Character Recognition in Natural Scene Images

Proposal: Pattern Recognition and Machine Learning Pipeline

Yash Sinha, Prateek Jain & Nirant Kasliwal

Department of Computer Science & Information Systems, BITS Pilani

Abstract

This paper presents an approach to character recognition in natural scene images. Recognizing such text is a challenging problem in the field of Computer Vision, more than the recognition of scanned documents due to several reasons.

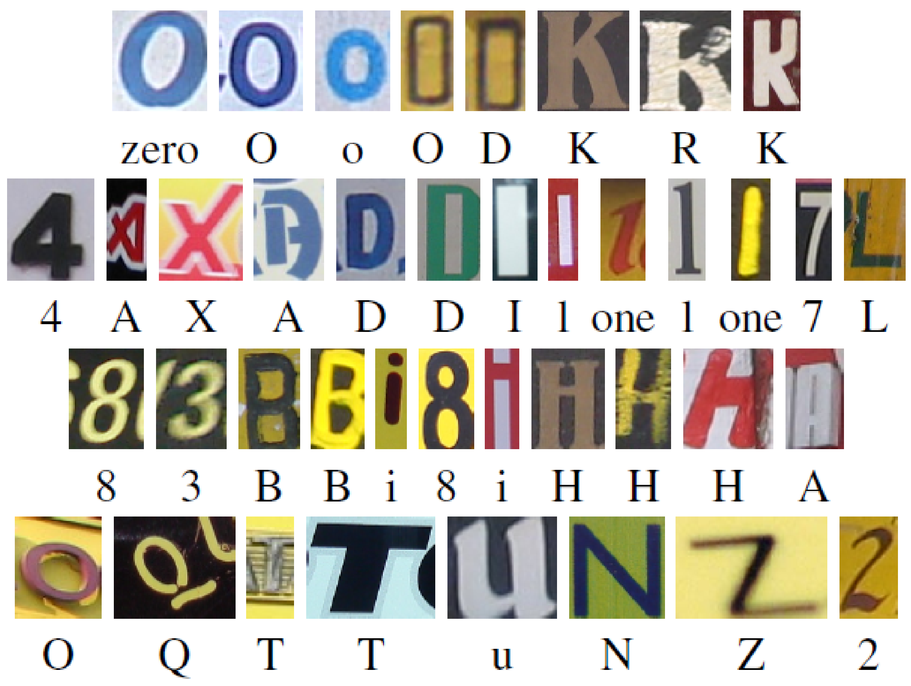
We propose a classification technique for classifying characters based on a pipeline of image processing operations and ensemble machine learning techniques. Our proposal comprises of a sequence of image processing operations: resizing, binarization, morphological opening and median filtering on the images to handle background clutter, noise, multi-sized and multi-oriented characters and variance in illumination.

Next, we use image pixels and Histogram of Oriented Gradients (HOG) as features to train three different machine learning models based on Nearest Neighbour, Random Forest and Extra Tree classifiers.

After pre-processing, HOG features were extracted and fed into Extra Tree Classifier, the model is an improvement over the known state-of-the-art accuracy by more than one percent on the complete Chars74k dataset.

Dataset

1. The data is taken from the standard Chars74K dataset, which consists of images of characters selected from Google Street View images.



Sample images from the Chars74k dataset

2. The data set differentiates between digits, uppercase and lowercase characters to give a total of 62 classes(0-9, a-z, A-Z).
3. The dataset has 12503 characters, of which 6283 (50.25%) are randomly sampled for training while rest 6220 (49.75%) are used as test data.
4. Handwritten and synthetic images were not used. Each character is segment

Experiments

We focus on improving the categorization accuracy of recognition of individual characters from natural scene images.

Proposed Methodology: The proposed methodology consists of four steps:

1. Image processing *tackles background clutter, fonts and illumination of images*: This step includes converting all images to a common standard size, binarization using Otsuś Thresholding followed by Morphological opening and median filter for removing salt and pepper noise.

Image Processing: Original image, Thresholding and Opening

2. Feature extraction *addresses variation in character size and orientation*: Histogram of Oriented Gradients is used as a method of feature extraction. The value of used cell size is calibrated experimentally to an optimal value. It also reduces the number of features in comparison to using direct pixel values.
3. Machine Learning *the learning for classifier models happens and classifies test images into classes*: kNN, Random Forest and Extra Tree classifier are tested and compared. The number of estimators are determined experimentally to an optimal value. The input parameters HOG features and pixel values both are tested and compared.
4. Performance testing *compares various classification and feature extraction techniques*: Cross-validation is performed and the classification model is tested on the test data set. Accuracy obtained from both the methods is mentioned.

Results

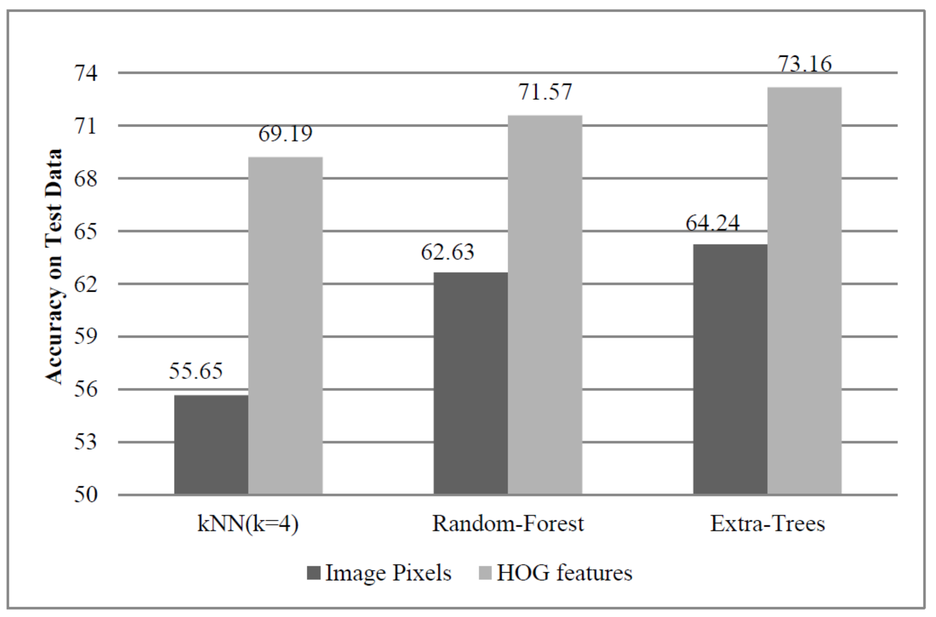
The proposed pipeline performs yields better categorization accuracy on the natural scene dataset from Chars74K than other known methods.

Method	Accuracy	Dataset	Classes
Proposed Method	**73.167%**	**Chars74K**	**62**
Zhang et al. [8]	72.68%	Chars74K(mixed)	62
Fraz et al. [2]	72%	Chars74K-15	49
Shi et al. [7]	69.9%	Chars74K-15	62
Newell et al. [5]	66.5%	Chars74K-15	62
Campos et al. [1]	55.26%	Chars74k	62

Table 1: Comparison with other state-of-the-art techniques

Comparative Study:

1. On cross validation, HOG gave 10% better accuracy than pixel data (using the the Extra Tree Classifier)
2. Otsu's Thresholding with image Pixels gave less accuracy than average thresholding with Image Pixels. In contrast, the accuracy obtained using HOG Features with Otsu thresholding was 0.4% more than HOG Features with average thresholding.
3. On using kNN, k=1 yielded the highest accuracy with image pixels as features. k=4 yielded the highest categorization accuracy with HOG.
4. Extra Trees classifier performed better than both Random Forest and kNN (on both pixel data and HOG features)



Features	kNN(k = 4)	Random Forest	Extra Tree
HOG Features	69.549%	71.575%	**73.167%**
Image Pixels	55.659%	62.636%	64.244%

Table 2: Accuracy (Features)

Future Prospects

Kernel Regression for Image Processing and Reconstruction is a less investigated yet intriguing idea. Shape context can be possibly used for better feature extraction.

In the machine learning pipeline, boosting and dimensionality reduction can be applied to improve accuracy. Convolutional neural networks is definitely a promising alternative.

References

[1] T. E. de Campos, B. R. Babu, and M. Varma. Character recognition in natural images. In *Proceedings of the International Conference on Computer Vision Theory and Applications, Lisbon, Portugal*, February 2009.

[2] Muhammad Fraz, M Saquib Sarfraz, Eran A Edirisinghe, and UK Loughborough. Exploiting colour information for better scene text recognition.

[3] Deepak Kumar and AG Ramakrishnan. Recognition of kannada characters extracted from scene images. In *Proceeding of the workshop on Document Analysis and Recognition*, pages 15–21. ACM, 2012.

[4] Lukas Neumann and Jiri Matas. A method for text localization and recognition in real-world images. In *Computer Vision–ACCV 2010*, pages 770–783. Springer, 2011.

[5] Andrew J Newell and Lewis D Griffin. Natural image character recognition using oriented basic image features. In *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*, pages 191–196. IEEE, 2011.

[6] Karthik Sheshadri and Santosh Kumar Divvala. Exemplar driven character recognition in the wild. In *BMVC*, pages 1–10, 2012.

[7] Cun-Zhao SHI, Chun-Heng WANG, Bai-Hua XIAO, Yang ZHANG, and Song GAO. Multi-scale graph-matching based kernel for character recognition from natural scenes. *Acta Automatica Sinica*, 40(4):751–756, 2014.

[8] Ziming Zhang, Paul Sturgess, Sunando Sengupta, Nigel Crook, and Philip Torr. Efficient discriminative learning of parametric nearest neighbor classifiers. 2012.

Acknowledgements

We thank Shreya Inamdar, Microsoft and Dr. Sundaresan Raman, BITS Pilani for review of the camera ready paper. We thank Dr. Navneet Goyal, BITS Pilani for the freedom in pursuing course project which later translated into this paper.