Scene Text Recognition in Multiple Frames Based on Feature Extraction

A.yuvaraj¹, Ms.N.Radhi⁵
¹ P.G Student, Computer Science Department, Prist University, TamilNadu , India
² Assistant Professor, Computer Science Department, Prist University, TamilNadu , India

Abstract

Recognition of text in natural scene images is becoming a prominent research area due to the widespread availability of imaging devices in low-cost consumer products like mobile phones. Detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of Computer Vision applications, such as computerized aid for visually impaired and robotic navigation in urban environments. Retrieving texts in both indoor and outdoor environments provides contextual clues for a wide variety of vision tasks. In this project, we implement two processes such as text detection and text recognition. In text detection, use contrast map is then binaries by median filter and combined with Canny’s edge map to identify the text stroke edge pixels based on feature extraction. The features extractors are Harris-Corner, Maximal Stable Extremal Regions (MSER), and dense sampling and Histogram of Oriented Gradients (HOG) descriptors. Then implement text recognition. The first one is training a character recognizer to predict the category of a character in an image patch. The second one is training a binary character classifier for each character class to predict the existence of this category in an image patch. The two schemes are compatible with two promising applications related to scene text, which are text understanding and text retrieval. Further we extend this concept with word level recognition with lexicon techniques with accurate results. And also recognition text in real time images, videos and mobile application images.

Keywords: character recognition, text detection, text recognition

I. Introduction

Automatic detection of interest regions is an active research area in the design of machine vision systems and is used in many applications such as tourist’s assistant systems, mobile robot navigation, vehicle license plate detection and recognition. Vision systems are mainly focused on constantly monitoring traffic and observe passing vehicles, extracting important features such as vehicle type, color and distinct marks. One feature that separates text from other elements of a scene is its nearly constant stroke width. This can be utilized to recover regions that are likely to contain text. An exhaustive search is then applied to group components into regions and a text level classifier is used for classification of these regions. As an important prerequisite for text recognition, text detection in natural scene images still remains an open problem due to factors including complex background, low quality images, variation of text content and deformation of text appearance. Recently, Maximally Stable Extremal Regions (MSERs) based text detection has been widely explored. The main advantage of these approaches over other component based approaches is rooted in the effectiveness of using MSERs as character/component candidates. It is based on the observation that text components usually have higher color contrast with their backgrounds and tend to be form homogenous color regions, at least at the character level. The MSER algorithm adaptively detects stable color regions and provides a good solution to localize the components without explicit binarization. Text detection and recognition in natural scene images has recently received increased attention of the computer vision community. Since text is a pervasive element in many environments, solving this problem has potential for significant impact.

II. Related Work

In this paper, introduce a new skeleton pruning method based on contour partitioning. Any contour partition can be used, but the partitions obtained by Discrete Curve Evolution (DCE) yield excellent results. The theoretical properties and the experiments presented demonstrate that obtained skeletons are in accord with human visual perception and stable, even in the presence of significant noise and shape variations, and have the same topology as the original skeletons. In particular, we have proven that the proposed approach never produces spurious branches, which are common when using the known skeleton pruning methods. Moreover, the proposed pruning method does not displace the skeleton points. Consequently, all skeleton points are centers of maximal disks. Again, many existing methods displace skeleton points in order to produces pruned skeletons.
The increasing market of cheap cameras, natural scene text has to be handled in an efficient way. Some works deal with text detection in the image while more recent ones point out the challenge of text extraction and recognition. We propose here an OCR correction system to handle traditional issues of recognizer errors but also the ones due to natural scene images, i.e. cut characters, artistic display, incomplete sentences (present in advertisements) and out-of-vocabulary (OOV) words such as acronyms and so on. The main algorithm bases on finite-state machines (FSMs) to deal with learned OCR confusions, capital/accented letters and lexicon look-up. Moreover, as OCR is not considered as a black box, several outputs are taken into account to intermingle recognition and correction steps. Based on a public database of natural scene words, detailed results are also presented along with detection and recognition to be used in a high accuracy end-to-end system.

We study the question of feature sets for robust visual object recognition; adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of histograms of oriented gradient (HOG) descriptors significantly outperform existing feature sets for human detection. We study the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good results. The new approach gives near-perfect separation on the original MIT pedestrian database, so we introduce a more challenging dataset containing over 1800 annotated human images with a large range of pose variations and backgrounds.

III. Character Recognition

Character Reader

OCR is the process of taking images or photographs of letters or typewritten text and converting them into data that a computer can interpret. A good example is companies and libraries taking physical copies of books, magazines, or other old printed material and using OCR to put them onto computers. While far from perfect, OCR is currently the best method of digitizing typed pages of text.

IV. Feature extraction

In this phase, each image is divided into 15 vertical zones and 15 horizontal zones, where size of each vertical zone is 2*30 pixels and the size of each vertical zone is 30*2 pixels. Then sum of all on pixels in every zone is determined as a feature value for the zone. Finally, 30 features are computed from all zones and are stored in to a feature vector X as described in the equations (1) to (5):

\[ X = \left( \begin{array}{c} V \text{Features} \mid H \text{Features} \end{array} \right) \]  \hspace{1cm} (1)

\[ V \text{Features} = [Vf_i] \hspace{1cm} 1 \leq i \leq 15 \]  \hspace{1cm} (2)

\[ H \text{Features} = [Hf_i] \hspace{1cm} 1 \leq i \leq 15 \]  \hspace{1cm} (3)

Where, \( Hf_i \) is a feature value of \( i \)th horizontal zone and it is computed as shown in (4).

\[ Vf_i = \sum_{x=1}^{d} \sum_{y=1}^{d} g(x,y) \]  \hspace{1cm} (4)

\[ Hf_i = \sum_{x=1}^{d} \sum_{y=1}^{d} g(x,y) \]  \hspace{1cm} (5)
Where, \( g_i \) is \( i \)th zone that encompasses the chosen region of the character image. The dataset of such feature vectors obtained from training samples is further used for construction of knowledge base.

V. Text Recognition

Since commercial OCR engines achieve high recognition performance when processing black and white images at high resolution, almost all the methods in the literature that addressed the issue of text recognition in complex images and videos employed an OCR system to finally recognize characters. However, this OCR software cannot be applied directly on regions previously extracted by a text localization procedure. To extend the recognition capability of the OCR for image and video text, the main research efforts focus on text segmentation and enhancement.

VI. Text Detection

As mentioned early, the proposed approach consists of 5 stages. The text region detection phase detects the text region from the input media. This approach extracts mainly 2 features- LBP(local binary pattern) and HOG(histogram of oriented gradients). The preprocessing stage is mainly employed to remove the skew of the text region and also to remove the noise. Script identification mainly focuses on identifying the script of the document so that it can be properly fed into the appropriate character segmentation algorithm. Character recognition phase recognizes the character thereby identifying the destination place. For each input image, connected component analysis is performed. The candidate regions obtained through the connected component analysis are subjected to feature extraction.

![Figure 1: Text detection](image)

**Figure 1: Text detection**

**Histogram of Oriented Gradients (hog)**

Histogram of Oriented Gradients (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. Local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The implementation of these descriptors can be achieved by dividing the image into small connected regions, called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then represents the descriptor. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination or shadowing.

VII. Conclusion

Text detection in natural scene images remains a challenging problem due to complex background, low image quality and/or variation of text appearance. In proposed presented a technique of scene text recognition from identify text regions, which is well-matched with mobile applications. It identifies text regions from image or video and distinguishes text information from the identify text regions. In scene text detection, describe analysis of color disintegration and horizontal alignment is performed to search for image regions of text strings. In scene text recognition, two methods, text understanding and text retrieval, are correspondingly proposed to take out text information from surrounding location. In future work, we will improve the accuracy rate of text detection, and add lexicon analysis to extend our system to word-level recognition.

**Reference**


