

A Comparative Analysis of Different Edge Based Algorithms for Mobile/Camera Captured Images

H.K.Chethan

Research Scholar, Department of
Studies in Computer Science,
University of Mysore, Mysore-570006, India

G.Hemantha Kumar

Professor, Department of
Studies in Computer Science,
University of Mysore, Mysore-570006, India

ABSTRACT

CBDA is an emerging field in Computer Vision and Pattern Recognition. In recent technology camera are attached to several equipments and are very interesting and playing a vital role by replacing scanner with hand held imaging devices like Digital Cameras, Mobile phones and gaming devices attached with the camera. Availability of High Resolution Camera has lead to new dimension in digital image processing. Mobile phones are ubiquitous and very powerful in nature due to their capability of multifunction. Camera were developed for the purpose to capture good and sad moments in life which can be remembered in the later stages of life but as the technology is being advanced new and new application are being developed for camera /Mobile devices. The goal of the work is to extract and recognize text from camera captured images based on edge based algorithms and compare the result with the existing system under different conditions. Precision and recall rates for each approach are analyzed to determine the success and limitations of each approach. The experimental results show the efficacy compared to the result of well known existing methods.

Keywords

CBDA, Edge Detection, Computer Vision, Precision, Recall and Pattern Recognition.

1. INTRODUCTION

Portable cameras are Ubiquitous. Either in standalone versions, or incorporated in cell phones, the quality of the images has risen at a fast pace while their price has dropped drastically. Such pervasiveness has given rise to unforeseen application such as using portable cameras for digitalizing documents by user of many different professional areas for instance, students and are taking photos instead of taking notes. This new research is evolving fast in many dimensions. Recent price -performance has given birth to several new application[1][2]. Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos[3]. With the help of digital cam-era we can capture characters and documents any-where in the 3D environment like signs and bill-boards, color, texture, shape, as well as the relation-ship between them. CBDA is required because we are no longer constrained to traditional 2D image of As stated by Jung, Kim and Jain[4] text data is particularly interested because text can be used to easily and clearly describe the contents of an image since text data can be embedded in an image or video in different font styles, sizes, orientations, colours and against a complex background. The goal of the work is to extract text from camera captured images and

compare the result with the existing system under different conditions.

Edge detection refers to the process of identifying and locating sharp discontinuities in an image[10][18]. There are many edge detection operators available, each designed to be sensitive to certain types of edges certain criteria involved in selection of an edge detection operator include edge orientation, Noise environment and edge structure. Camera based OCR has been emerging in the recent field due to its wide variety of application. In image processing and computer vision, edge detection treats the localization of significant variations of a gray level image and the identification of the physical and geometrical properties of objects of the scene. The variations in the gray level image commonly include discontinuities (step edges), local extreme (line edges) and junctions. Most recent edge detectors are autonomous and multi-scale and include three main processing steps [11]: smoothing, differentiation and labelling. The edge detectors vary according to these processing steps, to their goals, and to their mathematical and computational complexity.

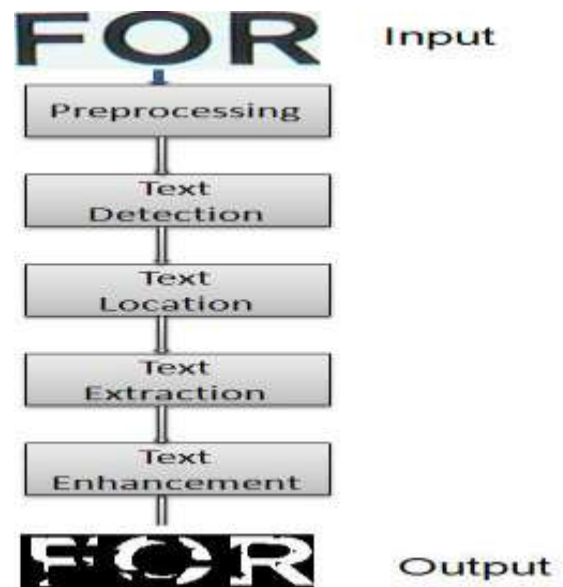


Figure 1. Proposed Block Diagram

1.1 Robert's Cross Operator

Roberts's method finds edges using the Roberts approximation to the derivative. It returns edges at those points where the gradient of I is maximum. Performs a simple, quick to compute, 2-D spatial gradient measurement on an image [12][19]. Pixel values at each point in the output represent the estimated absolute magnitude of the spatial gradient of the input image at that point. The operator consists of a pair of 2x2 kernels as shown in figure 1

$$G_X = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$G_Y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Figure 2 .Roberts cross gradient convolution kernels

1.2 Sobel Operator

The Sobel method finds edges using the Sobel approximation to the derivative. It returns edges at those points where the gradient of I is maximum [13]. The operator consists of a pair of 3X3 convolution kernels as shown in figure. 2. These kernels are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid.

$$G_X = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

$$G_Y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & -1 \end{bmatrix}$$

Figure 3.Sobel convolution

1.3 Prewitt's Operator

The Prewitt method finds edges using the Prewitt approximation to the derivative. It returns edges at those points where the gradient of I is maximum. It is similar to the Sobel operator and is used for detecting Vertical and horizontal edges in images.

$$G_X = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$

$$G_Y = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

Figure 4.Prewitts convolution kernels

1.4 Canny Edge detector

The Canny method finds edges by looking for local maxima of the gradient of I. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges [15][20]. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges.

$$G_X = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$G_Y = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$

Figure 5.Canny Edge detector convolution kernels

1.5 Laplacian of Gaussian

The Laplacian of Gaussian (LoG) is a combination of Laplacian and Gaussian filter where its characteristics is determined by the Parameter and the kernel size. The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image [14]. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection. The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian Smoothing filter in order to reduce its sensitivity to noise. The operator normally takes a single gray level image as input and produces another gray level image as output. The Laplacian $L(x,y)$ of an image with pixel intensity values $I(x,y)$ is given by:

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (1)$$

$$G_X = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$G_Y =$	-1	0	+1
	-2	0	+2
	-1	0	+1

$G_Z =$	-1	0	+1
	-2	0	+2
	-1	0	+1

Figure 5. Three commonly used LoG convolution kernels

2 RELATED WORK

The purpose of this project is to implement, compare, and contrast different edge-based Methods for text extraction and recognition. Various methods have been proposed in the past for detection and localization of text in images and videos. These approaches take into consideration different properties related to text in an image such as colour, intensity, connected-components, edges etc. These properties are used to distinguish text regions from their background and/or other regions within the image. The algorithm proposed by Wang and Kangas in [5] is based on colour clustering. The input image is first pre-processed to remove any noise if present. Then the image is grouped into different colour layers and a gray component. This approach utilizes the fact that usually the colour data in text characters is different from the colour data mostly Chinese and English characters in images; some false alarms occurred due to uneven lighting or reflection conditions in the test images. The text detection algorithm in [6] is also based on colour continuity. In addition it also uses multi-resolution wavelet transforms and combines low as well as high level image features for text region extraction. The text finder algorithm proposed in [7] is based on the frequency, orientation and spacing of text within an image. Texture based segmentation is used to distinguish text from its background. Further bottom-up ‘chip generation’ processes carried out which uses the spatial cohesion property of text characters [16]. The chips are collections of pixels in the image consisting of potential text strokes and edges. The results show that the algorithm is robust in most cases, except for very small text characters that are not properly detected. Also in the case of low contrast in the image, misclassifications occur in the texture segmentation.

3 PROPOSED METHODOLOGY

Approach

The goal of the project is to implement, test, and compare and contrast all the 5 edge based approaches for text region extraction in natural images, and to discover how the algorithms perform under variations of lighting, orientation, and scale transformations

of the text. Figure 1 shows the block diagram of the proposed method.

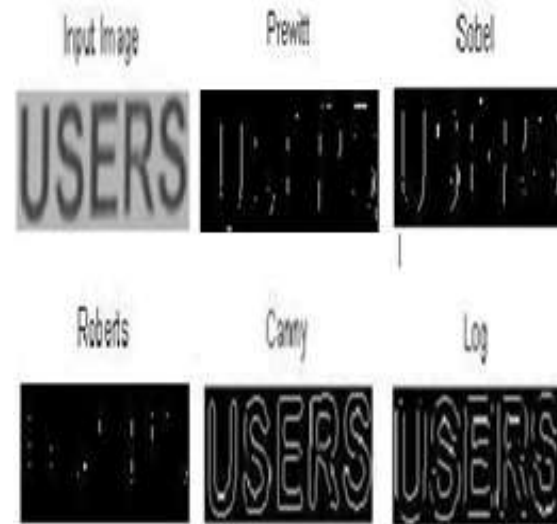


Figure 6 Different Edge based Methods

4 CAMERA-BASED ACQUISITIONS

The advantage of using a camera in alternative to scanner is that a camera have the capability to capture document in 3D environment, it can capture images easily which are at some distance by in the background. The potential text regions are localized using connected component based heuristics from these layers. Also an aligning and merging analysis (AMA) method is used in which each row and column value is analyzed [5]. The experiments conducted show that the algorithm is robust in locating Zooming which is not possible by scanner with camera we are able to capture images on moving objects but suffers from distortions and it involves more pre-processing steps to be done in order to extract the text.

5 PRE-PROCESSING

Camera captured images suffer from noise due to low brightness contrast and various illuminated en-ironmen, low resolution and broken characters are processed to extract text in document. In this step camera captured image is converted into a gray level image and image enhancement algorithm is applied to the image the enhanced image is binarized and then the noises are removed by salt and pepper algorithm [5].

Camera captured image is converted to gray scale image as below:

$$I_s(x, y) = [0, 1, 2, .255], 1 \leq x \leq I_x \text{ and } 1 \leq y \leq I_y \quad (2)$$

Where 0 corresponds to black

255 corresponds to white.

5.1 Image Enhancement

A variety of methods exist for removing image degradations and emphasizing important image information, and in computer graphics, digital images can be generated, modified, and combined for a wide variety of visual effects. Partly very low contrast of intensity on image exists because of illumination variation and photographing angle such as figure 2 (a). It causes misclassification of foreground characters from back- ground. The reduce operation is carried out by convolving the image with a Gaussian low pass filter.

$$f_1(x, y) = (L - 1) * f(x, y) - \frac{\min}{\max} - \min$$

$$\max = \max [f(x, y)]$$

$$\min = \min [f(x, y)]$$

$$\text{for } 1 \leq x \leq M \text{ and } 1 \leq y \leq m \quad (3)$$

Where $f(x, y)$, $f_1(x, y)$ denotes the gray level value at pixel (x, y) and the pixel level after image enhancement, respectively. L denotes gray level range of image to be converted and M denotes the height and width of image. Max and min are maximum value and minimum value among pixels on image

5.2 Binarization

Image binarization converts an image of up to 256 gray levels to a black and white image. Survey [6] have showed that global thresholding is not ideal for Camera-captured images due to lightning variations so We proposed locally adaptive thresholding method that is robust to variation of illumination. The simplest way to use image binarization is to choose a threshold value, and classify all pixels with values above this threshold as white, and all other pixels as black.

$$T_{rxr} = t * g_{high} - g_{low} \quad (4)$$

Where g_{high} and g_{low} are maximum and minimum Intensity value of pixels in rxr sub-window.

6 Text Detection

In detection phase Given an input image, the region with a possibility of text in the image is detected [7] [8]. Therefore one of the main problems in working with multiresolution representations is to develop fast and efficient techniques [9]. The original image is convolved with a Gaussian kernel A Gaussian pyramid is created by successively filtering the input image with a Gaussian kernel of size 3×3 and down sampling the image in each direction by half. Let $I(x, y)$ be the original image The Gaussian pyramid on image I is defined as:

$$I = G_o(x, y) \quad (5)$$

Hence reduce operation is carried out by convolving the image with a Gaussian low pass filter.

$$G_{i+1}(x, y) = \text{Reduce}(G_i(x, y)) \quad (6)$$

The Gaussian kernel is given by:

$$W(r, c) = W(r)W(c) \quad (7)$$

$$\text{Where, } W(r) = \frac{1}{4} - \frac{r}{8}, a, \frac{1}{16} - \frac{r}{2} \quad (8)$$

7 DATABASE

The experimentation of the proposed algorithm was carried out on a data set consisting of different images. We have collected our own database according to our requirement using Nokia 2 Mega pixel Camera with 640×480 resolution. We have differentiated our database into seven types, where type 1 include clear image, type 2 include blur image, type 3 is of image captured with different variance and orientation, type 4 image is capture at different lightning conditions, type 5 is mixture of both blur and clean image, type 6 include different orientation and illumination and finally type 7 includes all the above types.



Figure 6: Database Types

7.1 General Algorithm for Edge based Methods

The basic steps of the edge-based text extraction algorithm are given below, and in Figure 2 we have represented different edge based methods after applying the algorithm.

Algorithm

Step1. Create a Gaussian pyramid by convolving the input image with a Gaussian kernel and successively down-sample each direction by half. (Levels: 4)

Step2. Create directional kernels to detect edges at 0, 45, 90 and 135 orientations.

Step3. Convolve each image in the Gaussian pyramid with each orientation filter.

Step4. Combine the results of step 3 to create the Feature Map.

Step5. Dilate the resultant image using a sufficiently large structuring element (7×7 [1]) to cluster candidate text regions together.

Step6. Create final output image with text in white pixels against a plain black

8 EXPERIMENTAL RESULTS

Experiments have been carried out on large dataset .Two database have been utilized for the above experiments. One set is from ICDAR (2005) and the second set is created by ourselves which consists of both graphic text and scene text.

$$\text{Precision Rate} = \frac{CDW}{CDW+FDW} * 100 \quad (8)$$

$$\text{Recall Rate} = \frac{FDW}{CDW} * 100 \quad (9)$$

Correctly detected words are the block that contains text. False detected block does not contain any text. Recognition rate is ratio of correctly detected word to the sum of correctly detected words (CDW) added with false detected text. False rate is detected by the ratio of false detected text to the correctly detected block. The performance of each technique has been evaluated based on its precision and recall rates obtained. As explained in the earlier sections, precision and recall rates are calculated as mentioned in equation 8 and 9. Precision rate takes into consideration the false positives, which are the non-text regions in the image and have been detected by the algorithm as text regions. Recall rate takes into consideration the false negatives, which are text words in the image, and have not been detected by the algorithm. Thus, precision and recall rates are useful as measures to determine the accuracy of each algorithm in locating correct text regions and eliminating non-text regions

Comparison of different Edge based Methods

a) SOBEL

Sl.No.	Group	Type of Images	Recognition Rate (%)
1	Type 1	Clean	79.6
2	Type 1	Blur	68
3	Type 1	Orientation	79
4	Type 1	Illumination	64

b) Canny

Sl.No.	Group	Type of Images	Recognition Rate (%)
1	Type 1	Clean	99.2
2	Type 1	Blur	75
3	Type 1	Orientation	79
4	Type 1	Illumination	80

C) Log Transform

Sl.No.	Group	Type of Images	Recognition Rate (%)
1	Type 1	Clean	95
2	Type 1	Blur	72
3	Type 1	Orientation	65
4	Type 1	Illumination	76

a) Prewitt

Sl.No.	Group	Type of Images	Recognition Rate (%)
1	Type 1	Clean	76
2	Type 1	Blur	62
3	Type 1	Orientation	70
4	Type 1	Illumination	65

b) Robert's

Sl.No.	Group	Type of Images	Recognition Rate (%)
1	Type 1	Clean	72
2	Type 1	Blur	70
3	Type 1	Orientation	65
4	Type 1	Illumination	66

9 CONCLUSION

In this paper, We present an approach for comparing different edge based methods and finally recognize automatic text extraction from camera captured images, first we detect the presence of text using Gaussian Kernel, dilation and logical AND operation are applied for locating text blocks and finally edge based OCR is applied to extract text. Based on the different edge algorithm for each and individual edge algorithm recognition is performed and result is compared. Based on the result Canny Edge Algorithm Performs better for camera captured documents when compared to other algorithms because canny yields thin lines for its edges by using non-maximal suppression. Canny also utilizes hysteresis when thresholding. The proposed method outperforms the existing methods as shown in the result but in future measures should be taken such that the algorithm also works under different scale and lightning conditions.

10 REFERENCES

1. Jian Liang, David Doermann and Huiping Li: Camera-based analysis of text and documents: a Survey, Springer-Verlag 2004.
2. Majid Mirmehdi: Special issue on camera-based text and document recognition, Springer-Verlag.2005.
3. Palaiahnakote Shivkumara, Weihua Huang and Chew Lim Tan ,Efficient Video Text Detection using Edge Features, IEEE 978-1-4244-2175, 2008.
4. Keechul Jung, Kwang In Kim and Anil K. Jain: Text Information Extraction in Images and Video: a Survey, Pattern Recognition, 37 PP.977-997, 2004.
5. P.Shivkumara, G.H.Kumar: New Filter Based Unsupervised Rules for Boolean Metric ICCTA, Kolkata, India, Pp.611-617, 2007.
6. Y.Zhong and H.Zhang and A.K.Jain' Automatic Caption Localization in Compressed Video IEEE Trans. on Pattern Analysis and Machine Intelligence 22(4), pp.385-392, April 2000.

7. Q.Ye and Q.Huang and W.Gao and D.Zhao: Fast and Robust Text Detection in images and Video sequences Image and Vision Computing pp.565-576 23, 2005.
8. K.C.Kim and H.R.Byun ,Scene Text Extraction in Natural Scene Images using Hierarchical Feature Combining and Verification
9. W. Frei and C.-C. Chen. "Fast boundary detection: A generalization and a new algorithm". IEEE Trans. Comput., vol. C-26, no. 10, pp.988-998, 1977.
10. W. E. Grimson and E. C. Hildreth. "Comments on Digital step edges from zero crossings of second directional derivatives". IEEE Trans. Pattern Anal. Machine Intell, vol. PAMI-7, no. 1, pp. 121-129, 1985.
11. R. M. Haralick. "Digital step edges from zero crossing of the second directional derivatives," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, no. 1, pp. 58-68, Jan. 1984.
12. J. F. Canny. "A computational approach to edge detection". IEEE Trans. Pattern Anal. Machine Intell, vol. PAMI-8, no. 6, pp. 679-697, 1986.
13. J. Canny. "Finding edges and lines in image". Master's thesis, MIT, 1983.
14. M. H. Hueckel. "A local visual operator which recognizes edges and line". J. ACM, vol. 20, no. 4, pp. 634-647, Oct. 1973.
15. Y. Yakimovsky, "Boundary and object detection in real world images". JACM, vol. 23, no. 4, pp. 598-619, Oct. 1976.
16. D. Marr and E.Hildreth. "Theory of Edge Detection". Proceedings of the Royal Society of London. Series B, Biological Sciences., Vol. 207, No. 1167. (29 February 1980), pp. 187-217.
17. M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer. "A Robust Visual Method for Assessing the Relative Performance of Edge Detection Algorithms". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 12, pp. 1338-1359, Dec. 1997.
18. M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer. "Comparison of Edge Detectors: A Methodology and Initial Study ".Computer Vision and Image Understanding, vol. 69, no. 1, pp. 38-54 Jan. 1998.
19. M.C. Shin, D. Goldgof, and K.W. Bowyer. "Comparison of Edge Detector Performance through Use in an Object Recognition Task".Computer Vision and Image Understanding, vol. 84, no. 1, pp. 160-178,Oct. 2001.
20. T. Peli and D. Malah. "A Study of Edge Detection Algorithms".Computer Graphics and Image Processing, vol. 20, pp. 1-21, 1982.