

Text Localization and Recognition in Real-Time Scene Images

Ananta Singh¹, Dishant Khosla²

¹M.Tech Student, Department of Electronics and Communication CGC-College of Engineering, Landran

²Assistant Professor, Department of Electronics and Communication CGC-College of Engineering, Landran

Abstract: *An end-to-end text localization and detection is a much truer form of text extraction in image analysis if it can be harnessed efficiently. In this paper, a robust system to detect text from scene images is proposed. The text is extracted when image goes through multiple pre-processing stages. The method uses connected component (CC), Extremal Regions(ER) and sliding window concept. A set of rules and mathematical operations have been devised to extract actual text regions from background. Various techniques that are used to accurately localize text from scene images are implemented using HOG and stroke detection. Finally, text is recognized in an OCR stage of character recognition based on gradient direction feature and trained using synthetic fonts.*

Keywords: localization, stroke detection, recognition, HOG transformation.

1. Introduction

Nowadays, the wide use of digital images that are being captured by various capture devices such as mobile phone, digital cameras etc have gained more attention from the past years.. 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. For example, reading scene text can play an important role in navigation for automobiles equipped with street-facing cameras in outdoor environments and in assisting a blind person to navigate in certain indoor environments (e.g., a grocery store reading words in unconstrained images is a challenging problem of considerable practical interest.) [2] This widely use of digital image capturing devices helps to extract text localization/detection of text scene images and serves as an important part of the image text information extraction system. For this a Text Information Extraction (TIE) system is composed that consist of three parts:

- 1) Text detection.
- 2) Text localization.
- 3) Text character recognition. [1]

The first two are identical to each other as they both focus to detect the localized image region that consists of string and text characters. On the other hand Text Character Recognition allows automatic character recognition through optical mechanism which is also called as Optical Character Recognition (OCR). Though, it still remains a challenge to work on all the features of the text as they vary in texture, size, font, array and various other components.

Thus two existing methods can be classified separately under two classes: Region based method and connected components (CCs).Region based methods are based on observation used for text recognition in image analysis which differentiated the text on the basis of distinctive texture, structure and density gradient distribution. And CC-based methods are based on observation that text is viewed as a combination of separate connected component where each of them different intensity, color distribution have and linked edge contours. Comparing with CC-based methods, region-based methods are less sensitive to image noise but have similar localization accuracy

The real time performance evolution of the real-time scene text is achieved from useful subsequent set of Extremal Regions (ERs).The ER detector is sturdy in contrast to blur, color and texture variation, low contrast and illumination and is helpful for real time text images recognition. The main purpose of this paper is to review real-time text localization and recognition method proposed techniques which combines several feature detectors and scene text detection and recognition algorithms.

2. Text Localization and Recognition

Reading and detecting text from scene images is noted as scene text localization and recognition. It has nowadays seeked its help in various fields with interesting applications like helping visually imperfect people, rephrasing an unidentified script to an understandable language with an automatic text translator etc.

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].Text localization can be computationally very expensive because in an image of N pixels generally any of the 2N subsets can correspond to text [7]. This method can be divided into two groups: the first group accomplishes a sliding-window approach which localizes each character or full word. Sliding window based detectors has been very successful for challenging tasks, such as face and pedestrian detection [4]. Stability of such method includes robustness to noise and blur, 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) .The complication of the methods does not depend on the parameters of the text as characters of all scales and orientations can be detected in one pass and the connected component representation also provides character segmentation which can be exploited in an OCR stage [3]. Text Localization focus on single text character output from an image, for this a proposed algorithm modified connected component analysis (MCCA) given by K.Woraratpanya is applies to extracted feature from text-background

decomposition and filter all the unwanted objects. MCCA method consists of three steps: Text localization in lower and upper levels, boundary grouping and text localization in middle level.



Figure 1: a) Original image b) Localized Text [6]

The automatic detection of text is an effective research area that has a vital role in the design of machine vision systems nowadays. Scene text understanding is an endeavor to recognize text in an image of a natural setting. Recently, scene text recognition via handheld devices has had a great heap of care from many researchers. If a mobile camera is used to recognize scene text directly, it would lead for many applications and huge benefits for users [8]. Text Recognition of any script can be broadly broken down into three stages; text extraction, classification, post-processing [5]. The text extraction is a collection of various operations that apply consecutive transformations on an image. The process involved are taking in a raw image, removing noise, skewness, distortion and detect text by applying segmentation, feature extraction and connected component analysis techniques. The decision making stage is the classification stage of the text recognition system which uses all the features that are extracted in the earlier stage and analyze the text segment according to the rules. The final stage is the post-processing process, which enhances the recognition of the text and identifies the words using the context.



Figure 2: a) Original image b) Detected text [8]

3. HOG Transformation and Stroke Detection

After analyzing existing feature descriptor methods it is found experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for character detection and best suited for the proposed system [5]. Histogram of Oriented Gradient (HOG) and its various transformations have been used widely in computer vision and Optical Character Recognition (OCR) fields. This wide use is only due to their rigid ability that characterizes the strength and consistency of object figuration. In HOG implementation each pixel's gradient vector is calculated using Sobel operator, where the image is disintegrated into 4-orientations or 8 directions. Sobel operator uses two 3x3 masks vertical and horizontal respectively shown in fig below based on parallelogram law.

-1	-2	-1	-1	0	1
0	0	0	-2	0	2
1	2	1	-1	0	1

Figure 3: 3x3 horizontal and vertical Sobel masks [1]

Further, in a specified local region, each value of HOG is calculated by compiling all adjacent bin values on each pixel. To avoid reactivity to brightness of image, all HOG values are normalized by dividing intensity standard deviation (STD) value. The character structure in an image has a number of oriented strokes which are the basic elements of text. Stroke of a printed text is defined in a region that is surrounded by two parallel boundary segments, the distance between them is termed as stroke width and the orientation is regarded as stroke orientation. A stroke of direction α can be detected as two opposing ridges in the gradient perpendicular to the stroke direction, where the distance w between the two ridges corresponds to stroke width [13]

4. Related Work

As mentioned above text localization and detection involves various steps that are used to extract text from images. The first part includes two steps: preprocessing and region analysis. At pre processing step, the image is first converted to gray-level space from RGB as the system deals only with the gray-level image after this the image in gray-level is resized and an image pyramid is formed by nearest neighbor approach since there are various sizes of text to be detected in an image. At region analysis step, window sampling, integral feature map generation, window classification and feature extraction are adopted subsequently to detect text from image.

The second part also consists of two steps: text line generation and text extraction. At text line generation step, a window grouping approach is used to group the detected windows into candidate text lines, and then the incorrect lines around the correct ones are filtered out by text line competition analysis. At text extraction step, connected components are extracted from each text line region by local binarization, and then a connected component analysis approach based on MRF model is employed to filter out non-text components and localize text lines accurately [1].

Image segmentation is the process of locating region in the printed text. Segmentation of the text differs from figures and graphics in an image and when it is applied to text it isolates character or words. Niblack's binarization algorithm is adopted to segment connected components from the image. The formula to binarize each pixel is defined as

$$b(x) = \begin{cases} 0, & \text{if } \text{gray}(x) < \mu_r(x) - k \cdot \sigma_r(x); \\ 255, & \text{if } \text{gray}(x) > \mu_r(x) + k \cdot \sigma_r(x); \\ 100, & \text{other} \end{cases}$$

Where $\mu_r(x)$ and $\sigma_r(x)$ are the intensity mean and STD within a r radius window centered on the pixel x and the smoothing term k is set to 0.4 in practical. For a binarize image, components with 0 or 255 values are extracted as candidate

text components while 100 ones are not considered further [3].

After text character localization and image binarization, character feature extraction process starts; this is used to extract the features of symbols. Features are the characteristics of the text, in this, symbols are characterized and unused features are left behind. This technique is not used to match the character pattern but instead is used to analyze the features like intersection, lines, spaces, dots etc present in character. The main goal of feature extraction is to maximize the recognition rate with minimum no. of elements used. After analyzing existing feature descriptor methods it is found experimentally that grid of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for character detection and best suited for the proposed system [5].

5. Future Scope

In future, the system can be improved on two aspects: 1) more features can be added to enhance differentiating ability of the feature pool. 2) Modifying the region and component based analysis method to make the system more accurate and fast. Also an updated feature i.e. the translation of one language to another can be done which will help people from different countries to understand the local language.

6. Conclusion

Nowadays natural scene text detection is a challenging task. In this paper an overall review of real scene text detection and making the text understandable from an image is analyzed. Various types of methodology are applied for correct feature extraction of an image which may include several variations like uneven illumination, color variation, size, text alignment etc. Here, region based and cc-based methods are used to localize and detect text and character recognition is done using HOG method.

References

- [1] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu. A Robust System to Detect and Localize Texts in Natural Scene Images. s.l.: National Laboratory of Pattern Recognition.
- [2] Kai Wang, Boris Babenko and Serge Belongie. End-to-End Scene Text Recognition. s.l.: ICDAR, Department of Computer Science and Engineering, 2007.
- [3] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu. Text Localization in Natural Scene Images based on Conditional Random Field. s.l.: 10th International Conference on Document Analysis and Recognition, 2009.
- [4] Jawahar¹, Anand Mishra¹ Karteek Alahari² C. V. Top-Down and Bottom-up Cues for Scene Text Recognition. s.l.: INRIA - WILLOW/ École Normale Supérieure, Paris, France, 2002.
- [5] Deepak Chandra Bijalwan, Alok Aggarwal. Automatic text recognition in natural scene and its translation into user defined language. s.l.: International Conference on Parallel, Distributed and Grid Computing, 2014.
- [6] Mirko GUARNERAA Member, IEEE, Giuseppe MESSINA, Member, IEEE, Edoardo. Text Localization from Photos. s.l.: IEEE, 2009.
- [7] Alvaro González, Luis M. Bergasa, J. Javier Yebe, Sebastián Bronte. A Character Recognition Method in Natural Scene Images. s.l.: 21st International Conference on Pattern Recognition (ICPR 2012), 2012.
- [8] Aradhya.V.N, Vinod.H.C Niranjana.S.K Manjunath. An application of Fourier statistical features in scene text detection. s.l.: IEEE, 2014.
- [9] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu. A Hybrid Approach to Detect and Localize Texts in Natural Scene Images. s.l.: IEEE, 2011.
- [10] Yi-Feng Pan, Cheng-Lin Liu, Xinwen Hou. FAST SCENE TEXT LOCALIZATION BY LEARNING-BASED FILTERING AND VERIFICATION. s.l.: IEEE 17th International Conference on Image Processing, 2010.
- [11] Matas, Lukáš Neumann Jiří. Scene Text Localization and Recognition with Oriented Stroke Detection. s.l.: IEEE International Conference on Computer Vision, 2013.
- [12] Hyung Il Koo, Duck Hoon Kim., Scene Text Detection via Connected Component Clustering and Nontext Filtering. s.l.: IEEE, JUNE 2013.
- [13] B.H.Shekar, Smitha M.L.P. Shivakumara. Discrete Wavelet Transform and Gradient Difference based approach for text localization in videos. s.l.: IEEE, 2013.
- [14] Ajith VPillai*, Arun A. Balakrishnan, Rina Anna Simon! Renoh C Johnson and Padmagireesan S. Detection and Localization of Texts from Natural Scene Images using Scale Space and Morphological Operations. s.l.: IEEE, 2013.
- [15] Kuntpong Woraratpanya, Kitsuchart Pasupa., Text Background Decomposition for THAI Text Localization and Recognition in Natural Scenes. s.l.: IEEE, 2014.
- [16] Xiaodong Huang, Kehua Liu, Lishang Zhu. Auto Scene Text Detection Based on Edge and Color Features. s.l.: IEEE, 2012