan, "Scene text recognition using co-occurrence of histogram of oriented gradients," in the *Proceedings on 12th International Conference Document Analytical Recognition. (ICDAR),* Aug. 2013, pp. 912–916.

[7] L. Wang, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, "Regularization of Neural Networks using drop connect," in *Proceedings 30th International Conference Machine Learning (ICML),* 2013, pp. 1058–1066.

[8] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proceedings CVPR,* Jun. 2011, pp. 2879–2886.

[9] D. L. Smith, J. Field, and E. Learned-Miller, "Enforcing similarity constraints with integer programming for better scene text recognition," in *Proceedings CVPR*, Jun. 2011, pp. 73–80.

[10] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in *Proceedings Asian Conference Computer Vision*, 2011, pp. 770–783.

[11] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "PhotoOCR: Reading text in uncontrolled conditions," in *Proceedings IEEE International Conference Computer Vision*, Dec. 2011, pp. 785–792.

[12] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of- parts," in *Proceedings CVPR*, Jun. 2011, pp. 1385–1392.

[13] T. E. de Campos, B. R. Babu, and M. Varma, "Character recognition in natural images," in *Proceedings VISAPP*, 2009, pp. 273–280.

[14] C. Yi, X. Yang, and Y. Tian, "Feature representations for scene text character recognition: A comparative study," *in Proceedings IEEE 12th International Conference Document Analytics Recognition* Aug. 2011, pp. 907–911.

[15] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proceedings 21st International Conference Pattern Recognition (ICPR),* Nov. 2012, pp. 3304–3308.

[16] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition" in *Proceedings Computer Vision.*, vol. 61, no. 1, pp. 55–79, Jan. 2005.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings CVPR*, vol. 1. Jun. 2005, pp. 886–893.

[18] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 robust reading competitions," in *Proceedings ICDAR*, vol. 2. Aug. 2003, pp. 682–687.

[19] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: Benchmarking of state-of-the-art techniques," *Pattern Recognition.*, volume 36, no. 10, pp. 2271–2285, Oct. 2003.