IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 5, Ver. VI (Sep. - Oct. 2016), PP 63-65
www.iosrjournals.org

Visual Text Localization and Extraction for Devanagari Script

[1]Naresh Pal, [2]Vivek Verma
[1] *Department of Computer Science & Engineering, University of Quebec in Outaouais*
[2]*Dep. of Computer Science & Engineering, Manipal Univesity Jaipur, India*

***Abstract**: Text information in natural scene images is important for various kinds of applications. In this paper a novel method proposed to detect Devanagari text from visual components. Detecting text information from images plays an important role in a variety of applications such as image retrieval with text information, intelligent driving assistance, reading application for visually impaired people and visual component analysis, etc. Consequently, the research of text detection in scene images has received much attention from the computer vision community. Although numerous algorithms have been proposed earlier for scene text detection and it is still challenging research area. The most challenging problems occur due to difficulties like cluttered background, occlusion, blur and variations of font type, size, color, etc. We are proposing a robust system for Devanagari script detection from visual component.*
***Keywords:** Text localization, Text detection, OCR, Segmentation*

I. Introduction

The widespread availability of intelligent mobile devices equipped with cameras leads to the explosive growth of natural scene images. Existing methods for text detection in scene images can be classified into two categories: the sliding window based methods and the connected component based methods. Sliding window based methods exploit the textural property of texts by scanning the image with sliding windows at different scales. However, to cope with the large variation in appearance of the text in natural scene images, window descriptors have to be extracted and classified at numerous locations and scales. The high computational complexity of these methods is a major disadvantage compared with connected component base methods. Connected component based methods, on the other hand, detect text by analyzing the shapes of the edges detected or components extracted from the images. The computational complexity of connected component based methods are only subjected to the number of edge pixels detected, which makes it much faster than sliding window based methods [1]. However, these methods are very dependent on the quality of edge detection or component extraction. The stroke width based methods are a type of connected component based methods. These methods are mainly based on the assumption that in a single text group the stroke width values of the characters are always nearly constant, which is independent of the language type and computationally efficient.

Generally, text extraction in can be divided into the following stages: 1)Text detection, finding regions in a visual component that contain text; 2)Text localization, grouping text regions into text instances and generating a set of tight bounding boxes around all text instances; 3)Text tracking, following a text event as it moves or changes over time and determining the temporal and spatial locations and extents of text events; 4)Text binarization, binarizing the text bounded by text regions and marking text as one binary level and background as the other; 5)Text recognition, performing OCR on the binarize text image [2]. Occasionally binarization step is eliminated in favor of applying OCR on color/gray level images. Visual component Localization and Extraction for Devanagari Script System structure is shown in figure 1 below, which includes the above steps mentioned.

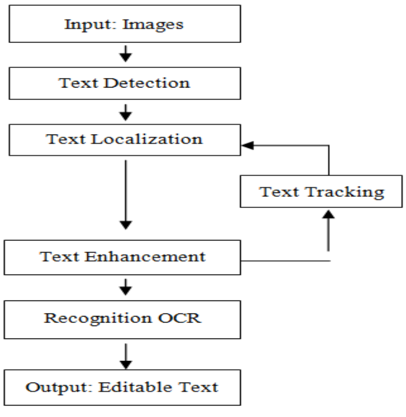

Figure 1: General architecture of the System

II. Literature Review

A variety of methods have been developed for scene text detection and recognition. Maximally stable external region stroke width transforms and HOG [3] have been successfully applied in scene text detection. The outputs of the detection algorithm are usually the bounding boxes of either characters or words. HOG templates have conjointly been accustomed match character instances in take a look at pictures with coaching examples [4]. Shi et al. [5] planned to use manually designed deformable part-based model to represent characters. Most of those ways ar chiefly involved with solely frontal text. exploitation rotation invariant feature like scale invariant feature remodel (SIFT) to explain characters has been tested flourishing. Phan [6] planned to use dense SIFT rather than traditional SIFT to explain individual characters. With the first SIFT, the descriptors ar solely extracted at thin interest points. Since scene characters suffer from deformations like blurring and uneven illumination, the quantity of detected interest points isn't comfortable. The dense SIFT outlined within the literature was designed for scene classification, that doesn't need rotation unchangeability [7]. AN extraction theme that fixes the position and size however permits the orientation of the interest points to vary was devised in their work, that provides rotation unchangeability. The work of Phan et al. [6] provides useful insights into perspective scene text recognition, and it conjointly introduces 2 datasets that are accustomed benchmark perspective recognition performance. SIFT aligns a neighborhood frame of reference to the dominant gradient direction at every detected interest purpose, that depends on the idea that such a dominant gradient orientation is on the market. SIFT does not work well for arbitrary positions or dense feature computation, and it is a main source of error in dense image alignment [8]. Most recent text recognition approaches skip this step and use the non-invariant dense HOG features with a sliding window classifier. CHOG, on the other hand, offers a well defined rotation behavior by rep-resenting circular HOG on Fourier domain. The window function of CHOG is isotropic, unlike rectangle spatial window; the descriptor rotates with respect to rotation of its underlying data without leading to any discrete binning artifacts in the histogram. Also, there is no need for interpolation. CHOG can be computed both densely and efficiently, while still being highly discriminative like HOG with rectangle spatial Window. Therefore, we use CHOG as the low level building block of our method.

III. Proposed Work

While image acquisition there is chances to noise in the image which need to identify and remove before the actual processing. Therefore few of the preprocessing phases are implemented.

Gray Conversion

System takes input image in gray tone having pixels intensity values between (0-255) and using a thresh-holding approach color image converts from RGB to Gray tone.

Dilation and Erosion

Dilation and erosion are 2 morphological operations use for the reconstitute the constituent parts of image. Primarily Dilation adds pixels to the boundaries of parts in a picture. While pictures capturing few of the constituent points incomprehensible and owing to noise and it got to recover for acceptable recognition. In alternative hand erosion is that the method that removes pixels on parts boundaries of the image. the amount of pixels further or aloof from parts|the weather} in a picture depends on the scale and form of the structuring element accustomed method the image [6]. Scene Text localization is a vital a {part of} the topic of Scene Text recognition; it limits the search house of the popularity part, thus it will be computed quicker. Scene Text recognition is a vital subject in these days life. There square measure today several automobiles on the road. it's helpful to spot the automobiles for lots of applications. as an example for machine-controlled tolling systems, traffic observance, public safety and security, speed management and road prizing. during this paper we have a tendency to projected methodology supported edge density. For locating region of interest i.e. candidate text portion the total image is broken to multiple blocks supported threshold worth. once dynamical a picture to grey toneit is split into little blocks to judge the part density. the fundamental geometric feature of the scene image is that it holds many of vertical edges[7]. Within the projected methodology one by one method of candidate image detection from the total image. each block of the image scan from left prime to the proper bottom for vertical stroke density and once this chosen blocks think about for the peak and dimension quantitative relation proportion. At the last chosen blocks tested on the geometrical options of Devanagari extracted from the image like intensity worth of background and foreground oftentimes modification. In many pictures there's over one candidate Text space detected. From the candidates verity Text space should be hand-picked on the second and third feature extraction part[8][9].

At the second a part of the work extracted scene text image ought to convert into editable text. it's used for the aim of conversion of pictures of text into corresponding editable characters. Optical Character Recognition system is employed for this purpose. Before the popularity of character it should be divided into individual character image*Character segmentation*

To isolate the characters of candidate text portion from one another, several techniques are used; the extracted Scene Text is resized into a typical threshold model size. At this model, all character positions area unit given ahead. when resizing, the precise location is calculated for the characters. This technique sturdy and it's high accuracy rate. just in case of shift in extracted vehicle plate, the ends up in background rather than characters.

Character Recognition

The character recognition approach has 2 basic components feature extraction and therefore the feature classifier. Feature analysis determines the descriptors, or the feature set accustomed describe all characters. Given a personality image, the feature extractor derives the options that the character possesses. The derived options ar then used as input to the character classifier.Template matching or matrix matching, is one in all the foremost common classification strategies. Here individual image pixels are used as options. Classification operation is performed by comparison input characters with a collection of templates from the dataset of every character. Every comparison leads to a similarity live between the input characters with a collection of templates. One live will increase the quantity of similarity once a constituent within the discovered character is the image of constant constituent within the example image. If the pixels disagree the live of similarity is also shriveled. Finally templates are compared with the discovered character image; the character's identity is allotted the identity of the foremost similar example. Example matching could be a trainable method as example characters are often modified.

IV. Result Analysis

To verify the effectiveness of our approach, a benchmark check is performed over the whole system module. All testing pictures area unit single text part pictures, that more segmental to the word wise for individual process of the module. Input pictures area unit extracted from completely different sources and scanned over 330 DPI scale. Our benchmark check results area unit summarized in single part that indicates the performance of projected system. the general testing method demonstrates that our approach has, total ninety four accuracy. This confirms the effectiveness of our approach for visual text localization and extraction for Devnagiri.

References

[1] Koo, HyungIl, and Duck Hoon Kim. "Scene text detection via connected component clustering and nontext filtering." Image Processing, IEEE Transactions on 22.6 (2013): 2296-2305.

[2] Zhang, Honggang, et al. "Text extraction from natural scene image: A survey."Neurocomputing 122 (2013): 310-323.

[3] Wang, Kai, Boris Babenko, and Serge Belongie. "End-to-end scene text recognition." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[4] Weinman, Jerod J., Erik Learned-Miller, and Allen R. Hanson. "Scene text recognition using similarity and a lexicon with sparse belief propagation."Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.10 (2009): 1733-1746.

[5] Shi, Cunzhao, et al. "Scene text detection using graph model built upon maximally stable extremal regions." Pattern Recognition Letters 34.2 (2013): 107-116.

[6] Shivakumara, Palaiahnakote, TrungQuyPhan, and Chew Lim Tan. "A laplacian approach to multi-oriented text detection in video." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.2 (2011): 412-419.

[7] Bosch, Anna, Andrew Zisserman, and Xavier Muoz. "Scene classification using a hybrid generative/discriminative approach." Pattern Analysis and Machine Intelligence, IEEE Transactions on 30.4 (2008): 712-727.

[8] Lin, Wen-Yan, et al. "Aligning images in the wild." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.

[9] Verma, Vivek Kumar, and Pradeep Kumar Tiwari. "Removal of Obstacles in Devanagari Script for Efficient Optical Character Recognition." In Computational Intelligence and Communication Networks (CICN), 2015 International Conference on, pp. 433-436. IEEE, 2015.