IRAQI Academic Scientific Journals
العراقية المجلات الاكاديمية العلمية
Al-Kut University College Journal
ISSN (E) : 2616 - 7808 II ISSN (P) : 2414 - 7419
www.kutcollegejournal.alkutcollege.edu.iq
Vol. 7 , Issue 1 , June 2022

Real - Time Video Text Detection Using Laplacian and SVM Classifier

Huda Dheyauldeen Najeeb [1]

Abstract

Text detection in the videos is still considered an interesting and challenging problem for researchers in the field of computer vision and image processing. In this paper, we have been suggested a robust method for real-time text detection in the videos through determining the candidate text lines by using Laplacian edge detection and Morphological Dilation, then using an SVM classifier to eliminate the false candidate text lines. The experimental results for 6 different types of videos show that the proposed method able to detect text lines in real-time with a 98.2% recall rate and a 0.86% false alarm rate as well as it is able to detect different text line languages (English, Chinese, and Arabic) with high accuracy of about 96.4%.

Keywords: Text detection, Laplacian Operator, Morphological Dilation, Support Vector Machine (SVM)

Affiliation of Author

[1] Al Iraqia University, College of Media, Department of Public Relations, Iraq, Baghdad, 10001

[1]111806@student.uotechnology.edu.iq

[1] **Corresponding Author**

Paper Info.

Published: June 2022

اكتشاف نص الفيديو في الوقت الفعلي باستخدام مصنف Laplacian وSVM

هدى ضياءالدين نجيب [1]

الملخص

لا يزال اكتشاف النص في مقاطع الفيديو يمثل مشكلة مثيرة وصعبة للباحثين في مجال الرؤية الحاسوبية ومعالجة الصور. ففي هذا البحث ، تم اقتراح طريقة قوية لاكتشاف النص في الوقت الفعلي في مقاطع الفيديو من خلال تحديد خطوط النص المرشحة باستخدام كشف الحواف Laplacian والتوسع Dilation ، ثم استخدام مصنف SVM للتخلص من سطور النص المرشحة الخاطئة. تظهر النتائج التجريبية لستة أنواع مختلفة من مقاطع الفيديو أن الطريقة المقترحة قادرة على اكتشاف خطوط النص في الوقت الفعلي بمعدل استدعاء 98.2٪ ومعدل إنذار كاذب بنسبة 0.86٪ بالإضافة إلى أنها قادرة على اكتشاف لغات سطر النص المختلفة (الإنجليزية والصينية والعربية) بدقة عالية بلغت حوالي 96.4٪.

الكلمات المفتاحية: كشف النص ، مشغل Laplacian، التوسع Dilation ، شعاع الدعم الآلي (SVM)

انتساب الباحثة

[1] الجامعة العراقية، كلية الاعلام، قسم العلاقات العامة، العراق، بغداد، 10001

[1]111806@student.uotechnology.edu.iq

[1] **المؤلف المراسل**

معلومات البحث

تأريخ النشر : حزيران 2022

1. Introduction

Digital videos are playing a significant role in education, amusement, and other multimedia applications. Video is an important source of information that contains tremendous information for analysis such as images, sound, and text [1].

The text in the video has played a more significant role than ever in contemporary society as an important instrument for contact and cooperation, the rich and precise high-level semantics embodied in a text could be helpful to understand the world around us. For instance, in a wide variety of real-world applications, text information can be used such as industrial automation, robot navigation, instant translation, geographic location, image search and scene analysis [2]. Therefore, text detection and recognition in computer vision have become an increasingly common and important research subject [3]. Generally, text that appears in images can be divided into two groups: graphics text and scene text. The graphic text, such as the name of a journalist during a news program, is created separately from the images and laid over it at a later date. while, scene text is a part of the image, and appears accidentally, such as in traffic signs, etc [1]. This work focuses on graphics tex.t There are several methods for text detection such as Coarse detection, wavelet histogram, MSER algorithm, Stroke Width Transform. We proposed in this paper, a robust and accurate method for text detection in videos through detecting a candidate text lines by using edge detection and mathematical morphology, then using the SVM classifier to eliminate the false candidate text lines. The present paper is organized as follows. In section 2 literature discussion about work related to text detection in the images or videos. In sections 3, 4 and 5, Laplacian edge detection, morphological dilation, and support vector machine are discussed. In section 6, the design of proposed method is described. Experimental results and the conclusion of the paper are shown in sections 7 and 8 respectively.

2. Related work

There are many researches concerned with Text detection; some of these researches are as follows:

Qixiang in 2005 [4], presented algorithm for detecting the text in both image and video which contains three phases: extracting the tines of text by using Coarse detection, wavelet histogram to find the features. Finally, using SVM to both training and classification. It achieved a high accuracy of about 93.2%. **Anila in 2016** [5], proposed technique for detecting the text in a natural image through two steps: first, by using MSER algorithm to find the similar region intensities in the image, then using Stroke Width Transform to generate and filter the candidates characters. This technique was tested with 50 images. The result was that the algorithm able to detect the text with good detect rate. **Zhe in 2016** [6], presented a text extraction method that includes two phases: extracting the text region by using Harris corner and classification; these text regions by using an SVM classifier. This method was tested with 395 images. Experimental results show that the method was able to detect the text with an accuracy of about 92.7%. **Wenhao in 2017**, designed a scheme to detect a multi-oriented scene text by using direct regression and deep convolutional neural network (CNN). This algorithm has been compared with the algorithm of [7] which is proposed to detect the texts in natural images. The result of this algorithm is better than [7] with an accuracy of about 93%. **Jianqi in 2018** [8] proposed the same previous schema where the difference is that used a Rotation Region rather than direct regression with a deep convolutional neural network (CNN) for detecting the text. Experimental results showed that the scheme was able to detect the text with an accuracy of about 95%. **Youssef in 2020** [4], presented an algorithm

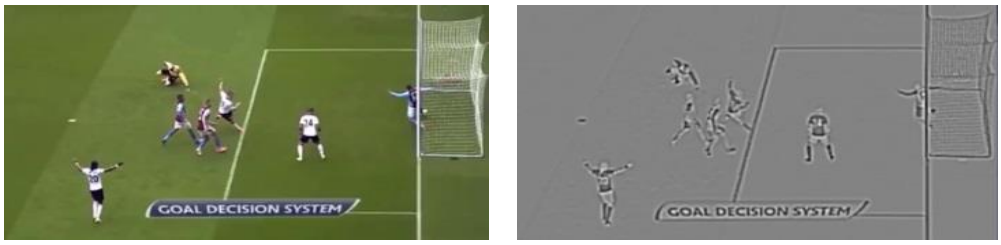
for detecting and recognition of the text in a video that was based on binarization by using the Baselines Information and Stroke Width Detection, and video optical character recognition (OCR) algorithm. The finding of the research is the binarization, which is an important step to detect the text and a hybrid binarization, which is better than a single binarization approach for dealing with various types of text in the video.

3. Laplacian Edge Detection

Edge detection is an image processing technique that recognizes the boundaries of objects within images. It operates by detecting brightness discontinuities which is shown in Fig.(1). Edge detection is used for data extraction and image segmentation in fields such as image processing, computer vision, and machine vision. Popular algorithms for edge detection include Canny, Sobel, and Laplacian methods [9].

Just one kernel is used in the Laplacian operator. It calculates second order derivatives in a single pass. Highlighting the edges is the end product of this filter. The operator typically takes as input a single gray level image and creates another binary image as output [10].

A kernel used in $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$



a. Original image b. Laplacian

Fig. (1): Edge detection

4. Morphological Dilation

Mathematical morphology is a technique and theory for processing and analysis the geometrical structures, based on set theory, topology, lattice theory , and random functions. Mathematical morphology is most widely used for digital images, but it can be employed as well on solids, surface meshes, graphs, and many other spatial structures[11]. Dilation is the most basic morphological operation uses to add pixels to objects boundaries in an image to remove the noise from the image or for image enhancing. The number of pixels added to the image objects depends on the shape of the structuring element and size used for the image processing that determines the shape of a pixel neighborhood over which the maximum is taken[12]. Dilation can be applied to several (iterations) times. Fig. (2) represents dilation that was applied 2 times. The mathematical definition of dilation operation of *A* by *B*, denoted $A \oplus B$, [13]

$$A \oplus B = \bigcup Ab,\ b \in B \qquad (1)$$

$$A \oplus B = B \oplus A \cup Ba,\ a \in A \qquad (2)$$

a. Original image b. Dilation

Fig. (2): Mathematical Morphology

5. Support Vector Machine (SVM):

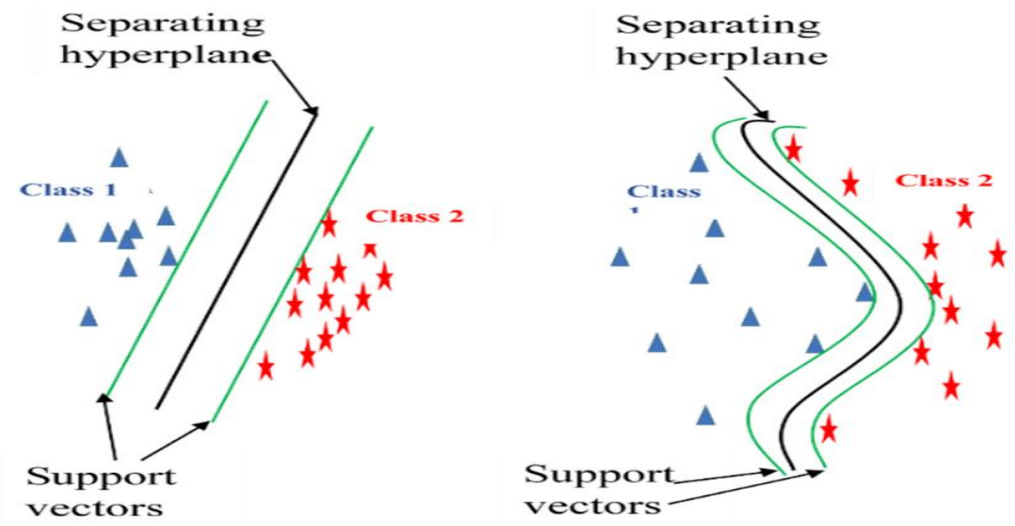
One of the most common machine learning methods is the Support Vector Machine which is mainly used for linear or nonlinear classification problems by giving a set of negative and positive training values which is shown in Fig.(3). It gives a robust solution with noise [14] and [15]. The SVM's fundamental concept is that to distinguish data from two different classes depending on the features of the data, it should find the best hyperplane such as that the distance between the two classes, i.e. maximized the margin. The basic SVM classifier is built from a simple classifier of the linear maximum margin. SVM classifier in the following manner: [16]

$$Đ = (x_1, y_1), (x_2, y_2),\ldots, (x_L, y_L) \quad (3)$$

$$Đ = (x_i, y_i), \forall i = 1, 2,\ldots, L \quad (4)$$

where x_i is input vector and $y_i \in \{+1,-1\}$ is the associated class label for every $i = 1, 2, \ldots, L$.

The positive values represent an interesting object which should be tracking while the negative values represent all the remaining things that should not be tracked [14].



(a) Linear classifier (b) Non-linear classifier

Fig. (3): SVM classifier [17]

6. Proposed Method

The block diagram of proposed method of text detection in videos is shown in Fig.(4). It includes many important steps.

Step 1: Pre-processing: The first step takes the video and splits it into a sequence of frames to make it suitable for the next step.

Step 2: Edge Detection: This step is used to identify the boundaries of objects inside frames. Usually, written text is positioned on a clear background which tends to create a high response to edge detection. Laplacian operator is used to create regions that are more probably to be a part of the text.

Step 3: Enhancing the Text image: The enhanced image is determined by aligning the various instances of a specific text region across frames and selecting the color corresponding to the minimum intensity value across frames for each pixel. Dilation used to add five pixels to regions boundaries in the frame to enhance the text image.

Step 4: Find Contour: Describes the shape of the candidate text to make it is possible for detection**.**

Step 5: Geometrical Constraints : By using rectangle geometry to crop the candidate text lines which are input to SVM Classification.

Step 6: SVM Classification: The final step takes the candidate text lines and classifies them depending on SVM classifier to delete the false detected text.

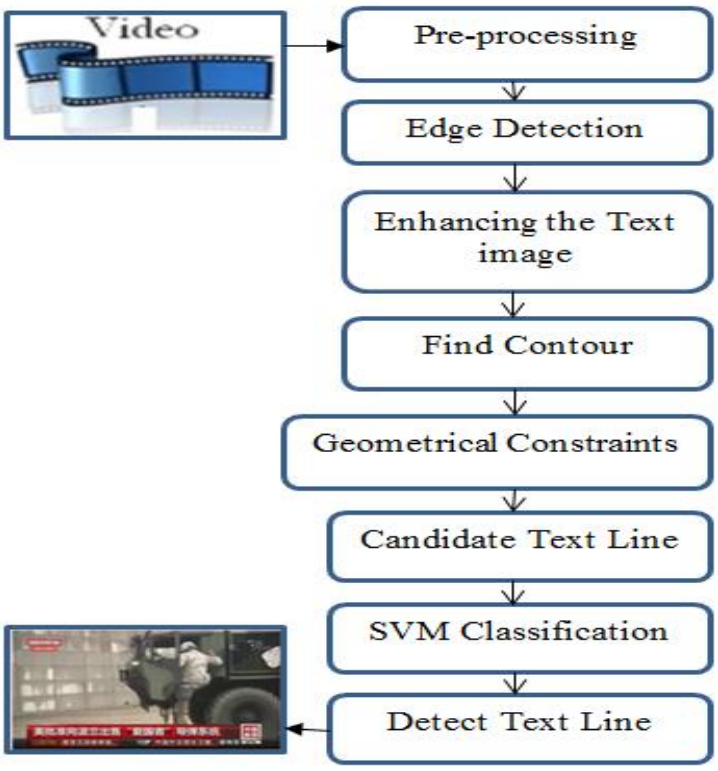


Fig. (4): Block Diagram of Proposed Method

7. Experimental Results

In Visual Studio c# 2013; this model was implemented. Six different videos from YouTube are used as a dataset, which is divided into two groups namely; 70% for training and 30% for tests. 2900 samples of different text and 6000 samples of non-text were extracted from these videos (from training group) and saved in the dataset (shown in Fig.5) which is used to train the SVM. The accuracy of the SVM classifier was calculated in the training stage according to equation (5) where the result was 96.4%.

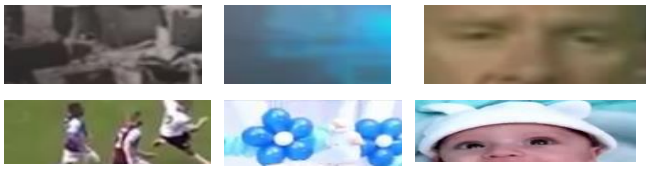
$$\text{Accuracy}= \frac{\text{Number of correctly classified samples}}{\text{Number of samples}} \qquad (5)$$

In the testing stage, the proposed method has been applied to six different videos from YouTube (on test group). These videos are two videos in Arabic, one in Chinese, and three ones in English. Recall rate and false alarm rate were calculated according to equation (6) and equation (7), respectively which are summarized in Table (1).

$$\text{Recall} = \frac{\text{Number of correctly detected texts}}{\text{Number of texts}} \times 100\% \qquad (6)$$

$$\text{False alarm rate} = \frac{\text{Number of falsely detected texts}}{\text{Number of detected texts}} \times 100\% \qquad (7)$$

The proposed method was able to detect text lines in real-time which is illustrated in Fig.(6) with a 98.2% recall rate and a 0.86% false alarm rate.



a) non-text data

b) text data

Fig. (5). Examples of the training data

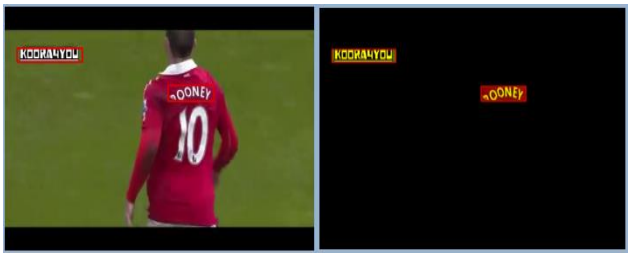
a. detected text lines in video4

b. detected text lines in video5

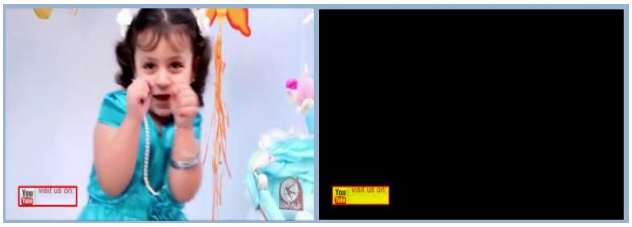
c. contains a false alarm in video4

d. detected text lines in video6

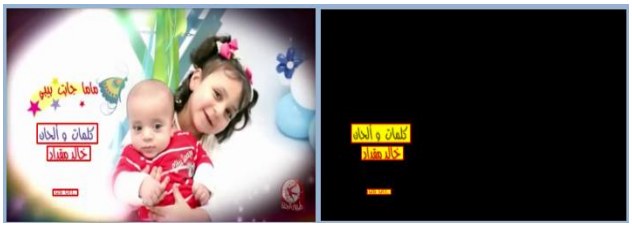
e. detected text lines in video1



f. detected text lines in video3



g. detected text lines in video2



h. missed text lines in video2

Fig. (6): Examples of Results in Proposed Method

Table 1 : Performance of Proposed Method

Videos	Total Number of frames	Number of detected text lines	Number of non detected text lines	Number of false detected text lines	Recall rate (%)	False alarm rate (%)
Video1	451	1052	0	3	99.7%	0.28%
Video2	4396	4915	44	12	98.8%	0.24%
Video3	6014	12105	206	141	97.2%	1.16%
Video4	2114	6303	57	76	97.9%	1.20%
Video5	1705	6799	42	88	98.1%	1.29%
Video6	3321	6602	65	69	97.7%	1.04%
Total	18001	37776	414	389	98.2%	0.86%

Our results was compared with the previously related works and is shown in Table 2.

Table 2 : Comparable our Result with Related Work

Year	Name of researcher	Method	Result
2005	Qixiang and others	Coarse detection, wavelet histogram, and SVM classifier.	93.2%
2016	Anila and others	MSER algorithm and Stroke Width Transform.	able to detect the text with good detect rate.
2016	Zhe and others	Harris corner and SVM classifier.	92.7%
2017	Wenhao and others	Direct regression and CNN	93%
2018	Jianqi and others	Rotation region and CNN	95%
2020	Youssef and others	binarization and video optical character recognition	a hybrid binarization better than a single binarization approach for detection the text in the video.
2020	Our proposal	Laplacian edge detection, Morphological Dilation, and SVM classifier.	96.4%

8. Conclusion

Text data provides valuable information for the study, indexing, and recovery of videos. This paper presented a robust method for detecting the text in the video in real-time. This method contains many steps: preprocessing, edge detection to identify the boundaries of texts inside frames, enhances the text image by using dilation method to add five pixels to regions boundaries in the frame, determines the candidate text by using contour method, cropping the candidate text lines by using geometrical constraints. The final step is to eliminate the false candidate text lines by using an SVM classifier.

The text lines were detected in 18001 frames by using six different video types including cartoon, sport, news, and a movie with true detection of approximately %96. The proposed method works efficiently which gets a high recall rate with an average is 98.2% and low false alarm rate with an average is 0.86%.

References

[1] P. Shivakumara, W. Huang, and C. L. Tan, "An efficient edge based technique for text detection in video frames," in *2008 The Eighth IAPR International Workshop on*

Document Analysis Systems, 2008, pp. 307–314.

[2] Y. Cao, S. Ma, and H. Pan, "FDTA: Fully Convolutional Scene Text Detection With Text Attention," *IEEE Access*, vol. 8, pp. 155441–155449, 2020.

[3] S. Long, X. He, and C. Yao, "Scene text detection and recognition: The deep learning era," *Int. J. Comput. Vis.*, pp. 1–24, 2020.

[4] Q. Ye, Q. Huang, W. Gao, and D. Zhao, "Fast and robust text detection in images and video frames," *Image Vis. Comput.*, vol. 23, no. 6, pp. 565–576, 2005.

[5] S. Anil and N. Devarajan, "Text Detection and Recognition using Stroke Width Transform," *Asian J. Inf. Technol.*, vol. 15, no. 21, pp. 4318–4324, 2016.

[6] Z. Guo, Y. Li, Y. Wang, S. Liu, T. Lei, and Y. Fan, "A method of effective text extraction for complex video scene," *Math. Probl. Eng.*, vol. 2016, 2016.

[7] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in *2012 IEEE conference on computer vision and pattern recognition*, 2012, pp. 1083–1090.

[8] J. Ma *et al.*, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Trans. Multimed.*, vol. 20, no. 11, pp. 3111–3122, 2018.

[9] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 6, pp. 679–698, 1986.

[10] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 886–890, 2007.

[11] S. Pawar and V. K. Banga, "Morphology approach in image processing," in *International Conference on Intelligent Computational Systems (ICICS'2012)(Dubai). Dubai*, 2012, pp. 148–150.

[12] S. N. Mohammed, "Emotion Detection Using Facial Image Based on Geometric Attributes," *Baghdad Coll. Sci. Univ. Baghdad*, 2016.

[13] R. S. Obied, "Tennis Game Events Recognition Using Digital Video," University of Technology, 2015.

[14] K. R. Reddy, K. H. Priya, and N. Neelima, "Object Detection and Tracking - A Survey," *Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN 2015*, no. May, pp. 418–421, 2016, doi: 10.1109/CICN.2015.317.

[15] D. C. Parvathi P1, "An Analysis of Short Text Detection and Classification Algorithms," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 8, no. 4, pp. 176–182, 2020.

[16] A. K. Nandi and H. Ahmed, *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*. John Wiley & Sons, 2020.

[17] H. Ahmed and A. K. Nandi, "Compressive sampling and feature ranking framework for bearing fault classification with vibration signals," *IEEE Access*, vol. 6, pp. 44731–44746, 2018.