International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.1 (2016), pp.43-50
http://dx.doi.org/10.14257/ijsip.2016.9.1.05

A Robust and Real Time Approach for Scene Text Localisation and Recognition in Image Processing

Er.Ananta Singh[1] and Er. Dishant Khosla[2]

M.tech Student, Department Of Electronics and Communication CGC-College of Engineering, Landran[1]
Asst Professor, Department Of Electronics and Communication CGC-College of Engineering, Landran[2]
anantasinghsambyal@gmail.com[1], cgccoe.ece.dk@gmail.com[2]

Abstract

The text localization and recognition in real time scene text images is still a big issue in current application. Mobile application and digitization in real world gives a vital and broad impact on real time scene text images. However, the efficiency of recognition rate depends upon the text localization, i.e., higher the purity of text background segmentation and decomposition, higher the rate of accuracy for the image recognition. In this paper, we present a new scene text detection algorithm based on Stroke detection and Hog Transform method. The method introduces an approach for character detection and recognition which combines the advantages of Hog Transform and Connected Component methods. Characters are detected and recognized on the basis of image regions which contain strokes of specific orientations in a specific relative position, where the strokes are efficiently detected by convolving the image gradient field with a set of oriented bar filters. The method was evaluated on a standard dataset consisting mostly real time images where it achieves state-of-the-art results in both text localization and recognition. The results clearly depict the higher bit of accuracy in terms of localization and recognition for a collected dataset.

Keywords*: Scene Text Localization, Preprocessing, Stroke detection, Segmentation, Hog Transformation*

1. Introduction

Nowadays, the wide use of digital images that are being captured by various application devices such as mobile phone, digital cameras *etc* have achieved more attention from the past decades. Of all the content in the images, the text-based information is of more interest as it gives more meaningful information that is understood by both human and computer and can be used in a variety of applications [1]. The major issues in this area are follows: i) the text intensity affected by lighting conditions such as shadows, low and high light which affects the resolution of natural scene text, ii) a variety of text fonts, colors, orientation *etc*, iii) language of text. Text characters and strings in natural scene can provide valuable information for many applications. Extracting text directly from natural scene images or videos is a challenging task because of diverse text patterns and variant background interferences [2]. Reading text from photographs is a challenging problem that has received a significant amount of attention. Two key components of most systems are (i) text detection from images and (ii) character recognition, and many recent methods have been proposed to design better feature representations and models for both [5]. Since text is a pervasive element in many environments, solving this problem has potential for significant impact. For example, reading scene text can play an important role in navigation for automobiles equipped with

ISSN: 2005-4254 IJSIP

street-facing cameras in outdoor environments and in assisting a blind person to navigate in certain indoor environments (*e.g.,* a grocery store)[4]. Text Localization is of fundamental importance in image understanding and content based retrieval. For instance the localization must always be achieved prior to Optical Character Recognition (OCR) [6].We propose effective algorithms of text localization from detected text regions in scene image. In scene text detection process we apply Stroke Detection and Hog Transformation method. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors and shape contexts that uses overlapping local contrast normalization for improved accuracy. The stoke width transform is used to calculate the width of the most likely stroke containing the pixel in an image. First, all pixels are initialized with ∞ as their stroke width. Then, we calculate the edge map of the image by using the canny edge detector. Morphological Edge Detection is used to detect extract the text from image. Pixel-based layout analysis is adopted to extract text regions and segment text characters in images, based on color uniformity and horizontal and vertical alignment of text characters. In this paper our total work is focused over localization and segmentation of image and finds the higher level of efficiency of the input image.

2. Related Work

Most of the scene text detection algorithms in the literature can be classified into Region-based and Connected Component (CC)-based approaches. Region-based methods adopted a sliding window scheme, which is basically a brute force approach which requires a lot of local decisions. Therefore, the region-based methods have focused on an efficient binary classification (text versus non text) of a small image patch. Text Localization is of fundamental importance in image understanding and content based retrieval. For instance the localization must always be achieved prior to Optical Character Recognition (OCR). Stability of such method includes robustness to noise and blurness because they accomplish features assembled throughout the region of interest. The second approach used is localizing the individual characters using the local parameters of an image (intensity, stroke-width, color, gradient *etc*) . Feature extraction also plays a vital role in image localization process. The main goal of feature extraction is to maximize the recognition rate with minimum number of elements used in it. After analyzing existing feature descriptor methods it is found experimentally Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for character detection and best suited for the proposed system. Many researchers have made research related to this but no technique is almost perfect and they found need to improve the work in more areas at different instants and techniques.

Some of the literature reviews are as follows:

Cong Yao Xiang Bai Baoguang, Shi Wenyu Liu (2014) [1] proposed a method for Multi-Scale Representation for Scene Text Recognition. Though extensively studied, localizing and reading text in uncontrolled environments remain extremely challenging, due to various interference factors. In this paper, they proposed a novel multi-scale representation for scene text recognition. The representation consisted of a set of detectable primitives, termed as stroke lets, which capture the essential substructures of characters at different granularities.

Lukas Neumann Jirı Matas (2012) [3] worked over Real-Time Scene Text Recognition System for the sequential selection from the set of Extremal Regions (ERs). In the first classification stage, the probability of each ER being a character was estimated using novel features calculated with firstly, complexity per region tested. Only ERs with locally maximal probability were selected for the second stage, where the classification

was improved using more computationally expensive features. A highly efficient exhaustive search with feedback loops which then applied to group ERs into words and to select the most probable character segmentation. Finally, text was recognized in an OCR stage trained using synthetic fonts.

Kai Wang, Boris Babenko and Serge Belongie (2011) [4] proposed a methodology for End-to-End Scene Text Recognition. They focused on the problem of word detection and recognition in natural images. The problem was significantly more challenging than reading text in scanned documents, and had only recently gained attention from the computer vision community. They fill this gap by constructing and evaluating two systems. The first, representing the de facto state-of-the-art, was a two stage pipeline consisting of text detection followed by a leading OCR engine. The second was a system rooted in generic object recognition. They showed that the latter approach achieved superior performance. The proposed research started with the objective of processing and refining image dataset that we are using a in the proposed framework and algorithm and in this process, following steps and processes evolved.

Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, Andrew Y. Ng (2011) [5] worked over Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning. They found that reading text from photographs was a challenging problem that had received a significant amount of attention. Two key components of most systems were (i) text detection from images and (ii) character recognition, and many recent methods had been proposed to design better feature representations and models for both.

3. Design and Implementation

The proposed research started with the objective of processing and refining image dataset that we are using in the proposed framework and algorithm . In this process, following steps and processes are evolved which lead to development of the research work. To find the proposed objectives the proposed work mainly works upon two algorithms.

Histogram of Oriented Gradients (HOG) are feature descriptors used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant, feature transform descriptors, and shape contexts but differs in a way that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

The Stroke Width Transform: A stroke in the image is a continuous band of a nearly constant width. The Stroke Width Transform (SWT) is a local operator which calculates for each pixel the width of the most likely stroke containing the pixel. First, all pixels are initialized with ∞ as their stroke width. Then, we calculate the edge map of the image by using the canny edge detector. We consider the edges as possible stroke boundaries, and we wish to find the width of such stroke. If p is an edge pixel, the direction of the gradient is roughly perpendicular to the orientation of the stroke boundary. Therefore, the next step is to calculate the gradient direction gp of the edge pixels, if the gradient direction gq at q is roughly opposite to gp, then each pixel in the ray is assigned the distance between p and q as their stroke width, unless it already has a lower value. There are various processes and implementation made to design our proposed method. Collection of Dataset- Selecting highly relevant Image Instances to collect datasets: As the work was done on domain specific area, it was important that only relevant images instance datasets

were picked. In this phase we collect dataset from different aspects and varieties like real time scene text images, logo images *etc*.

Preprocessing Phase: In order to conveniently compare with various different size images and consider the computer's speed, the size of all these images should be limited within appropriate pixels. First, it's necessary to choose a proper format of input images and their appropriate dimensions. After that the actual procedure of pre- processing phase is taken under process. Since, the proposed work is highly based over text localization in real life trends, for the sake of this, the various phases of pre- processing like conversion of color image into gray scale, binarization of image, gradients, morphology and dilation of image are done for horizontal and vertical axis both.

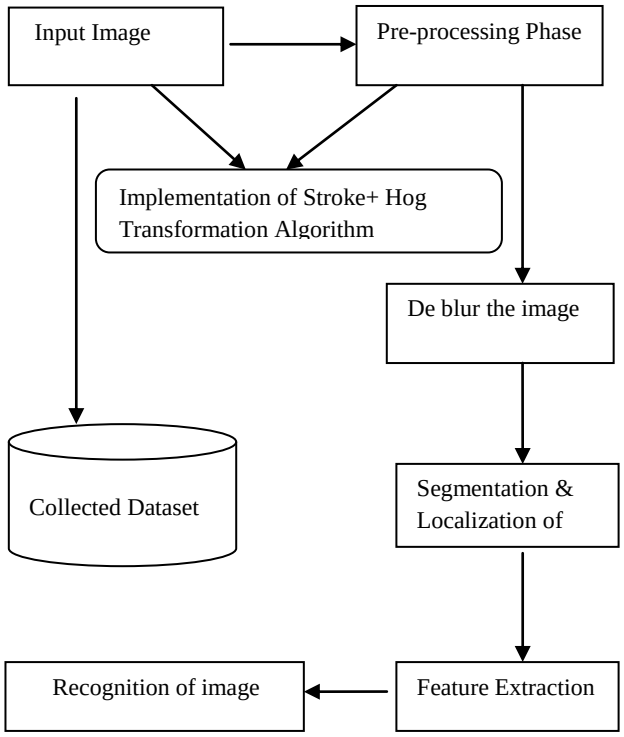


Figure 1. Steps of Methodology

Feature Extraction and Implementation of proposed algorithm: In the proposed algorithm we are using the combination of HOG transform and Stroke detection as a feature descriptor. Implementation of the HOG algorithm is as follows:

1. Divide the image into small connected regions called cells and for each cell compute a histogram of gradient directions or edge orientations for the pixels within the cell.
2. Discrete each cell into angular bins according to the gradient orientation.
3. Each cell's pixel contributes weighted gradient to its corresponding angular bin.
4. Groups of adjacent cells are considered as spatial regions called blocks. The grouping of cells into a block is the basis for grouping and normalization of histograms.
5. Normalized group of histograms represents the block histogram.

Implementation of Stroke Width Transform is as follows:

The SWT Text Detector application is designed to locate and mark the regions of an image that are suspected to contain text. It returns an image of the same size as the input image. The implementation of the application contains several parts:

1. The stroke width transform: edge detection and stroke width calculation.
2. Removing stray lines from the SW map.
3. Finding letter candidates: finding the connected components and detecting the components with the features of a letter.
4. Grouping the letters into regions of text.

Text Segmentation and Localization: In this phase, we have to segment the characters in the image by bounding box method or by bounding the characters by edges and after that finally we are in position to normalize the image to find out correct characters present in the input images and localize the text in the input image.

Recognition: On the basis of above all procedure of the proposed algorithm finally the recognition and text localization of input images will be done in this phase.

4. Results and Discussion

The input image is validated through a number of various pre-processing stages which help in actual text recognition from an image. These various steps are explained below:

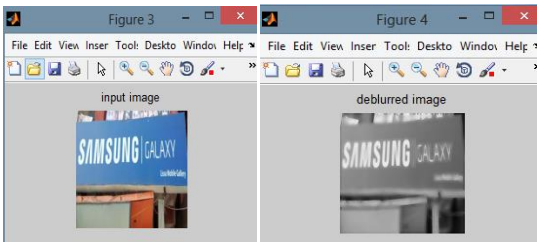
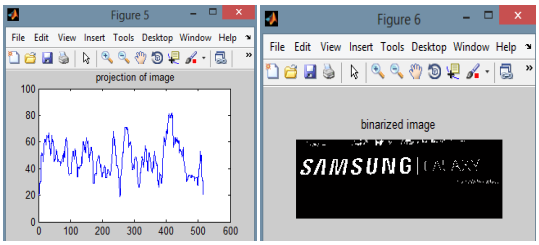
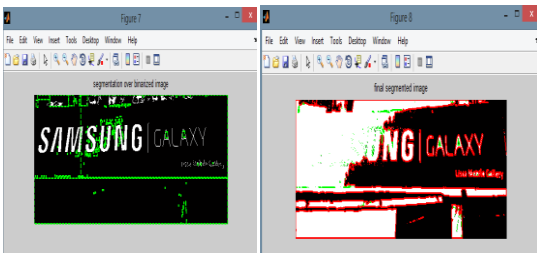


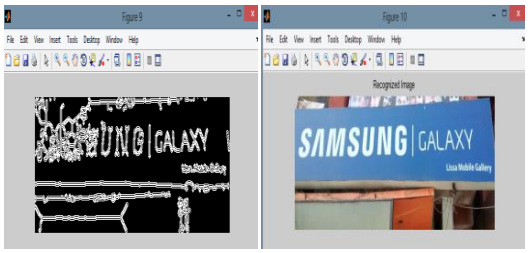
Figure 3. (a) input image (b) deblur image



(c) Projection of image (d) binarize image



(e) Segmented image (f) final segmentation



(g) HOG transformation (h) recognized image

We collected over 45 images from various environments (real time captured by camera, on line logo images) and created our own dataset. We used our proposed approach of real time scene text localization and recognition and found that the efficiency of finding and localization of characters, text and image is quiet higher for every input image at various instants and rotation. The proposed work mainly deals with real time approach in which dataset mainly captured by camera or mobile phones at various angle, environment and rotation. Here, Figure 2 is used to explain the outcomes of the proposed system.

a) Original Image b) Localized recognized image

(a) Original Image b) Localized recognized image

a) Original Image b) Localized recognized image

a) Original Image b) Localized recognized image

Figure 2. Some of the Images in which are Text is Recognized

For results evaluation we use efficiency as parameters and found that there is higher bit of efficiency for every input images from our own sets of dataset.

Efficiency: It is the total number of accuracy based on the features, segmentation and recognition of image and some other properties.

Efficiency = ∑ total number of favorable condition on the basis of features/Total number of conditions. Higher will be the efficiency higher will be the results accurate and effective.

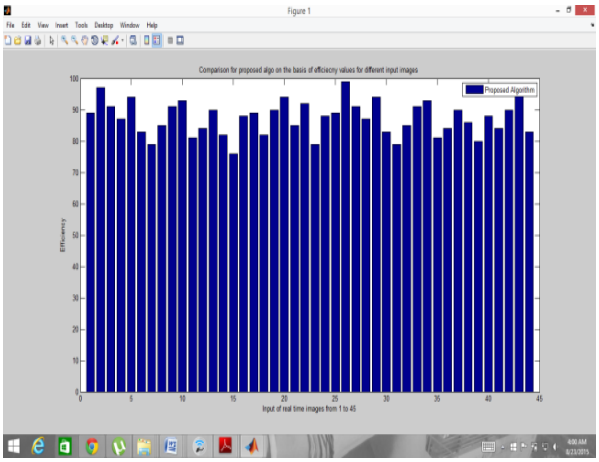


Figure 3. Efficiency of the Proposed Algorithm for All Images in the Dataset

The above Figure 3 clearly depicts the efficiency graphs of the proposed system for all input images taken in the dataset and it is found that the performance of proposed work is quite higher for every images and overall it's total average efficiency is around 86% which is quite efficient and optimist for any system.

5. Conclusion and Future Scope

The main objective of this paper is to localize and recognize real time scene text images. It has been found that each technique has its own benefits and limitations, no technique is best for every case. The proposed algorithm was made over each images stored in the set of data set of Real Time Scene Text images at various instants and rotation. Some of them captured by cameras or smart phone devices and some of them are trademark images. The proposed algorithm is the combination of HOG transformation and Stroke algorithm. Moreover, various other methods like preprocessing, binarization, noise removal, segmentation, localization and recognition would be done effectively. For the sake of above all entire process of the proposed algorithm the efficiency graph was calculated for every images in the dataset and found that there was higher bit of efficiency in terms of localization and recognition of real time scene text images. The results clearly depicts that the value of efficiency for all the images stored in the set of dataset is quiet high and approximately an average of 86 % efficiency for the entire process. Furthermore, there is need to improve the results more by introducing more feature extraction phase.There is also need to use of some more technique and maybe comparison of different technique. As a future scope, one may also choose some more parameters to compare the results.

References

[1] C. Yao, X. Bai, B. Shi and W. Liu, "Strokelets: A Learned Multi-Scale Representation for Scene Text Recognition", IEEE, **(2014)**.
[2] C. Yi and YL. Tian, "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration", IEEE, vol. 23, Issue 7, **(2014)**.
[3] L. Neumann and J. Matas, "Real-Time Scene Text Localization and Recognition", IEEE, **(2012)**.

[4] K. Wang, B. Babenko and S. Belongie, "End-to-End Scene Text Recognition", IEEE, **(2011)**.
[5] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu, A. Y. Ng, "Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning", IEEE, **(2011)**.
[6] L. Neumann and J. Matas, "Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search", ICDAR, **(2011)**.
[7] C. Yao, X. Zhang, X. Bai, W. Liu, Member and Y. Ma, "Detecting Texts of Arbitrary Orientations in Natural Images", IEEE, **(2012)**.
[8] H. Hase, T. Shinokawa, S. Tokai and C. Y. Suen, "A Robust Method of Recognizing Multi-font Rotated Character", IEEE, vol. 1, **(2004)**, pp. 1051-4651.
[9] R. Renuka, V. Suganya and B. A. Kumar, "Online hand written character recognition using Digital Pen for static authentication", IEEE, **(2014)**.
[10] V. J. Dongre and V. H. Mankar, "A Review of Research on Devnagari Character Recognition", International Journal of Computer Applications, **(2010)**, pp. 0975-8887.

Author

Ananta Singh is from Jammu, She was born on 4 May 1991. She has completed her B. Tech (Electronics And Communication Engineering) from ARNI UNIVERSITY Karthgarh Indora, India in the year 2013. She is pursuing M. tech (Electronics and Communication Engineering) from CGC College of Engineering, Landran, Mohali, India.

Dishant Khosla, He is currently working as Assistant Professor at Chandigarh Group of Colleges, Landran (Punjab), India. He has completed his M. Tech from UCOE, Punjabi University, Patiala (Punjab), India and B. Tech from SVIET, Ramnagar (Punjab), India. He has more than 3 years of teaching experience and has published papers in many international journals, national and international conferences. Area of interest is image and video compression and antenna design.

