A Novel Algorithm To Detect and Extract The Texts of Road-Sign Plates in Video Scenes

Dr. Abdul Munem S. Rahma
Computer Science Department, University of Technology/Baghdad
Email: monem.rahma@yahoo.com
Teaba Wala Aldeen Khairi
Computer Science Department, University of Technology/Baghdad
Email: twk232004@yahoo.com

Received on:10/6/2015 & Accepted on:12/11/2015

ABSTRACT

Traffic Road-Signs contain useful information for the road users; the operation of many of modern applications like the automatic or smart vehicle requires an automatic discrimination of the texts of the traffic road-sign. Discrimination of text compose of several stages, the first of these stages is detection and extraction of the texts. In this work, an algorithm is developed to detect, locate, and segment of the texts and the word in the video clips, that existing in the road signs in the city of Baghdad. The proposed approach includes two stages, the first one is processing the image to locate and extract images of the road-sign and neglect the rest of the image, and the second stage is processing the image of the road-sign plate to detect and extract the texts without symbols and shapes. The basic structure of the algorithm depends on the following functions: edge-detector, dilation, and filling-hole, morphological-opening. The total recall values of 89%, the total precision value of 93%.The algorithm is, then, tested based on video clips, implementation of the algorithm based on video clips confirms its ability to detect the texts which may appear in the video scenes, recall-rate(r) is excellent with a total value of 94.5% and a total precision value of 86.5%. The algorithm is tested to measure its validity to work under real-time operation, by processing one frame and exceed a set of next frames, the test appears that the algorithm is able to work under real-time operation with recall-rate(r) of (93%).

Keywords: Text-detection, Road-sign plate, Matlab, Algorithm, Edge-detection.

INTRODUCTION

Development in the field of adapting electronic devices to amend and suggest alternatives to be adopted as a substitute for the human element in making decisions. There were several methods to transfer the information to a computer (the alternative to the human element) some of which are sensors and cameras. Being directed to a computer system work inside the car to be a guide for special signs hints that distinguish roads specially road signs hints, that motivate an attitude for the researcher to distinguish road signs through photos of portable cameras that are to be set later on in the car. Text detection is defined as the task that localizes text in complex background without recognizing the characters individually [1]. Text plays a dominate role in video understanding as a text caries valuable important information related to the video contents. Text

https://doi.org/10.30684/etj.34.2B.1
2412-0758/University of Technology-Iraq, Baghdad, Iraq

detection pertains to the segmentation of text region in a video frame by identifying exact boundaries of text lines. Text detection plays a fundamental role in subsequent steps, such as, text localization, text word segmentation, and text recognition [2]. The large changeability of the appearance of the text like (a complex background, different font styles, and variable size),make the text extraction from images a very rough duty [3,4].A good examples of objects exists in natural environments scenes are Road-Signs, which have a content rich of information. [5]

Road Sign Features

A general study has been made on the road signs in Baghdad city; the road-sign can be categorized into two types: Side way road-sign and Upper panel road-sign. Generally, road sign is of rectangular shape and it may be horizontal or vertical rectangle. The dimensions of the road sign plate depend on the amount of traffic information that is displayed in it and some other factors. It has been observed that there is a variety of standard of the dimensions (height, width) of the road sign in Baghdad city. Background color of the road-sign plate may be one of the following forms:

a. *White*: This color is used as a background in the road-sign of the main and subset streets.
b. *Green*: this color is used for plates that refer to areas outside Baghdad.
c. *Blue*: this color is for road signs of highways lines.

It is possible to find a panel containing more than one background color, for example white and green. The information that may be appeared in road sign is:

- Words in Arabic language, usually, placed in the top of the plate.
- Words in English language, usually, placed under the Arabic words
- Numbers.
- Symbols such as denotation arrow and shapes indicate to airport or mosque, etc.

The font color of those items is white for a plate of green or blue background while for white background the font color will be black. It is important to mention that the front face of the road-sign has intense ability to reflect the light .Therefor, in the dark, once a direct light is applied on the front face of the road sign, all the words and symbols can be seen clearly.

The Proposed Algorithm for Text Detection and Extraction

In this work, an algorithm has been developed, the task of this algorithm is to detect, localize and extract the texts of the road signs which may appear in a video clip. The proposed program is developed as an m-file by using Matlab under Microsoft Windows 7- 64bit Professional operating system. In general the program consists of the following main four stages:

a) Preprocess.
b) Detection and locate the region of the road sign (or road signs).
c) Detection and locate the text (or texts) inside the road sign.
d) Post-processing (display the results).

Figure (1) shows a general graphical chart of the proposed algorithm.

Preprocess Stage

The videos which are processed in this study were shouted by a digital camera that mounted on a moving vehicle. The algorithm is adapted to be suitable for most of Video format like (AVI (*.avi), MPEG (*.MPG), MP4).The program can be adapted to process video of any resolution. Preprocessing stage consists of the following sub-stages:

- Converting the video into frames.
- Pre-Process the frames: the source video is framed into a number of frames, each frame of the source video is saved, with (Frame-x) label, in jpg format. This image converted into gray-scale then the boundaries of the objects in the image are obtained by using Canny Edge Detector (CED). Two thresholds have to be specified to adjust the operation of canny-function which are:

A. 2-elements sensitivity thresholds vector, those 2-elemnts are the low threshold (T_1) and the high threshold (T_2).
B. Sigma, which is the standard deviation of the Gaussian-filter (GF).

Figure (2) shows frame number-27 of the source video while Figure (3) and Figure (4) show the gray-scale and CED of this image respectively. The objects, appear in Figure (4) in term of its boundaries, thus, it is possible to extract the features of those objects like: shape, dimensions, area, etc. Also it is possible to note the presence of meaningless, small objects. To reduce the complexity of the image, the small unwanted objects are removed by using the morphological-opening function. Opening function cleans all connected components (objects) that have less than P pixels, so these functions have two inputs which are:

a. The binary image (to be cleaned).
b. The threshold P_1, which is a function of image resolution.

Figure (5) shows the image after applying morphological-opening function.

Detection and Extraction of the Road-Sign Plate

To detect the Road-Signs (RS) there are several steps that will be explained in the next sections:

Detection of Complete RS Plate

As mentioned before the road signs have a rectangular shape therefore, the first step of this stage is to find the rectangular objects. Normally, after applying CED, the boundary of the RS appears as a complete and continuous rectangle. Discontinuity may arise in the RS shape because of presence of defects in the RS itself, long shooting distance or other reasons which may deformation the image. It is important, here, to enhance the image to ensure appearing all the rectangles fully without any defects, to do this, morphological-dilation is applied. The morphological-dilation uses a structure element (SE) object for expanding shapes contained in the input binary image. There are two characteristics that must be identified: 1-Shape of the SE and 2-Size of the SE. To complete the blanks of the border for any rectangles, it is needed two dilation operations; the first dilation applied on the pair of vertical lines, line structure element is used with angle of 90̊. The second dilation is applied on the pair of horizontal lines, also, line SE is used but with angle of 0̊. Figure (6) shows the image after applying the two dilations. The next step in this stage is to fill all holes in the image. As a result of this process, all closed area will turn into solid object as shown in Figure (7). It is obvious from Figure (7) that all the road sign areas are included in this action (that is required), in addition to some objects that don’t mean anything. The strategy, that used to discriminate the road-sign comprises:

1) Labeling the connected pixels objects where the background is labeled 0 and label 1 dedicated to the first object and label 2 to the second object and so on.
2) Measuring properties of the objects, among many features provided by region props-function, it will be relied on only the required features to determine whether the object is RS or not. These features and their descriptions are listed in Table (1).

The object can be considered as a RS plate if the following conditions are true:

i. Extent value has to be close to unity which means the object is rectangle.
ii. The area of the object has to be larger than a specific limit, this condition leads to exclude all small and meaningless objects also small and far RS plates that have tiny and unreadable texts. The threshold or the limit of the area is determined to be larger than 1 percent of the image area.
iii. Ratios of object width to object height have to approximately equal one of the known ratios of plate width to plate height. This condition is based on the ratios not on the dimensions because the dimensions are change with shooting distance while ratios remain constant as shown in Algorithm (1).

Algorithm (1) Preprocess the Frame
Input: Image (frame) of source video. Output: Edge-Detection of (IM) with applying Opening Function.

Step1: Start.
Step2: Save frame with (Frame-x) label and in jpg format.
Step3: Read image IM=frame[x].jpg.
Step4: Measure the size of image IM (height, width).
Step5: Set MS=zero matrix (same dimension of IM (h, w)).// MS is a gray-scale of IM//
    For Ih= 0 to high
      For Jw=0 to width
        Get Red(R), Blue (B), Green (G); MS(Ih,Jw)= apply equation(2.1);
      End Jw.
    End Ih.
Step6: Find(g3)= binary-image with edge by applying CED on image (MS) according to T_1, T_2 and sigma thresholds.
Step7: Remove all connected objects of less than P_1 pixels by using Morphological Opening-function of image (g3).
Step8:End.

Detection of incomplete RS plates

In same video clips, it is possible to appear an uncompleted RS (reminder of the plate is outside the scope of photography). Example of such cases is shown in Figure (8). Reviewing and analysis of many video clips, gives the following notes:

1. Some of the RS are incomplete in all frames of the video clip.

2. Some of RS are complete in a number of frames and incomplete in the others.

By applying the same procedure which described in Section (3.2.1), incomplete plate will not be detected at all because for those RSs, and by applying the edge detector, only three sides of the framework will appear and the fourth side will be outside the scope of the image. Therefore, the solid object cannot be achieved, and then the road sign cannot be detected. An algorithm is proposed. The idea of this algorithm can be summarized as:

a. Find the incomplete RS by searching a pair of parallel lines which perpendicular to one of the image edges.

b. If the length of these two lines is taller than a specific value and the distance between them is more than a specific value, it could be considered that these two lines are an opposite sides of the rectangle.

c. To complete the fourth side of this rectangle, a line align to the image edge will be added between the two opposite sides.

The conditions, which determine whether the solid object is incomplete RS or not, are different from the conditions of the complete RS, because there are differences between the two cases in term of areas, dimensions and locations.

The object can be considered as an incomplete RS plate if the following criteria’s are true:

- Extent value has to be close to unity (this condition is common for the cases of complete and incomplete RS plate).
- The solid object has to be tangent to one of the edges of the image.
- Generally the area of incomplete RS is less than the area of complete RS. Neglecting the incomplete the RS which has exposed portion of less than 30% of the total sign area. On this basis,

the threshold or the limit of the area is determined to be more than 0.3 percent of the total area of the image as shown in Algorithm(2,3,4).

Algorithm(2) Complete RS Which Severed from the Top of Image
Input: bw=binary image after morphological dilation, Image Width, and Image Height Output: Binary image (bw) with closed rectangle from top of image.

Step1: Start.
Step2: Set Right Line=0;Flag=0;
For J=1 to ImageWidth-5
Step3: Evaluate Sum Top=number of white pixel in the slice of bw(1:100,J:J+5);
Step4: IF Sum Top>50 and J-Right Line>20 Then //distance between 2 rectangles have to be at least20pixels//
Flag=Flag+1;//there is a nominee line there//
Step5: IF Flag =1 Then
LeftLine=J;//specifying location of left side of the rectangle//
Step6: IF Flag=2 and J-Left Line>25 Then //distance between 2 sides have to be at least 25 pixels//
Right Line=J;//specify position of right side of the rectangle//
bw (1:5,LeftLine-5,RightLine+5)=1;//create a line, align to the top edge of the image, between the two sides of the rectangle, with thickness of 5 pixel width//
Flag=0;//reset flag value for next rectangle//
End IF. End For.

Algorithm(3) Complete RS Which Severed from the Left Side of Image
Input: bw =binary image after morphological dilation, image width, and image height Output: Binary image (bw) with closed rectangle from left side of image.

Step1:Start.
Step2: Set Lower Line=0;Flag=0;
For I=1 to ImageHeight-5
Step3: Evaluate Sum Left=number of white pixels in the slice of bw(I:I+5,1:100);
Step4: IF Sum Left>50 and I>LowerLine+20 Then
Flag=Flag+1
Step5: IF Flag =1 Then
Upper Line=I;//specifying location of Upper side of the rectangle//
Step6: IF Flag=2 and I>UpperLine+25 Then
Lower Line=I;//specify position of Lower side of the rectangle//
bw(UpperLine-5:LowerLine+5,1:5)=1
Flag=0;//reset flag value for the next rectangle//
End IF.
End For.

Algorithm(4) Complete RS Which Severed from the Right Side of Image
Input: bw=binary image after morphological dilation, image width, and image height Output: Binary image (bw) with closed rectangle from right side of image.

Step1:Start.
Step2: Set Lower Line=0;Flag=0;
For I=1 to Image Height-5
Step3: Evaluate Sum Right = number of white pixel in the slice of bw(I:I+5,ImageWidth:ImageWidth-100);
Step4:IF Sum Right>50 and I>LowerLine+20 Then
Flag=Flag+1;
Step5:IF Flag =1 Then Upper Line=I;//specifying location of Upper side of the rectangle//
Step6:IF Flag=2 and I>UpperLine+25 Then
Lower Line=I;//specify position of Lower side of the rectangle//

```
        bw(UpperLine-5:LowerLine+5,ImageWidth-5:ImageWidth)=1; Flag=0;
        End IF.
      End For.
```

Crop the RS Plates

The solid objects which satisfy the set of criteria mentioned in Section (3.2.1) and the objects which satisfy the set of criteria mentioned in section (3.2.2), these objects, often represent RS plates. By using imcrop-function the RS are cropped, where this function is used to crop a portion of the original image (IM), this portion is appointed by the location and the area of the RS. The cropped portions will be saved as an image and passed to the next stage of the algorithm and the rest of the original image will be ignored. It is important to mention here, in this work, the proposed algorithm has the ability to crop all the RS plates in the image whatever the number. Examples of images with their cropped RSs are shown in Figure (9).

Detect and Segment the texts Inside the RS Plate

For each plate, which is detected and cropped by using the previous steps of the algorithm, the following procedure will be applied in order to detect the texts inside this RS plate:

a) The image is converted to grayscale binary image as shown in Figure(10a) and Figure(10d).

b) The edge of the objects in the image is obtained by using CED with T_1=0.04, T_2=0.1 and δ=3 as shown in Figure (10b) and Figure (10e).

c) Noise and un-wanted edges are removed by using the morphological open-function with threshold P_2 as shown in Figure (10c) and Figure (10f).

d) Connecting all components and letters of the word to alter it to one, continuous and interrelated piece. English words consist of a number of isolated letters, so that dilation function is used with line SE at angle of 0̊, to connect letter with each other. The small letters (i) and (j) contain a point, again dilation function is used to connect the points with the rest of word, the SE, here, is line with 90̊ angle. Arabic words consist of one section, or two sections or even more, dilation function with line SE at angle 0 is used to link these sections as happened with English words. The letters(ب،ت،ث،ج،خ،ذ،ز،شن،ض،ظ،غ،ف،ق،ن،ي) contain points and the letter(ك) contain Hamza " ء", as previously mentioned dilation function will connect the points with the rest of the word. In some cases, such as that letter (ک) comes after the letter (ر) or after the letter (ز), the need is arisen to use dilation function with disk-shaped structure element to achieve the connection between the letters. It is important to determine appropriate values of the length of the line SE and the radius of the disk SE, for low values, the required connection between the components of one word will not be achieved. At high values, it is possible to overlap adjacent words with each other. These wrong and suitable dimensions of the SE lead to inaccurate detection. The dimensions of the SE have to be a function of word size, or in other words a function of the size (Area) of the RS. Studying and analysis of a variety of RSs and the texts inside them, the following empirical formulas are developed for the dimensions of the SEs:

a. For line SE (at angle =0̊): Dimensions=$C_1 \times a \times resolution$

b. For line SE (at angle =90̊): Dimensions= $C_2 \times a \times resolution$

c. For Disc SE: Dimensions=$C_3 \times a \times resolution.$

Where a=Area of the RS, and C_1,C_2,C_3= constants, resolution= Resolution of the image. The image after applying the dilation function is shown in Figure (11).

e) Morphological opening function is applied, again here, to remove all the unconnected small object the threshold (P_3) is a function of plate area. P_2 and P_3 are a function of both of area (a) and the image resolution, but P_3>>P_2. Figure (12) shows the image of the plates after applying morphological opening-function.

f) Labeling all the objects in the image by using bwlable-function.
g) Measuring the properties of all objects in the image by using regionprops- function. These properties are: Area, Bounding-Box, Locations (upper corner of the bounding-box), and Dimensions (plateheight, platewidth).

Post processing: Extraction and Display the Results

Each object will be considered as a text if it satisfies the following conditions:
1] The area of the object has to be greater than 1.5 percent of plate area.
2] Object width has to be more than 1.2 times the object height. This condition is based on the fact that all Arabic and English words are written horizontally. By these two criteria's, the texts will be segregated from the rest symbols and signs in the RS plate. The last step of the Algorithm, in this work, is to display the results i.e. detection and extraction of all the texts in the RSs which
3] appear in a video as shown in Algorithm (5). For each frame in the video clip, the results may be displayed according to one of the following forms:
- Locating the text on the image by drawing a rectangle around each text, as shown in Figure (13).
- Crop the texts from the image and display them as shown in Figure (14)

Algorithm (5) Detection the Text and Display the results
Input: Cropped RS Image (IM_2), platestartX, platestartY, plateArea Output: Display the results

Step 1: Start;
Step 2:Convert(IM_2) to gray-scale image(Imsb);
Step 3:Find t3=CED of (Imsb)with T1=0.04,T2=0.1,δ=3;
Step 4:Apply morphological opening on(t3) with threshold P_2 to find image (g3sb);
Step 5:Apply dilation function with line SE and angle 0°;
Step 6: Apply dilation function with line SE and angle 90°;
Step 7: Apply dilation function with disc SE;
Step 8: Label all the objects in image (Idsb).
Step 9:Measure the properties of all objects in image(Idsb);
Step 10: Measure the number of labels (nsb).
For isb=1:nsb
IF text Width>=1.2 $\times$ $text\ Height$ and text Area >0.015 $\times$ $plateArea$ Then
Step 11.1 crop object(isb) from image(IM_2);//coordinates of upper-left corner of object(isb) = (textstartX,textstartY)width of object(isb)= textWidth height of object(isb) = textHeight//
Step 11.2 Draw a rectangle with the following parameter:
• Coordinate of upper-left corner =(platestartX +textstartX,platestartY+textstartY)
• Rectangle width= textWidth
• Rectangle height= textHeight
End IF;
End For;
Step 12: End.

Arrangement of the Main Algorithm

Majority of the proposed program is described in a form of sub-algorithms. The main algorithm of the program includes all these sub-algorithms taking the following points into account:

A. The following steps are repeated for each frame of the source video: Pre-process the frame, Algorithm which solves the problem of incomplete road-sign plates. Detection of road-sign plates [as described in Section (3.2.1)]. The resultant is detection of the road-signs in the frame.

B. For each road-sign, the steps (Detection the Text and Display the results) is repeated and the final results are obtained.

C. Execution time of the program depends on the following factors: Duration of the video clip (in second). Number of frames per one second. Execution time for one frame.

If all frames is processed the execution time will be long. In video processing applications it is not required to process all the consecutive frames, because the variation of the scene, from frame to next frame is trivial. In order to reduce the execution time, processing can be carried out on a single frame pre a number of frames equals to (frame step). By choosing an appropriate value of (frame step see Table 5), the objective of the program can be achieved with an acceptable execution time as shown in Algorithm (6).

Algorithm(6) Main Algorithm
Input: Source Video Output Detected Text Regions in the Road Signs

```
Step 1: Start.
Step 2: Call Reading Video File
Step 3:fstep=1; //frame step//
For x=1:fstep:number of frames
Step4.1: Call pre-process the frame;
Step4.2: Display image(IM);
Step4.3: Apply dilation function with line SE at angle 90̊.
Step4.4: Apply dilation function with line SE at angle 0̊.
Step4.5:Call complete RS which severed from the top of image;
Step4.6:Call complete RS which severed from the left side of image;
Step4.7:Call complete RS which severed from the right side of image;
Step4.8:Fill holes in the image;
Step4.9:Label all the objects in the image;
  Step4.10:Measure the properties of all objects;
        //plate Area, plate Width, plate Height, plate start X, plate start Y, plate Extent//
  Step4.11:Count the number of objects(n object);
For ii=1:nobject
IF (plate Area>0.01 image area and plate Extent>0.8 and plate Width/plate Height is approximately
equal to one of the following values[2.25, 3.6, 1.25, 1, and 1.36]) //set of criteria for complete road-sign//
or (plate Area >0.003×image area and plate Extent>0.8 and (object upper corner locate on left or top
edge of image or upper corner + plate Width locate on right edge of image))//set of criteria for
incomplete road sign//
Then
Step5.1:Crop the object(ii) from image(IM) and save it as an image(IM2);
// IM is one frame of the source video and IM2 is the image of one of road-signs in IM, coordinates of
upper-left corner of the object(ii)                =(plate start X, plate start Y),
  width of object(ii) =plate Width,
  height of object(ii)=plate Height //
Step5.2:Call Detection of The text and Display the Results;
// Algorithm(3.6), The output of this algorithm is one of the following forms:
-      Display rectangle surrounding the text[Step11.2 in Algorithm(3.6)] on image(IM) [Step4.2].
-      Crop and display the texts and save it as an image.//
End IF;
End For;
End For;
Step6: End.
```

Experimental Results and Implementation Analysis

Terminology of the Evaluation

In order to evaluate the performance and measure the robustness of the present algorithm, the following terms and ratios are used [6, 7]:

1- Recall(r): divide the correctly detected texts to the sum of false-negative and correctly detected texts. False-negative means the region which is indeed a text, but the algorithm has not detected it.

$$r = \frac{C_t}{C_t + M_t} \times 100\% \qquad \text{………..... (1)}$$

Where Ct = Correct Detected Text and Mt= Missed Text (false-negative).

2- Precisions (p): divide the correctly detected texts to the sum of false-positive and correctly detected texts. The term false-positive means the region which is indeed not a text, but the algorithm has detected it as a text.

$$p = \frac{C_t}{(C_t + F_t)} \times 100\% \qquad \text{…………. (2)}$$

Ft= False Positive.

3- F-Measure (F_s): is the weighted average of the recall rate and precision rate.

$$F_s = 2 \times \frac{r \times p}{r+p} \quad \text{………... (3)}$$

Evaluation Based on Image Tests

The algorithm is subjected to several tests, taking into account the different variables and effects. As a first step, to assess the application of the algorithm on video clips, it is necessary to carry out tests on single frames of these videos. Many cases study are analyzed and evaluated and the results are tabulated in Table (2).

Evaluation Based on Sequential Fames Tests

Realistic and effective evaluation of the algorithm is achieved by measure its ability to detect texts in video clips. All the sequential frames of a video are tested and evaluated; the eventual assessment of the algorithm is the resultant of the evaluations of all frames in that video. The values of r and p are changed from frames to another where these values are affected by the following factors: Vibration of the vehicle and the camera, Shooting distance and Shooting angle. Many video clips are tested and the results are summarized in Table (3).

Evaluation Based on Video Tests

For each tested video, recall and precision rates is determined based on the total number of words, the number of detected words and the number of false-positive detection during that video. According to the results of Table (4), the values of recall, precision rate are perfect and this demonstrates and confirms the validity and the accuracy of the proposed algorithm for various operating conditions.

Measure the Performance Based on Real-time Operation

In addition to evaluate the ability of the algorithm to detect the texts in the video clips, there is an important factor which must be measured and discussed, that factor is the time required to execute a video clip. It is noted that this time depends on the following parameters:

a. Specification of the video clip.

b. Number of frame per second.

c. Resolution of the video clip.

d. Nature of video clip according to: number of road-signs plates, number of texts, complete or incomplete plate and complexity of background.

By testing the algorithm on a variety of video clips, average execution time of one frame is ranged between (1.2-2.8 sec), the algorithm may operate under real-time condition. To evaluate the suitability of the algorithm to work under real-time conditions, the following case study is tested in table below the description of the case study:

Video Duration	Video Format	Total No. of Frames	Total No. of text	Frame Rate	Day time	Weather condition	Speed of Vehicle	Resolution
17 Sec.	Avi(.avi)	422	28	24 fps.	noon	Weather condition	60 Km/hr	(1920×1080),(1280 ×720), (768×575)

The results of execute this video with the three different resolutions are tabulated in Table (5). From Table (5), it can be concluded the following notes:

I. When all frames of the video are processed, the detection rate is excellent and this confirms the results obtained in Section (4.3).

II. To reduce the execution time, the number of processed frames is reduced to insure that the execution time is equal or less than the duration time of the video clip. In this situation, it could be said that the algorithm is able to work under real-time operation.

III. When the algorithm operates under real-time condition i.e. reduced number of processed frame the detection rates are, still, good and acceptable.

So it could be said that the proposed algorithm has the ability to detect the texts in videos with an acceptable-high detection rate and under real-time condition.

CONCLUSION

The significant conclusions of this work may be outlined into the following points:

1. Evaluation based on a single image test indicates promising results, where r is ranging from 70%to100% and p is ranging from 86%to100%.minimum values of text-detection rates appeared at long-distance shooting evaluation based on sequential frames confirmed the success of the algorithm to achieve very good detection-rate.

2. It is necessary to detect all text of all frames, if the algorithm is able to detect all the words which may appear during the video clip, this is considered as a perfect indicator. Implementation of the algorithm based on video clip gives an excellent recall-rate(r) with a minimum value of 86%.

3. The algorithm is tested to measure its validity to work under real-time operation, by processing one frame and exceed a set of next frames; the test appears that the algorithm is able to work under real-time operation with excellent values of detection-rate(r).

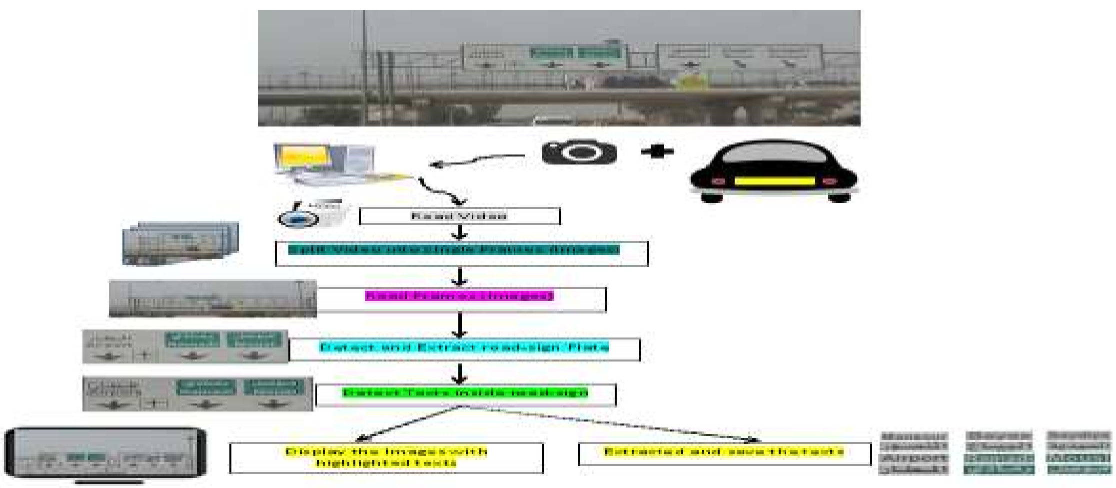
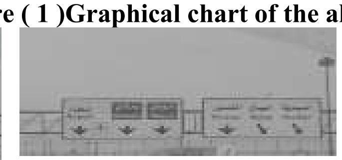


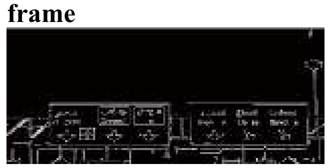
Figure (1)Graphical chart of the algorithm

Figure (2) Frame of the source video



Figure(3) Gray-scale of frame

Figure(4) Canny edge detection of the frame



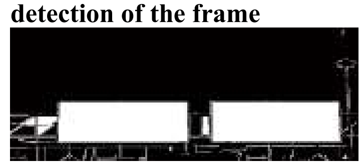


Figure (5) morphological opening of the frame | **Figure (6) Dilated image** | **Figure (7) image with filled holes**

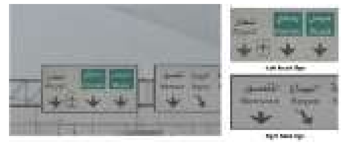
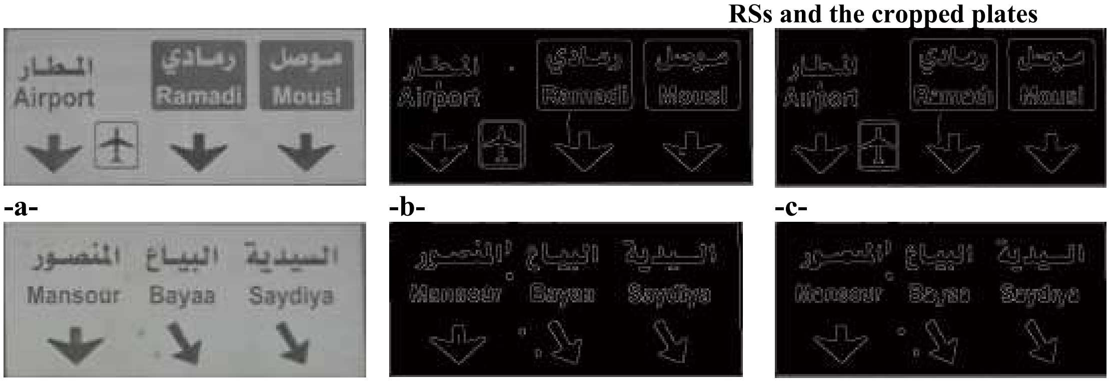


Figure (8) Incomplete RS plate | **Figure (9-a) Image has 2-RS and the cropped plates** | **Figure (9-b) Image with complete and incomplete RSs and the cropped plates**



-a- -b- -c-

-d- -e- -f-

Figure(10 a) First Road-Sign plate. b) First Road-Sign plate after applying canny edge detector. c) First Road-Sign plate after removing small objects. d) Second Road-Sign plate. e) Second Road-Sign plate after applying canny edge detector. f) Second Road-Sign plate after removing small objects

-a-

-b-

Figure(11 a) Dilated image for the first plate. b) Dilated image for the second plate

-a-

-b-

Figure (12) Applying morphological opening for a) the first road-sign plate. b) Second road-sign plate

Figure (13) Locating the text by bounding box.

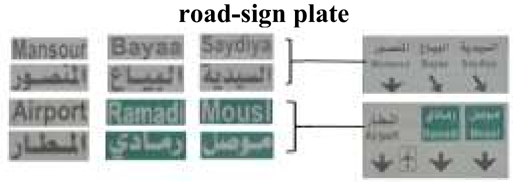


Figure (14) Extracted text of the two road-sign

Table (1) required features to detect RS

Property	Description
Area	The number of pixels in the binary object
Bounding Box	The smallest rectangle containing the object

Upper Corner	Specifies the left upper corner of the bounding box in form(x,y)
Dimensions	Specifies the dimensions of the bounding box in form (horizontal dimension, vertical dimension)
Extent	Specifies the ratio of area of the object to area of the total bounding box

Table (2) Results of cases study

Resolution	Image Format	Number of Plates	RS. Position	Total No. of Text	Ct	Mt	Ft	r %	p %	Time and weather conditions
1920x1080	JPEG (.jpg).	1	upper	8	7	1	1	88	88	Morning with shadow
1920x1080	JPEG (.jpg).	2	upper	12	12	0	2	100	86	Noon with
1920x1080	JPEG (.jpg).	2	upper	10	7	3	1	70	88	Noon with clouds
1920x1080	JPEG (.jpg).	1	side way	18	14	4	1	78	93	Night
1920x1080	JPEG (.jpg).	1	upper	8	8	0	1	100	89	Noon
1920x1080	JPEG (.jpg).	1	upper	5	5	0	0	100	100	Noon
1920x1080	JPEG (.jpg).	1	upper	10	9	1	0	90	100	Night
1920x1080	JPEG (.jpg).	1	upper	4	4	0	0	100	100	Night
1920x1080	JPEG (.jpg).	1	upper	6	6	0	0	100	100	Noon

Table (3) Evaluation Based on Sequential Fames Tests

Number of Frames	Frame with Text	Video Duration In Sec.	Frame Rate Per Sec.	Weather Description	Day Time	Total No. of text	Ct	Mt	Ft	Recall (%)	Precision (%)	Execution Time (sec.)	FS
41	41	1.7	24	Noisy	Night	918	512	406	22	56	96	61.5	71
68	50	2.8	24	Noisy	Night	400	281	119	191	70	60	91.7	65
66	66	2.2	30	Shadow	Morning	660	563	97	24	85	96	121.4	90
29	29	1.16	25	Shadow	Morning	232	207	25	85	89	71	67.3	79
37	30	1.4	25	Cloudy	Noon	390	319	71	53	82	86	65.4	84
14	14	0.4	29	sunny	noon	121	86	53	12	71	88	22.7	79

Table (4) Evaluation Based on Video Tests

Case No.	Number of Frames	Video Duration In Sec.	Descriptions of the video	Day timey	Total No. of word	ct	mt	ft	r ≈%	p≈%	ER %	Fs
1	134	4	sunny	Noon	8	8	0	5	100	62	0	76
2	306	10	Shadow	Night	12	12	0	2	100	86	0	92
3	29	1	shadow	Morning	7	6	1	5	86	55	14	67

4	236	8	RS slope	Night	28	27	1	5	96.4	84	3.6	90
5	575	19	Vogue vision	Night	18	16	2	6	89	73	11	80
6	333	11	shadow	Morning	10	10	0	2	100	83	0	91
7	43	2	Cloudy	Noon	12	12	0	0	100	100	0	100
8	15	1	Sunny	Morning	4	4	0	4	100	50	0	67
9	47	2	Cloudy	Noon	18	17	1	6	94	74	5.5	83
10	29	2	Cloudy	Noon	11	11	0	0	100	100	0	100

Table (5) Measure the Performance Based on Real-time Operation

	Process all Frames of the Video			Process a reduced Number of Frames		
Resolution	Total no. of Frames	Required Time in Sec.	Detection Rate%	No. of Processed Frames (percentage)	Required Time in Sec.	Detection Rate%
1920×1080	422	360.0	100	9(1:48)	15	93
1280×720	422	209.7	100	18(1:24)	14	89
768×575	422	114.7	96.2	33(1:13)	15	93

REFERENCES

[1] H. anoual, D. Aboutajdine, S. Elfkihi Ensias, A. Jilbab Enset, “Features Extraction for Text Detection and Localization”, IEEE, 2003.

[2] T. Lu ,S. Palaiahnakote, C.L. Tan, and W. Liu ,“Video Text Detection” ,Springer, 2014.

[3] Jung K, In Kim K, Jain A. K., “Text information extraction in images and video: a survey”, Elsevier, Volume 37, Issue 5, Pages 977–997, 2004.

[4] Y. Zhong, H.J. Zhang, and A. K. Jain,” Automatic Caption Localization in Compressed Video”, IEEE, Vol. 22, No. 4, 2000.

[5] Qixiang Ye, Wen Gao, “a New Text Detection Algorithm in Images/Video Frames “, Pacific Rim Conference on Multimedia, Tokyo, Japan, 2004.

[6] Thomas B. Moeslund, “Introduction to Video and Image Processing Building Real Systems and Applications”, Springer, 2012.

[7] Hayder Hadi Abbas, “A proposed Algorithm for Interactive Geometric Shapes Recognition”, Eng. & Tech. Journal, Vol.32, Part (A), No.5, 2014.

[8] Sneha Sharma, “Extraction of Text Regions in Natural Images”, Masters Project Report, Department of Computer Science, Rochester Institute of Technology, 2007.

[9] Faten Hassan AL_Kathy , “Digital Video Automatic Segmentation Algorithms Using Edge Detection”, Eng. & Tech. Journal, Vol.28, No.12, 2010