CORRELATION BASED CAR NUMBER PLATE EXTRACTION SYSTEM

[1]PHYO THET KHIN, [2]LAI LAI WIN KYI

[1,2]Department of Information Technology, Mandalay Technological University The Republic of the Union of Myanmar
E-mail: [1]phyothetkhin1986@gmail.com, [2]laelae83@gmail.com

Abstract- Car License Plate Recognition (CLPR) is a key research area of computer vision and image processing. And it can be used in various application areas such as security control of restricted areas, traffic law enforcements, surveillance systems, toll collection and parking management systems. In order to recognize the license plate efficiently, location and extraction is the most important part. It determines the accuracy of the whole process. The objective of this work is to detect and extract car number plate area from video clip. So, the input key frames are first generated from the video clip by using key frame generation algorithm. Science this algorithm can reduce the number of frames, the system reduce the processing time. The background region of the plate (Black) and text (White) are usually assigned with opposite colors. By using these connected features of text, this text area is detected and extracted from background with the help of Connected Component Correlation method. The proposed system is tested on 670 samples video clips. These video clips are collected from the internet and public high way lane in Myanmar. The experiments are carried out by using MATLAB programming language. This system can be easily adopted to the system requiring character recognition.

Index Terms- connected component, DWT, horizontal and vertical projection, text extraction, video clip.

I. INTRODUCTION

Automatic text detection in video is an important task for efficient and accurate indexing and retrieval of multimedia data such as events identification, events boundary identification etc. Among them, Automatic Number Plate Recognition (ANPR) has gained much popularity in recent years. It is a technique aimed at the detection and extraction of license plate using image processing technique. The need of more secure, automated security system, toll plaza, over speed detection systems have greatly evaluated the need for an automatic system which can detect the vehicles. Connected components are becoming increasingly popular in image processing techniques. These are essentially a group of pixels in contact with one another. Most of the previous approaches, the detecting and extracting 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. Clearly, the detection of text is a challenging task due to varying lighting, complex movement and transformation.

Text detection and extraction in video is usually addressed by the following methods:

1) Method based on edge extraction: it can quickly locate the text area; there is relatively high accuracy if video frame contains strong edge information.

2) Method based on texture: it is more versatile than the other methods, but it usually performs Fast Fourier Transform (FFT), wavelet transform, so it must be time-consuming, poor in real-time performance.

3) Method based on time-domain characteristics: it use the appearance and disappearance of video caption text to detect text area because the appearance and disappearance of video caption text can cause the change of the gray value in text area, but no change in the non-text area. In this paper, a method based on edge extraction, namely connected component correlation, is proposed. There are two main steps in this system. The first step is background subtraction which can be distinguished between texts and complex background (scenes).

The second step is text (number plate region) extraction. In this step, characters are segmented as a single line. And they are grouped as a tight bounding box around all text instances. Then the bounding box is cropped to verify and recognize the result images. The rest of this paper is organized as follows. Section II reviews the related work. Section III describes the key frame generation. It also discusses the number plate region detection and extraction using connected component correlation method. The experimental results are shown in Section IV. Section V concludes the research.

II. RELATED WORKS

Most of the published methods for text detection can be categorized as either texture-based or component-based. Texture-based approaches, such as the salient point detection and the wavelet transform, have also been used to detect the text regions. Bertini et al. [1] detect corner points from the video scene and then detect the text region using similarity of corner points between frames. Zhong et al.[2] detect text in JPEG/MPEG compressed domain using texture features from DCT coefficients. They first detect blocks of high horizontal spatial intensity variation as text candidates, and then refine these candidates into regions by spatial constraints. The potential caption text regions are verified by the vertical spectrum energy. But its robustness in complex background

may not be satisfying for the limitation of spatial domain features. After the text detection step, the text extraction step should be employed. The text extraction methods can be classified into color-based [3] and stroke based methods [4], since color of text is generally different from that of background, text strings can be extracted by thresholding. Otsu [3] 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. On the other hand, some filters based on the direction of strokes have also been used to extract text in the stroke-based methods. The four-direction character extraction filters [4] are used to enhance the stroke like shapes and to suppress others. However, since the stroke filter is language-dependent, some characters without obvious stripe shape can also be suppressed. For component-based methods, text regions are detected by analyzing the geometrical arrangement of edges or homogeneous color/grayscale components that belong to characters. For example, Smith and Kanade [5] located text as horizontal structures of clustered sharp edges. Similarly, Lienhard [6] identified text as connected components which are of same color and have corresponding matching components in successive video frames. Jain and Yu [7] decomposed the video frames into sub images of different colors and then examined if each sub image contains text components which satisfy some specified heuristics.

Edge-based approaches are also considered useful for video text detection since text regions contain rich edge information. The commonly adopted method is to apply an edge detector to the video frame and then identify regions with high edge density and strength. This method performs well if there is no complex background and it becomes less reliable as the scene contains more edges in the background. Lyu et al. [8] 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 generality.

III. PROPOSED METHOD

The proposed method is the extraction of texts which are car number plate from traffic video clips. A video clip is composed of a large numbers of video frames. Each video frame is actually an image. Furthermore, there are at least thirty video frames for a second of video scene. Thus, within a video scene, consecutive video frames are similar image. For the consideration of processing speed, all video frames included in a particular video clip are not need to put to the proposed method. There are two steps in this method such as key frame generation and connected component based text region extraction. Hence, the first step of the proposed method is to select which video frame should be processed. For this step, DWT based video key frame generation approach is proposed. There are four sub-bands (LL, LH, HL and HH) in DWT. Among them, LL and LH are used in this approach. For example, there are five video scene changes in a video clip, there are at least five key frames in this video to represent the scenes. Proposed key frame generation algorithm can be described as follow.

A. Key Frame Generation Algorithm

Input : Video V, contains N frames

Output : Key frame for input video

Algorithm of Key Frame Extraction

Begin

Read Video V;

Step 1

For loop of each video frame k = 1 to N

1. Read video frame Vk and Vk+1

2. Eliminate RGB frame to Gray image

Gk = gray image of Vk and Gk+1 = gray image of Vk+1

3. Gray image is transformed by level 1 of DWT to obtain four channel sub bands

cL1 = LL band of Gk , cV1 = LH band of Gk, and cL2 = LL band of Gk+1, cV2 = LH band of Gk+1

4. Find difference total of each band

$$D1(k)= \sum_{i=1}^{m}\sum_{j=1}^{n}\left(cL2(i,j)-cL1(i,j)\right),$$

$$D2(k) = \sum_{i=1}^{m}\sum_{i=1}^{n}\left(cV2(i,j)-cV1(i,j)\right)$$

Where, m and n are numbers of row and column. D1andD2are difference of cL and cV bands. End of for loop.

Step 2

Compute Mean & Standard Deviation

1. Mean (M) calculation

$$M1 = \frac{\sum_{i=0}^{N-1}(D1(i))}{N-1} \quad \text{and} \quad M2 = \frac{\sum_{i=0}^{N-1}(D2(i))}{N-1}$$

2. Standard Deviation (STD)

$$STD1 = \sqrt{\frac{\sum_{i=0}^{N-1}(D1(i)-M1)^2}{N-1}},$$

$$STD2 = \sqrt{\frac{\sum_{i=0}^{N-1}(D2(i)-M2)^2}{N-1}}$$

Step 3

Estimation of Threshold

Th1 = M1 + □STD1, Th2 = M2 + □STD2

Where, □ is a constant to tune numbers of key frame.

Step 4

For k =1 to N-1

If ((D1(k)>th1 AND D2(k)>Th2)

Write Vk+1 as output key frame

End of for loop

End of algorithm

In the above algorithm, two sub-band matrixes of DWT, LL and LH are used. LL band can give details

of an image and LH channel give edge features. If consecutive video scenes of a video clip are different, the features include in the two sub-bands are also different. Hence, by finding the different features of adjacent video frames, key frame which represent corresponding video scene can be detected. In this algorithm, if the difference feature is larger than a threshold, second video frame of the two can be recorded as key frame. This key frame generating approach is slower than manual in which video frames that are selected in periodically. Here, a key frame is a special image which can represent a video scene. For example, let select 2nd video frame, 7th video frame, 12th video frame and so on which are used in this process. This is very easy and so fast but which have to know the finite video length, time duration of every video clip. It is very difficult to satisfy the constraints. Hence, key frame generation method is suitable for most of video processing.

B. Connected Component Based Text Region Detection and Extraction

To detect and extract the car number plate, the input key frame image is first converted into YUV color space. In this color space, Y channel represent brightness and intensity of an image so the channel is used for later processes. And Canny Edges detector is used to extract edges of the image because the car number is very contrast comparing with its background. By extracting edges, unwanted background objects also have been removed. Unwanted lines and noises are removed by using median filter with larger kernel size such as 7x7 or 9x9. After finishing this steps, edges of the number region can be subtracted from background and ready for counting gradient pairs. In calculating gradient value, negative gradient is found at the entering edge of number region and positive gradient is found at the outgoing edge.

The negative edge gradient and positive edge gradient are defined as a gradient pair. Hence an image, the numbers of gradient pair have to estimate and determine the texts of number plate region contains or not. In this step, the horizontal and vertical projection profiles for the candidate text regions are analyzed. The projection is more efficiency way to find such high density areas. The vertical projection is defined the sum of pixels present in each column of the intensity. Similarly, the horizontal projection is defined the sum of pixels present in each row of the intensity image. The average value of maximum and minimum of the vertical projection is taken as threshold and in the same way, the maximum and minimum of the horizontal projection is taken as threshold. In case of the vertical projection, the rows which exists sum of pixel intensities above the threshold level are taken. Similarly, in case of horizontal projection only the columns which exists sum of pixel intensities above the threshold are taken.

If the numbers of gradient pair is larger than predefined threshold, the image contains text. The threshold value is received from practical tests. And, if the number gradient pairs are accepted, regions in the image are labeled and taken each regional property. In this situation, almost of all consecutive regions are the numbers because all objects are already removed in above process. These consecutive regions are grouped and collected by boundary box. By cropping this boundary box area, the car number plate area can be extracted from its complex background.

According to the above processing steps, this system can easily extracted car number plates from a recorded video clips. The processing time is depending on the length and frame rate of given video clips. The accuracy and step by step results can be seen in next section. The following figure shows the car number plate region detection and extraction steps carried out in the key frames.

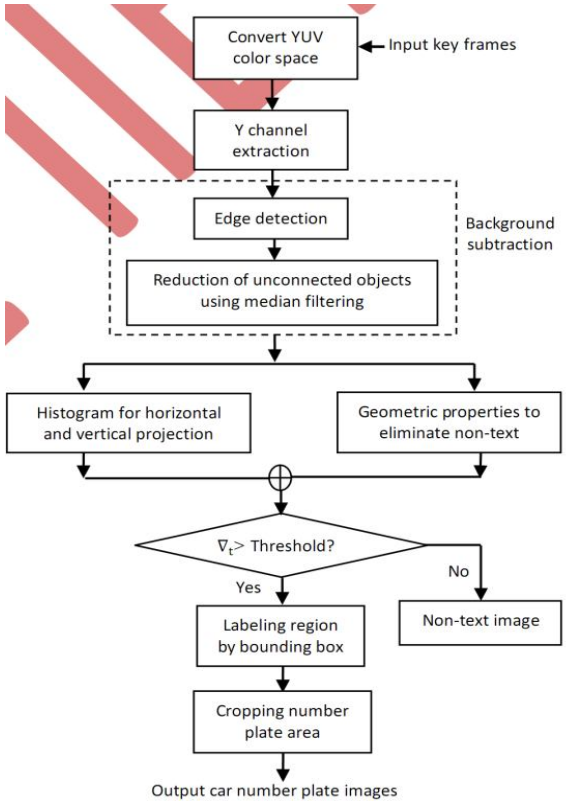

Fig. 1: Block Diagram of Car Number Plate Region Detection and Extraction

Figure 1 illustrates step-by-step procedure for the detection and extraction of car number plate system, after achieving the input key frames. On the other hand, this step has to detect whether texts are include or not in the key frames by finding connected components. If there are no connected components, the key frames will be skipped. If connected components are found, detection is performed again whether the gray values of area which is around the components and the gray values of components are opposite or not. The number plate has opposite color

of background. Hence, if the two gray values are not opposite, the components are skipped.

IV. EXPERIMENTAL RESULTS

The proposed method is implemented by using Matlab2014. In this paper, "use1.avi" and "use2.avi" are presented as samples. The "use1.avi" has more than 500 video frames with frame rate 30. "use2.avi" is 670 frames and also 30 frame rate. As the proposed method has two main sections, key frame generation and number plate region extraction, the experimental results can be described as two results according to this sections. The following results are some key frames of the given two video clips.



Fig. 2: Video scene of the car in day view



Fig. 3: Key frame image of the Car in day view



Fig. 4: Video scene of the car in night view



Fig. 5: Key frame image of the Car in night view

In the number text extraction process, parameters of Canny Edge detector is 0.3 to 0.4 is used. Median filter size is 7x7 and extraction process is performed with the processing steps which is expressed in the above section. The result of the extraction process can be described as follows.



Fig. 6: Number plate extraction results for day and night view

According to the experimental results, the proposed method can achieve precise extraction of number plates in day and night. However, it can be confused with some other closed connected edges which are not number texts. It can be fixed by using larger edge detection parameter. Furthermore, as more than one key frame images are generated for video scenes, the number plate texts can be captured in another key frame. Hence, the proposed method can extract at least two cropped results for a car number text.

CONCLUSION

The proposed system has presented the key frame generation and connected components approach for the extraction of car number plate from a video clip. This system can reduce the key frame calculation time by using the difference features of adjacent video frame instead of calculating the length of all frames in video clip. But the key frame generation approach is slower than manual in which video frames that are selected in periodically. Using geometrical features of the plate region we filter out clutter regions from the candidates. Final extraction of the plate region is performed using the correlation between candidate regions and a plate. As edges of texts are connected objects, the proposed method has achieved outperformed results. This system is

focused on the extraction of car number plate written in English language, not written in Myanmar language. In the future researches, this system will go on the image security area trying to achieve more precise and better results. It can also extend to apply in real time application.

REFERENCES

[1] 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.

[2] Y. Zhong, H.J. Zhang, A.K. Jain, Automatic caption localization in compressed video, IEEE Transactions on PAMI 22 (2000) 385–392.

[3] N. Otsu, "A threshold selection method from gray-levl 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] M. A. Smith and T. Kanade, Video skimming and characterization through the combination of image and language understanding techniques, Technical Report CMU-CS-97-111, Carnegie Mellon University, Feb. 1997.

[6] R. Lienhart, Indexing and retrieval of digital video sequences based on automatic text recognition, Technical Report 6/96, University of Mannheim, 1996.

[7] A.K. Jain and B. Yu, Automatic text location in images and video frames, Pattern Recognition, vol. 31, no. 12, pp 2055-2076, 1998.

[8] 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.

★★★