Video image caption location based on FAST corner detection

Huibai Wang[1, a], Wen Yuan[2,b]

[12]North China University of Technology , Beijing, China

[a]wanghb@ncut.edu.cn ,[b]yuanwen313@126.com

Keywords: FAST corner detection, Caption location, Caption detection

Abstract. It is well known that the caption information is significantly important upon video and image retrieval analysis. A caption location method based on FAST (Features From Accelerated Segment Test) corner detection algorithm is illustrated in this paper. This method use FAST corner detection algorithm to get the corners information, then position the row and column respectively through the analysis of the horizontal integral projection and connected region, finally get validation by heuristic rules and locate the caption region. The experimental results show the performance of our method.

Introduction

The number of online video is growing geometrically with the rapid progress of global information society, therefore making the video retrieval and review become necessary in terms of social justice and internet health. Hence, Video caption, being an important part of the video which contains a large amount of information, provides a reference for the video semantic tracking and video content review. There is a strong link between video extraction effectiveness and video caption positioning accuracy since caption positioning is the basis of video extraction. Generally speaking, caption location can be typically classified into four methods, shown below:

A. Method based on connected component

Kim et al [1] use gray level variance and color variance, by color clustering method to extract the text area. The disadvantage to this approach is robustness in complex underground and therefore unfeasible to broadcast.

B. Methods based on edge

Srivastav and Kumar [2] used to apply Canny edge detection method to identify the original text region and non-text area according to the generated adaptive threshold and then remove the obvious non-text region by referencing heuristic rules. In addition, a similar edge-based detection method was introduced by Shivakumala [3] to locate text. To summarize, methods based on edge is fast but sensitive to noise, and sometimes it is unfeasible to distinguish between text area and background area.

C. Methods based on texture

Ye et al [4] used to use local binary pattern (LBP) extraction and neural network classifier, respectively, to clarify the texture of specific region by drawing feature histogram and identify text region using neural network classifier. It is noted that the methodology mentioned above might be less feasible when the line structure of some non-text regions have similarities with text texture.

D. Methods based on machine learning

Hanif et al [5] use edge, gradient and other features to form a mixed feature set, use the Adaboost algorithm to construct a strong classifier for training and testing, to get text area. The method based on machine learning is robust but the efficiency is low.

It takes for granted that text always comes with a large number of corners as it has more strokes. Therefore, we can use the corner detection algorithm to obtain the information of these corners and locate relevant text based on the angular distribution, angular density and other relevant information. In many cases, Harris detection algorithm is generally used to detect corners. For instance, Kai-Hsiang [6] uses a Harris based method to obtain corner information and then extract related text area by implementing morphological transformation toward corner image. This paper provides a caption location method based on FAST corner detection algorithm. Under this method, corner information is detected by FAST corner detection algorithm and processed by horizontal integral projection &

connected region analysis. Such processed information provides criteria for text region location, and then proved by heuristic rules. Compared with former methods, the advantages of ours are shown below:

•Compared with Harris corner detection algorithm, our method is highly efficient coupled with high accuracy.

•Compared with edge detection method, our method has a higher robustness, regardless of low resolution and complex background.

•Compared with solo morphological expansion method, the text region detection based on our method is not only more accurate, but also more practical for the future process.

The detection and location of caption area

The caption location method based on FAST corner detection algorithm consists of three steps:

1) Initial Image Process. We use FAST corner detection algorithm to detect the corners of the image and mark the relevant corners, then threshold processing of the image and get the binary image which have corners information.

2) Targeting Accurately. We use the horizontal integral projection method towards binary image to get the vertical coordination, which is used as the basis to get the horizontal coordination through applying connected region analysis. The text region (W1) is defined by the vertical and horizontal coordination above.

3) In the caption area W1, we use heuristic knowledge to tell the text region from non-text region and mark the text region as W2 by drawing a blue rectangle, which is considered as the result of caption location (shown in Figure 1).

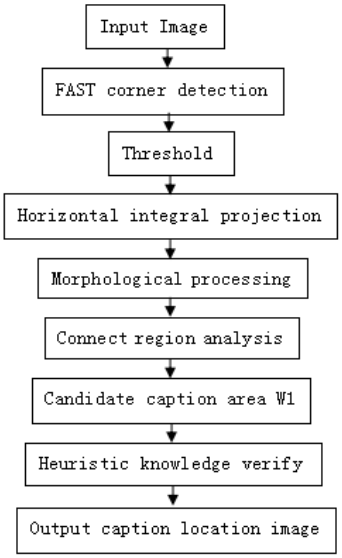


Fig. 1 Flow chart of the text location algorithm

A. Corner detection

The corner is an important local feature in the image which always conveys extensive information as it has high curvature in two-dimensional region edge.

Traditional caption detection method based on the corner detection is mostly adopted by Harris corner detection algorithm. As the horizontal and vertical gradient calculation is time-consuming and complex, Harris detection method is considered as inefficient even though corner information detected by Harris is extensive. Aimed at compensating such shortage, this paper determines to use FAST corner detection algorithm instead.

Before the image corner detection, it is necessary to convert original image into gray image using gray-scale transformation, the transformation rule is shown in equation 1.

$$Y(x,y) = 0.299R + 0.587G + 0.114B \quad (1)$$

Based on the gray image above, we use median filtering method (equation 2) to implement noise reduction.

$$g(x,y) = \underset{W}{Med}\{f(x,y)\} = Med\{f(x+m,y+n) \mid W_{nm} = 1, (x,y) \in I^2\} \quad (2)$$

f(x,y) is the gray scale value of the input image. W is the filter window. G(x,y) is the output image.

In the FAST corner detection algorithm, the corner is defined as a point where a number of its surrounding pixel points are not in the same region with it. The corner response of FAST corner detection algorithm is described in equation 3.

$$N = \sum_{x \in circles(p)} |I(x) - I(p)| < e_d \quad (3)$$

The FAST corner detection algorithm is applied to the gray level image after filtering. We scan the entire image pixel points, find the point which there are a large number of its surrounding points that the gray scale value are either bigger or smaller than it. If the number exceeds the threshold ,the point is the corner and make a mark. In this paper ,threshold is 12. The schematic diagram of circular template of a certain point is illustrated in Figure 2.

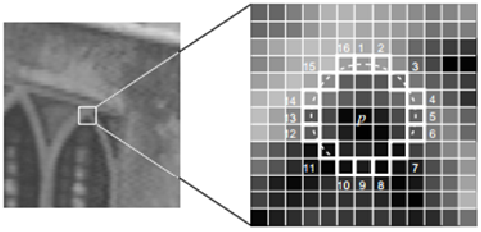


Fig. 2 Schematic diagram of circular template

As Chinese and English characters always come with numerous strokes and cross, the existence of gray scale difference is more common between a stroke pixel and its surrounding background, which gives FAST corner detection method a better performance. Moreover, FAST corner detection algorithm has a less calculation intensity and a higher efficiency compared with Harris corner detection algorithm since it only focus on the gray scale differences of pixel points . Figure 3 is used as a demonstration, where 3a is the original images and 3b is the FAST detection image. Compared with 3a, the caption is isolated from its background in 3b. More processing steps including noise reduction will be introduced in the future.

(a) (b)

Fig. 3 Original images and Corner detection image

B. Accurate Location

Due to the extensive pixel information in the corner image after threshold processing, the caption zone can be instantly positioned via integral projection transform. The integral projection method describes the feature of image's projection on a range of given direction, mainly including horizontal projection H(x) and vertical projection V(y). Given that G(x,y) is the gray scale value of the image (x,y), the horizontal integral projection H(x) of the image region[y1,y2] and the vertical integral projection V(y) of the [x1,x2] region are described in equation4,5:

$$H(x)=\frac{1}{x_2-x_1}\sum_{x=1}^{x_2}G(x,y) \tag{4}$$

$$V(y)=\frac{1}{y_2-y_1}\sum_{y=1}^{y_2}G(x,y) \tag{5}$$

The corner detection image and horizontal integration projection image are shown in Figure 4.

(a) (b)

Fig.4 Corner detection image and horizontal integration projection image

From the integral projection image, it is noted that the caption line has a considerable number of pixel information. Aimed at extracting the coordinates of specific information effectively, interpolation method is introduced to conduct smoothing toward integral projection image.

$$temp=\frac{1}{step}\sum_{i=0-\frac{step}{2}}^{\frac{step}{2}}f(x+i) \tag{6}$$

In equation 6, X is the input value of a certain point in the integral projection while temp is the value after smoothing. It is well studied the smoothing effect is the best when the step size is set to 5.

The horizontal coordinates of potential information, described as (y1,y2…), is tracked by conducting position of the wave trough in the integral projection after smoothing . As to the vertical coordinates, someone used to apply the vertical integral projection of the image to conduct location (according to the literature [7]) . It is necessary to point out that, however, such application has an obvious shortage, which is that parts of the caption section can be missed when there are more than one section existing. Based on such consideration, the segmentation of the vertical projection built on former processed horizontal projection is used (according to the literature [8]). Even though such solution is effective in many cases, nevertheless, text split phenomenon might occur when the text has few strokes (shown in Figure 5).

Fig. 5 Split simple

In order to avoid such mistake, this paper determines to use the connected region analysis method to track the vertical coordinates of potential information. First of all, the local information after horizontal location is processed through morphological method and thresholding sequentially and then becomes a binary image with a number of connected regions. Secondly, the horizontal coordinate of every connected region got from connected region analyzed method is considered as the horizontal coordinate, described as (x1, x2), of the given caption zone. Compared with full morphological analysis, the morphological analysis toward local image avoids both the caption adhesion and split phenomenon (shown in Figure 6). Combining the horizontal coordinate with its related vertical coordinate, the potential caption zone, described as R(r1, r2, r3), is defined . It is reasonable to claim that the methodology mentioned above works smoothly for the location of multi-caption zones and guarantees the completeness and the accuracy simultaneously.

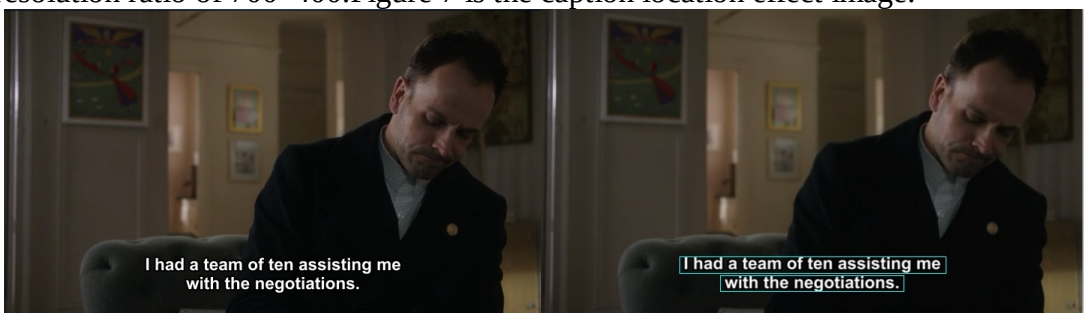
Fig. 6 Split repair simple

C. Heuristic Knowledge

It is necessary to identify and keep the corrected caption zone while remove the non-caption zone from the potential caption zone. In order to identify the caption zone correctly, various characteristics of text including text size, arrangement, color, corner density are taken into account. More specifically, the length-width ratio of texts is always in a reasonable range and the arrangement of caption is either in horizontal or vertical dimension. In addition, contrast ratio of caption and its background is generally high, and the corner information of text always appears in cluster. Hence, the caption zone can be accurately and correctly located by referencing the characteristics above.

Experimental results and discussion

The methodology proposed in this paper has been proved to be effective in video and images, including television, movies, and news clips, which contain both Chinese and English captions with multi lines and a resolution ratio of 700*400.Figure 7 is the caption location effect image.



(a) (b)

Fig. 7 Caption location image

Three benchmarks, namely recall, accuracy and time, are introduced to assess the methods for detecting caption. Recall refers to the ratio of number of correct regions to number of total caption regions. Accuracy refers to the ratio of number of correct regions to total of detected regions. Time refers to the time used to deal with an image.

Method	**Recall**	**Accuracy**	**Time**
Method of literature [6]	90.59%	90.76%	45ms
Method of literature [7]	91.64%	92.22%	36ms
Method of this paper	91.88%	95.69%	32ms

Table 1 experimental results

The experimental results are listed in the table 1, showing that the FAST corner detection algorithm used in this paper has a significant improvement in terms of speed compared with the Harris corner detection algorithm mentioned in literature [6]. Given that the vertical coordinate location method in literature [7] is vertical projection, a method which is generally considered to have a high processing speed, the value of speeds in both methods is therefore similar (32ms VS 36ms) . It is noted that, however, the accuracy of vertical projection is negative due to the inappropriate caption zoning mentioned before.

The benefits of the method in this paper are concluded below. First of all, the corner feature of text has a relatively lower noise compared with other text features, which makes the method significantly effective regarding caption location. Secondly, FAST based corner detection has a higher efficiency

and a lower calculating intensity compared with Harris method. Last but not least, the positioning accuracy of connected region analysis combined with integral projection is of course higher than that of solely connected region analysis. It is noted that all the experimental tests are under VS2010 and OPENCV programming environment.

Conclusions

This paper provides a caption location method based on FAST corner detection algorithm, consisting of following steps: Firstly, corner information of an image is detected and tracked by FAST corner detecting algorithm. Secondly, the image is under thresholding process and then transformed into a corner map based on the former corner information. Thirdly, the horizontal and vertical coordinates of captions are calculated by horizontal integral projection and connected region analysis, respectively. Lastly, heuristic knowledge is introduced to conduct detecting verification. Our method in terms of caption location and verification has been strongly proved by extensive actual tests.

Acknowledgements

This work was supported by Fund as the Key Project of Science and Technology Plan of Beijing "The Research of Intelligent Retrieval and Data Mining Technology Oriented Video Promotion Service"(D16110100520000).

References

[1] Kim K C, Byun H R, Song Y J, et al. Scene text extraction in natural scene images using hierarchical feature combining and verification[C]. In: Proc of the 17th International Conference on Pattern Recognition. Cambridge: IEEE CS, 2004, 679-682.

[2] Srivastav A, Kumar J. Text Detection in Scene Images using Stroke Width and Nearest neighbor Constraints [C]. 2008 IEEE Region 10 Conference on TENCON, 2008, 1-5.

[3] Shivakumara P,Trung Quy Phan,Chew Lim Tan.Video Text Detection based on Filters and Edge Features[C]. IEEE Ineternational Conference on Multimedia and Expo,2009:514-517.

[4] Ye J, Huang L L, Hao X L. Neural Network Based Text Detection in Videos Using Local Binary Patterns. In: Proc of Chinese Conference on Pattern Recognition. Nanjing: Beijing Section, 2009, 1-5.

[5] Hanif S M, Prevost L. Text Detection and Localization in Complex Scene Images Using Constrained AdaBoost Algorithm[C]. In: Proc of the International Conference on Document Analysis and Recognition Catalonia, Spain,2009, 1:1-5.

[6] Zhao X, Lin K H, Fu Y, et a1. Text from comers: a Hovel approach to detect text and caption in videos[J]. IEEE Trans. Image Processing, 2011, 20(3): 790-799.

[7] Shi Jian-yong, Luo Xi-ling, Zhang Jun. An Edge-based Approach for Video Text Extraction[C]. 2009 International Conference on Computer Technology and Development, 2009: 331-335.

[8] Liao Jia, Wang Yun-fei, Wang Hong-mei. A Simple and Fast Text Location Method in Color Images[J].Computer Knowledge and Technology, 2010(6), 8075-8