

Text Localisation from Natural Scene Images:- Comparison of Various Pre-Processing and Edge Detection Methods

Deepthi M Pisharody
Dept of Computer Science
Prajyoti Niketan College
Thrissur, India

Abstract — Text extraction and recognition from natural scene images like traffic images, digital photographs, images of sign board is of great interest now a days. This paper makes a review of various text localization methods and compares various pre-processing and edge detection methods. The objective of study is to focus various image binarization methods especially various global thresholding methods used for text recognition methods with higher performance.

Keywords — *Text Extraction, Text Recognition, Image binarization, Classifiers*

I. INTRODUCTION

The field of computer vision and pattern recognition shows a great interest in content retrieval from images and videos. This content can be in many forms such as objects, color, texture, shape, text as well as the relationships between them. The schematic information provided by an image can be useful for content based image retrieval, as well as for indexing and classification purposes. Natural scene images like digital photographs, speed camera images, sign board images contain text information which needs to be automatically recognized and processed. The textual data in an image or video can be of varying font style, sizes, orientations, color or even varying languages. Therefore the text extraction from images is a challenging problem. Text recognition systems have to undergo three phases:- **text detection and localization, text extraction from background and text recognition**. Text detection and localization is the step of identifying “is there any text in image and where it is”. Text extraction phase challenges extracting text from background with uneven and complex lighting backgrounds. Since text is a meaningful object text extraction is a critical step as it sets up the quality of text recognition. Text extraction aims at isolating text pixels from background. Another challenge of text extraction phase is that no hypothesis can be text (font style, size, orientation, language) and so data is huge. The existing methods can be categorized into two classes:- region based methods and connected component based ones. The application for text recognition lies in various areas like factory automation, text based image indexing, keyboard based image search, automated processing and reading documents, intelligent transport system, camera based

document analysis, robotics, object classification, multimedia processing etc.

Many methods have been tested and implemented for text detection and localization. All approaches consider various properties of text in images such as color, intensity, edges, connected- components etc. These properties help to distinguish background or other regions in image from textual data.

II. REVIEW ON BINARIZATION

Binarization is the process of converting gray scale image to its binary format. System thresholding techniques can be categorized as either global or local. Global thresholding algorithm used as single threshold, while local thresholding algorithm compute separate threshold for each pixel (or group of pixels) based on neighborhood of pixels. Where $f(x,y)$ is the gray level of the point, $p(x,y)$ is some local property of the point and x and y are the spatial coordinates. A thresholded image $g(x, y)$ is defined as

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{otherwise} \end{cases}$$

When T depends only on $f(x, y)$, all pixels will have same value and this operator is called global thresholding. When T depends on both $f(x, y)$ and $p(x, y)$ it is called local. If T also depends on the spatial coordinates the thresholding is called adaptive or dynamic.

Kapur. et al [2] proposed an automatic threshold selection method for picture segmentation. They use the entropy of the gray level histogram. A picture can successfully be thresholded into a two-level image by a prior maximization of entropy.

Abutalib extended the entropy based thresholding algorithm to the 2-dimensional histogram. In his approach, the gray level value of each pixel as well as the average value of its immediate neighborhood is studied. The threshold is a vector that has two entries: the gray level of the pixel and the average gray level of its neighborhood.

The vector that maximizes the 2-dimensional entropy is used as the 2-dimensional threshold.

Otsu's [4] suggested an histogram based global thresholding method which was one of the oldest and most accepted one.

Minetto et al. [5] proposed a method using toggle mapping for character segmentation in a multiresolutional way since natural scene images have large character size variations and strong background clutter.

Burian et. al. [3] in their method proposed an adaptive scheme to find the threshold for each pixel in the image. They use (i) the moving average, (ii) the maximum and (iii) the minimum of the correct moving average to determine the threshold at each pixel.

Niblack's algorithm [6] is a local thresholding method based on the calculation of the local mean and of local standard deviation. The threshold is decided by the formula

$$T(x,y) = m(x, y) + k \bullet s(x, y) \text{ -----(1)}$$

where $m(x, y)$ and $s(x, y)$ are the average of a local area and standard deviation values, respectively. The size of the neighborhood should be small enough to preserve local details, but at the same time large enough to suppress noise. The value of k is used to adjust how much of the total print object boundary is taken as a part of the given object.

Zhang and Tan [7] proposed an improved version of

$$T(x,y) = m(x, y) \bullet [1 + k \bullet ((1 - s(x,y))/R)] \text{ -----(2)}$$

where k and R are empirical constants. The improved Niblack method uses parameters k and R to reduce its sensitivity to noise.

III. REVIEW ON EDGE DETECTION METHODS

There are many ways to perform edge detection. However, the most may be grouped into two categories:

A. Gradient or Search Based

The search-based methods detect edges by first computing a measure of edge strength, usually a first-order derivative expression such as the gradient magnitude, and then searching for local directional maxima of the gradient magnitude using a computed estimate of the local orientation of the edge, usually the gradient direction. The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image.

B. Laplacian or Zero Crossing Based

The zero-crossing based methods search for zero crossings in a second-order derivative expression computed from the image in order to find edges, usually the zero-crossings of the Laplacian or the zero-crossings of a non-linear differential expression, as will be described in the section on differential edge detection following below. This essentially captures the rate of change in the intensity gradient. Thus, in the ideal continuous case, detection of zero-crossings in the second derivative captures local maxima in the gradient.

As a pre-processing step to edge detection, a smoothing stage, typically Gaussian smoothing, is almost always applied. The edge detection methods that have been published mainly differ in the types of smoothing filters that are applied and the way the measures of edge strength are computed. As many edge detection methods rely on the computation of image gradients, they also differ in the types of filters used for computing gradient estimates in the x - and y -directions

The Canny [9] edge detection algorithm is known to many as the optimal edge detector. In his paper, he followed a list of criteria to improve current methods of edge detection. The first and most obvious is low error rate. It is important that edges occurring in images should not be missed and that there be no responses to non-edges. The second criterion is that the edge points be well localized. In other words, the distance between the edge pixels as found by the detector and the actual edge is to be at a minimum. A third criterion is to have only one response to a single edge. This was implemented because the first two were not substantial enough to completely eliminate the possibility of multiple responses to an edge.

IV. STEPS OF TEXT LOCALIZATION

Almost all followed the following steps for text localization

A. Preprocessing step:-Image binarization

Binarization is a process of converting a grayscale input image to a by-level image by using an optimal threshold. The purpose of binarization to extract those pixels from image which represents an object (either text or other line image data such as graphs, maps). Though the information is binary, the pixels represents a range of intensities, thus the objective of binarization is to mark pixels that belong to true foreground, regions with a single intensity and background regions with different intensities.

B. Text localization is done in 2 steps:

❖ Edge detection

The Canny [11] edge detection algorithm is known to many as the optimal edge detector.

❖ Image Segmentation

- Horizontal cropping to remove rows which are not required
- Vertical cropping to remove columns which are not required

V. CONCLUSION

Global techniques are most effective in character recognition because most of the character images have relatively constant contrast. Global threshold typically work well when the text occupies a large part of the picture and is well contrasted from background. For a natural scene images Otsu's method found to be more effective. It does not produce good results when the background intensity is high and there is low contrast in the image. In case of noisy shaded image a good result is obtained is after subdividing images into six and applying Otsu's method in each subimage individually[1]. Kapur Entropy et al. - Performed

well if the document images had good contrast. Local adaptive shareholding techniques are appropriate in some special instances, especially for the documents with locally varying foreground and background intensities such as engineering drawings, maps, newspapers and mail envelopes. When work with noisy images with textual data, No single global thresholding algorithm can work well due to the varying characteristics of the image like background/foreground intensities, contrast. Local adaptive thresholding algorithms have not proven effective with this type of image [8]. Since traditional thresholding finds a threshold value in one stage, s a special case of the multi-stage thresholding approach (i.e. one stage thresholding) can do better.

REFERENCES

- [1] Rafel C Gonale, Richard E Woods "Digital Image Processing", pp. 756-758, Perason, 2008
- [2] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," Computer Vision, Graphics, and Image Processing, Vol. 29, pp. 273-285, 1985.
- [3] Adrian Burian, Markku Vehniläinen, Mejdi Trimeche, and Jukka Saarinen "Document Image Binarization Using Camera Device in Mobile Phones", *Proc. Intl. Conf. on Image Processing*, 2005, (ICIP 2005), Volume 2, pp:546-548.
- [4] N. Otsu, "A threshold selection method from grey level histogram", *IEEE Trans. Syst. Man Cybern.*, vol. 9 no. 1, 1979, pp. 62-66
- [5] R. Minetto, N. Thome, M. Cord, J. Stolfi, F. Precioso, J. Guyomard, and N. J. Leite, "Text detection and recognition in urban scenes," in *ICCV Workshops*, 2011, pp. 227-234
- [6] . W. Niblack, *An Introduction to Digital Image Processing*, pp. 115-116, Prentice Hall, 1986. R. Minetto, N. Thome, M. Cord, J. Stolfi, F. Precioso, J. Guyomard, and N. J. Leite, "Text detection and recognition in urban scenes," in *ICCV Workshops*, 2011, pp. 227-234
- [7] Z. Zhang and C. L. Tan, "Restoration of images scanned from thick bound documents", *Proc. Int. conf. Image Processing.*, vol. 1, 2001, pp.1074-1077.
- [8] Graham Leedham, Saket Varma, Anish Patankar and Venu Govindaraju. "Separating text and background in degraded document images – a comparison of global thresholding techniques for multi-stage thresholding", *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)* 0-7695-1692-0/02 2002 IEEE
- [9] Canny "A Computational Approach to Edge Detection".