Multiple Object Detection in Outdoor Videos: Issues and Methodology

P.P.Halkarnikar, Associate Prof.
Dept. of Comp. Sc. Engg.
D.Y.Patil College of Engg.,Kolhapur.

Dr. S.N.Talbar, Professor,
Dept. of Electronics,
S.S.G.G.S. College of Engg., Nanded

***Abstract*— In many computer-vision applications identifying moving objects from a video sequence is a fundamental and critical task. Applications such as surveillance, auto navigation and robotics demand detection of multiple moving objects from the frames captured by video devices. While identifying the objects it is essential to identify the objects individually. Accurate detection of Individual object is very critical and important in many application areas of surveillance, military equipments and medical sciences. The object separation is difficult because of clutter of the objects in the given image frame. Cluttering of the objects is due to overlapping of objects or their cast shadows. In videos when moving object is to be tracked it is observed that cast shadow of the object creates clutter which disturbs the individual object detection in multiple object detection system. Shadows appear in the detection stage as moving objects itself because of pixel variations of shadow along with object. In this paper we have discussed the effect of object's cast shadow in object merging in multiple object detection system. Outdoor videos are selected as cast shadow has prominent effect on object detection. For detection of moving object background subtraction technique is used. Gaussian mixture model is used to separates foreground and background pixel per frame. This method updates the changes in foreground pixel per frame. The separated foreground pixels form the mask. In order to reduce the full frame scanning for shadow pixel only masked pixels are used. The colour intensity of pixels is done to find shadow pixels. Probable Shadow pixels are removed from frame and objects are detected individually. Such a technique improves the object detection success rate considerably.**

Keywords -- object localization, multiple object detection, object merging, Gaussian mixture model, shadow detection.

I. INTRODUCTION

Object detection in any video system start with localization of object in given frames. The object location may be at one place or multiple places in same frame. Once target object is localized then it is passed to recognition system for recognition of system with known objects [1]. This object localization may get affected by many conditions. Such as scaled objects, rotation, occlusion by other objects, different illumination condition, clutter of objects either by proximity or cast shadow and environmental noise. The moving object detection and tracing is important in many applications like surveillance, medical diagnosis, monitoring ect. It is essential to locate region of interest (ROI) for any object detection system. Many moving object detection systems uses pixel based analysis for localization of object. They try to discriminate foreground and background pixels. Foreground pixels are analyzed for object of interest. While localizing such object cast shadows create unfavorable result by merging the pixels. These blobs of pixels merged are treated as one object. Many techniques exist for object detection and shadow detection. Optimum performance can be achieved by properly selecting object detection and shadow detection algorithms. In this paper we have demonstrated the effect of cast shadow to create clutter situation in videos, we have selected Gaussian mixture model for fast detection of foreground pixels. These pixels are analyzed by statistical methods for shadow removal; such pixels are passed to form the object blobs, which are tested for object of interest. To further reduce the processing time foreground mask is used while passing the pixels to shadow detection. Colour information is used for fast detection of shadow pixels out of detected foreground pixels. Removal of shadow pixels improves the multiple object detection system.

II. RELATED WORK

Now a day's Computer vision systems are commonly employed in many applications. Surveillance system, traffic monitoring, robotics are some of the common areas where computer vision systems are employed for automation. Detection of moving objects is fundamental and crucial task in these systems. Much of the research work is carried out for correct detection and tracking of such objects. Most of the task is carried under controlled environmental conditions. Generally background pixel and foreground pixels are separated for object detection. Background subtraction or motion detection techniques are employed for object detection. But these systems detect moving object along with their shadow as object. [2]. Many surveillance systems employ human detection system. In these systems shape of the object plays important role for distinguishing human and non human objects. [3]. Human detection may get failed due to closeness of human and non human objects. Traffic monitoring systems are also based on moving object detection. The environment for such systems are outdoor hence conditions are not under strict control. Illumination changes affect the performance of the system. Changing shadow, cast by objects detoriate the performance of traffic monitoring system. [4]. Due to cast shadow of moving object shape get changed. This shadow also merges the object which gives false object detection. To reduce the effect of cast shadow use of colour information instead of only illumination information is necessary. [5] Robust system for moving object detect is still a challenge under various condition in videos.

III. PROPOSED SYSTEM

Our aim is to design the robust multiple object detection system. For such a system design proper selection of moving object detection system and shadow detection system is required. Combination of these two systems gives different performance under different environmental conditions.

1.1 Moving object detection:

Moving object detection system has got two parts. First part localized the object and in second part reorganization by discrimination takes place. Therefore it is essential to exactly locate the object. Localizations of object become difficult due to illumination changes, size of the object, cast shadow merging the objects. In this paper we have attempted to solve the effect of cast shadow to avoid the merging of objects. A common approach to identifying the moving objects from video is background subtraction, where each video frame is compared against a reference or background model.[6] Pixels in the current frame that deviate significantly from the background are considered to be moving object. The background model learns only the stationary parts of the scene and ignores the moving foreground. Our system uses the Gaussian Mixture Model for modeling the background adaptively. Gaussian Mixture Model is popular technique for background detection where reference background is not available. [7].This model discriminate foreground and background pixels under dynamic changes in video scenes. Hence the motion regions are identified in the frame, which constitute the regions of interest (ROI). The probability of observing the current pixel value is considered given by the following formula:

$$P(X_t) = \sum_{i=0}^{k} \omega_{i,t}\, \eta\left(x_t, \mu_t, \sum_{i,t}\right) \quad \ldots (1)$$

Where the parameters are K is the number of distributions, $\omega_{i,t}$ is a weight associated to the i^{th} Gaussian at time t with mean $\mu_{i,t}$ and standard deviation $\sigma i, t$. η is a Gaussian probability density function:

$$\eta\left(X_t, \mu, \sum\right) = \frac{1}{\left((2\pi)^{\frac{n}{2}} \left| \Sigma \right|^{\frac{1}{2}}\right)} e^{-\frac{1}{2}(X_t - \mu) \sum^{-1} (X_t - \mu)} \quad \ldots\ldots\ldots (2)$$

Each pixel is characterized by a mixture of K Gaussians. Once the background model is defined, the different parameters of the mixture of Gaussians must be initialized. The parameters of the MOG's model are the number of Gaussians K, the weight $\omega_{i,t}$ associated to the i^{th} Gaussian at time t, the mean $\mu_{i,t}$ and the covariance matrix $\sum_{i,t}$. K determined the multimodality of the background and by the available memory and computational power. Stauffer and Grimson (1999) proposed to set K from 3 to 5. Once the parameters initialization is made, a first foreground detection can be made and then the parameters are updated. Criterion used is the ratio $r_j = \omega_j / \sigma j$ and ordered the K Gaussians following this ratio. This ordering supposes that a background pixel corresponds to a high weight with a weak variance due to the fact that the background is more present than moving objects and that its value is practically constant. The first B Gaussian distributions which exceed certain threshold T are retained for a background distribution:

$$B = argmin_b \left(\sum_{i=1}^{b} \omega_{i,t} > T\right) \quad \ldots\ldots (3)$$

The other distributions are considered to represent a foreground distribution. When new frame comes then match test is carried out for each pixel. If match is found with one of the K Gaussians, pixel is identified as a background one else the pixel is classified as foreground. If no match is found with any of the K Gaussians the pixel is classified as foreground. At this step, a binary mask is obtained. To make the next foreground detection, the parameters must be updated. The parameters are maintained per frame and foreground pixels are detected.[6].

These foreground pixels forms the blobs. The blobs less than the threshold are rejected as small objects or noisy blobs. The remaining blobs are detected as objects. Theses blobs get merged and detected as single large object due to shadow or overlapping of objects. Overlapping of objects are difficult to remove as both objects fall in region of interest (ROI).

1.2 Shadow detection:

Shadow pixels can be detected and removed by different shadow detection techniques in order to solve the problem of object clutter. The foreground pixels detected by Gaussian mixture model forms the mask for object detection. In order to reduce processing time pixels within the mask are tested for shadow. The algorithm analyzes the foreground pixels and detects those that have similar chromaticity but lower brightness to the corresponding region when it is directly illuminated. The adaptive background reference image provides the needed information. Chromaticity can be represented using the standard RGB colour space or some of its transformations such as Lu *v*, La*b* or HSV. In RGB colour space, chromaticity is represented by two values r and g:

$$r = \frac{r}{r+g+b}, \quad g = \frac{g}{r+g+b} \qquad \ldots. (4)$$

Where r,g,b are the intensity level of red, green, blue in RGB colour space. Because these values are independent, a small chromaticity change provokes a small change of r or g or both of them. Considering these small changes shadow pixels can identified. The pixel is said to be shadow if following criteria is met

$$T_1 < \frac{BG}{r_i} < T_2 \qquad \ldots\ldots\ldots (5)$$

$$T_3 < \frac{BG}{g_i} < T_4 \qquad \ldots\ldots\ldots (6)$$

$$T_5 < \frac{BG}{b_i} < T_6 \qquad \ldots\ldots\ldots (7)$$

$$abs_r\left(\frac{FR_i}{r_i+g_i+b_i} - \frac{BG_r}{BG_r+BG_g+BG_b}\right) < T_7 \qquad ..(8)$$

$$abs_g\left(\frac{FR_i}{r_i+g_i+b_i} - \frac{BG_g}{BG_r+BG_g+BG_b}\right) < T_8 \qquad ..(9)$$

$$abs_b\left(\frac{FR_i}{r_i+g_i+b_i} - \frac{BG_b}{BG_r+BG_g+BG_b}\right) < T_9 \qquad ..(10)$$

Where FR = foreground values, BG= values in reference frame, i is current frame values. The threshold values have been defined as T_1=4, T_2=1.5, T_3=2.8, T_4=1.3, T_5=2.05, T_6=1.14, T_7=0.129, T_8=0.028 and T_9=0.145. If all these conditions are mate then the corresponding pixel is marked as shadow pixel. These shadow pixels detected are removed from the current frame and again pixel blobs are formed. These blobs show isolated objects. Such individual objects can be tracked independently for different applications.

IV. RESULTS & DISCUSSION

The algorithm and steps discussed above are implemented using MATLAB and tested on standard traffic video available. Main objective is concentrate on frames with object merge due to cast shadow. These video are outdoor videos containing cars with object merging in many frames due to shadow. Outdoor videos are selected as cast shadow has got significant effect in such videos. Profiles of the videos used are as in table 1

Table 1. Video profile.

Video Name	Total object merge	Successful detection by proposed system	failed detection in proposed system
Highway_1	40%	73%	27%
Traffic	34%	80%	20%

The results are as shown below before and after the algorithm steps. Group images A are original shadow clutter. Group images B are Gaussian mixture model mask. Group images C are shadow pixels detected. Group images D are multiple objects detected by proposed system.

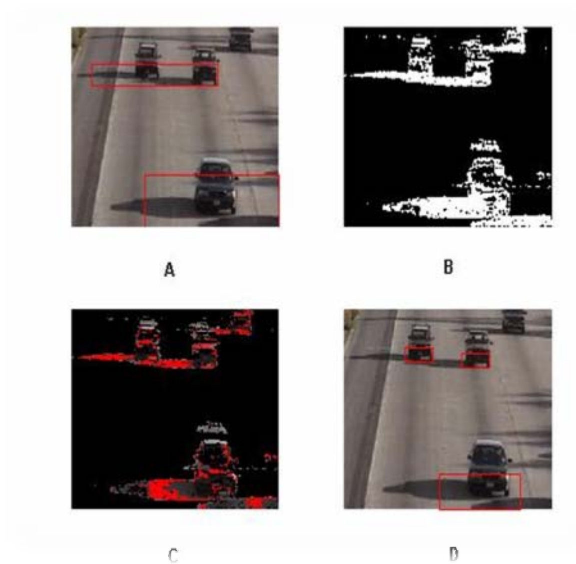

Fig. 1. Output for Frame # 74 A) showing clutter effect B) foreground pixel separation using GMM C) shadow pixels D) clear segmentation of objects.

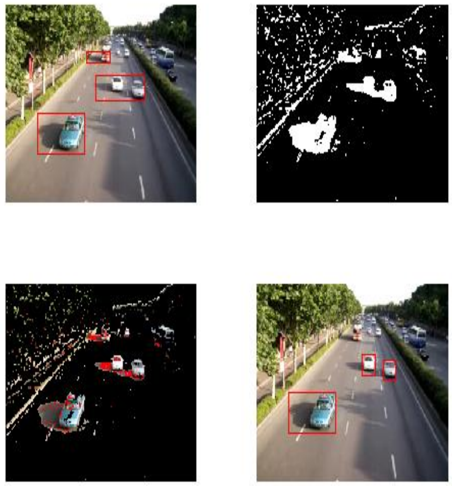

Fig. 2. Output for Frame # 128 A) showing clutter effect B) foreground pixel separation using GMM C) shadow pixels D) clear segmentation of objects

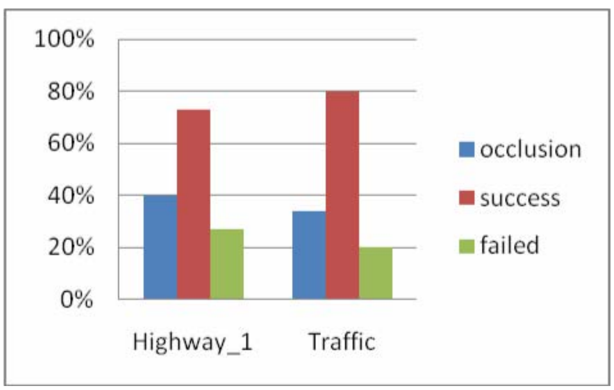

Fig. 3 The graph shows Improvement of object detection in frames where objects are merged due to cast shadow.

Frames with object merged are only considered for finding the success rate of the algorithm. Failures are due to size of the object or cast shadow coverage of one object to another which leads to the situation like object occlusion. If number of objects are more then, foreground pixels detected gives wrong pixel values after application of static shadow detection algorithm which add to failure rate.

V. CONCLUSION

In this paper we have discussed about the problem of object merging found in many object detection algorithms. The object merging may take place due to object overlap or overlap of object cast shadows. This may give unexpected result in the case of application like surveillance, object counting, tracking etc. To increase robustness of such software, it is necessary to remove object clutter due to shadow. Combination of efficient object detection system with fast shadow detection and removal gives better performance. Aim of such combination should be to reduce the processing time for both the stages. Gaussian Mixture Model gives good discrimination between foreground and background pixels. Use of foreground pixel mask for next stage of shadow detection reduces processing time considerably. Above experiment shows better results for removal of object clutter due to shadow. Process pipeline of object detection and shadow detection may further improve the system of multiple object detection.

REFERENCES

[1] Dimitris Tzikas, Aristidis Likas, and Nikolas Galatsanos, *"Large Scale Multikernel RVM for Object Detection",* Proceedings of the 4th Helenic conference on Advances in Artificial Intelligence, Springer-Verlag Berlin, Heidelberg,pp 389-399,2006

[2] Pankaj Kuxrrar, Kuntal Sengupta, Adrian Lee, "*A Comparative Study of Different Color Spaces for Foreground and Shadow Detection for Traffic Monitoring System"*, The IEEE 5fi International Conference on Intelligent Transportation Systems, Singapore, September 2002.

[3] Prithviraj Banerjee and Somnath Sengupta, *"Human Motion Detection and Tracking for Video Surveillance",* National Conference on Communication,IIT Bombay, February 2008.

[4] Sen-Ching S. Cheung and Chandrika Kamath, "*Robust techniques for background subtraction in urban traffic video*", 2007.

[5] C. Wren, A. Azabayejani, T. Darrel, and A. Pentland, " Real-time tracking of the human body," IEEE Transactions on Pattern Analysis and Machine Intelligence 19, pp. 780-785, July 1997.

[6] Marko Heikkila,Matti Pietikainen, *"A Texture-Based Method for Modeling the Background and Detecting Moving Objects"*, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 28, No. 4, April 2006, pp 657-662.

[7] Fida El Baf, Thierry Bouwmans, Bertrand Vachon, *"Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling"*, International Symposium on Visual Computing, ISVC 2008, Las Vegas, 2008.

[8] Habib Ullah, Mohib Ullah, Muhammad Uzair, Fasih ur Rehman, *"Comparative Study: The Evaluation of Shadow Detection Methods",* International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS Vol:10 No:02,2008

AUTHORS PROFILE

Pratap P. Halkarnikar received B.E. from Government College of Engineering, Pune in 1986. M.E. from Walchand College of Engineering, Sangli in 1993. Presently he is working as Associate Professor in Department of Computer Science And Engineering at D.Y. Patil College of Engineering, Kolhapur. He is visiting Professor at P.G**.** center, Department of Technology, Shivaji University, Kolhapur. He is consultant to many industries for development of microcontroller based products. His interest lies in microcontroller based instrumentation, computer vision, data mining and web technology. He is member of ISTE and IEI.

S. N. Talbar has received his B.E and M.E degrees from SGGS Institute of Technology, Nanded, India in 1985 and 1990 respectively. He obtained his PhD. from SRTM University, India in 2000. He is recipient of the Young Scientist Award by URSI, Italy in 2003. He had Collaborative research programme at Cardiff University Wales UK. Currently he is working as Professor in Electronics & Telecommunication Department of SGGS Institute of Technology Nanded, India. He is a member of many prestigious committees in academic field of India. His research interests are Signal & Image processing, Pattern Recognition and Embedded System Design. He has published about 40 papers in Journals and about 105 papers in Conferences.