Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 106 (2017) 46683-46688

Awakening to Reality

Elixir

ISSN: 2229-712X

A Comparative Approach for Overlay Text Detection and Extraction from Complex Video Scene

Anupama S. Budhewar and Ravindra C.Thool
Department of Computer Science and Engineering

ARTICLE INFO

Article history:
Received: 5 April 2017;
Received in revised form:
7 May 2017;
Accepted: 17 May 2017;

Keywords
Optical Character Recognition (OCR)

ABSTRACT

Overlay text brings important semantic clues in video content analysis such as video information retrieval and summarization, since the content of the scene or the editor's intention can be well represented by using inserted text. Most of the previous approaches to extracting overlay text from videos are based on low-level features, such as edge, color, and texture information. However, existing methods experience difficulties in handling texts with various contrasts or inserted in a complex background. In this paper, we propose a novel framework to detect and extract the overlay text from the video scene. Based on our observation that there exist transient colors between inserted text and its adjacent background, a transition map is first generated. Then candidate regions are extracted by a reshaping method and the overlay text regions are determined based on the occurrence of overlay text in each candidate. In this paper represents comparative approach between edge and change in intensity models that are computationally fast and invariant to basic transformations like horizontal, Vertical and scaling. We demonstrate that change in intensity context can be used to detect overlay text regions. Intensity context describes all boundary points of the shape with respect to any single boundary point.

1. Introduction

Overlay text gives important semantic clues in video information retrieval and summarization. With the development of video editing technology, overlay text is inserted into video contents to provide viewer better understanding. If we extract text information from the subtitles, it will be very helpful to establish a database of video's content that includes annotations and indexes. Most of the broadcasting videos tend to increase the use of overlay text to convey direct summery of semantics and deliver better viewing experience [1]. It is a super imposed text.

A vast variety of techniques are proposed in literature for video analysis ranging from extraction of low-level features to high-level semantic features and all these techniques are based on color, texture, shape, sound, text, and objects. Of all the available techniques of the video annotation, only text analysis is useful for the high-level semantic directly, while other techniques require an extra effort to produce high-level semantics [6].

Text displayed in the videos can be classified into scene text and overlay text [5]. Scene text occurs naturally in the background as a part of the scene, such as the advertising boards, banners, and so on. In contrast to that, overlay text is superimposed on the video scene and used to help viewers' understanding. Since the overlay text is highly compact and structured, it can be used for video indexing and retrieval [6]. Overlay text brings important semantic clues in video content analysis such as video information retrieval and summarization, since the content of the scene or the editor's intention can be well represented by using inserted text [8]. Most broadcasting news videos tend to increase the use of overlay text that usually represent names of anchors persons, places, persons, or description of news in crisp. Moreover in sports news it may be name of player, type of sport, location, score and many more.

With the development of video editing technology, there are growing uses of overlay text inserted into video contents to provide viewers with better visual understanding. Most broadcasting videos tend to increase the use of overlay text to convey more direct summary of semantics and deliver better viewing experience. For example, headlines summarize the reports in news videos and subtitles in the documentary drama help viewers understand the content. Sports videos also contain text describing the scores and team or player names [1]. In general, text displayed in the videos can be classified into scene text and overlay text [2]. Scene text occurs naturally in the background as a part of the scene, such as the advertising boards, banners, and so on. In contrast to that, overlay text is superimposed on the video scene and used to help viewers' understanding. Since the overlay text is highly compact and structured, it can be used for video indexing and retrieval [3]. However, overlay text extraction for video optical character recognition (OCR) becomes more challenging, compared to the text extraction for OCR tasks of document images, due to the numerous difficulties resulting from complex background, unknown text color, size and so on.

There are two steps involved before the overlay text recognition carried out, i.e., detection and extraction of overlay text. First, overlay text regions are roughly distinguished from background. The detected overlay text regions are refined to determine the accurate boundaries of overlay text strings. To generate a binary text image for video OCR, background pixels are removed from the overlay text

Tele:
E-mail address: budhewar.anupama@gmail.com

strings in the extraction step. Although many methods have been proposed to detect and extract the video text, a few methods effectively deal effectively with different color, shape, and multilingual text.

Digital video libraries and archives of immense size are becoming accessible over data networks. Efficient video retrieval and browsing has become crucially important. Understanding the semantic contents of the video and using them for indexing is inevitable. Automatic indexing and retrieval of video information based on content is very challenging research area [1].

A vast variety of techniques are proposed in literature for video analysis ranging from extraction of low-level features to high-level semantic features and all these techniques are based on color, texture, shape, sound, text, and objects. Of all the available techniques of the video annotation, only text analysis is useful for the high-level semantic directly, while other techniques require an extra effort to produce high-level semantics [6].

Overlay text will be very helpful to establish a database of video's content that includes annotations and indexes.

Applications of Overlay Text Detection and extraction are as follows:

1. There are growing uses of overlay text inserted into video contents to provide viewers with better visual understanding.
2. Headlines summarize the reports in news videos and subtitles in the documentary drama help viewers to understand the content.
3. Sports video contains text describing the scores and team member or player's names.
4. Text extraction from images with complex backgrounds remains a challenging problem in character recognition applications, such as document search on the World Wide Web, image indexing and retrieval, video indexing and retrieval, etc. The goal of text extraction is to convert the grayscale or color text images, which have been detected and localized in images and videos, into the OCR-ready binary images.
5. Information retrieval and data mining are very important and widespread techniques in computer applications. Among these techniques, most of them are focused on text data because it is convenient for plentiful, colorful, and real than the text. News videos bring us every day's stories and more meaningful information than other types of videos. News video retrieval is an essential issue in information retrieval.
6. To provide news video indexing, text localization and detection for news videos is necessary for building a complete information retrieval system. In indexing with the help of overlay text we can maintain an index and from that particular index entry we can directly move to that entry details. In this way we can reduce the time of searching.
7. If we can extract text information from the subtitles, it will be very helpful to establish a database of video's content that includes annotations and indexes.

2. Literature Review

Now see methods to detect and extract overlay text *using low*-level features such as edge, color and texture information and experience difficulties in handling text with various contrasts. The commonly adopted method is to apply an edge detector to the video frame and then identify regions with high edge density and strength. Some employ the high-frequency wavelet coefficients and connected components to detect the text regions.

Most of the existing methods are based on approaches such as edge, color, and stroke. Few of them are as follow:

M.R. Lyu, J. Song, and M. Cai implemented a method, there are two main approaches the first is language dependent and other language independent[2]. In that character recognition is based on language or character dependent English language is giving more accurate results. It is elaborated more as follows

According to the linguistic classification, English, French, and Spanish belong to alphabetic literal, whereas Chinese, Japanese, and Korean belong to ideograph. Their differences in the following four aspects affect the video text processing.

Zhong Y. et. al .[4] extracted text as those connected components that follow certain size constraints. In [7] stroke model based character extraction from gray level document images are proposed. The method extracts the text with known stroke width but fails to distinguish noise in linear form from true strokes. Liu Y. et.al.[1] proposed a hybrid approach to detect the text regions and expectation maximization algorithm to binarize the text regions. Kasar et.al. [3] Proposed an edge based connected component approach to detect the characters in camera based document images and used a specialized binarization to separate the characters from the background. The survey indicates that all the above methods fail to isolate the foreground text in document images having complex background.

N. Otsu implemented another approach [11]in 1979. It is color based text extraction method. It is simple and efficient method. It is not robust for text extraction with similar color background. Otsu method [14] is a widely used color-based text extraction method due to its simplicity and efficiency of the algorithm. However, Otsu method is not robust to text extraction with similar color of background due to the use of global thresholding. To solve this problem, the detected text regions are divided into several blocks and then Otsu method is applied locally to each block, such as the adaptive thresholding introduced in [7], where a dam point is defined to extract text strings from background.

T. Sato implemented another method in Jan.1998 [4] .It considers stroke of text in horizontal, vertical, up-right, and up-left directions and generate the edge map along each direction. Apply an interpolation filter based on vertical, horizontal, left diagonal, right diagonal directions to enhance the performance of text extraction.

The concept of text composed of uniform color was introduced by L. Agnihotri and N. Dimitrova [5] in Jun.1999. Color-based approaches assume that the video text is composed of a uniform color. In the approach by Agnihotri *et al.* [4], the red color component is used to obtain high contrast edges between text and background. In [5], the "uniform color" blocks within the high contrast video frames are selected to correctly extract text regions.

K. C. K. Kim *et al* introduced the concept of cluster color based on Euclidean distance in the RGB space and used 64 cluster color channels for text detection [9]. Color-based approaches assume that the video text is composed of a uniform color. In the approach by Agnihotri [5], the red color component is used to obtain high contrast edges between text and background. In his approach, the "uniform color" blocks within the high contrast video frames are selected to correctly extract text regions.

C. Liu, C. Wang, and R. Dai presented an approach based on edge [13].

Edge based approach is very useful for overlay text detection since text region contains rich edge information. This method performs well if there is no complex background. Use a modified edge map with strength for text region detection and localize the detected text regions using coarse-to-fine projection. They also extract text strings based on local thresholding and inward filling

X. Liu and J. Samara bandu presented[14] based on the silent point detection and the wavelet transformation has also been used to detect the text regions. Use multiscale edge detector to detect the text regions. They compute the edge strength, density, and orientation variance to form the multiscale edge detector. Texture-based approaches, such as the salient point detection and the wavelet transform [8]; have also been used to detect the text regions. Bertini *et al.* [10] detect corner points from the video scene and then detect the text region using similarity of corner points between frames. Sato *et al.* [11] apply an interpolation filter based on vertical, horizontal, left diagonal, right diagonal directions to enhance the performance of text extraction. Gllavata *et al.* [2] employ the high-frequency wavelet coefficients and connected components to detect the text regions. However, since it is almost impossible to detect text in a real video by using only one characteristic of text, some methods take advantage of combined features to detect video text [12], [13].

In this paper we are using the color models for finding out the intensity of each pixel. With the help this we can remove the minimum intensity pixels and highlight only high intensity pixels. To find out the intensity following color models are used.

3. Overlay text detection using change in intensity

In the last section represents the different methods for overlay text detection and extraction based on low level features. Overlay text brings important semantic clues in video content analysis such as video information retrieval and summarization. The implemented method is robust to different character size, position, contrast and color. Text in images and video frame carries important information for video content understanding and video retrieval.

In general, text displayed in the videos can be classified into scene text and overlay text. Scene text occurs naturally in the background as a part of the scene, such as the advertising, boards, banners, and so on. In contrast to that, overlay text is superimposed on the video scene and used to help viewers' understanding.

Since the overlay text is highly compact and structured, it can be used for video indexing and retrieval. However, overlay text extraction for video optical character recognition (OCR) becomes more challenging, compared to the text extraction for OCR tasks of document images, due to the numerous difficulties resulting from complex background, unknown text color, size and so on.

There are two steps involved before the overlay text recognition is carried out, i.e., detection and extraction of overlay text. First, overlay text regions are roughly distinguished from background. The detected overlay text regions are refined to determine the accurate boundaries of overlay text strings. To generate a binary text image for video OCR, background pixels are removed from the overlay text strings in the extraction step. Although many methods have been implemented to detect and extract the video text, few methods can effectively deal with different color, shape, and multilingual text.

To solve this problem, the detected text regions are divided into several blocks and then Otsu [14] method is applied locally to each block, such as the adaptive thresholding introduced in, where a dam point is defined to extract text strings from background. On the other hand, some filters based on the direction of strokes have also been used to extract text in the stroke-based methods.

In this approach a method for new overlay text detection and extraction method using the transition region between the overlay text and background is represented. First, it generates the transition map based on our observation that there exist transient colors between overlay text and its adjacent background. Then the overlay text regions are roughly detected by computing the density of transition pixels and the consistency of texture around the transition pixels. The detected overlay text regions are localized accurately using the projection of transition map with an improved color-based thresholding method to extract text strings correctly. It generates the transition map and refines the detected text regions in this chapter. The overlay text extraction from the refined text regions is explained in the next chapter.

The overall procedure for overlay text detection & extraction is based on overlay text detection model as shown in Fig. 1 [1], where each module is applied on n no. of frames. The different steps are as follows

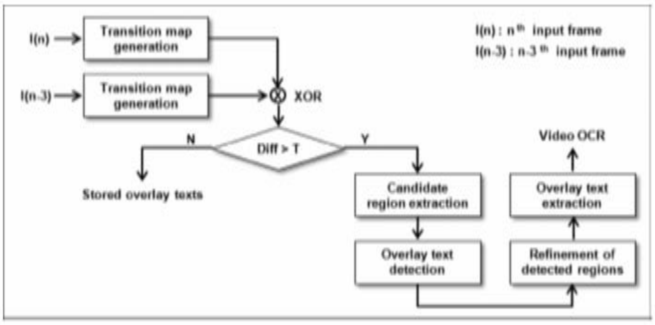
I(n)
Transition map generation
I(n-3)
Transition map generation
XOR
I(n) : n th input frame
I(n-3) : n-3 th input frame
Diff > T
N
Y
Stored overlay texts
Video OCR
Candidate region extraction
Overlay text detection
Refinement of detected regions
Overlay text extraction

Fig 1. Overall procedure of the text detection method.

3.1 Transition Map Generation

To effectively determine whether a pixel is within a transition region, the modified saturation is first introduced as a weight value based on the fact that overlay text is in the form of overlay graphics. The modified saturation is defined as follows:

$$S(x, y) = 1 - \frac{3}{(R+G+B)[\min(R, G, B)]} \qquad \text{---------} (3.1)$$

$$\check{S}(x, y) = \frac{S(x, y)}{\max(S(x, y))}$$

Where

$$\max(S(x, y)) = \begin{cases} 2 * \left(0.5 - \check{I}(x, y)\right), & \text{if } \check{I}(x, y) > 0.5 \\ 2 * \check{I}(x, y), & \text{otherwise} \end{cases} \qquad \text{---}(3.2)$$

S(x, y) & max S ((x, y)) denote the saturation value and the maximum saturation value at the corresponding intensity level respectively. $\check{I}(x, y)$ denotes the intensity at the (x, y), which is normalized to [0, 1].

Based on the conical HSI color model [15], the maximum value of saturation is normalized in accordance with $\check{I}(x, y)$ compared to 0.5(equ^{n}3.2). The transition can thus be defined by combination of the change of intensity and the modified saturation is as follows:

DL(x, y) = (1+dSL(x, y))*|I(x-1, y)-I(x, y)|

DH(x, y) = (1+dSH(x, y)*|I(x, y)-I(x+1, y)

Where dSL(x, y) =|S□(x-1, y)-S□(x, y)| and dSH(x,y)=|S(x,y)-S(x+1,y) ..………….. (3.3)

Since the weight dSL(x, y) and dSH(x, y) can be zero by the achromatic overlay text and background, add 1 to the weight. If a pixel satisfies the logarithmical change constraint given, three consecutive pixels centered by the current pixel are detected as the transition pixels and the transition map is generated

$$T(x,y) = \begin{cases} 1, \text{if } DH > DL + TH \\ 0, \quad \text{otherwise} \end{cases} \quad \text{------------------- (3.4)}$$

The thresholding value *TH* is empirically set to 80 in consideration of the logarithmical change.

3.2 Candidate Region Extraction

If a gap of consecutive pixels between two nonzero points in the same row is shorter than 5% of the image width, they are filled with 1s. If the connected components are smaller than the threshold value, they are removed. The threshold value is empirically selected by observing the minimum size of overlay text region.

Then each connected component is reshaped to have smooth boundaries. Since it is reasonable to assume that the overlay text regions are generally in rectangular shapes, a rectangular bounding box is generated by linking four points, which correspond to (min_x, min_y),(max_x, min_y) , (min_x, max_y),(max_x ,max_y) taken from the link map.

3.2 Overlay Text Region Determination

The next step is to determine the real overlay text region among the boundary smoothed candidate regions by some useful clues, such as the aspect ratio of overlay text region. Since most of overlay texts are placed horizontally in the video, the vertically longer candidates can be easily eliminated. The density of transition pixels is a good criterion as well. LBP is a very efficient and simple tool to represent the consistency of texture using only the intensity pattern[16].

$$LBP(P,R) = \sum_{i=0}^{P-1} s(gi - gc)2^{i}, \quad \text{-------------------- (3.5)}$$

Where

$$s(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases} \quad \text{-------------------- (3.6)}$$

$LBP_{6,4}=29(=2^4+2^3+2^2+2^0)$

Now we define the probability of overlay text (POT) using the operator as follows: The LBP operator is first applied to every transition pixel in each candidate region. Then, we compute the number of different LBPs to consider the intensity variation around the transition pixel. Since we use the 8 neighbor pixels to obtain the LBP value, the total number of potentially different LBPs is 2^8=256. Although the number of different LBPs is generally increasing as the candidate region includes more transition pixels, it may not be guaranteed since transition pixels can have same local binary pattern. The number of different LBPs and transition pixels in each candidate region is as shown in following table [17][25].

Table 1.Number of different LBP’s and overlay text pixels in each candidate region .

	Candidate 1 C1	Candidate 2 C2	Candidate 3 C3
# of different LBP’s	74	55	78
# of transition pixels	404	125	381

Let ωi denote the density of transition pixels in each candidate region and can be easily obtained from dividing the number of transition pixels by the size of each candidate region. POT is defined as follows:

$$POTi = \omega i * NOLi, \quad i = 1,2 \dots . N \quad \text{--------------------- (3.7)}$$

Where N denotes the number of candidate regions as mentioned. NOL_i denotes the number of different LBPs, which is normalized by the maximum of the number of different LBPs (i.e., 256) in each candidate region.

If POT of the candidate region is larger than a predefined value, the corresponding region is finally determined as the overlay text region. The thresholding value in POT is empirically set to 0.05 based on various experimental results [18].

3.3 Overlay Text Region Update

$$d\,(T_n, T_{n-3}) = \sum_{(x,y)\in T} Tn(x,y) \otimes Tn - 3\,(x,y) \quad \text{-------- (3.7)}$$

if (d (T_n, Tn_{-3}) <th) TRn= TR_{n-3}

Otherwise, find new TR_n -------------------- (3.8)

Where T_n & T_{n-3} denote the transition map obtained from the n^{th} frame & the (n-3) th frame respectively. TR_n & TR_{n-3} denotes the detected overlay text regions in the n^{th} frame & the (n-3) th frame respectively. In other words, if the values on the nth frame and the (n-3) th frame transition map are same, the result of XOR between two values is set to be 0. Otherwise, the result of XOR between two values is set to be 1. The overlay text region update method can reduce the processing time efficiently [19].

3.3.1Color Polarity Computation

Overall procedure of extraction method discussed. And two opposite scenarios, in which either the overlay text is darker than the surrounding background[20].

3.3.2Adaptive Thresholding

Since it is confirmed that the overlay text is always bright in each text region, it is safe to employ Lyu’s method to extract characters from each overlay text region. First, each overlay text region is expanded wider by two pixels to utilize the continuity of background. This expanded outer region is denoted as ER. Then, the pixels inside the text region are compared to the pixels in ER so that pixels connected to the expanded region can be excluded. We denote the text region as TR and the expanded text region as ETR, i.e. Next, sliding-window based adaptive thresholding is performed in the horizontal and the vertical directions with different window sizes, respectively[21].

3.3.3 Modified dam point labeling

Compared to the Lyu’s method, the height of expanded text region is not normalized in our method. Let and denote gray scale pixels on ETR and the resulting binary image, respectively. All are initialized as “White”. The window with the size of is moving horizontally with the stepping size 8 and then the window with the size of is moving vertically with the stepping size. If the intensity of is smaller than the local thresholding value computed by Otsu method in each window, the corresponding is set to be “Black”. The process of applying the sliding windows for adaptive thresholding[22].

3.3.4 Inward Filling

Authors assume that the background pixels in TR are generally connected to ER in terms of intensity. They use filling from ER to the connected pixels in TR to remove the background pixels. However, since the text pixels might be connected to the background pixels in TR, the unwanted removal can occur when the filling task is conducted. Performance measurement is done on the basis of precision and recall, probability of error parameters. The recall and precision is used to detect overlay text with higher efficiency. Recall is used to detect pixels other than overlay text pixels. Precision is used to detect the overlay text region[23].

Recall=Card (P∩T)/Card (T) -------------------- 6.1

Table 2. Results of Wonjum (CHANGE IN AN INTENSITY) method on more videos.

Input Video	No. Frame	Size of frame	LBP		No. of pixels		No. Candidate region	FPS		Recall	Precision	PE
Video1	127	240,320	44	87	3618		2	34.37		0.5470	0.2480	0.090
Video2	135	240,320	63	99	21	4312	3	35.63		0.5638	0.3124	0.120
Video3	100	240,320	56		1450		1	32.24		0.7356	0.4260	0.078
Video4	110	240,320	99	25	28	38	5480	4	35.89	0.5894	0.3828	0.128
Video5	127	240,320	58				1250	1	29.35	0.7352	0.4259	0.072
Video6	135	240,320	48	56			3456	2	34.38	0.5468	0.2479	0.089

Precision=Card (P∩T)/Card (P) --------------------- 6.2

Where

P- The set of detected pixels by each method.

T -The no of pixels belonging to overlay text

Recall= Used to detect overlay text region.

Precision= Used to detect empty space.

Probability of error is used find out false error and evaluated as follows.

PE=P (T) P (B|T) +P (B) P (T|B) -------------------- 6.3

Where

P (T) = the probability of overlay text.

P (B) = the probability of background pixels.

P (B|T) =the error probability to classify overlay text pixels as background pixels.

P (T|B) =the error probability to classify background pixels as overlay text.

Small PE means that the accuracy of overlay text is high [24].

Similarly we have applied the method on some more videos and the results of them are as shown in following table.

Conclusion

A novel method for overlay text detection and extraction from complex videos is implemented in this approach. The detection method is based on the observation that there exists a transient color between inserted text and its adjacent background.

The method is useful for real time application. An experimental result shows that change in intensity based approach gives 99% hit rate and 1% false rate. Whereas in edge approach implemented an effective video text detection method based on line features. This method exploits an improved canny edge detector to detect text pixels. By considering spatial distribution of edge pixels stroke information is incorporated in text region generation and filtering. This approach represents the distinguishing features of edge based. To detect the text canny edge detector is used because it acts as an optimal edge detector. From experimental results it is observed that edge based approach is having 93 % edge detection where is change in intensity approach is having 99% correct overlay text detection rate. The 7% false rate is present in edge based approach where as in change in intensity approach only 1 to 2 % false rate is present.

Then line map generated to detect continuous lines and remove isolated lines which are placed horizontally. In this way we detect overlay text using edge based approach. An edge based method increases the computational cost whereas change in an intensity based approach reduces the computational cost.

Future Work

The method is very useful in real time application. The future work is to detect and extract the moving overlay text to extend the algorithm for more advance and intelligent application.

References

1. Wonjun Kim, Changick Kim, “A New Approach for Overlay Text Detection and Extraction from Complex Video Scene”, in IEEE Trans. on image processing, Vol. 18, no. 2, Feb.2009.

2. Lyu, M.R., Song, J. and Cai, M., 2005. A comprehensive method for multilingual video text detection, localization, and extraction. IEEE transactions on circuits and systems for video technology, 15(2), pp.243-255.

3. N.Ostu, “A threshold selection method from gray-level histograms,” IEEE Trans. Syst., Man, Cybern, vol.9, no.1, pp.62-66, Mar. 1979.

4. T. Sato, T. Kanade, E.K. Hughes, and M.A. Smith, “Video OCR for digital news archive,” in Proc. IEEE International Workshop on Content-Based Access of Image and Video Libraries, Jan.1998,pp.52-60

5. L. Agnihotri and N. Dimitrova, “Text detection for video analysis,” in Proc. IEEE Int. Workshop on Content-Based Access of Image and Video Libraries, Jun. 1999, pp. 109–113.

6. R.C.Gonzalez and R.E.Woods, Digital Image Processing, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

7. C.G.M. Snoek and M.Worring, “Time interval maximum entropy based event indexing in soccer video,” in Proc. Int. Conf. Multimedia and Expo, Jul. 2003, vol. 3, pp. 481–484.

8. J.Gllavata, R.Ewerth, and B.Freisleben, “Text detection in images based on unsupervised classification of high-frequency wavelet coefficients,” in Proc. Int. Conf. Pattern Recognition, Aug. 2004, vol. 1, pp. 425–428.

9. J. Cho, S. Jeong, and B. Choi, “News video retrieval using automatic indexing of Korean closed-caption,” Lecture Notes in Computer Science, vol. 2945, pp. 694–703, Aug. 2004.

10. X.S.Hua, P.Yin, and H.J.Zhang, “Efficient video text recognition using multiple frame integration,” in Proc. Int. Conf. Image Processing, Sep. 2004, vol. 2, pp. 22–25.

11. K.C. K. Kim et al., “Scene text extraction in natural scene images using hierarchical feature combining and verification,” in Proc. Int. Conf. Pattern Recognition, Aug. 2004, vol. 2, pp. 679–682.

12. M.R.Lyu, J.Song, and M.Cai, “A comprehensive method for multilingual video text detection, localization, and extraction,” IEEE Trans. Circuit and Systems for Video Technology, vol. 15, no. 2, pp. 243–255, Feb. 2005.

13. C.Liu, C.Wang, and R.Dai, “Text detection in images based on unsupervised classification of edge-based features,” in Proc. Int. Conf. Document Analysis and Recognition, Sep. 2005, vol. 2, pp. 610–614.

14. X.Liu and J. Samarabandu, “Multiscale edge-based text extraction from complex images,” in Proc. Int. Conf. Multimedia and Expo (ICME), Jul. 2006, pp. 1721–1724.

15. M.Bertini, C.Colombo, and A.D.Bimbo, “Automatic caption localization in videos using salient points,” in Proc. Int. Conf. Multimedia and Expo, Aug. 2001, pp. 68–71.
16. J. Wu, S.Qu, Q.Zhuo, and W.Wang, “Automatic text detection in complex color image,” in Proc. Int. Conf. Machine Learning and Cybernetics, Nov. 2002, vol. 3, pp. 1167–1171.
17. Y.Liu, H.Lu, X.Xue, and Y.P.Tan, “Effective video text detection using line features,” in Proc. Int. Conf. Control, Automation, Robotics and Vision, Dec. 2004, vol. 2, pp. 1528–1532.
18. T.Ojala, M.Pierikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.
19. S. Antani, D. Crandall, and R. Kasturi, “Robust extraction of text in video,” in Proc. Int. Conf. Pattern Recognition, Sep. 2000, vol. 1, pp.
20. J. M. Pike and C. G. Harris, “A combined corner and edge detector,” in Proc. Alvey Vision Conf., 1988, pp. 147–151.
21. S.U. Lee, S.Y.Chung, and R.H.Park, “A comparative performance study of several global thresholding techniques for segmentation,” Comput. Vis., Graph., Image Process., vol. 52, pp. 171–190, 1990.
22. R.Lienhart and Axel Wernicke, "Localizing and Segmenting Text in Images and Videos", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 12, No.4. pp.256-268, April 2002
23. Y.Zhong, K.Karu, and A.X.Jain," Locating text in complex color images", Pattern Recognition, 28(10): 1523-1535, 1995
24. H.Li, D.Doermann, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. on Image Processing, vol. 9, pp. 147-156, Jan. 2000.
25. Min Cai, Jiqiang Song, and M.R. Lyu, "A new approach for video text detection", IEEE Conf. on Image Processing (ICIP'2002), vol. I, pp. 11117 -11120. 2002.