Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online)
Volume 30, Issue 1, 2019
DOI: http://doi.org/10.23851/mjs.v30i1.531

Research Article **Open Access**

Text Detection in Natural Image by Connected Component Labeling

Zamen A. Ramadhan*, Dhia Alzubaydi
Department of computer sciences, College of Science, Mustansiriyah University, IRAQ
*Correspondent author email: zamen927@gmail.com

ArticleInfo

Received
25/05/2018

Accepted
07/07/2018

Published
15/08/2019

Abstract

The process of detecting the text from the natural image is a complex and difficult process because the variance by the devises that take the images and different the texts that found in images in the orientation, size and style. Given the importance the texts in images in the several of application of computer vision. In this paper dependent on the spatial natural images and on the spatial data set for the street sign that include the texts by the different size and different orientation. In this paper detected the texts in images by using robust method by using several algorithms, at the first stage making preprocessing for the image to blur the image and reduce the nose on it by Gaussian blur, second stage making processing that include canny edge detection to detect the edges and dilation, third stage applying connected component to filling all objects in image then applying stroke width transform(SWT) to detect the letter candidate and applying the system on the several images that include different types of texts, we got the best result when applying the proposed system on the several images that related on the sign street images and detected the most of the letters in these images.

Keywords: Gaussian blur, edge detection, Canny edge detection, Connected component, SWT.

الخلاصة

عملية تحديد النص في صورة طبيعيه هي عملية معقدة وصعبه بسبب التباين الموجود بين الاجهزة التي تاخذ الصور واختلاف في النصوص الموجودة في الصور من الناحية الاتجاه والحجم. تعزى اهمية النصوص في الصور وذلك في العديد من تطبيقات computer vision. في هذا البحث اعتمدنا على صور طبيعيه خاصه على مجاميع بيانات خاصه لاشارات الشوارع التي تتضمن النصوص بمختلف الاحجام والاتجاهات وتحتوي على العديد من المشاكل الت عالجناها بالتقنيه المستخدمه في البحث.فس هذا البحث حددت النصوص وذلك باستخدام طريقه متينه باستخام عدة خوارزميات في البدايه عملنا تهيأة ماقبل المعالجه للصور لعمل blur والتقليل من الضوضاء الموجود على الصورة باستخدام فلتر Guassian وفي المرحلة الثانيه عملنا معالجه التي تتضمن تحديد الحواف باستخدام canny edge detection وتقنية dilation وفي المرحلة الثالثه طبقنا connected component labeling لملأ الاشياء المحددة في الصورة وبعدها طبقنا (stroke transform width (SWT لتحديد الحروف المرشحه وقد حصلنا على نتائج جيدة وذلك من خلال المقارنه مع الدراسات السابقه بحيث تم تحديد كل حروف النص في الصور وعلى انواع مختلفه من الصور التي تتضمن انواع مختلفه من النصوص

Introduction

Recognition and text detection used in images real life display acquired wide awareness in past ten years. Unlike text recognition used in the documents, that is being addressed in a satisfactory manner by the state of the art optical character recognition systems. The text detection method is a challenging task to rapidly increase the digitization of all the material. It is complex because needing to find out where the text actually located in the image and how much part is not relevant to the recognition process [1].

Text detection in natural scenes are widely researched field also the significant part of various tasks in Computer Vision. Almost all the published procedures for text recognition and localization are established on sequential pipeline processing depending on 3 steps: the 1'st step is the text-localization step, the 2nd step is the text-segmentation and the last one utilizing OCR for processing printed

documents. When using these methods, the method's total rate for success is calculated by taking the result of the success rate for each one of the stages as the decisions that were made in the former stages cannot be refined [2].

Related work

1. **Pratik Yadav and Prabhudev Irabashetti. (2015),** "Robust Text Detection and Extraction in Natural Scene Images using Conditional Random Field Model and OCR" [3]**.** In this paper, they are not focused the detection of largely blurred text in low -resolution natural scene image. They proposed system by using MSER that include several stages character candidate extraction, text candidate extraction, text candidate elimination and text candidate classification, they used conditional Random field (CRF) for analyzing Connected Component. The approach also implements text extraction utilizing OCR. It is accurate and fast procedure.
2. **S.Anila and N.Devarajan. (2016),** "text detection and recognition using stroke width transform (SWT)" [4]. They proposed technique to detect and recognize the texts of arbitrary directions in complex natural scenes image. This proposed algorithm include: first step detect Maximally Stable Extremal Region(MSER),second step filter character candidate using connected component analysis, third step filter character candidates utilizing stroke width image, after that define bounding box enclosing text region and applying optical character recognition on text region, This algorithm has been assayed for different kinds of images which has various scales orientations, color and size, the testing has been commenced for several images which are stored in the database(horizontal and non-horizontal).

Image Enhancement

The process of image enhancing is one of the image preprocessing methods. The image is a robust medium to carry visual information. Digital images now and then unintentionally bungled by unwanted signals, called noise. In digital Image Processing, eliminating of noise is a greatly demanded area of research. Images are most of the times ruined by noise through the processes of apprenticing and acquisition, images subjected to noise may be found in a great deal of existing time imaging applications. Television images are bungled as a result of atmospheric intrusions and diminution in image closeting, Furthermore, noise is adduced in digital art works during the process of scanning corrupted surfaces of the original, darting noise corruption often happen in digital image acquiring or dispatching process as a result of photo-electronic sensor failings or channel bit errors, there are various noise types and various sorts of filter for removing the noise[5]

Color image to Gray Scale

Grayscale is a domain gray shade in the absence of obvious color. The darkest shade possible is black, which is the entire nonappearance of light that is reflected or transmitted. White is the lightest shade possible, the whole light reflection or transmission at all the apparent wavelengths. average gray shades are demonstrated by equal numbers of the 3 main pigments (C, M, Y) for the reflected light or equal levels of brightness of the main 3 colors (R, G, B) for the transmitted light. While color image has three colors, any pixel is represented by (R, G, B) of intensities for the colors red, green, and blue[6].

Gaussian Blur

The process of image blurring across gauss function called Gaussian blurring. In commoners, this method is utilized to decrease the image detail and noise; it is widely utilized in graphics software and image processing applications. The Gaussian smoothing is utilized as preprocessing phase in the algorithms concerned with computer vision for the purpose of enhancing the image structure at various scales and enormous useful as detectors for edge detection [7]. In this technique the central idea is to update the pixel's value with

the average of neighboring pixels. Instead of calculating the average of all neighbor pixels, will compute a weighted the average of them. If this value congregates on related blur pixel, this displays average weight of each pixel has the biggest value. Basically, Gaussian blurring method helps to find the weight of each neighbor pixel [8]. This filtering algorithm can be considered as one of the widely wiled and most consequential algorithms in image processing (non-liner filter) Gaussian filter (G) is characterized in equation (1).

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-x^2+y^2}{2\sigma^2}} \quad (1)$$

That G is Gaussian mask at the position with x,y coordinates, and σ is a parameter that defines standard deviation of Gaussian.
The image smoothing effect will be higher when the value of σ is large, the equation of smoothing is:

$$\mathbf{f(x,y)} = \sum_{\mathbf{i=0}}^{\mathbf{h-1}} \sum_{\mathbf{j=0}}^{\mathbf{w-1}} \mathbf{G(i,j)I(x-i,y-j)} \quad (2)$$

Canny edge detection

The Canny edge detector is an operator that detects the edges that used several stages algorithm to detect the hard edges in image. It is suggested by John F. Canny in 1986. Canny edge detector or include five steps [9],[10].

- Applying Gaussian blur to clear any dotted on the image and reduced the noise from image.
- Canny edge detector used gradient operator for obtaining the gradient for the edges and direction.
- Canny edge detector determines if the pixel in the image is better edge candidate than its neighbors.
- Double thresholds: to detect the strong edges.
- Final edges are circumscribed by oppressing all the edges that are not related to the strong edges by hysteresis.

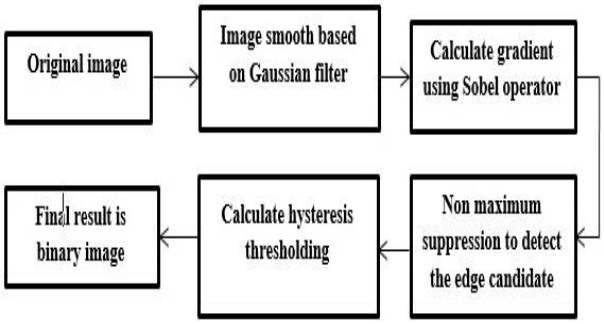

Figure 1: Canny edge detection steps.

Canny edge detector is determined based on three parameters:

- The width of the Gaussian mask wields in the smoothing step and the larger and smaller thresholds used for the edge tracker.
- When the increase the width of the Gaussian kernel will be debase sensitivity the edges to noise.
- When the increase the width of Gaussian then the localizing error increase.

Dilation technique

This technique is one of four significant operators in mathematical morphology field, a concept for analyzing spatial structure. The techniques of mathematical morphology make it feasible for big number of powerful techniques related to image analysis, thus operators and their applications are of important practical and theoretical benefit to a lot of concerned with image analysis and processing. Nearly all algorithms regarding morphology are founded on these 4 operators. Dilation is exemplarily paten to the binary image; however there are some types which can work on gray scale images. The major impact of the binary image's operator is incrementally extending regions boundaries of foreground pixels (i.e. typically white pixels). So, the areas of foreground pixels become big, at the same time the holes in these regions become small, it enlarges the width of maximum regions and increase the valleys, thus eliminating negative impulsive noises but do tiny on positives ones [11].

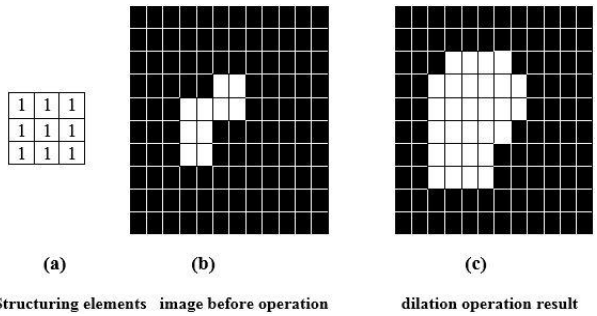

Figure 2: Dilation technique

Two date pieces is taken as input by the operator of dilation. First data piece is the image that need to be dilated; The second data piece is a (generally small) group of coordinate points recognized as structuring element or (kernel), the exact dilation impact on the input image is determined by the kernel. One of the simplest applications of dilation is for bridging gaps. The kernel has been utilized for the purpose of gap repairing bridged gaps [12], [13].

Connected component labeling algorithm

This method is a fundamental step in many implementations, like the community detection in social networks and coherent structure identification in highly parallel science simulations [14].

The value of a pixel in a binary image is either 0 or 1. 1's will be utilized for implying foreground pixels and 0's for implying background pixels. In image processing the value of a certain pixel is important, in addition to values of its neighbors while processing that pixel itself. A pixel at index (P(x, y)) has 4 immediate neighbors denoted as N_4 (P) and 4 diagonal ones. N_4 (P) of P(x, y) is made up of P(x-1, y), P(x+1, y) P(x, y-1) and P(x, y+1) identified as West, East, North and south respectively. N_D (P) of P(x, y) is made up of P(x-1, y-1), P(x+1, y-1), P(x-1, y+1) and P(x+1, y+1) identified as North-west, North-east, South-west, South-east respectively. The N_4 (P) and N_D (P) are together referred to as 8-neighbors of the pixel P, represented as N_8 (P) .Labeling connected elements in a black-and-white image is the basic operation of image analysis, machine intelligence, pattern identification, and computer vision. This process is necessary when independent objects (i.e. connected components) need to be identified in a black-and-white image [15].

Stroke Width Transform

It predefine a stroke as a contiguous part of an image that formulae a band of roughly constant width. It is a local for the image laborer that computes for every pixel the width of the bulk possibly the stroke that containing the pixel. It is a general process for the job of distinguishing texts from natural images because the characters have the extended shape of closely uniform width. The output image of the SWT has the size that equal to the size of the input image where every element has the width of the width of the stroke related with the pixel. The basic impetus of the stroke width algorithm is that stroke width approximately in the text or in the single character remnants the same of this; though there is the important change in the stroke width in the region of the non-text non-text [16]. The stroke as a portion of the image that shapes a band of roughly width, as shown in Figure 3.

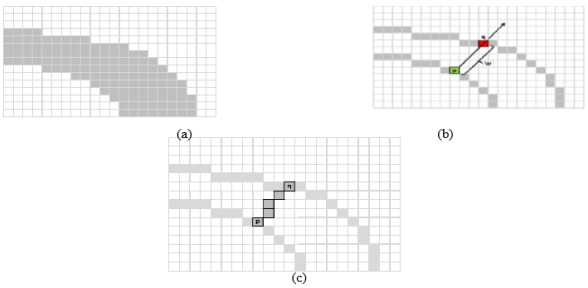

Figure 3: Execution of the SWT.

The first step, every pixel is set with the value ∞ as the stroke width. The second step, consider the edges by way of bearable stroke borders and find the width of the stroke. If p is the edge pixel, the path of the gradient of the pixel is coarsely vertical to the declination of the stroke border so the following step is to direction g_P compute the gradient of the every edge pixels, follow the ray $=p+n*g_P$ ($n> 0$) till find edge pixel q another. If the gradient direction g_q at q is coarsely opposite to g_q, then every pixel is on the ray is given the distance between p and q assigned the stro.ke width for this, except it actually has a lesser value. If g_q is not opposite to g_q or an edge pixel q is not found, the ray is ostracized. It is applying the

algorithm twice for accommodating both bright text on a dark background and dark text on a bright background: once with the ray direction g_q and once with $-g_q$.

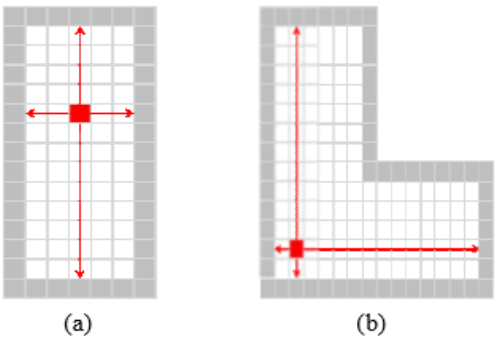

Figure 4: SWT values in the pixels.

In extra difficult situations SWT values, alike corners as shown in figure 4-b after the first step that shown above will not true the stroke width. So, it pass along every non-discarded ray all over again, compute median 'SWT' value m for every pixel, then assigned every pixel of the pixels that on the ray with SWT values that above 'm' to be equal to m [2].

Proposed Method

The proposed method consists of five phases; the first phase preprocessing (gray scale, Gaussian blur), second phase processing(X Gradient, Y Gradient, Canny, merge and dilation), third phase connected component, fourth phase find object and fifth phase stroke width transform(SWT) as shown in Figure 4.

Preprocessing for the proposed algorithm

In this step will make preprocessing for the image to remove it the noise and blur for image, the preprocessing consists of two steps:

- Step one: convert input image to gray scale.

In this step will take a natural image and converted it to the gray scale image.

- Step two: applying Gaussian blur.

In this step will take output image for step one and apply Gaussian blur for reducing the noise that finds on the image and blur for the image that the mask is on the several type (3*3, 5*5, 7*7..).

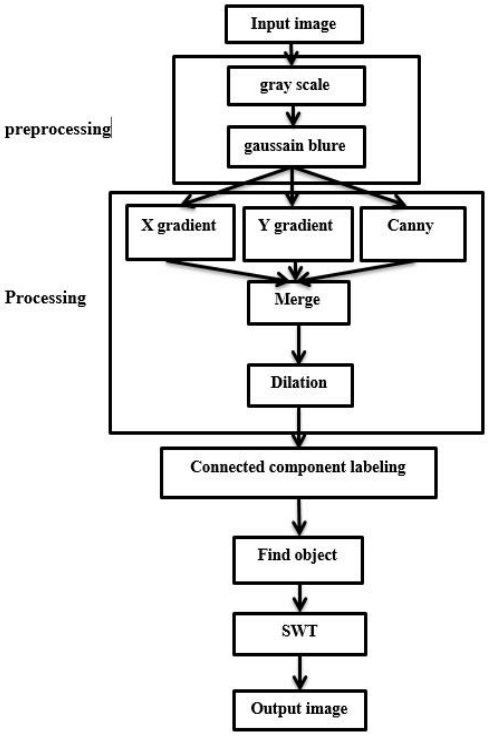

Figure 5: Proposed algorithm.

Smoothed of this image can be completed by the convolve the input image I(x, y) with a Gaussian kernel G(x, y) that achieved by previous Equ(1) as shown in Equ(2), which obtain by computing the sum of products the pixels among the input image and a Gaussian kernal of the size (3×3) or (5×5) Then we will choose the best from the image to determine the kernel size and the value of sigma, then found that the best image when used size of kernel is 7×7 and the value of sigma is 0.1.
As shown in the following algorithm:

Algorithm 1: Preprocessing.

Input: image , sigma , size of kernel **Output**: image
Begin **Step 1**:convert input image to gray scale image **Step 2**: apply Gaussian blur on the gray scale image **End**

Processing

The processing consists of four steps:

- Step one: X gradient, Y gradient.

Find X gradient, Y gradient for the image by computing the gradient and intensity for each pixel in the image. The image is convolved

with both kernels to approximate the derivatives in horizontal and vertical change.

- Step two: Canny edge detector.

Apply canny edge detector by applying sobel filter to the X, Y gradient for edge detection. Then apply Gaussian operator to detect the edges and find two thresholds to detect accurate edges for the image.

- Step three: merge of image.

In this step will merge the resulting image that produced from X, Y gradient and canny edge detector by addition equation to detect the edges by the best manner to produce the best edges.

- Step four: dilation technique.

In this step used dilation technique for the appearance the edges by the best result and filling the space that found in edges in the image that produce from the previous step, applied this technique that include using structuring element and data of the image. When applying dilation technique on the merge image, all edges appear in the best of these cases. As shown in the algorithm (2):

Algorithm 2: Processing.

Input: image
Output: image

Begin
Step 1:find X, Y gradient for image
Step 2: find canny edge detection
Step 3:merge of the image
Step 4: apply dilation technique
End

Connected component labeling

After edge detection for the object in the image, in this step, wanted to fill the objects that include in this image and for making this took every pixel and checked the neighbors of these pixels for connected and made two steps:

Step one: included several passes:

- First pass: checked the currentness pixel if it was background (its value 0) meaning must be ignored and going for the comer pixel, if it was not background, meaning must be staying on the pixel and relocates to the next pass.
- Second pass: gave a label for the currentness pixel and before this checked several cases: 1) If the pixels that above or on the left corner or on the right corner or on the left for the currentness pixel are background, the meaning is giving the currentness new label. 2) If one of the neighbor's pixels that previous shown in case 1 has label then will the same label for the currentness pixel. 3) If more of pixel from neighbor's pixels that previous shown in the case 1 have labels then must be move to the next pass.
- Third pass: checked the labels for the neighbor's pixels that previous shown in the case 1 for the curentness pixel then choose the small label and assigned it to the curentness pixel.
- Fourth pass: moved to the next pass.
- Fifth pass: after the finished the first row, checked second row and continue for every row insomuch finished every pixel in the image, went to the next step.
- Step two: Applied algorithm union find data structure to rearrange the labels of the pixels, in the pixel that connected circle between two pixels above and left made the big label child for the small label then converted every pixels that had big labels to the small labels insomuch converting all the cases that finding in the image then filled all the objects in the image on this method.

SWT

In this step applied SWT algorithm on the all objects that appear in the image that produced from previous step for acquiring of the width for each pixel in these objects, these widths that consider one of the properties that benefit in finding the required conditions for finding the letters candidate to extinguish the text from non-text and extinguish the words in the texts.
As shown in the algorithm 3:

Algorithm 3: Connected component **.**

Input: image
Output: image

Begin
Step1:apply connected component algorithm on the image
Step2:compute stroke width transform (SWT) on the image
End

The following figure shows all the steps of the proposed system:

Figure 5: Explain the proposed method.

Conclusions

In this paper that detects the text in the natural image, the system used connected component and stroke width transform that upon using these technique to detect the text in the street sign images and this proposed system is propertied by using technique that low complexity compared with the previous study and using X gradient and Y gradient before and within the canny edge detection to detect all the edge by the best result and canny evidenced is the best detector for external and internal lines of object figuration edges and has best protection to noise than Robert, Prewitt and sobel operator. The proposed algorithm observed that the hybrid algorithm achieved the precision in detection of edges, applied the algorithm at nearly 20 images and appears the best result; using dilation technique is the best morphological technique to fill the gaps that found on edges. The proposed algorithm for connected component to fill the objects that found in the sign is the best for shown the best results, SWT algorithm is preferable to specify the conditions that are used to find the candidate characters, the proposed system achieved the best result for detecting most of text's characters in image; even though they have any problems in shadow and different characters in size and orientation, the proposed system shows the best result for detecting the text in images that are close in distance.

References

[1] Tajinder Kaur and Nirvair Neeru, "Text Detection and Recognition from Natural Scene", IJARCET Volume 4 Issue 7, July 2015.

[2] Lukas Neumann and Jiri Matas," A method for text localization and recognition in real-world images", ACCV, November 8-10-2010.

[3] Pratik Yadav and Prabhudev Irabashetti," Robust Text Detection and Extraction in Natural Scene Images using Conditional Random Field Model and OCR", International

Journal of Current Engineering and Technology, Vol.5, No.6, 14 Dec 2015.

[4] S.Anila and N.devarajan,"Text detection and recognition using stroke width transform", Asian journal of information technology 15(21): 4318-4324, 2016.

[5] Mr. Mandar D. Sontakke and Dr. Mrs. Meghana S. Kulkarni," Different types of noises in images and noise removing technique", International Journal of Advanced Technology in Engineering and Science Volume No.03, Issue No. 01, January 2015.

[6] Nikki Burnette, spellbinding images: A grayscale fantasy coloring book: Beginners Edition (volume 3), October 31, 2016.

[7] Sonu Jain, Akhilesh Dubey, Diljeet Singh Chundawat and VPrabhat Kumar Singh." Image Deblurring from Blurred Images", International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014), Vol. 2, Issue 3 (July - Sept. 2014).

[8] Leila kabbai, Anissa Sghaier, Ali Douik and Mohsen Machhout," FPGA implementation of filtered image using 2D Gaussian filter", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 7, 2016.

[9] Hemil A. Patel and Kishori S. Shekokar," Detecting Text or character in Natural Scenes with Stroke Width Transform", International Journal of Innovative Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2015.

[10] Hyung I Koo and Duck Hoon Kim, "Scene Text Detection via Connected Component Clustering and Nontext Filtering", in IEEE transactions on image processing, vol.22, no.6, june 2013.

[11] Naser Jawas and Nanik Suciati," Image Inpainting using Erosion and Dilation Operation", International Journal of Advanced Science and Technology Vol. 51, February, 2013.

[12] Prof. sagar B Tambe, Prof. Deepak Kulhare, M. D. Nirmal and Prof. Gopal Prajapati," Image Processing (IP) Through Erosion and Dilation Method", International Journal of Emerging Technology and Advanced Engineering,(ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 7, July 2013).

[13] R. Gonzalez and R. Woods, "Digital Image Processing, Addison-Wesley Publishing Company", 1992, pp. 518 - 519, 549.

[14] J.iverson, C.kamath and G.kaypis,"evaluation of connected component labeling algorithms for distributed memory systems", parallel computing 44(2015)53-68.

[15] Li-Feng He, Yu-Yan Chao and Kenji Suzuki," An Algorithm for Connected-Component Labeling, Hole Labeling and Euler Number Computing", Jouranl of Computer Science and Technology 28(3): 468–478 May 2013.

[16] Kethineni Venkateswarlu and Sreerama Murthy Velaga ,"Text Detection On Scene Images Using MSER ", International Journal of Research in Computer and Communication Technology, Vol 4,Issue7, July-2015.