



REVIEW OF RESEARCH



VOLUME - 6 | ISSUE - 9 | JUNE - 2017

DESIGN AND DEVELOP METHODOLOGY FOR DETECTION AND RECOGNITION OF TEXT FROM NATURAL SCENE IMAGES

Prof. Apoorva Raikar



ABSTRACT

Detecting text regions in natural scene images has become an important area due to its various applications. Text Information Extraction (TIE) System involves detecting text regions in a given image, localizing it, extracting the text part and recognizing text using OCR. This work basically concentrates on the detection and extraction of text in natural scene images. In this work, the test image will be pre-processed using RGB to Gray conversion, binarization, Edge Detection method and Geometric based Noise removal method. The features from the pre-processed image are extracted. The extracted features are used by the trained SVM classifier to detect the text regions. After detecting text regions, characters are extracted and finally displayed.

KEYWORDS: Text Information Extraction (TIE); Natural Scene Images; Text Detection; Text Extraction; Support Vector Machine (SVM).

I. INTRODUCTION

Images are the most convenient means of conveying or transmitting information as they quickly convey information about positions, sizes and inter-relationships between the objects. They depict spatial information such that it can be recognized as an object. Human beings are very good at deriving information from the images, because of our instinctive visual and mental abilities. About 75% of the information received by humans is in the pictorial form. The analyses of pictures that employ an overhead perspective, including the radiation invisible to human eye are considered.

Textual part of still images or video frames of a scene does not have a priori knowledge of environment, objects supporting text, lighting, acquisition parameters and finally text itself. It can be easily viewed as text in real-world conditions without any constraints or assumptions.

Text detection and text extraction in natural scene images is challenging because of the varied conditions under which the image is taken. Natural scene text understanding aiming at extracting text from daily images is the main concern of this text. There are several information sources for text information extraction in images (Ex color, texture, motion, geometry, shape etc.)

It is always advantageous to merge various information sources to enhance the performance of text information extraction system. Automatic text recognition from the natural scene images receives a growing attention because of potential applications in image retrieval, robotics and intelligent transport system.

II. RELATED WORKS

An earlier method proposed by Basilio Gatos et al. based on an efficient binarization and enhancement technique followed by a suitable connected component analysis procedure. Image binarization successfully processes natural scene images, which is having shadows, non-uniform illumination, low contrast and large signal dependent noise. Connected component analysis is used to define the lastly binary images that mainly consist of text regions. The proposed methodology results in increased success rates for commercial OCR engines. Experimental results based on public database of natural scene images prove the efficiency of the proposed approach.

Another method proposed by Nobuo Ezaki et al. describes the system design and evaluates several character extraction methods. System tries to find in the image areas with small characters present in the test images. Then it zooms into found areas to retake higher resolution images necessary for character recognition. Four character-extraction methods based on connected components are proposed. They are sobel edge detection, Otsu binarization, connected component extraction and Rule-based connected component filtering. Since four methods are used its time consuming method.

Hideaki Goto et al. proposed that some spatial frequency-based features used for text region detection in natural scene images are analyzed, and the DCT-based feature are redefined. Fisher's discriminant analysis was in use to improve the DCT-based feature and to achieve higher accuracy. An unsupervised thresholding method for discriminating text and non-text regions is introduced and tested. Experimental results show that a wide high frequency band, covering some lower-middle frequency components, is generally more suitable for scene text detection despite the original definition of the DCT-based feature.

Yi-Feng Pan et al. proposed a hybrid approach that robustly detected and localized texts in natural scene images. A text region detector is designed to estimate the text existing confidence and scale information in image pyramid, which help segment candidate text components by local binarization. To efficiently filter out the non-text components from the image, a conditional random field (CRF) model considering unary component properties and binary contextual component relationships with supervised parameter learning is proposed. Later, text components are grouped into text lines/words with a learning-based energy minimization method. Since all the three stages are learning-based, there are very few parameters requiring manual tuning. This system robustly detects and localizes texts in natural scene images.

Jung-Jin Lee et al. Proposed a novel text detection algorithm that extracts different class features of text and uses the Modest AdaBoost with multi-scale sequential search. Experiments showed that the algorithm could detect the text regions with a $f = 0.70$, which include images with text of various fonts, sizes, colors, alphabets and scripts. Features within 64×32 pixels windows extracted by applying the different types of extraction strategies - X-Y derivatives, local energy of Gabor filter, statistical texture measure of image histogram, measurement of variance of wavelet coefficient, analysis of connected components and edge interval.

Xu-Cheng Yin et al proposed an accurate and robust method for detecting texts in the natural scene images. A fast and effective pruning algorithm was designed in order to extract Maximally Stable Extremal Regions (MSERs) as character candidates using the strategy of minimizing regularized variations. Character candidates are grouped into text candidates by the single-link clustering algorithm, where distance weights and threshold of the clustering algorithm are learned automatically by a novel self-training distance metric learning method. The posterior probabilities of text candidates corresponding to non-text are estimated with a character classifier; text candidates with high probabilities are then eliminated. Finally texts are identified with a text classifier.

III. SYSTEM OVERVIEW

The objective of this paper is to develop an automated system that will be able to detect the location of text in the Natural Scene Images and then extract the text from the images. As mentioned above,

different methods have different its own advantages and followed by its disadvantages. Region based method lower computational cost but more false positive result.

Texture based detection method can detect and localize accurately but speed is problem. Learning based methods give more accurate results but difficult to realize and storage is problem. Clustering based methods are faster but computational complexity is the bottleneck.

In this paper, Text understanding systems include three main phases: text detection, text localization and text extraction. The overall system is shown in Fig 2.

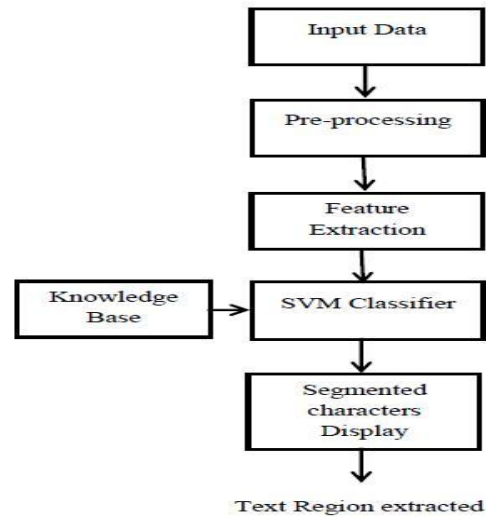


Fig 1: Architecture of Text Detection and Extraction System using SVM classifier

A. Pre-processing

In order to detect the text in the image, given image required to be preprocessed. First converting image into gray scale, then binarization of image is done and later noise is removes and edge detection method is applied on the images and shown in Fig 3.



Fig 2: Stepwise pre-processing of first image

B. Feature Extraction

Feature extraction is a type of dimensionality reduction that efficiently represents interesting parts of an image and those parts are called as compact feature vector. A reduced feature representation is required to quickly complete tasks such as image matching and retrieval. Feature detection, feature

extraction and matching are combined to solve common computer vision problems such as object detection and recognition, texture classification and face detection

The statistical texture measures[6] of image histogram to differentiate text from non-text regions are used. Defining μ as the average of the intensity, Z_i a random variable indicating intensity, $p(Z)$ the histogram of intensity level in the image and L number of possible intensity level, the following are five features.

Average Entropy

$$f1 = - \sum_{i=0}^{L-1} p(Z_i) \log_i p(Z_i)$$

.....(eq.1)

Variance of Histogram

$$f2 = \sum_{i=0}^{L-1} (Z_i - \mu)^2 p(Z_i)$$

.....(eq.2)

Relative Smoothness

If σ is the standard deviation, Relative smoothness is given by

$$f3 = 1 - \frac{1}{1 + \sigma^2}$$

.....(eq. 3)

Skewness

$$f4 = \sum_{i=0}^{L-1} (Z_i - \mu)^3 p(Z_i)$$

.....(eq. 4)

Kurtosis (Peakedness)

$$f5 = \sum_{i=0}^{L-1} (Z_i - \mu)^4 p(Z_i)$$

.....(eq. 5)

Gabor Filter

Text includes letters of a variety of sizes, shapes and orientations. It tends to have higher spatial frequency components compare to non-text

Local energy is being used to extract these high frequency components in four orientations ($\theta = 0, \pi/4, \pi/2$ and $3\pi/4$) with three different radial frequency f (0.2, 0.8 and 0.9) and σ channels ($\sqrt{3.5}, 1$ and $\sqrt{2.5}$).

Horizontal and vertical projections

The projection $H[i]$ along the rows and the projection $V[j]$ along the columns of a binary image are given by

$$V[j] = \sum_{i=0}^{n-1} B[i, j]$$

C. SVM Classifier

Support vector machines (SVMs)[9][10] are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification. The architecture of SVM training has shown in Fig 4.

In training SVM classifier all features of the candidate region are stored in a file. Training of SVM is done for classification. The datasets will have text images and non-text images. Those will be pre-processed and after that features will be extracted and stored in the files. Later this knowledge will be used to test the input test image.

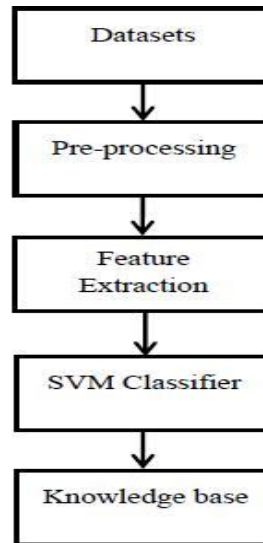


Fig 3: Architecture of training SVM classifier with Text and Non-text data

CONCLUSION

Detection of text in Natural Scene Images is challenging for complex background. There are many methods available to perform the text detection and character extraction in natural scene images. In this work, SVM classifier is used to detect the text. For training SVM eight features such as Average Entropy, Variance of Histogram, Relative smoothness, skewness, kurtosis, Local Energy of Gabor Filter and Horizontal and vertical projections. Experimental results demonstrated the effectiveness of the method by locating most text regions in test images. The selected images are pre-processed, features are extracted and classified using SVM classifier and characters are extracted. Extracted texts are displayed in a window. This method is working well for large size text only.

REFERENCES

1. Beaufort and C. Mancas-Thillou, 'A weighted finite-state framework for correcting errors in natural scene OCR', in Proc. 9th Int. Conf. Document Anal. Recognit., pp. 889–893, 2007.
2. 3.X. Chen, J. Yang, J. Zhang, and A. Waibel, 'Automatic detection and recognition of signs from natural scenes', IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, 2004.
3. Coates et al., 'Text detection and character recognition in scene images with unsupervised feature learning', in Proc. ICDAR, pp. 440–445, 2011.
4. N. Dalal and B. Triggs, 'Histograms of oriented gradients for human detection', in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 886–893, 2005.
5. T. de Campos, B. Babu, and M. Varma, 'Character recognition in natural images', in Proc. VISAPP, 2009.

-
6. Smith, 'An overview of the tesseract OCR engine', in Proc. Int. Conf. Document Anal. Recognit., pp. 629–633 2007.
 7. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, 'ICDAR 2003 robust reading competitions', in Proc. Int. Conf. Document Anal. Recognit., pp. 682–687, 2003.
 8. Zhang and R. Kasturi, 'Extraction of text objects in video documents: Recent progress', in Proc. 8th IAPR Int. Workshop DAS, , pp. 5–17, 2008. 10.Q. Zheng, K. Chen, Y. Zhou, G. Cong, and H. Guan, 'Text localization and recognition in complex scenes using local features', in Proc. 10th ACCV, pp. 121–132, 2010.