A Survey On Thresholding Operators of Text Extraction In Videos

Lahouaoui LALAOUI *laalaoui58@yahoo.fr*
Faculty technology/ Departmen of electronicst/ Laboratory the LGE
University Mohamed boudiaf M'sila
Ichbilia, 28000, Algeria

Abdelhak DJAALAB *zino4525@yahoo.fr*
Faculty Faculty technology/ Department of electronics
UniversitySetif1
Maabouda ,19000, Algeria

Abstract

Video indexing is an important problem that has interested by the communities of visual information in image processing. The detection and extraction of scene and caption text from unconstrained, general purpose video is an important research problem in the context of content-based retrieval and summarization. In this paper, the technique presented is for detection text from frames video. Finding the textual contents in images is a challenging and promising research area in information technology. Consequently, text detection and recognition in multimedia had become one of the most important fields in computer vision due to its valuable uses in a variety of recent technical applications. The work in this paper consists using morphological operations for extract text appearing in the video frames. The proposed scheme well as preprocessing to differentiate among where it as the high similarity between text and background information. Experimental results show that the resultant image is the image with only text. The evaluated criteria are applied with the image result and one obtained bay different operator.

Keywords: Thresholding, Sequence Video, Segmentation, Operators.

1. INTRODUCTION

Text detection in images is an active and exciting research field on account of its valuable implementations in many technical branches. The challenge text extraction in video consists of three steps. The first one is to find text region in original images. Then the text needs to be separated from the background. And finally, a binary image has to be produced. Or it's the same we present the tree steps given here, first text detection, the sequential multi-resolution paradigm, the background-complexity-adaptive local thresholding of edge map, and the hysteresis text margin recovery.

In second text localization, the coarse-to-fine localization scheme, the Multilanguage-oriented region dividing rules, and the signature-based multi-frame verification. Finally, text extraction the color polarity classification, the adaptive thresholding of gray scale text images, and the dam-point-based inward filling. In the new globalized world, digital contents are developing rapidly over both The Internet and digital media which resulted in generating numerous electronic recourses. These resources are available in a variety of multimedia forms such as images, video and audio frames.

Accordingly, the old fashion paper-based products such as books, letters, journals, and newspapers are converting recently towards digitizing; as a result, one might find several digital images including some textual contents. Images that contain texts might be either seen images

where the text appears naturally as a part of the scene or caption text images where the text overlays on the image [1].

Working with natural (scene) images with complex backgrounds can be valuable in some computer and robotics applications. Moreover, Text detection in images such as advertisements, street or traffic signs might help in vehicle license plate recognition or text reading programs for visually impaired person. Detecting texts in images, as well, helps in various technical applications such as keyword-based image search, automatic video logging, and text-based image indexing [2]. Several algorithms and studies have been proposed and developed in text detection researchers.

The best algorithms and methods were mainly based on edge detection, projection profile analysis [3],[4] texture segmentation, In 2[], “Improved adaptive Gaussian mixture model for background subtraction,” Proc. ICPR, 2004.] proposed a method for adapting the scene to light changes, by adding new samples and discarding the old ones a reasonable period. Color quantization and histogram techniques, artificial neural network and wavelet transforms [5][6] authors proposed a text detection algorithm that uses thresholding morphological operators for edge detection and text candidates' labeling. Similarly, Authors in [7] applied some morphological operators on the image; however they also studied wavelet-based features to label text candidates.

A grayscale image segmentation algorithm was applied in the proposed technique in [8],[9], authors employed Haar wavelet transform (DWT) on the resulted image and then studied the connected components' features to find candidate text areas. In [10] aimed to detect hand written texts by using a Gaussian window for blurring and enhancing text line structures. The technique proposed in [11] employs multi-scale edge detection and an adaptive color modeling in the neighborhood of candidate text regions, it then performs a layout analysis as the final step for labeling text candidates. On the other hand, the technique in [12] detects the text in an image by using morphological operations, and then it labels the connected components by electing some criteria to filter non-text regions.

Finally, it imprints the text from the original image by removing the detected text. In [ ] text detection was based on evaluating and analyzing the stroke width generated from the skeleton of the text candidates. [13] use Stroke generation also in the chip generation step after segmenting texture; strokes are connected into chips if they compose connected components with certain conditions, where each of which will be tested and classified into a text or non-text regions. In [14], the proposed system is based on the Modest AdaBoost algorithm which constructs a strong classifier from a combination of weak classifiers. However, in [15], the algorithm applies some recently developed methods in machine learning which uses unsupervised feature learning for text detection and recognition. The technique in[16],[17] applied the wavelet transform to detect edges, then text candidates were obtained and refined by using two learned discriminative dictionaries, adaptive run-length smoothing algorithm and projection profile analysis.

The extracted text itself may include significant information about the image contents which may help in media elucidation. Caption texts in the news are examples of artificial texts that were added to the video frames to present a clearer understanding for the viewers. Text detection and extraction, however, is a tricky issue due to many reasons. The text detection process as well as the complexity in the image background [18].

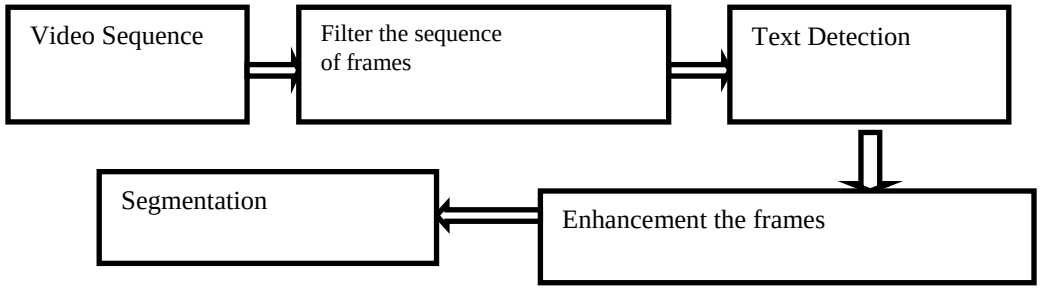
In this paper, we improved a text detection technique that applies image enhancement to improve the edge detection along with some well known morphological operations to locate more accurate edges on the scene image as a first step. In the second step, the technique uses the Hough transform for each connected component, and then the text is extracted by selecting those connected components whose number of peaks is greater than a discriminating threshold value. The paper is organized as follows: Section 2 reviews some previous research on text detection

field. Then, in section 3, the proposed technique is presented. Section 4 presents an analysis and discussion of the technique results. Finally, the conclusion of the paper is presented in section 5.

2. MATERIALS AND METHODS

An obvious step towards video segmentation is to apply image segmentation techniques to video frames without considering temporal coherence [19].

These methods are inherently scalable and may generate segmentation results in real time. However, lack of temporal information from neighboring frames may cause jitter across frames.



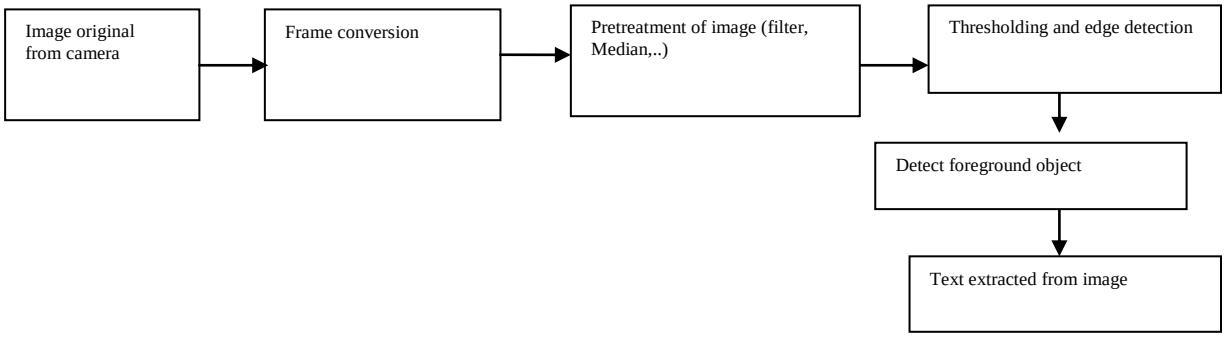
Presentation of The Difficulties Obtained

Background and text may be ambiguous; text color may change, or the text can have arbitrary and non-uniform color. Background and text sometimes reversed, text possibly will move, unknown text size, position, orientation, and layout: captions lack the structure usually associated with documents. Unconstrained background: The background can have colors similar to the text color. The background may include streaks that appear very similar to character strokes. Lossy video compression may cause colors to run together, and the low bit-rate video compression can cause loss of contrast.

In this paper, we improved technique for automatic text detection in scene images. The proposed technique divided into four phases which are: Image Enhancement. Edge Extraction, labeling the candidate text regions, text extraction. The flowchart for the proposed technique is shown in figure1. Flowchart for the proposed technique.

2.1 Pretreatment

This phase takes account of enhancing the visual appearance of scene images to improve the detectability of objects and edges that will use in the Edge Extraction step. The image enhancement phase includes the following steps: Step 1: Get (Input) the colored scene image, see figure 1. Step 2: Convert the input image into gray scale.



Step 3: Apply two filtering techniques: Winner filter and Median filter to remove the additive white noise. Step 4: Apply intensity adjustment followed by filter. Step 5: Show (Output) an enhanced gray scale image. For areas extraction there is an effective morphological edge

detection scheme is applied on the enhanced image in order to find the edges and lines. This step scheme is given in the below following : Sobel filter on the enhanced image resulted and a set of morphological operations open operation on the edge map. Apply thresholding on the resulted image to remove minor non-text regions. Apply the dilation operation with the suitable structured element to connect isolated edges of each detail region. Apply open operation on the dilated image. Apply close operation on the dilated image. Find the average image of the two images resulted. The output is given. In labeling the text candidate regions phase, a Two-dimensional image convolution is applied on the edge extracted image from phase 2; then the connected components in the resulting image are labeled.

Text extraction phase involves extracting relevant text from the image. Hough transform is a technique which can be used to isolate features of particular shapes within an image. In the proposed technique we apply Hough transform on each connected component, and then the number of peaks for each connected component is computed. Through this research, we applied many experiments on text images; most of them showed that the number of peaks in non-text regions is higher than any text region since text regions are more homogenous.

Accordingly, all connected components with number of peaks greater than a certain threshold are eliminated from the image since they are assumed to be non-text regions. The final step in the text extraction phase is the (filling) operation which results in producing the extracted image.

An outdoor image from HueSangChannel dataset in all phases using the proposed technique. (a) The outdoor image. (b) The enhanced image after the image enhancement phase (c) The image after the edge extraction phase. (d) The image after labeling the text candidate regions. (e) The extracted image.

3. QUANTITATIVE EVALUATION METRICS QUANTITATIVE

BPR and VPR satisfy important requirements (cf. [26]): i. Non-degeneracy: measures are low for degenerate segmentations. ii. No assumption about data generation: the metrics do not assume a certain number of labels and apply therefore to cases where the computed number of labels is different from the ground truth.

Inconsistency among humans to decide on the number of labels is integrated into. The metrics, which provides a sample of the acceptable variability.

Segmentation outputs addressing different coarse-to-fine granularity are not penalized, especially if the refinement is reflected in the human annotations, but granularity levels closer to human annotations score higher than the respective over- and under-segmentations; this property draws directly from the humans, who psychologically perceive the same scenes to a Different level of detail.

The metrics allow the insightful analysis of algorithms at different working regimes: over-segmentation algorithms decomposing the video into several smaller temporally consistent volumes will be found in the high precision VPR Area and correspondingly in the high recall BPR area. More object-centric segmentation methods that tend to yield few larger object volumes will be found in the VPR high recall Area, BPR high precision area. In both regimes, algorithms that trade off precision and recall in a slightly different manner Can be compared in a fair way via the F-measure.

Both metrics allow comparing the results of the same algorithm on different videos and the results of different algorithms on the same set of videos. The VPR metric additionally satisfies the requirement of vii. Temporal consistency: object labels that are not consistent over time get penalized by the metric.

All previously proposed metrics do not satisfy all these constraints. The one in [4] is restricted to motion segmentation and does not satisfy (iii) and (v). The metrics in [28] dovnot satisfy (iii). The boundary metric in [1] is designed for still image segmentation and do not satisfy (vii). The region metrics in [1] have been extended to volumes [27, 10] but do not satisfy (i) and (v).

We propose to benchmark video segmentation performance with a boundary oriented metric and with a volumetric one. The effectiveness and the robustness of the proposed text detection technique are measured quantitatively by calculating four metrics:

3.1 Boundary Precision Recall (BPR)
The boundary metric is most popular in the BSDS benchmark for image segmentation [18, 1]. It casts the boundary detection problem as one of classifying boundary from non boundary pixels and measures the quality of a segmentation boundary map in the precision-recall framework: where S is the set of machine generated segmentation boundaries and f Gig M i=1 is the M sets of human annotation boundaries. The so-called F-measure is used to evaluate aggregate performance. The intersection operator solves a bipartite graph assignment between the two boundary maps. The metric is of limited use in a video segmentation benchmark, as it evaluates every frame independently, i.e., temporal consistency of the segmentation does not play a role. Moreover, good boundaries are only half the way to a good segmentation, as it is still hard to obtain closed object regions from a boundary map.

We keep this metric from image segmentation, as it is a good measure for the localization accuracy of segmentation boundaries. The more important metric, though, is the following volumetric metric.

3.2. Volume Precision Recall (VPR)
VPR optimally assigns spatio-temporal volumes between the computer generated segmentation S and the M human annotated segmentations and measures their overlap. A preliminary formulation that, as we will see, has some problems is The volume overlap is expressed by the intersection operator \ and j:j denotes the number of pixels in the volume. A maximum precision is achieved with volumes that do not overlap with multiple ground truth volumes. This is relatively easy to achieve with an over-segmentation but hard with a small set of volumes. Conversely, recall counts how many pixels of the ground truth volume are explained by the volume with maximum overlap. Perfect recall is achieved with volumes that fully cover the human volumes. This is trivially possible with a single volume for the whole video.

Obviously, degenerate segmentations (one volume covering the whole video or every pixel being a separate volume) achieve relatively high scores with this metric. The problem can be addressed by a proper normalization, where the theoretical lower bounds (achieved by the degenerate segmentations) are subtracted from the overlap score:

For both BPR and VPR we report average precision (AP), the area under the PR curve, and optimal aggregate measures by means of the F-measures: optimal dataset scale (ODS), aggregated at a fixed scale over the dataset, and optimal segmentation scale (OSS), optimally selected for each segmentation. In the case of VPR, the F-measure coincides with the Dice coefficient between the assigned volumes.

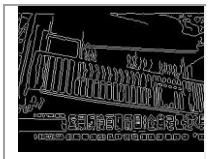
These metrics are calculated based on computing the number of corresponding matched text between the ground truth and detected text area in the image. Therefore, we need to calculate three measurements firstly which true positive are tp. True positive tp represents the number of pixels that are truly classified as text, and false positive fp represent the number of pixels that are falsely classified as text while it is a background. False Negative FN represents the number of pixels that are falsely classified as background while it is a text. Depending on these measurements, Precision, Recall, and

F-Score is calculated as follows:

$$F = \frac{2*Recall*Precesion}{Recall+Precesion} \quad (1)$$

4. RESULTS AND ANALYSIS

An evaluation of our proposed text detection algorithm is presented in this section. We presented the results of different operators as follows: Sobel, Prewitt, Robert, and Canny, etc. **Table.4.1:** Results of the different operators for to frame in the sequence video.

original color image	gray-level image	Robert Detection	Sobel Detection	Prewitt Detection	Canny Detection
					
					

At present the values of evaluation criteria, we have set in each case an operator from the operators used in our work. The results obtained for the various operators are present in the following figures.

That we set in each case an operator from the operators used in our work. The results for individual operators are shown in the following figures. We used MATLAB tool to process the image and apply different text in the image sensing operators, processing results shown in Figure 4.1 is the operator of the image histogram with Sobel shown in Figure IV.2.Starting here with a result of representation of the Sobel operator and histogram applied to an original image.

a b c

FIGURE 4.1: (a) Image original, (b) the image with Sobel detected, Histogram of the image image (b).

In this section of text we presented the original image with their resultant obtained by the Prewitt operator and histogram.

a b c

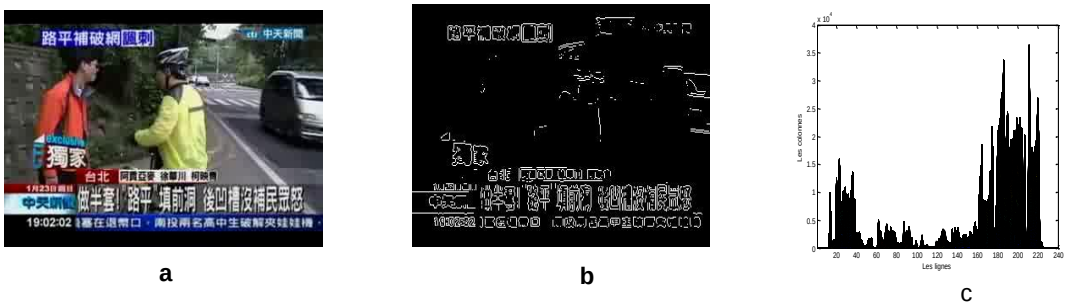
FIGURE 4.2: (a) Image original,(b) Prewitt detection Image, and (c)Histogram image.

In this section, we presented the original image with their resultant obtained by the canny operator and histogram.

a b c

FIGURE 4.3: (a) Image original, (b) Canny detector image, image Histogram (c).

In this section we presented the original image with their resultant obtained by the Robert operator and histogram.



a b c

FIGURE 4.4 : (a) Image original, (b) Image détection de Robert, image Histogram (c).

Is given here the original image with their resultant obtained by the Gray level operator and histogram.

a **b** **c**

FIGURE 4.5: (a) Image original, **(b)** Gray level image. **,** image Histogram(c).

Criteria Results

In this section we used, in each case the resulting image for each operator is considered a reference image. Now we displayed the values of the evaluation criteria, which we fixed in each case an operator from the operators used in our work. The results obtained for the various operators are present in the following figures.

4.2.1 We Fixed Sobel

In this first case we consider that the Sobel operator as a reference image and calculate the different criteria as shown in the table in Figure 4.6. (a)

Criteria operator	PSNR	VIFP_MSCALE	MI2	YASNOFF
PREWITT	23.0710	0.7648	1.2166	0.1924
CANNY	10.7251	0.1791	0.9364	0.4890
ROBERT	12.4206	0.2625	0.9214	0.3040
Level of GRAY	8.0414	0.1191	0.9214	0.9798

FIGURE 4.6.a: Values of different criteria for each case to image segmentation with Sobel image reference.

In figure 4.6.b present a cure for canny operator image reference.

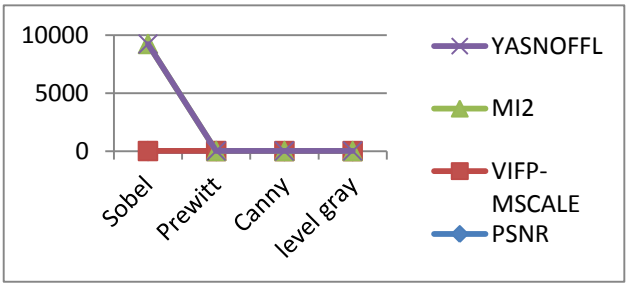


FIGURE 4. 6. B : Curve for operator for Sobel reference.

4.2.2 We have Fixed The Image Prewitt

In the second case we consider that the operator of Prewitt as an image reference (ground through) and calculate the different criteria as shown in the table in Figure 4.7.a.

Criteria operator	PSNR	VIFP_MSCALE	MI2	YASNOFF
SOBEL	12.4206	0.2625	0.9214	0.3040
PREWITT	12.4172	0.2622	0.9211	0.3040
CANNY	9.8466	0.0747	0.9190	0.4889
Level of gray	8.1227	0.11 41	0.9743	0.9814

FIGURE 4.7.a: Values of different criteria for each case to image segmentation, with Robert is an image reference.

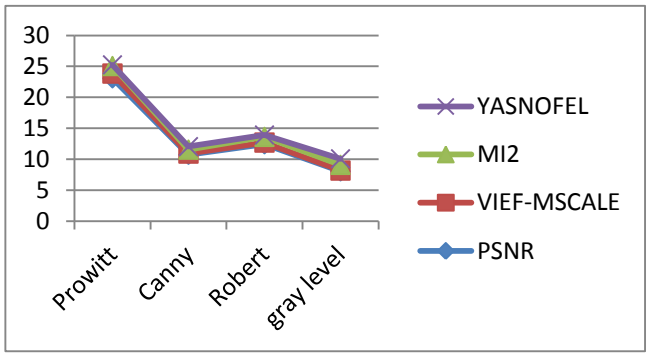


FIGURE 4.7.b: curve for Robert operator reference.

4.2.3 We Fixed The Canny Picture

In this first case we consider that the operator of Canny as a reference image and calculate the different criteria as shown in the table in Figure 4.8.a.

Criteria operator	PSNR	VIFP_MSCALE	MI2	YASNOFF
SOBEL	10.7251	0.1791	0.9364	0.4890
Prewitt	10.7473	0.1784	0.9365	0.4889
ROBERT	9.8466	0.0747	0.9190	0.4889
Gray level	7.7612	0.0943	0.9838	0.9781

FIGURE 4.8.a: Values of different criteria for each case to image segmentation, with canny is an image reference.

In figure 4.8.b present a cure for canny operator image reference.

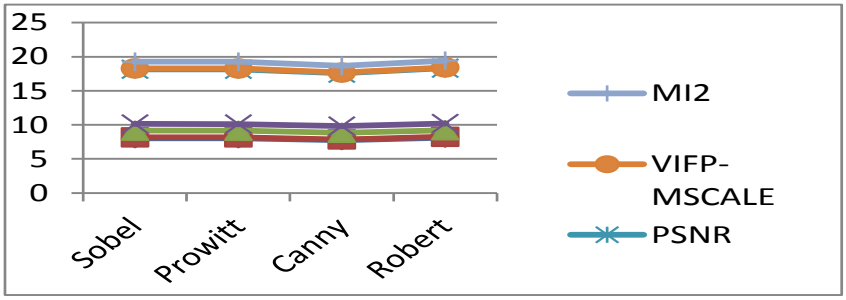


FIGURE 4.8.b: Curve for canny operator image reference.

4.2.4 We set the picture Robert

In this first case we consider that the operator of Robert as a reference image and calculate the different criteria as shown in the table in Figure 4.9.a.

Criteria operator	PSNR	VIFP_MSCALE	MI2	YASNOFF
SOBEL	23.0710	0.7648	1.2166	0.1924
CANNY	10.7473	0.1784	0.9365	0.4889
ROBERT	12.4172	0.2622	0.9211	0.3040
NIVEAU DE GRAY	8.0385	0.1169	0.9793	0.9817

FIGURE 4.9.a: Values of the criteria for each case to segment the image gray level is image reference.

In figure 4.9.b present a cure for gray level operator image reference.

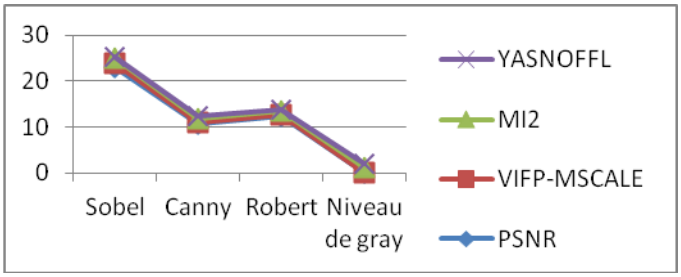


FIGURE 4.9.b: Curve for gray level operator reference.

4.2.4 We Fixed The Image of Gray Level

In this first case we consider that the gray level image as a reference image and calculate the different criteria as shown in the table in Figure 4.10.a.

Criteria operator	PSNR	VIFP_MSCALE	MI2	YASNOFF
SOBEL	8.0414	0.1191	0.9797	0.9798
PREWITT	8.0385	0.1169	0.9793	0.9817
CANNY	7.7612	0.0943	0.9838	0.9781
ROBERT	8.1227	0.11 41	0.9743	0.9814

FIGURE 4.10.a : Values of different criteria for each case to image segmentation, with prewitt is an image reference.

In figure 4.10.b present a cure for canny operator image reference.

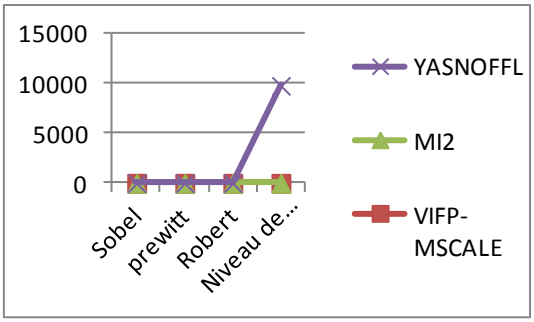


FIGURE 4.10.b: curve for canny operator image reference.

5. DISCUSSION

This paper presents efficient, survey algorithms to detect the text in video frames. These algorithms are capable of detecting scrolling text and skipping still text through the adoption of spatial and temporal processing. This article aims is to locate all text boxes in artificial images extracted from videos. We presented the various technical segmentations by the various operators.

The results we have considered each time a result for each operator as an image ground truth. Evaluate the results obtained are a few criteria of validity. We presented a text extraction method that combines the use of morphological operators and spatial coherence criteria adaptable to all regions in various orientations.

The approach presented in this article is a contribution to evaluate with tools the result of video frames.

Indeed, detection, segmentation and text recognition are essential tasks in many applications such as segmentation of newscasts. Based on the used material and experiments, we will discuss the performed measurements and the results analysis.

The experiment performed on multiple images and videos and the positive conclusions from the comparison with results from different methods. We presented in this section different algorithms used to segment a video picture and the validation criteria.

The results show that the criterion of Yasnoff is best in all cases we apply these criteria. The proposed technique proved to be robust in detecting the text accurately from these images

original color image gray-level image, Robert Detection, Sobel Detection Prewitt Detection Canny Detection.

6. CONCLUSION

This paper presents an comparison detection the Text in video frames.. Pre-filtering is used to enhance text information, and adaptive temporal differential computation is used to identify scrolling text. Post-processing is used to remove noise-related pixels and expand the range of scrolling text. The efficacy of the algorithm was verified using an extensive database of videos obtained from actual TV programs. The results demonstrate the accuracy of gray level is the best in the differentiation of scrolling text, without false detections or missed detections. In the future, the proposed scrolling text detection algorithm could also be used to hide scrolling text without jeopardizing background video information. The pixels of the scrolling text can also be hidden using interpolation techniques. Finally, the scrolling text could be used to produce an index for the classification and archiving of videos.

7. REFERENCES

[1] C.P. Sumathi, CT. Santhanam and G.Gayathri International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012.

[2] Min Cai, Jiqiang Song, Michael R. Lyu(),"A New Approach For Video Text Detection",Proceedings International Conference On Image Processing, Volume 1, pp: I-117-I-120, 2002.

[3] C. Gopalan,"Text Region Segmentation From Heterogeneous Images", International Journal of Computer Science And Network Security, Vol.8 No.10, pp.108-113, 2008.

[4] Uday Modha, Preeti Dave, "Image Inpainting-Automatic Detection and Removal of Text From Images", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622 Vol. 2, Issue 2, 2012.

[5] Aria Pezeshk and Richard L. Tutwiler, "Automatic Feature Extraction and Text Recognition from Scanned Topographic Maps", IEEE Transactions on geosciences and remote sensing, VOL. 49, NO. 12, 2011.

[6] Xiaoqing Liu and Jagath Samarabandu, "Multiscale Edge-Based Text Extraction From Complex Images", IEEE Trans., 1424403677, 2006.

[7] Yassin M. Y. Hasan and Lina J. Karam, "Morphological Text Extraction from Images", IEEE Transactions On Image Processing, vol. 9, No11, 2000.

[8] Audithan,,R.M.Chandrasekaran, "Document Text Extraction From Document Images Using Haar Discrete Wavelet Transform", European Journal Of Scientific Research , Vol.36 No.4 , pp.502-512, 2009.

[9] S.-W. Lee, D.-J. Lee, and H.-S. Park, A New Methodology for Gray-scale Character Segmentation and Recognition, IEEE Transactions on Pattern Recognition and Machine Intelligence, 18 (10),1045-1050,1996.

[10] Y. Li, Y. Zheng, D. Doermann, S. Jaeger, "A New Algorithm for Detecting Text Line in Handwritten Documents", 10th International Workshop on Frontiers in Handwriting Recognition, La Baule (France), pp. 35-40, 2006.

[11] X. Liu And J.Samarabandu, "Multiscale Edge-Based Text Extraction From Complex Images," Proc. International Conference Of Multimedia And Expo,pp.1721-1724, 2006.

[12] Yassin M. Y. Hasan and Lina J. Karam, Morphological Text Extraction from Images, IEEE Transactions on Image Processing, 9 (11) (2000) 1978-1983.

[13] K. Jung, Kim, K.I., Jain, A.K.: “Text information extraction in images and video: a survey”. Patt. Recognit.37(5),977–997,2004.

[14] R. Lienhart., Kuranov, A., Pisarevsky, A.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. The 25th Pattern Recognition Symposium (2003) 297–304.

[15] K. Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Sparse Distance Learning for ObjectRecognition Combining RGB and Depth Information. In Proc. of IEEE International Confer- ence on Robotics and Automation (ICRA), 2010.

[16] S. Wenchang, S. Jianshe, Z. Lin, “Wavelet Multi-scale Edge Detection Using Adaptive threshold,” IEEE, 2009.

[17] Davod Zaravi, Habib Rostami, Alireza Malahzaheh, S.S Mortazavi,” Journals Subheadlines Text Extraction Using Wavelet Thresholding And New Projection Profile”, World Academy Of Science, Engineering And Technology .Issue 73, 2011.

[18] K. Jung, K. I. Kim, and J. Han, Text Extraction in Real Scene Images on Planar Planes, Proc. of International Conference on Pattern Recognition, Vol. 3, pp. 469-472, 2002.

[19] H.Winnemoller, S. C. Olsen, and B.Gooch. Real-time video abstraction. ACM SIGGRAPH, 25(3):1221–1226, 2006.