Advances in Aeronautical Science and Engineering
ISSN: 1674-8190

Edge Detection and Multiscale Approaches for Text Extraction in Complex Image Scenarios: A Review

Clara Hernandez*[1] & Dr. Marcello Rossi[2]
[1]M. Sc. Student, Department of Digital Communication, University of Barcelona, Barcelona, Spain
[2]Professor, Department of Digital Communication, University of Milan, Milan, Italy

ABSTRACT

Content that shows up in pictures contains significant and helpful in-arrangement. Location and extraction of content in pictures have been utilized in numerous applications. In this paper, we propose a multi-scale edge-based content extraction calculation, which can naturally distinguish and remove message in complex pictures. The proposed technique is a broadly useful content location and ex-footing calculation, which can bargain with printed record pictures as well as with scene content. It is powerful concerning the text dimension, style, shading, direction, and arrangement of content and can be utilized in a huge assortment of utilization fields, for example, versatile robot route, vehicle permit discovery and acknowledgment, object identification, archive recovering, page division, and so forth.

1. INTRODUCTION

The programmed identification of Region of Interests (ROI) is a functioning examination region in the structure of machine vision frameworks. Content inserted in pictures contains huge amounts of helpful semantic data which can be utilized to completely get pictures. Content shows up in pictures either as doc-uments, for example, examined CD/book spreads or video pictures. Video content can extensively be classified into two classifications: over-lay content and scene content. Overlay content alludes to those characters created by realistic titling machines and superimposed on video outlines/pictures, for example, video subtitles, while scene content happens normally as a piece of scene, for example, message in data sheets/signs, nameplates, sustenance compartments, and so forth.

Programmed identification and extraction of content in pictures have been utilized in numerous applications. Archive content restriction can be utilized in the utilizations of page division, docu-ment recovering, address square area, and so on. Content-based picture/video ordering is one of the commonplace uses of overlay content restriction. Scene content extraction can be utilized in versatile robot route to distinguish content based milestones, vehi-cle permit discovery/acknowledgment, object identification, and so on.

We are investigating calculations that can perform universally useful content confinement. Be that as it may, because of the assortment of text dimension, style, direction, arrangement just as the intricacy of the foundation, planning a strong general calculation, which can successfully distinguish and extricate content from the two sorts of im-ages, is loaded with difficulties.

Wang et al. [1] proposed an associated segment based technique which consolidates shading bunching, a dark nearness diagram (BAG), an adjusting and-combining investigation plot and a lot of heuristic standards together to identify message in the utilization of sign acknowledgment, for example, road pointers and announcements. As the creator referenced, uneven reflections result in incom-plete character division which expands the bogus caution rate in this technique. Kim et al. [2] executed a hierarchi-cal include blend technique to actualize content extraction in regular scenes. Be that as it may, creators concede that this technique couldn't deal with huge content very well because of the utilization of neighborhood includes that speaks to just nearby varieties of picture squares. Gao et al. [3] built up a three layer hierachical versatile content location calculation for characteristic scenes. This strategy has been connected in a model Chinese sign interpretation framework which for the most part has a flat or potentially vertical arrangement. We star represented an insights based strategy [4] to recognize and confine content based highlights by computing the spatial power variety. This technique is

basic and quick. In any case, in genuine scenes, because of uneven light, reflections and shadows, an im-age foundation may contain zones with high spatial force variety that don't contain content. Our investigations demonstrated that this calculation did not perform well under some situa-tions. This prompted the advancement of a progressively powerful calculation dependent on edges, a solitary scale edge-based content locale extrac-tion calculation [5] for indoor scene pictures, which is strong regarding text dimensions, styles, shading/force, directions, impacts of enlightenment, reflections, shadows, and viewpoint bending. In this paper, we propose a multiscale edge-based content extraction calculation, a universally useful strategy, which can rapidly and adequately limit and concentrate content from both record and indoor/open air scene picture

3. PROPOSED METHOD

The proposed method is based on the fact that edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Edge strength, density and the orientation
and variance of orientations than those of non-text regions.
We exploit these three characteristics to generate a feature map which suppresses the false regions and enhances true candidate text regions. This procedure is described in Eq.1.

M==nxcxc…

variance are three distinguishing characteristics of text em- bedded in images, which can be used as main features for detecting text. The proposed method consists of three stages: candidate text region detection, text region localization and character extraction.

Candidate Text Region Detection
This stage aims to build a feature map by using three impor- tant properties of edges: edge strength, density and variance of orientations. The feature map is a gray-scale image with the same size of the input image, where the pixel intensity represents the possibility of text.

4. MULTI-SCALE EDGE DETECTOR

In our proposed method, we use magnitude of the second derivative of intensity as a measurement of edge strength as this allows better detection of intensity peaks that normally characterize text in images. The edge density is calculated based on the average edge strength within a window. Consid- ering effectiveness and efficiency, four orientations (0o, 45o, 90o, 135o) are used to evaluate the variance of orientations, where 0o denotes horizontal direction, 90o denotes vertical direction, and 45o and 135o are the two diagonal directions, respectively. A convolution operation with a compass oper- ator (as shown in Fig. 1) results in four oriented edge in- tensity images E(θ), (θ 0, 45, 90, 135), which contain all the properties of edges required in our proposed method. Edge detector is carried out by using a multiscale strategy,

0º kernel 45º kernel 90º kernel 135º kernel

Fig. 1. Compass operator

where the multiscale pictures are delivered by Gaussian pyra-mids which progressively low-pass filter and down-example the first picture diminishing picture in both vertical and even bearings. In our proposed strategy, those acquired multiscale pictures are all the while handled by the compass administrator as individual sources of info.

Feature map generation
As we mentioned before, regions with text in them will have significantly higher values of average edge density, strength

$$fmap(i,j) = \mathrm{N}\{ \quad E(s,\theta,i+x,j+y) \times W(i,j)\} \quad (1)$$

$s=0 \quad \theta \qquad x=-c \; y=-c$

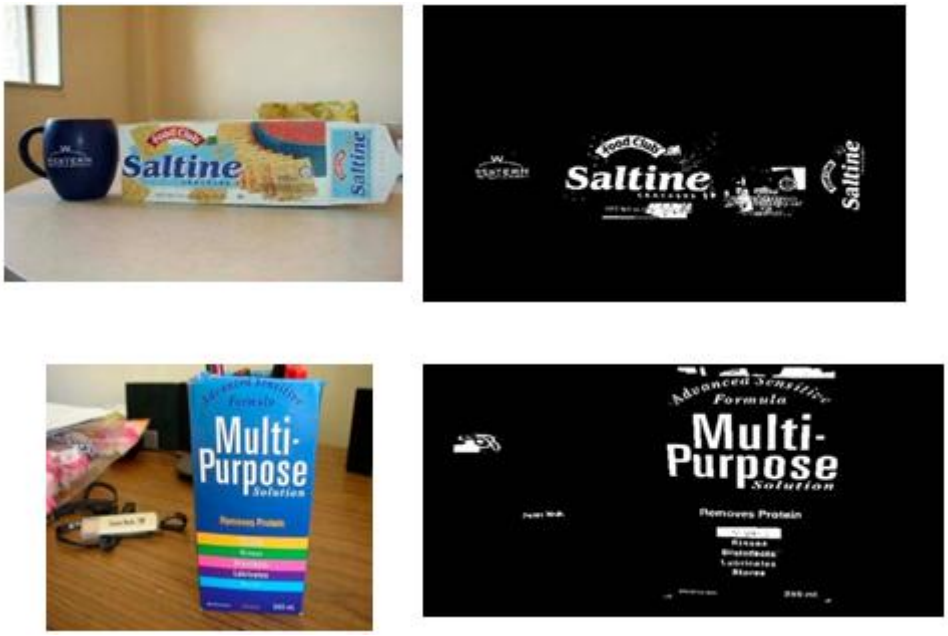
In the above condition, fmap is the yield highlight map, is an over scale expansion task, which utilizes the scale combination. n is the most abnormal amount of scale, which is controlled by the goals (measure) of the info picture. For the most part talk ing, the higher the goals is, the more scales can be utilized. In our usage, we utilize two scales for pictures with goals of 640 480. θ 0, 45, 90, 135 are diverse direction and N is a standardization activity. (I, j) are co-ordinates of a picture pixel. W (I, j) is the weight for pixel (I, j), whose worth is dictated by the quantity of edge ori-entations inside a window. The window size is dictated by a consistent c. To be specific, the more directions a window has, the bigger weight the middle pixel has. By utilizing this non straight weight mapping, the proposed strategy distin-guishes content districts from surface like areas, for example, win-dow outlines, divider designs, and so on.

5. TEXT REGION LOCALIZATION

Typically, content installed in a picture shows up in bunches, i.e., it is masterminded minimally. Hence, qualities of grouping can be utilized to limit content districts. Since the force of the component guide speaks to the likelihood of content, a straightforward worldwide thresholding can be utilized to feature those with high content plausibility districts bringing about a parallel picture. A mor-phological expansion administrator can without much of a stretch associate the exceptionally close districts together while leaving those whose position are far away to one another disconnected. In our proposed strategy, we utilize a morphological widening administrator with a 7 square structur-ing component to the past got parallel picture to get joint zones alluded to as content masses. Two requirements are utilized to filter out those masses which don't contain content [5], where the first imperative is utilized to filter out all the little disconnected masses though the second limitation filters out those masses whose widths are a lot littler than comparing statures. The holding masses are encased in limit boxes. Four sets of directions of the limit boxes are dictated by the most extreme and least arranges of the top, base, left and right purposes of the relating masses. So as to abstain from missing those character pixels which lie close or outside of the underlying limit, width and tallness of the limit box are cushioned by a modest quantities.

Character Extraction

Existing OCR (Optical Character Recognition) engines can only deal with printed characters against clean backgrounds



and can not handle characters embedded in shaded, textured or complex backgrounds. The purpose of this stage is to ex- tract accurate binary characters from the localized text regions so that we can use the existing OCR directly for recognition. In our proposed method, we use uniform white character pix- els in a pure black ground by using Eq.2

$$T = \bigcup_{i=1\ldots} |SUB_i|_z \qquad (2)$$

In the above equation, T is the text extracted binary output image. is an union operation. SUB_i are sub-images of the original image, where i indicates the number of sub-images. Sub-images are extracted according to the obtained boundary boxes in stage two. $_z$ is a thresholding algorithm which segments the text regions into white characters in a pure black background.

6. EXPERIMENTAL RESULTS AND DISCUSSION

So as to assess the presentation of the proposed strategy. We utilize 75 test pictures of four kinds including book covers, object marks, indoor lab nameplates and open air data signs, in which content has diverse text dimensions, hues, orienta-tions, arrangements, point of view projection under various light-ing conditions.

Fig. 2 ~ 5 demonstrate a portion of the outcomes.

Fig. 2. Book cover image (a) Original images (b) Extracted text, cover images are collected from google/yahoo web sites
From Fig. 2 5, we can see that the presentation of our proposed strategy on a wide assortment of picture set is amazing in general. Subsequently, we can reason that the proposed technique is a vigorous and viable way to deal with distinguish content based highlights in complex pictures.

Table 1 demonstrates the exhibition correlation of our ace presented technique with a few existing strategies, where our genius presented strategy demonstrates a reasonable improvement over existing meth-ods. In this table, the exhibition insights of different techniques are refered to from distributed work. Considering a 95% confi-dence interim, it creates the impression that Wang et al. [1], Xi et al. [6] and Gllavata et al. [7] have a comparative execution as the professional

Fig. 3. Object label image with different font sizes, col- ors and orientational alignments (a) Original images (b) Ex- tracted text posed method. However, our proposed methods are suitable for more types of images.

Table 1. Performance Comparison

Method	Image Source No.	Image Type	Precision Rate (%)	Recall Rate (%)
Proposed method	75	Four types	91.8	96.6
Wang et al.[1]	325	Outdoor scene	89.8	92.1
Kim et al. [2]	–	Outdoor scene	63.7	82.8
Agnihotri et al.[8]	293	Text captions	85.8	85.3
Xi et al.[6]	90	Text captions	88.5	94.7
Wolf et al.[9]	60	Text captions	–	93.5
Gao et al.[3]	–	Outdoor scene	–	93.3
Gllavata et al. [7]	326	Text captions	83.9	88.7
Messelodi et al. [10]	100	Document	–	91.2

The overall average computation time for 75 test images (with 480 640 resolution) using unoptimized matlab codes on a personal laptop with Intel Pentium(R) 1.8GHZ processor and 1.0G RAM is 14.5 seconds

(*stddev*. = 0.156), which in- cludes entire run time including image reading, computation as well as image display.

7. CONCLUSION

In this paper, we present a powerful and strong broadly useful content location and extraction calculation, which can automati-cally recognize and remove content from complex foundation im-ages. Our primary future work includes utilizing an appropriate existing OCR method to perceive the extricated content. The contribu-tions of the proposed technique are: can deal with both printed report and scene content pictures. Not touchy to picture shading/power, vigorous concerning text style, sizes, directions, arrangement, uneven illumina-tion, point of view and reflection impacts.

Fig. 4. Indoor nameplate images with different font sizes, per- spective distortion, colors and strong reflections (a) Original images (b) Extracted text

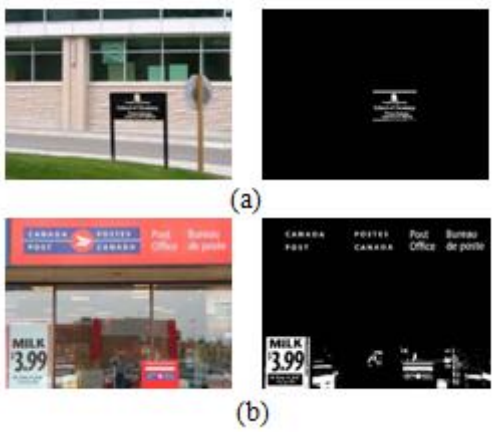


Fig. 5. Outdoor sign image (a) Original images (b) Extracted text

Dissimilar to generally utilized associated segment based meth-ods which investigate each and every character, the proposed strategy just examines content squares. Accordingly, it is com-putationally efficient, which is basic for ongoing applications.

Recognizes content locales from surface like districts, for example, window outlines, divider designs, and so on., by utilizing the vari-ance of edge directions.

Double yield can be legitimately be utilized as a contribution to a current OCR motor for character acknowledgment with no further preparing.

REFERENCES

[1] Kongqiao Wang and Jari A. Kangas, “Character location in scene images from digital camera,” Pattern Recogni- tion, vol. 36, no. 10, pp. 2287–2299, 2003.

[2] K. C. Kim, H. R. Byun, Y. J. Song, Y. M. Choi, S. Y. Chi, K. K. Kim, and Y. K. Chung, “Scene text extraction in natural scene images using hierarchical feature com- bining and verification,” in Pattern Recognition, 2004, Aug. 2004, vol. 2 of ICPR 2004. Proceedings of the 17th International Conference on, pp. 679–682.
[3] Jiang Gao and Jie Yang, “an adaptive algorithm fot text detection from natural scenes,” in Computer Vision and Pattern Recognition, 2001. CVPR 2001, 2001, Proceed- ings of the 2001 IEEE Computer Society Conference on, pp. II–84–II–89.
[4] X. Liu and J. Samarabandu, “A simple and fast text localization algorithm for indoor mobile robot naviga- tion,” in Proc. of the SPIE- IS&T Electronic Imaging 2005, San Jose, California, USA, Jan. 2005, vol. SPIE vol. 5672, pp. 139–150.
[5] X. Liu and J. Samarabandu, “An edge-based text re- gion extraction algorithm for indoor mobile robot navi- gation,” in Proc. of the IEEE International Conference on Mechatronics and Automation (ICMA 2005), Niagara Falls, Canada, July 2005, pp. 701–706.
[6] Jie Xi, Xian Sheng Hua, Xiang Rong Chen, Liu Wenyin, and Hong Jiang Zhang, “A video text detection and recognition system,” in Multimedia and Expo, 2001. ICME 2001, 2001, IEEE International Conference on, pp. 873–876.
[7] J. Gllavata, R. Ewerth, and B. Freisleben, “A robust algorithm for text detection in images,” in Image and Signal Processing and Analysis, 2003. ISPA 2003, 2003, Proceedings of the 3rd International Symposium on, pp. 611–616.
[8] L. Agnihotri and N. Dimitrova, “Text detection for video analysis,” in Content-Based Access of Image and Video Libraries, 1999. (CBAIVL ’99), 1999, Proceed- ings. IEEE Workshop on, pp. 109–113.
[9] C. Wolf, J. M. Jolion, and F. Chassaing, “Text localiza- tion, enhancement and binarization in multimedia docu- ments,” in Pattern Recognition, 2002, Aug. 2002, vol. 2 of Proceedings. 16th International Conference on, pp. 1037–1040.
[10] S. Messelodi and C. M. Modena, “Automatic identifica- tion and skew estimation of text lines in real scene im- ages,” Pattern Recognition, vol. 32, no. 5, pp. 791–810, 1999.