International Journal of Future Generation Communication and Networking
Vol. 9, No. 11 (2016), pp. 215-228
http://dx.doi.org/10.14257/ijfgcn.2016.9.11.20

# News Video Text Area Positioning Based on an Improved Trajkovic Corner Detector

Hua Huo[1, 2] and Di Li[1, *]

[1]*Information Engineering College, Henan University of Science and Technology, Luoyang, Henan, 471023, China;*
[2]*School of Software, Henan University of Science and Technology, Luoyang, Henan, 471023, China*
[*]*Corresponding author: 327017465@qq.com*

***Abstract***

*Comparing to other videos, text in news video contains more semantic information which are mostly descript the news event. So, the extraction of text plays an important role for obtaining the high-level semantic information from the video. In this paper, the improved Trajkovic corner detection algorithm has been studied based on the characteristics of text in news video. According to the characteristics of text, the adaptive thresholding method of scale in combination with standard deviation (SCSD) can be applied to determine the corners, which aims to accurately obtain the text-related corners. The breadth-first clustering algorithm was also been used to distinguish and plan the detected corners which are in same range to determine the text area in video frame. The accuracy of recognition and semantic exploration of the extracted text can be used to quickly understand news video that can effectively improve the efficiency to summarize the news videos. The experimental results show this method with high pertinence which can better achieve the targets. The detection accuracy of text-related corners can reach to 82.1% which is higher than other methods. Under the same experimental condition, recall ratio and precision ratio of the text area are also improved dramatically.*

***Keywords****: localization, corner detection, adaptive threshold, clustering etc.*

## 1. Introduction

The news videos are everywhere which is one of the good ways to obtain the real-time information. The readers will firstly attracted by the titles to find the most interesting information in the multifarious news video. There are various ways to find the interesting titles[1-2], but the title of most news video just stay in the subjective thought of the up-loader. The misunderstandings cannot be avoided if just focused on the titles and the most useful information may also be ignored. Exploring video semantic meaning[3] can make the readers better and faster to learn the subject of news. For example, in a lengthy financial news if we just want to know today's stock market that can explore the semantic information to position rapidly; Or if we need to find out the video fragment which is only remembered from the numerous video, the exploration of video text can also works. Currently, there are some application-systems summarized the semantic information of the paragraph based on the time nodes which can effectively assist the readers to browse the videos. Because there are specifications and standards for expressing news in videos[4], in this paper, the analysis of text information in the news video has been applied because news video comes with the formatted title, subtitles, and a clear aim to describe a particular event. All of the texts can express the high-level semantic information. Through the study of texts in news video, we can position the location of the text to achieve rapid orientation of the demand and improve the efficiency of watching.

ISSN: 2233-7857 IJFGCN

In view of the text, corner[5-6] is an important factor of text. Corner is a very important characteristic of the image which plays the important role on understanding of graphics and analysis. [7]Corner contains the pixel generally accounts for 5% of the total image pixel points. So, using corner point to process images can effectively reduce the amount of data information and keep the important characteristic of image at the same time which can effectively improve the computing speed. It can be the largest local curvature point on the contour or the relatively sharp change point of grey value. The text areas in news video come with many concentrated corners which are significantly useful for us to extract and analysis texts[8].

The most common corner extraction algorithm is Harris algorithm[9] which is it based on the differential and autocorrelation matrix to detect corners. The characteristics of extracted corners by this method are reasonable distributed, but a given specific threshold is compulsory which leads to the positioning precision is not very accurate. Moreover, the load of calculation is huge which can be the reason for lower time efficiency. In this research, the main corners focused are text corners. Therefore, in this paper, the 8-neighborhood Trajkovic corner detection algorithm has been used. The introduction of the improved cornerness calculating method and the corners local non-maxima suppression method can significantly improve news video text positioning effect, time efficiency and accuracy. Corners of text in video frame are one part of the corners. The objectives of this paper are including the text clustering based on the characteristics of corners, select text corners and calculate the text blocks in complex background.

To sum up, there is the huge potential demand on the video text localization and detection. After the detection of text from the videos that can greatly facilitate video retrieval and management which is very helpful for the further semantic analysis[10] of the videos.

## 2. Previous Researches

Currently, text location in diverse backgrounds is still a difficult problem. This is because the changes of different factors including the contract, size, location, and the font between texts colours in the picture and the background which is hard to find a common way to separate the background from them completely[11]. Text location includes scene text location[12-13] and caption positioning[14-15]. Scene texts are normally irregular, so it is difficult to position. This research is only focused on the caption added manually. Since the most of captions are horizontal or vertical distribution, and the colour of each block position is consistent with high background contrast. The stability of testing will be relatively high. The close attention to timeliness of the text in news video is necessary. News video is one kind of video with strong timeliness, so a good algorithm can make texts be positioned very fast and to be tested to accomplish the targets. Because the news video text-related corner has certain features, we decided to use the corner detection method to judge the text areas in the news video. Compared with the general text detection method, our method is more targeted, which can better complete text localization, and will help the work such as text extraction and semantic summarization.

Currently, the text detection methods are based on the region and texture[16].

### 2.1. Method Based on Region

The text extraction method based on region is mainly depends on the differences of colours or grey scale between text area and other areas which is normally based on connected component method and the method based on edge or corner. The method based on connected component is generally use text connectivity and location relationships to filter positioning text area. The method based on edge information is due to the abundant edge information in the text area. The text can be positioned by finding the law of the edge information.

Kim and Kim [17] proposed a novel framework to detect and extract the overlay text from the video scene. They used a transition map to distinguish transient colours between inserted text and its adjacent backgrounds. After, they used a reshaping method to extract the candidate regions. Based on the occurrence of overlay text in each candidate they determined the overlay text regions. Finally, they used the projection of overlay text pixels in the transition map to process text extraction. Lyu and Song *et al*. [18] proposed a method which emphasized the multilingual capability over the whole process. The text detection was carried out by edge detection, local threshold and hysteresis edge recovery. Lastly, they used coarse-to-fine localization scheme to accurately identify text regions.

Zhang and Liu *et al*. [19] generated candidate text regions by applying morphologic operation based on corner points detected in different scale and filter non-text regions by combining proposed stroke width feature with some simple geometric properties. Then they proposed a new multi-instance semi supervised learning strategy considering the unknown contrast parameter in stroke width extraction. Cai and Song *et al*. [20] also propose a coarse-lo-fine detection to local text regions by using invariant features, such as edge strength, edge density and horizontal distribution.

### 2.2. Method Based on Texture

The method based on texture is mainly focused on the differences between texture of text and the background. The text texture and background are assumed with some regularity. This method has been applied in text location based on machine learning of the classification through the filter, Fourier transform or others ways to extract the texture. Shivakumara and Phan *et al*. [21] used the wavelet single level decomposition LH, HL and HH sub-bands for computing features. Then, the computed features are fed to k-means clustering to classify the text pixel from the background of the image. The outcomes from this research help them to classify true text pixel in the image and to detect the text blocks.

Mosleh and Bouguila *et al*. [22] investigated the text locations via an unsupervised clustering performed on the connected components produced by the stroke width transform (SWT) and developed a novel edge detector which benefits from the geometric features revealed by the bandlet transform to build an accurate edge map. Then the motion patterns of the text objects are analysed to localize video texts. Ye and Huang *et al*. [23] proposed a novel coarse-to-fine algorithm by using multi-scale wavelet features to locate text lines. They separated the research into two parts which are coarse detection and fine detection. Firstly, using the wavelet energy feature to locate all possible text pixels and connect these pixels into regions. Then, represent the texture pattern of a text line with the extracted four kinds of texture features and the search algorithm is applied to select the most effective features. Finally, use the SVM classifier to identify true text from the candidates.

### 2.3. Method Based on Region and Texture

Shivakumara and Phan *et al*. [24] proposed a method based on the Laplacian in the frequency domain for video text detection. The input image is firstly filtered with Fourier-Laplacian. Then, use K-means clustering to identify candidate text regions based on the maximum difference. The skeleton of each connected component helps to separate the different text strings from each other. Finally, text string straightness and edge density are used for false positive elimination.

Yang and Quehl *et al*. [25] designed an edge-based multi-scale text detector to identify potential text candidates with high recall rate. Then, use an image entropy-based filter to refine the results. Finally, Stroke Width Transform (SWT) and Support Vector Machine (SVM) based verification procedures are applied to eliminate the false alarms.

In general, the existing text location methods have certain selectivity. Each method is only suitable for a particular type of writing to attain good extraction effect. In different videos or scenarios, the text alignment, background and size are not unified all the time. The text area detection under complicated background effect is still not effective, so the method focused on one character is difficult to achieve the desired effect. To solve this issue, this research was based on the news video text region extraction and simultaneously adopted the consideration of multiple aspects to achieve better detection results. Also, detection efficiency has been increased to achieve the timeliness requirement of news. This research put forward the method to control the local maximum value under certain to extract the text corner information. After that the method of text clustering angular point information to detect text which can significantly improve the text detection results under multi-scales.

## 3. Positioning And Extraction of Text Area

### 3.1. Caption Frame Detection

To analyse the video, the first step is to detect the caption frame of text[26]. When the caption appears in MPEG video, P and B frames in the code block will be increased [27]. In this paper, this method accompany with inter-frame difference to detect text frame. Frame difference is commonly used to test the shot change. The method used in this research is about frame difference of intercropping within a shooting to judge the grey-level histogram. For the frame contains text frame, the grey histogram is apparently different from other frames. So, the combination of P and B frames in the code block information can be used to determine the text.

According to the special character of news video and text frame, the texts are generally staying for more than 3 seconds to convenient for people to read. So, eliminating operation is necessary for the detected text frames. In time series, text display is 3 seconds. The detected framed after 100 will be ignored to reduce the load of calculation without any influences of detection results.

### 3.2. The Improved 8-neighbors Trajkovic Corner Detector

After the acquisition of detected text frames, corner detection and optimization can be taken for the detected text frames to explore the text area. Harris corner detection algorithm is the most common and popular corner detection algorithm. In general, it is defined the corner as the grey values for both of mutually perpendicular directions change significantly. The test results are in high accuracy. But, in view of the text in news video the most of corners are geometric. So, lots of texture featured indiscriminately detected corners will increase the complexity.

The improved Trajkovic corner detector [28] can better identify geometric corners which the detection accuracy is more stable than Harris, higher time efficiency and even more real-time requirement can be achieved. Since the subtitles in news video extraction depend more on time efficiency, so more suitable for this experimental project. In order to detect text areas are more accurate, the 8- neighbours Trajkovic algorithm is adopted which with enhanced ability of resisting noise and reduce the size of the diagonal side reaction compared to the Trajkovic algorithm based on 4-neighbours.

Following are the details of corner detection algorithm. The corner was defined based on the cornerness at certain pixel. The cornerness is used to measure the degree of a pixel point close to the corner. The higher amount of corners at a pixel point is more close to the corner. Trajkovic cornerness algorithm is calculated in a small round window to consider all the lines passed through the centre of the circle and then compare the grey-values between two intersecting points of the straight line and the centre of the circle. Thus, the change degree of grey value at every possible degree can be calculated. This

algorithm is different from others, such as Moravec[29] only calculate the grey value change in certain direction. The calculation for all of the directions can more accurately determine the grey value differences between the candidate points and related points.

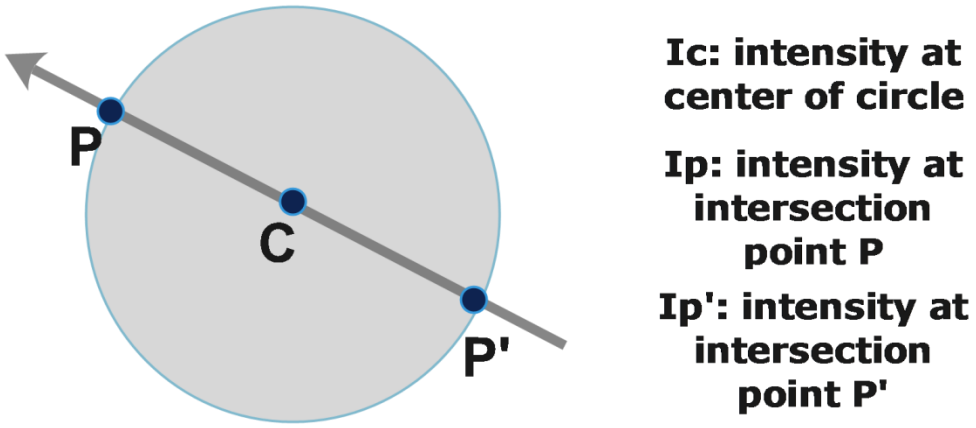



**Figure 1. The Basic of Corner Quantity Algorithm**

As shown in Figure 1. is the basic of cornerness algorithm, "**C**" is the centre of the circle, the straight line cross the circle and has two crossing points are P and P'. The (1 is the simple cornerness algorithm equation.

$$C(x,y) = \min((I_P - I_C)^2 + (I_{P'} - I_C)^2), \forall_{P,P'} \quad (1)$$

To start the calculation, the minimum value is assumed which is the threshold to define corners. The equation measure the degree of change in all directions in the image.

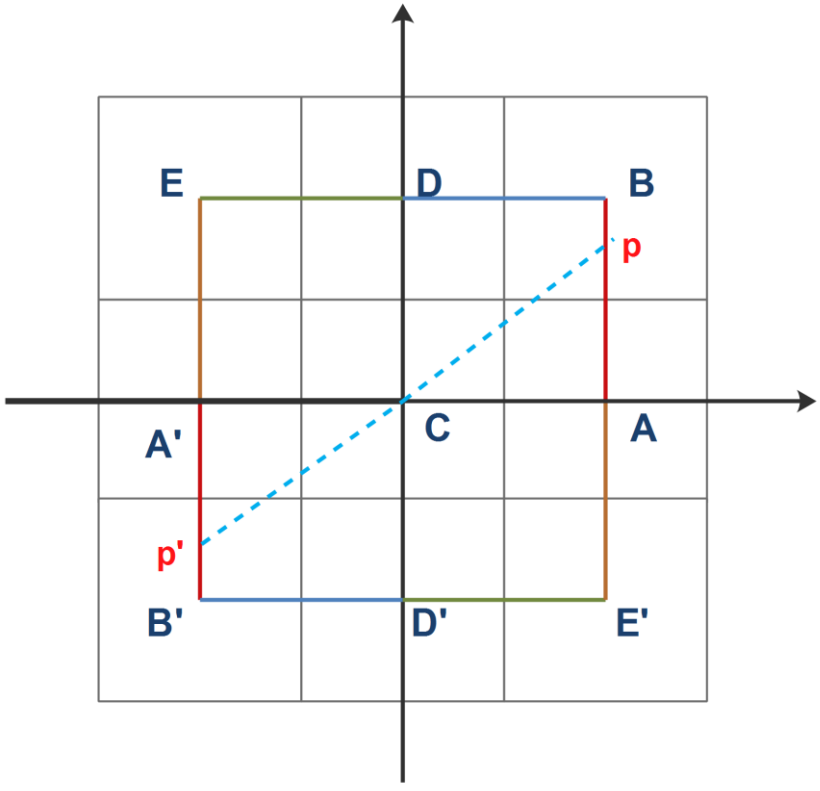



**Figure 2. The Example of Target Pixel is in 8-Neighbours**

As Figure 2 shows, the corners within the 8 neighbours centred by "**C**" point can be defined as (2)

$$C_{simple}(x,y) = \min(r_A, r_B, r_C, r_D) \tag{2}$$

The $\mathbf{r_A}$, $\mathbf{r_B}$, $\mathbf{r_C}$ and $\mathbf{r_D}$ are represent the degree of grey values change in four directions within 8 neighbours. For example in direction **A**, the equation is

$$r_A = (I_A - I_C)^2 + (I_{A'} - I_C)^2 \tag{3}$$

The interpolation method was introduced to statistic in all directions. In Figure 2, **P** and **P'** are the points between line **AB** and line **A'B'** respectively. There are corresponding four lines within 8 neighbours. For example line **PP'**

$$r_P = (I_P - I_C)^2 + (I_{P'} - I_C)^2 \tag{4}$$

The change of grey value in the image has the common rules, the closer pixels has small difference of grey values. Therefore, the equations based on the position relationship between **PP'** and **AB**, **PP'** and **A'B'** can obtained. As **Error! Reference source not found.** shows.

$$\begin{aligned} I_P &= (1-x)I_A + xI_B \\ I_{P'} &= (1-x)I_{A'} + xI_{B'}, x \in (0,1) \end{aligned} \tag{5}$$

By the analogy, the cornerness can be expressed and extended to 8 neighbours. The simple cornerness after interpolation method can be represented as **Error! Reference source not found.** which based on the interpolation method within 8 neighbours.

$$C_{Pixel} = \min_{x \in (0,1)} (r_1(x), r_2(x), r_3(x), r_4(x)) \tag{6}$$

The next step of calculation is to calculate the minimum values based on the equations shown in **Error! Reference source not found.**

$$C_{Pixel}(x,y) = \left\{ \begin{array}{l} C_{simple}(x,y), if\ (B_i \geq 0 \,\|\, (A_i + B_i) \leq 0),\ forany, i = 1,2,3,4 \\ \min(r_i - \frac{B_i^2}{A_i}), if\ (B_i < 0 \,\&\&\, (A_i + B_i) > 0),\ forall, i = 1,2,3,4 \end{array} \right\}$$

$$(r_1 = r_A, r_2 = r_B, r_3 = r_C, r_4 = r_D) \tag{7}$$

To able to process the project, the more detected geometric corners are needed but not texture featured corners. So, to reduce the texture featured corners by using the multi-grid algorithm. The texture featured corners are normally located densely and the grey value changes within a small threshold. So corner detection started after shrink the original images, for example: $C_{MG} = \frac{C_{ORH} + C_{ORV}}{4}$ , $\mathbf{C_{MG}}$ is the pixel after the multi-grid algorithm, $C_{ORH}$ and $C_{ORV}$ are the pixels in horizontal and vertical direction respectively which indicate the average of pixel of original image and four pixels in four directions after multi-grid algorithm. In this way can eliminate the change of grey value within threshold which can improve the efficiency of text detection by better geometric corners detection.

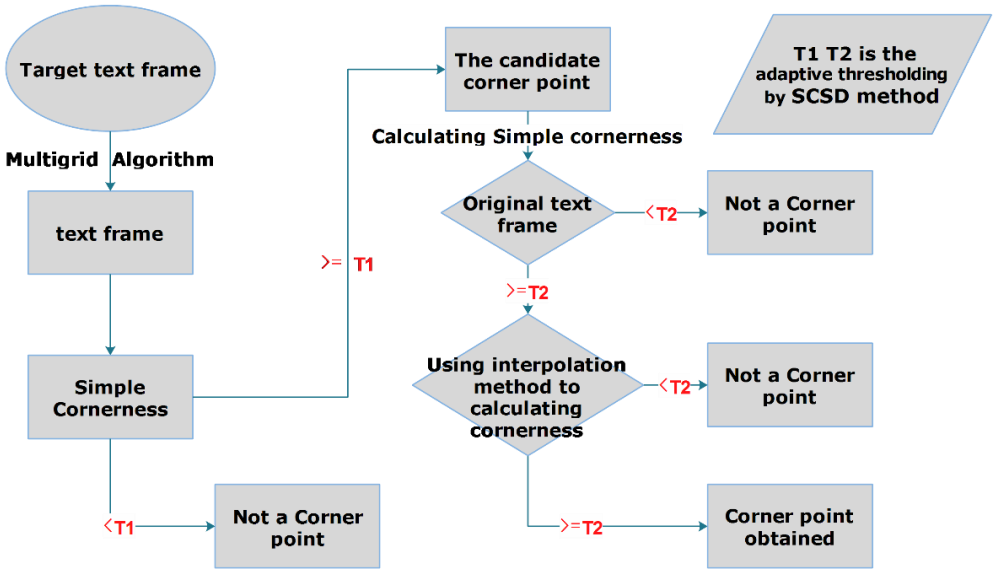



**Figure 3. The Flow Chart of Corner Detection in This Project**

In general, the corner detection method in this project can be summarized as Figure 3. Since the video frame sizes and text sizes are different from different video fragments in the process of the threshold value calculation. So, in order to make the threshold can be widely used to improve computational efficiency the adaptive threshold algorithm of scale in combination with standard deviation (SCSD) was used to control the threshold. The SCSD algorithm is set for determination of corner quantity T1 which aims to identify the corners in different sizes of images and texts.

$$t = \lambda * \sqrt{\frac{\sum_{i=1}^{8}(I(r) - I(r_i))^2}{8}}, \lambda \in (0,1) \tag{8}$$

$$t = \begin{Bmatrix} t & if \quad t \le T1 \\ T1 & if \quad t > T1 \end{Bmatrix}$$

As shows, T1 is the threshold of simple cornerness, "t" is the amount of new threshold for SCSD algorithm. So "λ" was introduced as a coefficient to control the size of the threshold T1 which is related to the scale of the image. If the threshold value of T1 ≥ t, then use the 't' replace the 'T1' as a new simple corner threshold, if the amount of T1 < t, use T1 as threshold. The introduction of SCSD adaptive threshold algorithm can improve the corner detection result more close to the ideal value and can control the quality of detected corners without reducing time efficiency under the condition of more accurate identification of corners.

### 3.3. Local Non-maximum Suppression of Corners and Text Area Positioning

**3.3.1. Local Non-maximum Suppression:** After the completion of the corner detection, the final corners can be obtained by local non-maximum suppression of detected corners. Ordinary local non-maximum suppression is not suitable for the corner detection in text frame for this research. The suppression method based on corner point range has been applied in this project to rank the obtained cornerness. In extensive research on text in video, text, the surrounding pixels and noise level, the candidate corner within 8*8 were investigated to determine its location. The corners out of this region have been eliminated in case to avoid the same location with multiple corners overlap that could influence to the judgment of the text region.

To cornerness, the traverse has been taken for all of the corners within the region. The point with more corners is treated as an important corner within the region. The selected other corners are eliminated which the corner value is set to 0 in the sequence of test results. This can effectively prevent the text part of the overlapping useless corners which can reduce the amount of calculation and increase the accuracy of text area positioning by corner clustering.

**3.3.2. Text Area Positioning:** Through the experiment results, the corners around text are obtained. The corners are only in view of the image characteristics in the text which also need to be further processed to acquire the text area. Due to the text area is angular and dense, so the corner Euclidean distance is used to determine its classification which is based on distance Breadth - First clustering algorithm to cluster the text corners. Different from other algorithms such as K-means which is necessary to set the number of cluster categories that cannot be better support clustering based on distance.

Algorithm used in this research with advantages which including there is no necessary to set the number of categories and also based on the characteristics of news video text: text box length value will be larger than the height value, we set formed by a weighted coefficient to control the shape of a category, combined with the pixel density, the clustering results in the horizontal direction than the more wide, so that we can make the clustering results more accurate and more close to the text box.

The Breadth-First clustering algorithm is shown in Figure 4. The distance between any two points within the corner set is calculated based on the .

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{9}$$

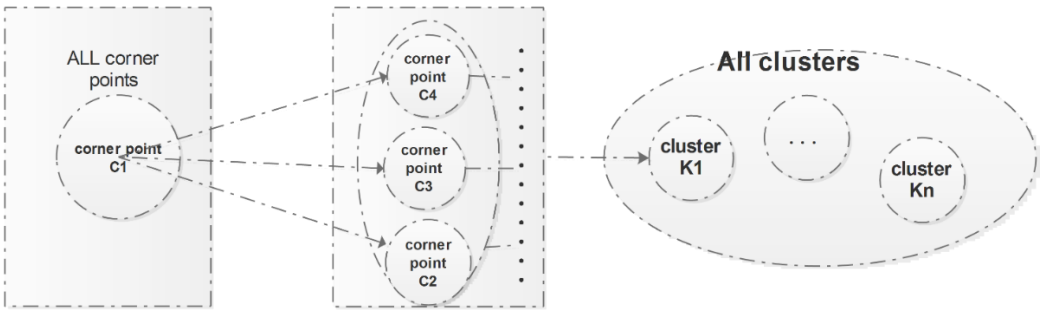



**Figure 4. Breadth-First Clustering Algorithm**

A set threshold value is needed. The distance from any point $C_1$ less than the threshold value is added into the set and assume there are $C_2$ and $C_3$ and $C_4$ three points meet the conditions. Then, assume $C_2$, $C_3$ and $C_4$ as basis points in the search for eligible points until there is no point in conformity with the conditions. The point set is classified as cluster K1 and repeat this step in the points of the remaining points until all points are

classified. To determine the point set, the point with fewer corners is noise point sets which are abandoned. Marking on the sets meet the requirements which are named as minimum circumscribed rectangle. Finally, the text areas of the image frames in news video are obtained.

## 4. Experimental Analysis

Text detection methods are various especially in recent year. The classification of the methods based on region or texture are insufficient which has become increasingly blurred image on text detection. So, the comprehensive methods have the potential to attain widely application. Because there is no standard video image database is used to test, this is largely affects the performance of different methods for evaluation. In this project, a lot of news video had been chosen to do experiments and compared with other methods to obtain the text corners detectable rate. Also, the recall and precision ratios are compared in the text area positioning to make sure the results more convincing.

During experimental analysis, hundreds of news video with 3-5 minutes length were investigated. The text frame corner detection and text localization compared with other algorithms have achieved significant effect. The typical experimental results are shown below.

### 4.1. Corner Detection Results

Figure 5. Compared Results between Harris Algorithm and the Method Used in This Project of Corner Detection Harris corner detection algorithm can detect there are many texture featured corners not related with the text. The large amount of testing is not conducive to text localization. The Trajkovic 4-neighbors method can effectively reduce the texture featured corners detection, but there are many overlapping corners around the text which can affect the text area positioning and clustering.


**Figure 5. Compared Results between Harris Algorithm and the Method Used in This Project of Corner Detection**

The Table 1 clearly show the method used in this research with high time efficiency and accuracy compared with other methods which can effectively increase the ratio of text corners detection and eliminate the useless corners.

**Table 1. Corner Detection Result**

| Method | Speed (msec/frame) | Number | Text-related Number | Accuracy |
|---|---|---|---|---|
| Harris | 980.49 | 657 | 286 | 43.5% |
| Trajkovic Operator (4-Neighbours) | 75.8 | 1610 | 1169 | 72.6% |
| Our method | 43.8 | 515 | 423 | 82.1% |

1) Speed. The speed is indicated by the average processing time per frame for the Corner detection.
2) Number. The number is the number of corner point in each text frame.
3) Text-related Number is the text-related Corner point in each text frame.
4) Accuracy. The detection accuracy evaluates how many presents of the detected text regions are correct.

### 4.2. Text Area Positioning Results

In Figure 6 is about the clustering algorithm in the text area positioning results and Figure 7 is about text enhancement and binarization results. The representative examples from numerous experimental videos are selected to prove the results. The result shows this method can effectively identify the text area and discard the unrelated and interfered text which has nothing to do with the video semantic extraction in the background.


**Figure 6. The Example of Text Area Positioning**

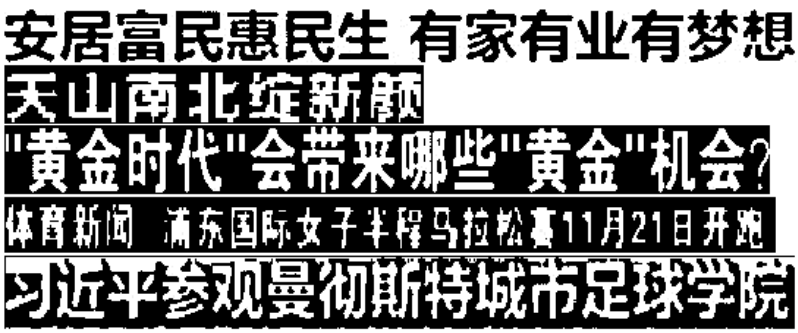



**Figure 7. Text Enhancement and Binarization Results**

The recall ratio **R**, precision ratio **P** and **F** index are used as indicators to measure the performance of text localization. More accurate results can be obtained by comparing with other news video text localization algorithms in the same video and operating conditions. The equations for recall and precision ratios are shown below which $\boldsymbol{n}_c$ is indicates the number of accurately detected text area; $\boldsymbol{n}_m$ is indicates the number of missing detected text areas in the frame; $\boldsymbol{n}_f$ is indicates the number of text areas being mistaken for text area.

$$R = \frac{n_c}{n_c + n_m} 100\% \tag{10}$$

$$P = \frac{n_c}{n_c + n_f} 100\% \tag{11}$$

The
Table 2. Comparison of Different Scene Text Detection **Algorithm**s which is obtained based on a large number of experiments. The method in this research has advantage on recall ratio **R**, precision ratio **P** and **F** index which can help to achieve a better text detection results.

**Table 2. Comparison of Different Scene Text Detection Algorithms**

| Method or author | P | R | F |
|---|---|---|---|
| Hua[30] | 0.730 | 0.600 | 0.660 |
| Chen[31] | 0.674 | 0.697 | 0.685 |
| Shivakumara[32] | 0.731 | 0.647 | 0.687 |
| Our method | 0.826 | 0.779 | 0.728 |

## 5. Conclusion

This paper dedicated to the study of news video text localization which adopts the method based on corner detector. The improved Trajkovic corner detector was used to extract corners. At the stage of setting threshold value, adaptive threshold algorithm with SCSD can meet requirements to select threshold in the direction suitable for news video in order to obtain more corners. Then process local non-maximum suppression for the detected results to obtain the final corners. Using Breadth-first clustering algorithm combining with the image scale, transverse weighted diagonal points and calculate the text area to better obtain text areas. The experimental results achieved significant effect which indicates the reasonability and effectiveness of this method.

Although this method has achieved significant results, but there are still disadvantages. Corner detection of adaptive threshold value method still has potential to be improved which can try to combine corner density calculation. Also, corner text clustering algorithm comes with clustering deviation that need to be further perfected. The applicability of this method can also be further improved and it can be a reference for other types of text detection. Completion of the text area positioning and extraction can lay the foundation of subsequent news video semantic information extraction, semantic annotation of time frame rank and automatic summarized.

## Acknowlegements

This study was supported by the National Natural Science Foundation of China under the grant 60743008, the Key Scientific and Technological Projects of Henan Province in China under the grant 142102210045 and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

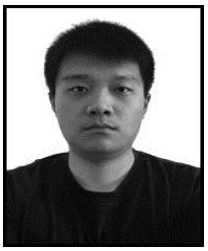
[1] X. Tang, X. Gao, J. Liu and H. Zhang, "A spatial-temporal approach for video caption detection and recognition", Neural Networks, IEEE Transactions, vol. 13, no. 4, **(2002)**, pp. 961-971.
[2] Y. Zhong, H. Zhang and A. K. Jain, "Automatic caption localization in compressed video", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 22, no. 4, **(2000)**, pp. 385-392.
[3] I. I. Fu, H. Hu, R. Chen and H. Ren, "Video Semantic Analysis", Information Technology Journal, vol. 11, no. 10, **(2012)**, pp. 1381-1390.
[4] Y. Nakamura and T. Kanade, "Semantic analysis for video contents extraction—spotting by association in news video", in Proceedings of the fifth ACM international conference on Multimedia, ACM, **(1997)**.
[5] X. Clady, S. H. Ieng and R. Benosman, "Asynchronous event-based corner detection and matching", Neural Networks, vol. 66, **(2015)**, pp. 91-106.
[6] D. Li, B. Zhong and K. K. Ma, "Multi-scale corner detection based on arithmetic mean curvature", in Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference, **(2015)**.
[7] P. R. Possa, S. A. Mahmoudi, N. Harb, C. Valderrama and P. Manneback, "A multi-resolution fpga-based architecture for real-time edge and corner detection", Computers, IEEE Transactions, vol. 63, no. 10, **(2014)**, pp. 2376-2388.
[8] T. Lu, S. Palaiahnakote, C. L. Tan and W. Liu, "Video Text Detection Systems", in Video Text Detection, **(2014)**, pp. 169-193.
[9] C. Harris and M. Stephens, "A combined corner and edge detector", in Alvey vision conference, **(1988)**.
[10] C. L. Huang, H. C. Shih and C. Y. Chao, "Semantic analysis of soccer video using dynamic Bayesian network", Multimedia, IEEE Transactions, vol. 8, no. 4, **(2006)**, pp. 749-760.
[11] C. S. Chen, Y. P. Hung and J. B. Cheng, "RANSAC-based DARCES: A new approach to fast automatic registration of partially overlapping range images", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 21, no. 11, **(1999)**, pp. 1229-1234.
[12] K. Kim, H. Byun, Y. Song, Y. W. Choi, S. Chi, K. K. Kim and Y. Chung, "Scene text extraction in natural scene images using hierarchical feature combining and verification", in Pattern Recognition, 2004, ICPR 2004, Proceedings of the 17th International Conference, **(2004)**.
[13] C. Li, X. Ding and Y. Wu, "Automatic text location in natural scene images", in icdar, IEEE, **(2001)**.
[14] H. J. Zhang and Y. Zhong, "Automatic caption text detection and processing for digital images", Google Patents, **(2001)**.
[15] H. Li, D. Doermann and O. Kia, "Automatic text detection and tracking in digital video", Image Processing, IEEE Transactions, vol. 9, no. 1, **(2001)**, pp. 147-156.
[16] Y. Hao, Z. Yi, H. Zeng-guang and T. Min, "Automatic text detection in video frames based on bootstrap artificial neural network and ced", **(2003)**.
[17] K. Wonjun and K. Changick, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene", Image Processing, IEEE Transactions, vol. 18, no. 2, **(2009)**, pp. 401-411.
[18] M. R. Lyu, J. Song and M. Cai, "A comprehensive method for multilingual video text detection, localization and extraction", Circuits and Systems for Video Technology, IEEE Transactions, vol. 15, no. 2, **(2005)**, pp. 243-255.
[19] B. Zhang, J. Liu and X. Tang, "Multi-scale video text detection based on corner and stroke width verification", in Visual Communications and Image Processing (VCIP), **(2013)**.
[20] M. Cai, J. Song and M. R. Lyu, "A new approach for video text detection", in Image Processing. 2002. Proceedings. 2002 International Conference, **(2002)**.

[21] P. Shivakumara, T. Q. Phan and C. L. Tan, "A robust wavelet transform based technique for video text detection", in Document Analysis and Recognition, 2009, ICDAR'09, 10th International Conference, **(2009)**.
[22] A. Mosleh, N. Bouguila and A. B. Hamza, "Automatic inpainting scheme for video text detection and removal", Image Processing, IEEE Transactions, vol. 22, no. 11, **(2013)**, pp. 4460-4472.
[23] Q. Ye, Q. Huang, W. Gao and D. Zhao, "Fast and robust text detection in images and video frames", Image and Vision Computing, vol. 23, no. 6, **(2005)**, pp. 565-576.
[24] P. Shivakumara, T. Q. Phan and C. L. Tan, "A laplacian approach to multi-oriented text detection in video", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 33, no. 2, **(2011)**, pp. 412-419.
[25] H. Yang, B. Quehl and H. Sack, "A framework for improved video text detection and recognition", Multimedia Tools and Applications, vol. 69, no. 1, **(2014)**, pp. 217-245.
[26] C. L. Huang and B. Y. Liao, "A robust scene-change detection method for video segmentation", Circuits and Systems for Video Technology, IEEE Transactions, vol. 11, no. 12, **(2001)**, pp. 1281-1288.
[27] U. Gargi, S. Antani and R. Kasturi, "Indexing text events in digital video databases", in Pattern Recognition, 1998. Proceedings, Fourteenth International Conference, **(1988)**.
[28] M. Trajković and M. Hedley, "Fast corner detection", Image and Vision Computing, vol. 16, no. 2, **(1998)**, pp. 75-87.
[29] H. P. Moravec, "TOWARDS AUTOMATIC VISUAL BBSTACLE AVOIDANCE", in International Conference on Artificial Intelligence (5th: 1977: Massachusetts Institute of Technology), **(1977)**.
[30] X. S. Hua, L. Wenyin and H. J. Zhang, "An automatic performance evaluation protocol for video text detection algorithms", Circuits and Systems for Video Technology, IEEE Transactions, vol. 14, no. 4, **(2004)**, pp. 498-507.
[31] D. Chen, J. M. Odobez and H. Bourlard, "Text detection and recognition in images and video frames", Pattern recognition, vol. 37, no. 3, **(2004)**, pp. 595-608.
[32] P. Shivakumara, A. Dutta, C. L. Tan and U. Pal, "Multi-oriented scene text detection in video based on wavelet and angle projection boundary growing", Multimedia tools and applications, vol. 72, no. 1, **(2014)**, pp. 515-539.

## Authors

**Hua Huo**, he is a professor with a Ph.D. in Information Engineering College and School of Software, Henan University of Science and Technology. His research interests in intelligent information processing and video semantic extraction.



**Di Li**, he is a postgraduate student of Information Engineering College, Henan University of Science and Technology. His research interests is intelligent information processing.