ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Structure Guided Scene Text Recognition

Balaji Pawade [1], Dr. Jayashree R. Prasad [2]

P.G. Student, Department of Computer Engineering, Sinhgad College of Engineering, Pune, India[1]

Associate Professor, Department of Computer Engineering, Sinhgad College of Engineering, Pune, India[2]

ABSTRACT: Scene text recognition has inspired great interests from the computer vision community in recent years. We propose a novel scene text recognition method integrating structure guided character detection and linguistic knowledge. We use part-based tree structure to model each category of characters so as to detect and recognize characters simultaneously. Since the character models make use of both the local appearance and global structure information's, the detection results are more reliable. For word recognition, we combine the detection scores and language model into the posterior probability of character sequence from the Bayesian decision view. The final word recognition result is obtained by maximizing the character sequence posterior probability via Viterbi algorithm. Experimental results on a range of challenging public data sets demonstrate that the proposed method achieves state-of-the-art performance both for character detection and word recognition.

KEYWORDS: Character recognition, Scene text recognition, Part-Based Tree-Structured Models (TSMs), Posterior probability, Word recognition.

I. INTRODUCTION

With the rapid growth of camera based applications readily available on smart phones and portable devices, understanding the pictures taken by these devices semantically has gained increasing attention from the computer vision community in recent years. Among all the information contained in the image, text, which carries semantic information, could provide valuable cues about the content of the image and thus is very important for human as well as computer to understand the scenes.

In an experimental study, Judd et al. [1] found that given an image containing text and other objects, viewers tend to fixate on text. This further demonstrates that text recognition is very important for humans to understand the scenes. In fact, text recognition is indispensable for a lot of applications such as automatic sign reading, language translation, navigation, and so on. Generally speaking, to understand the information carried by text in the image, we need to recognize the text.

Fig. 1. Some scene text images from ICDAR 2003 [2] and SVT [13].

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

The characters in these images have different fonts, shadows, distortions, deformations, low resolutions, and occlusions. The ICDAR 2003 robust reading challenge [2] published the first data set to highlight the problem of detecting and recognizing scene text. In this benchmark, the organizers identified four subtasks: 1) text localization; 2) robust character recognition; 3) robust word recognition; and 4) robust reading. Since text detection is the premise for recognition, a lot of methods have been proposed to address the problem of detecting and localizing text in scene images [3]–[2] and some have reported promising localization performance. For scene text recognition some previous methods [4], [8], used off-the-shelf Optical Character Recognition (OCR) for subsequent recognition, whose performance was unsatisfactory. Although many commercial OCR systems work well on scanned documents under controlled environment, they perform poorly on scene text images due to the unsatisfactory binarization results of text images of low resolution, unconstrained lighting, distortion, and complex background. Some recent methods [3]–[9] propose to recognize scene text without binarization. They adopt multiscale sliding window strategy to detect characters and directly extract features from the original image to recognize the characters.

The performance of some recently proposed method [5]–[7] is quite promising. As we can see, due to the high degree of intra-class variation of scene characters as well as the complexity of background, recognizing these text images is quite challenging even for state-of-the-art OCR methods. In fact, characters are designed by humans and each category of characters has unique structure representing itself. Therefore, no matter how the background changes or the character degrades, as long as the structure remains unchangeable, we could recognize them by detecting the unique structure from cluttered background. The proposed scheme scene text-recognition method; combine the structure guided character detection and linguistic knowledge. As character detection and recognition is the premise for word recognition, it plays an important role for the overall performance. Different from conventional multi scale sliding window character detection strategy, we use part-based TSMs each category of characters so as to jointly detect and recognize characters. Since the purpose of word-recognition is to understand the words and the case insensitive results could satisfy this need, we only report case insensitive word recognition results. We evaluate our method on a range of challenging data sets. Experimental results show that our method achieves state-of-the-art performance both for character detection and word recognition.

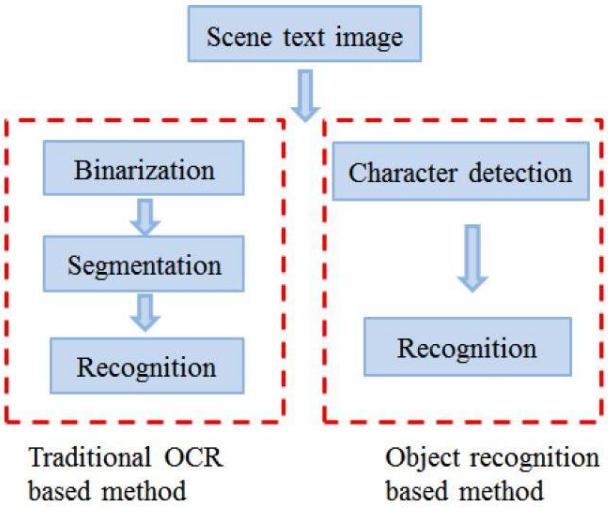

Fig. 2. The traditional OCR-based method and object recognition-based method.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

II. 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.

Cong Yao et al. [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.

L. Neumann and J. Matas [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.

K. Wang, B. Babenko and S. Belongie [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.

Adam Coates et al. [5] worked over Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning (UFL). 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.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

III. 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): HOG is 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 (SWT): 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.

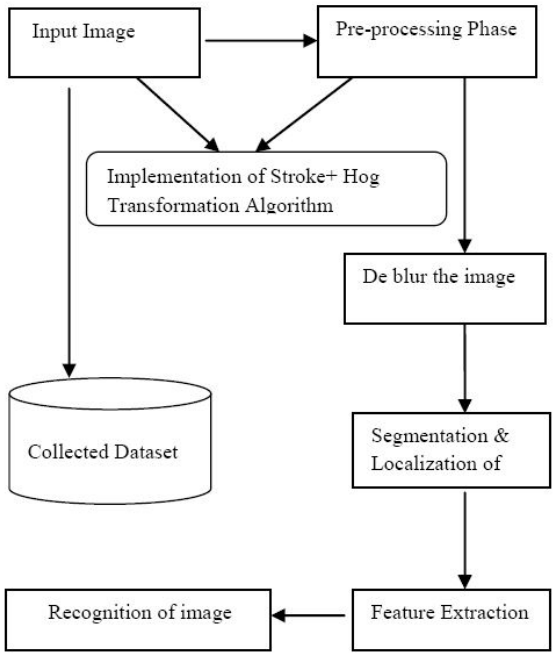

Fig. 3. Steps of Methodology.

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.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

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.

IV. SYSTEM OVERVIEW

As mentioned above, a good scene character-recognition method should make use of both the local appearance and global structure information. Motivated by the recent progress in object detection [6], [7] using part-based TSM, we find that we could adopt these models to use both the global structure and local appearance information of characters.

The flowchart of the proposed method is shown. Given an input text image, first, we use part-based TSM for all the categories of characters to detect the character-specific structures, based on which we get the potential character locations. Then, we convert the candidate detection scores to posterior probabilities via confidence transformation. For word recognition, we combine the detection scores and language model into the posterior probability of the character sequence from the Bayesian decision view. Bigram, trigram, and even higher order language model could be incorporated. The final word-recognition result is obtained by finding the most probable character sequence via Viterbi algorithm.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

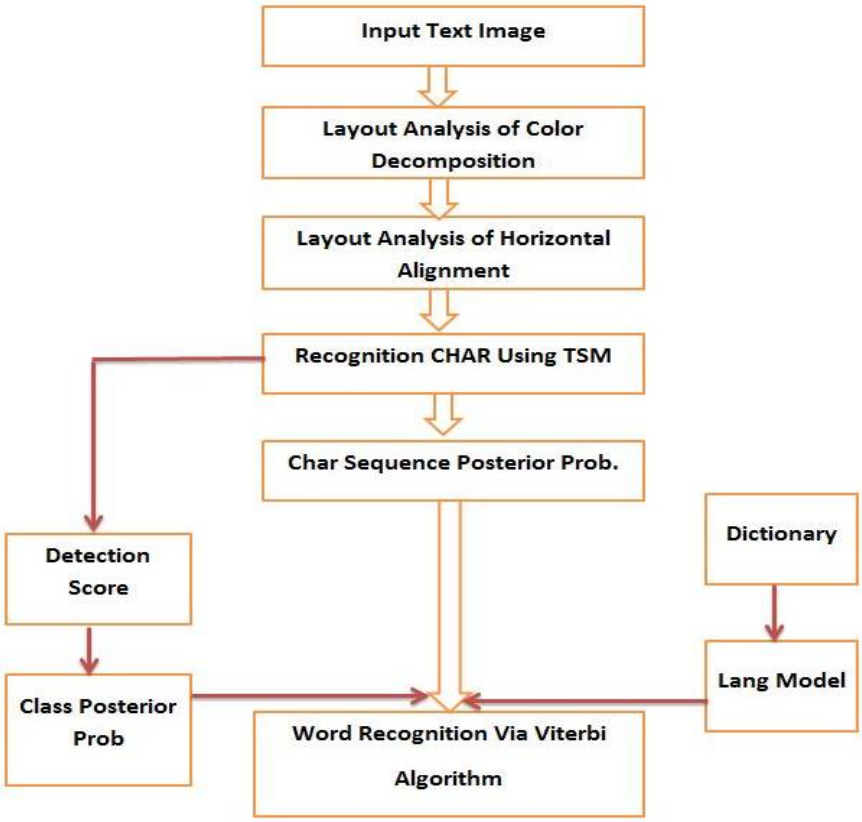

Fig. 4. Flowchart of the Proposed System.

V. CHARACTER DETECTION USING PART-BASED TSM

We propose to recognize characters by detecting part based tree-structures, which seamlessly combines detection and recognition together. Briefly shows how to train the TSM for character it shows how to recognize the characters using the trained character models. Next, we will give details about the model, the inference of the character specific structures and the learning of the parameters of the models.

A. Model

To make use of both global structure and local appearance information of characters, we build a part-based TSM for each category of characters.

1) Model for Characters: We represent each category of characters 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*. Then, the score of the configuration of all the parts-

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

$$S(L, I, k) = S_{\text{App}}(L, I, k) + S_{\text{Str}}(L, k) + \alpha_k \quad (1)$$

where

$$S_{\text{App}}(L, I, k) = \sum_{i \in V_k} w_i^k \cdot \phi(I, l_i) \quad (2)$$

and

$$S_{\text{Str}}(L, k) = \sum_{ij \in E_k} w_{ij}^k \cdot \psi(l_i \quad l_j). \quad (3)$$

As we can observe, the total score of a configuration L for model k consists of the local appearance score in (2), the structure or shape score in (3), and the bias α_k . Here, α_k is a scalar bias or prior associated with character k. Next, we will give details about the appearance model and the shape model.

2) Local Appearance Model: Equation (2) is the local appearance model, which reflects the suitability of putting the part-based models on the corresponding positions. w_i^k Represents the filter or the model for part i, structure k, and $\varphi(I, l_i)$ denotes the feature vector extracted from location l_i. Thus, the score of placing part w_i^k on position l_i is actually the filter response of template w_i^k. We choose HOG [9] as the local appearance descriptor due to its good performance on many computer vision tasks. For color image, we choose the color channel with the largest gradient magnitude to calculate the HOG features.

3) Global Structure Model: Equation (3) is the structure or shape model, which scores the character specific global structure arrangement of configuration L. Here, we set $\psi(l_i - l_j) = [dx\ dx2\ dy\ dy2]$, where $dx = x_i - x_j$ and $dy = y_i - y_j$ 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.

VI. EXPERIMENTAL SETUP AND DATA SET

Experimental Setup:
The system is built using Java framework (version jdk 8) on Windows platform. The Eclipse-jee-indigo-SR2 is used as a development tool. The system doesn't require any specific hardware to run any standard machine is capable of running the application.

Data Set:
In this Paper, we use Chars74k and ICDAR 2003 robust character recognition data set (ICDAR03-CH) [2], and for word recognition uses the challenging public data sets street view text (SVT) [13], ICDAR 2003 robust word recognition (ICDAR03) [2] and ICDAR 2011 word recognition (ICDAR11) data sets to evaluate the performance of the overall word recognition method.

VII. 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:

Fig 5(a): Input Image

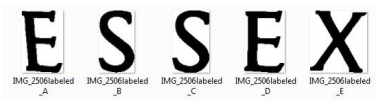

Fig 5(b): Extracted Character Set

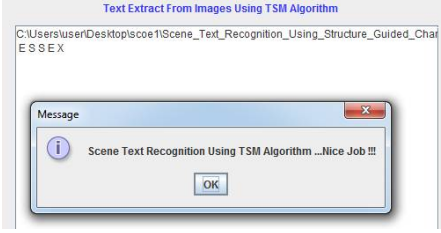

Fig 5(c): Scene Text Recognition

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

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.

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

$$\text{Efficiency} = \sum \frac{\text{Total number of favorable condition on the basis of features}}{\text{Total number of conditions.}}$$

Higher will be the efficiency higher will be the results accurate and effective.

VIII. CONCLUSION AND FUTURE SCOPE

CONCLUSION:

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 are best for every case. 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 image 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 82% efficiency for the entire process.

FUTURE SCOPE:

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.