Frame based Object Detection - An Application for Traffic Monitoring

Chin Hong Low, Ming Kiat Lee, Siak Wang Khor
Faculty of Information and Communication Technology
Universiti Tunku Abdul Rahman
Petaling Jaya, Malaysia
e-mail: chlow1986@hotmail.com, LeonLee0049@hotmail.com, khorsw@utar.edu.my

***Abstract*—In this paper, we describe a system that is capable of detecting and segmenting objects from video frames which helps in traffic surveillance. Shadow is one of the problems faced by most of the object detection systems. It will affect the result of object detection and segmentation. Hence a shadow removal method is applied in the preprocessing phase of the system to amplify the accuracy of detection. MATLAB is the major platform for developing this system. By differentiating the background and foreground of the video scenes using MATLAB embedded functions, the foreground regions are segmented as the objects. The segmented objects are saved for further development of computer vision applications such as object recognition and classification. The proposed system is tested with four traffic video scenes and the experimental results show that the system works well with an accuracy of approximately 90% achieved.**

Keywords-background subtraction; object detection; object segmentation; shadow removal; surveillance; traffic flow monitoring

I. INTRODUCTION

Object detection is an important field of research in computer vision. Before pattern recognition or object classification can be performed, finding the objects of interest from input image or video is vital. Once the objects of interest are detected and segmented, further processing can be done according to system's application. Therefore, as the foundation of many computer vision applications, object detection is crucial in affecting the accuracy of the final output.

Generally, there are two types of object detection methods. The first type is by doing background subtraction to locate the foreground pixels, followed by object segmentation to extract the objects of interest. Usually, image preprocessing techniques are required in this type of detection. The second type of object detection methods is more direct. It does not require much of the image preprocessing techniques. It exploits the relevant features of object in training classifiers to directly detect the objects of interest from input images or videos.

Frame based object detection belongs to the first type. Most of the frame based object detection systems encounter shadow problem during detection. Shadows create problems that make the system difficult to segment individual objects when they are too close to each other. When more than one objects move closely together, their shadow may stack or overlap, causing the objects joined together. The system will detect them as one object and thus lower down the accuracy. Hence, an approach is introduced to counter this problem, which is, using shadow removal in the preprocessing phase. This research focuses on the development of a system that works efficiently on traffic flow monitoring with the help of shadow removal algorithm.

II. RELATED WORKS

Many research works have been done in the field of object detection, hence there are many ways to detect object of interest in a video or image scene. Object detection methods have been classified as point detectors, segmentation, background subtraction and supervised classifiers [1]. For instance, a technique is proposed to detect bicycle and to determine bicycle flow based on video processing [2]. Skin-color is a useful technique to locate and track faces [3]. Skin-color model is introduced to segment the connected blocks. A shadow suppression algorithm [4] based on the HSV color space also used to reduce the connected blocks. Block 1 in Fig. 1 indicates a single person cycling across the detection line. Block 2 is two persons cycling side by side and it is connected by the shadow. Block 3 is two persons passing the detection line one after another. Block 2 and block 3 are each counted as one object passing by, if a normal subtraction method is used in the detection system.

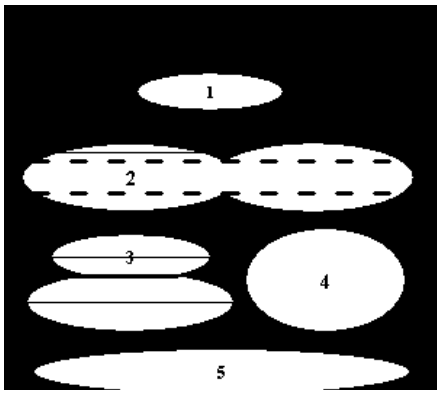

Figure 1. Some common connected blocks in binary image.

978-0-7695-4043-6/10 $26.00 © 2010 IEEE
DOI 10.1109/ICCRD.2010.63

In the proposed system, the skin-color model is used to remove the connected part of a vertical connected block (Block 3 in Fig. 1). It has been connected because the two bicycles are close to each other. The horizontal connected block (Block 2 in Fig. 1) is normally caused by the shadow. Shadow suppression method based on HSV color space is used to remove the shadow and to reduce the connected block.

Background subtraction algorithm is an algorithm to detect object by building a background scene model to find the deviations from the incoming frames of a video sequence. A threshold is used to find the difference between the background and the moving foreground. Several approaches for automatically adapting a background model to dynamic scene variations have been proposed in literature.

Such methods differ mainly in the type of background model used and in the procedure used to update the model. In order to learn gradual changes in time, Wren et al. [5] proposed modeling the color of each pixel of a stationary background with a single 3D (Y, U, and V color space) Gaussian. Once the background model is derived, for every pixel in the input frame, the pixels that deviate from the background model are labeled as foreground pixels.

However, a single Gaussian is not a good model for outdoor scene since multiple colors can be observed at a certain location due to repetitive object motion, shadow or reflectance. A substantial improvement in background modeling is achieved by using multimodal statistical models to describe per-pixel background color, such as using a mixture of Gaussians to model the pixel color [6].

III. METHODS

The method that we proposed in this paper uses frame based object detection to detect the interesting moving objects in a scene. A background image has to be defined at the beginning of the detection before an object can be detected. The shadow of an object will also be detected and removed in this method.

The first step in this method is to convert the video frames to images. The first image frame extracted from the video is stored as the background image. This background image is used to compare with the next sequential frames of the video. The second step is subtracting the background from a particular frame. The result of the background subtraction is an image that only consists of foreground objects. However, this result might be affected by the change of light source, image noise and some other factors.

To get a better result, a threshold should be obtained for the subtracted result. The third step is finding a threshold value to convert the subtracted image to binary image. The threshold value is a midpoint value to segment the foreground and background. As discussed earlier, the result of background subtraction may be affected by some factors like the illumination changes. Hence, instead of an exact value, a range of values are used to differentiate the object and the background. The value that is greater than the threshold will be taken as an object, otherwise it is decided to be the background. The result from this step is a binary image where the background pixels are set to 0 (black pixels) and the objects pixels are set to 1 (white pixels).

The result from the previous steps may include the shadows of objects in the foreground segments. The shadow will affect the accuracy of the object detection and segmentation result. Hence in our proposed system, the forth step is to detect and remove the shadows. The method that we used to remove shadow is based on HSV color space. Firstly, the HSV (Hue Saturation Value) values of each pixel from both the background image and also the next frame image are obtained, and then applied into (1), (2) and (3). *I(x, y)* and *B(x, y)* are the image and background pixel values, whereas *v*, *s*, and *h* are the resultant values from (1), (2) and (3) respectively. These values will be compared against several thresholds as indicated in (4), to decide whether the particular pixel in the image is a shadow pixel or a non-shadow pixel. Pixels that are detected as shadow will be removed from the binary image by setting them to 0.

$$v = \frac{I^V(x,y)}{B^V(x,y)} \tag{1}$$

$$s = I^S(x,y) - B^S(x,y) \tag{2}$$

$$h = \left| I^H(x,y) - B^H(x,y) \right| \tag{3}$$

$$I(x,y) = \begin{cases} shadow & if\ (\alpha \le v \le \beta)\ \&\ (s \le \tau_S)\ \&\ (h \le \tau_H) \\ non-shadow & otherwise \end{cases} \tag{4}$$

The resultant binary image now consists of foreground objects with the shadows eliminated, but still it may have some noise in it. In the fifth step, morphological noise filters are applied to remove the noise and the holes that occur in the foreground objects are filled to obtain a complete silhouette. After this, the white blocks on the image which are the objects, are being highlighted.

Step six is to count the number of objects and show the result to the users. By highlighting all the objects, they can be clearly shown to the users.

IV. RESULTS

The proposed detection system is developed and tested using Matlab. A total of four video sequences which consist of traffic scenes, are used in our experiments. These videos are captured using a static digital camera. The image frames are sampled at a rate of 30 fps and with the size of 320 x 240 pixels. Table 1 provides a summary of the detection result obtained in our experiments, and Fig. 2 illustrates the error rate of our system.

TABLE I. DETECTION RESULTS IN 4 VIDEOS

Dataset	Object Count by Our System	Object Count in Real	Error (%) (Difference between System and Real Object Count)
Video 1 With shadow removal Without shadow removal	 60 68	 65	 7% 12%
Video 2 With shadow removal Without shadow removal	 79 75	 85	 7% 11%
Video 3 With shadow removal Without shadow removal	 146 139	 163	 10% 14%
Video 4 With shadow removal Without shadow removal	 43 48	 45	 4% 6%

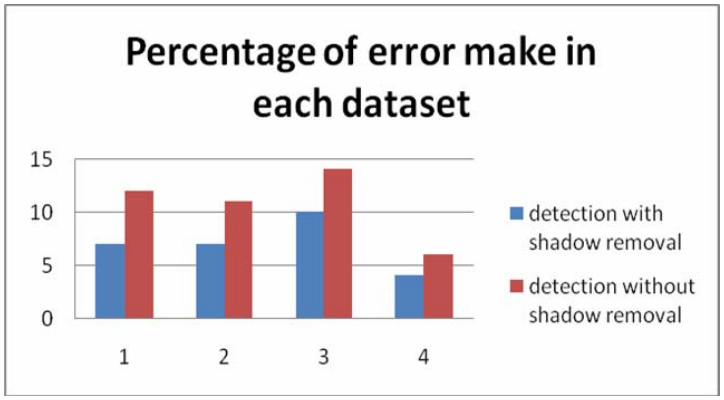

Figure 2. Error rate of our system.

As can be seen from Table 1, for detection with shadow removal, there are less than 10% of detection errors made in all tested videos. For all four datasets, the object counts by our system are less than the real object counts. This is due to due to some of the objects are connected with another object and each of the connected objects are counted as one. As a comparison, the detection without shadow removal shows more error than the detection with shadow removal. This indicates the successful of the shadow algorithm to improve system accuracy.

V. CONCLUSION

In our experiments, approximately 90% of system accuracy has been achieved. With the ability to remove shadows from the foreground objects, the system error rate has been reduced by 4-10%. To get a high accuracy of the application, the angle of the camera is also important. To avoid overlapping regions of vehicles, the camera should be set above the scene with a high angle. Other than that, the lighting condition will also affect the system accuracy, since traffic monitoring systems are usually implemented at outdoor area. There is one limitation of our system. The method that we proposed here can only be used to detect object during day time. If there is too much change of the light source, the application will not be accurate. Besides, a defined background frame is needed for the detection process. So, based on the datasets that we collected, the first frame of the video must be a background that do not consist vehicle or any other moving object. If there are sudden changes of the fixed background, it will cause error in the application. Hence the camera must be mounted at a static and stable position.

The proposed system is tested based on the recorded videos. In the future, it can be implemented in real time applications to improve its practicability. Currently, this system is sensitive to the background changes and illumination changes. Hence, a background update method

can be incorporated in the future to enhance the robustness of the system. Besides, a threshold training method can be explored and implemented in the shadow removal algorithm to help in improving the system accuracy and practicability.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," in ACM Journal of Computing Surveys, 2006, vol. 38, no. 4.

[2] C. Li, L. J. Jia, and H. Tao, "Image Processing Technique to Determine Bicycle Flow," 2003.

[3] T. Darell, G. Gordon, M. Harville, and J. Woodfill, "Intergerated Person Tracking Using Stereo, Color and Pattern Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recongnition, 1998, pp. 610–608.

[4] R.Cucchiara, C.Grana, M.Piccardi, and A. Prati, "Detection Objects, Shadows and Ghosts in Video Streams by Exploiting Color and Motion Information," Proceedings of the IEEE Conference on Image Analysis and Processing, 2001, pp. 360–365.

[5] C. R. Wren, A. J. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real Time Tracking of the Human Body," in IEEE Transaction on Pattern Analysis and Machine Intelligence, July 1997, 19(7), pp. 780-785.

[6] P. Spagnolo, M. Leo, G. Attolico, A. Distante, "A Supervised Approach in Background Modeling for Visual Surveillance," in AVBPA, 2003, pp. 592-599.