INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS…..

MULTIPLE OBJECTS TRACKING FOR INTELLIGENT SURVEILLANCE

K.Saranya[1], M.Kalaiselvi Geetha[2] , J.Arunnehru[3]

[1,2,3] Dept.of Computer Science and Engineering, Annamalai University.
[1]saranya.logon@gmail.com, [2]geesiv@gmail.com, [3]arunnehru.aucse@gmail.com

***Abstract**: Surveillance video analysis is an interesting issue in computer vision in which many methods have been actively studied. In video surveillance systems, the recorded or streamed video is used to monitor the designated areas from any suspicious or interested object activities. The moving object tracking is an important issue in video system, such as surveillance sports reporting, video annotation, and traffic management system. Difficulties in tracking objects can arise due to the abrupt object motion, changing appearance patterns of the object and the scene, non-rigid object structures, object-to-object and object-to-scene occlusions. Tracking is usually performed in the context of higher-level applications that require the location or shape of the object in every frame. Also, it can contain regions that correspond to two or more objects, which is called merged detection. These both events split and merged detections make the trajectory estimation and object counting a challenging task. For intelligent surveillance related applications, multiple objects and occlusion are major difficulties. The aim of this paper is to detect the motion and hence to track the movement of multiple objects in a scene for efficient and intelligent surveillance.*

***Keywords:** Video Surveillance, Object Detection, Background Subtraction, Shape tracking.*

1. INTRODUCTION

Visual surveillance has been a very active research topic in the last few years due to its growing importance in security, law enforcement, and military applications. Most of the surveillance cameras are installed in security sensitive areas such as banks, railway stations, highways, and borders. The large amount of data involved makes it infeasible to guarantee vigilant monitoring by human operators for a long period of time due to monotony and fatigue. In order to assist human operators with identification of important events in videos an “intelligent” visual surveillance system can be used.

Intelligent Surveillance system requires fast and robust methods for detecting multiple objects movement and tracking. Multiple object tracking has been a challenging research topic in computer vision. It has to deal with the difficulties existing in single object tracking, such as changing appearances, non-rigid motion, dynamic illumination and occlusion, as well as the problems related to multiple objects tracking including inter object occlusion, multi-object confusion. There has been much work on multiple object visual tracking. This work proposes a background subtraction method to track multiple objects based on object detection. The detection results recognize the tracking targets in each image. This work applies a background subtraction method to detect multiple objects. Multiple objects tracking method is reliable to deal with occlusions, irregular object movements, changing appearances by postponing the decision of object trajectories until sufficient information is accumulated over time.

1.1 Challenges in object tracking:

Major challenge in object tracking is change in shapes of the target object due to the various camera view position and lighting conditions. Other difficulties to track the objects are blurred motion and occlusion. Some algorithms deal with the problems of abrupt appearance change, leaving out from scenes and drifting, etc. The tracking results may be affected due to occlusion in some instance. A dynamic model facilitates reinitialization of a tracker after partial or full occlusions to reduce the search space of states. When a tracking algorithm is designed to account for in-plane motion and scale change with the similarity transform, a static appearance model may be an appropriate option.

1.2 Requirements of object tracking approaches:

To build a robust tracking system, following requirements should be considered. *Robustness:* Robustness means that even under complicated conditions, the tracking algorithms should be able to follow the interested object. The tracking troubles may be cluttered background, partial and full changing illuminations, occlusions or complex object motion. *Adaptively:* In addition to various changes of the environment that an object is located in, the object also undergoes changes. This requires a steady adaptation mechanism of the tracking system to the actual object appearance. *Real-time processing*: A system needs to deal with live video streams must have high processing speed. Thus, a fast and optimized implementation as well as the selection of high performance algorithms is required.

2. LITERATURE REVIEW

Chandrasekhar D. Badgujar, Dipali P. Sapkal [1] says that the computer vision technique, which is used to count peoples and track a particular object. The object should be detected effectively and measure the object position and motion using sensor. The task of target tracking is a key component of video surveillance and monitoring systems. It provides input to high level processing such as recognition, access control or re-identification is used to initialize the analysis and classification of human activities. It contains three key steps for tracking 1) detection of interesting moving objects 2) Tracking of such objects from frame to frame 3) Analysis of tracks to recognize their behavior. Shilpee A. Dave, Dr. M. S. Nagmode, Aditi Jahagirdar [2] says that the current technologies for each stage of the object detection and different algorithms to detect an object. The points, lines or blobs are the features, which is used to detect the object. The other methods to detect objects using shape based model, finding centroid of the target. The first step of the object detection is background subtraction. It subtracts the background image from foreground image. Classifying the object based on the motion and shape. In video surveillance system, the difficult task is to extract the moving object accurately. Sreepathi B, Halaharvi Keerthi [3] says that a method for accurately tracking persons in indoor surveillance video stream obtained from a static camera with difficult scene properties including illumination changes and solves the major occlusion problem. Person tracking is based on moving blobs. A computer vision system is used to monitor the movements. The background subtraction method uses different method of the current image and background image to detect moving objects, with simple algorithm, but very sensitive to the changes in the external environment. To track the object from frame to frame and matching the color information between motion blocks in the current frame and motion blocks in the previous frame. Also find correspondence few feature of previous frame are stored and compared with the next frame. If the features matches then same object found in both frames. M. Vinod, T. Saravanthi, Brahma Reddy [4] says that a new morphological technique for object tracking and counting. To obtain a more accurate people count, the two stage segmentation is developed for extracting each person from a crowd. The object segmentation is done by three approaches 1) frame differencing 2) background subtraction 3) optical flow. In object tracking, four methods are used. They are 1) Region based tracking 2) Active contour based tacking 3) Feature based tracking 4) Model based Tracking. Temporary occlusion can be handled successfully and if the same person re-enter the field of view after a reasonably short period of time, the system can recognize re-entry and assigns the same label to the person. R. O'Malley, E. Jones, M. Glavin [5] says Track Two-Stage object, a method based on Kernal and Active Contour. They can locate an object effectively in complex condition with camera motion, Partial occlusions, clutter etc the diffusion snake is used to evolve the object contour in order to improve the tracking precision. In the first object localization stage, the initial target position is predicted and evaluated by the Kalman filter and the Bhattacharyya coefficient, respectively. In the contour evolution stage, the active contour is evolved on the basis of an object feature image generated with the color information in the initial object region. In the process of the evolution, similarities of the target region are compared to ensure that the object contour evolves in the right way. They have following disadvantages1. This method is time consuming; 2. Method can't effectively track the object when the color feature of the object is very similar to that of the background.

3. PROPOSED WORK

In this work, initially foreground objects are extracted from the background. Next, the foreground object motion between the current and previous frames is obtained. In tracking interested object, shadows affect the performance of tracking and leads to false tagging. To avoid this problem, this work applies Morphological operations to remove noise. In this paper, propose a shape tracking technique to track moving multiple objects accurately and handling the occlusion based on setting unique ID for each object.

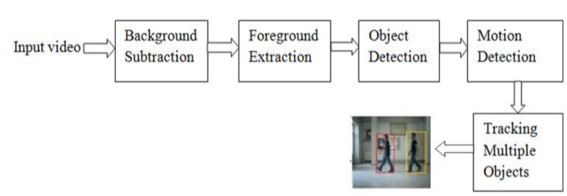


Figure 1. General overview of object detection and tracking

4. OBJECT DETECTION

4.1 Background subtraction:

Moving object detection is always the first step of a typical surveillance system. Moving object detection aims at extracting moving objects that are interested out of a background which are dynamic. Since subsequent processes are greatly dependent on the performance of this stage, it is important that the classified foreground pixels accurately correspond to the moving objects of interests. Background subtraction is widely used for moving object detection especially for cases where the background is relatively static because of its low computational cost. In Background Subtraction, the initial frame is set as a reference frame and it was done by subtracting reference frame from current frame as shown in fig 2. Background Subtraction delineates the foreground from background in the images.

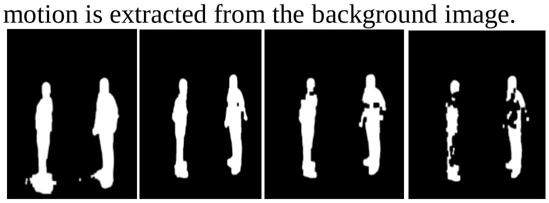
$$R_{k(x,y)} = \begin{cases} 1 & if \left| I_{k(x,y)} - B_{k(x,y)} \right| > T \\ 0 & otherwise \end{cases} \quad (1)$$

Where $I_{k(x,y)}$, represents the incoming video frame, $B_{k(x,y)}$ represents the background frame, k represents the frame number, T is the threshold value.

(a) (b)

Figure 2. Results of Background Subtraction
(a) Reference frame
(b) Background subtracted frame

In simple implementations, the segmentation is determined by a single parameter called as the intensity threshold. In a single pass, each and every pixel in the image is compared with this threshold value. If the pixel's intensity is greater than the threshold value, the pixel is set to, say, white in the output. If it is less than the threshold value, it is set to black. In this Threshold of t = 30 has been used in the experiments. To capture the dynamic information, motion is extracted from the background image.



t=20 t=30 t=50 t=80

Figure 3. Results of various threshold values

At segmentation stage, morphological operations are used. Dilation and erosion are the two morphological operations used for segmenting the object edges here. Mathematical morphology is used for examine the object shape characteristics such as size and connectivity, which are difficult to accessed by linear approaches. The advantages of morphological approaches over linear methods are direct geometric interpretation, simplicity and efficiency in hardware completions. Basic operation of a morphology-based approach is the translation of a structuring element over the image and the erosion and/or dilation of the image content based on the shape of the structuring element.

Dilation:

$$X \oplus S = \bigcup y \epsilon S \, X_y \quad (2)$$

Erosion:

$$X \Theta S = \bigcap y \epsilon S \, X_y \quad (3)$$

In equations (2) and (3), X and S to denote sets in n-dimensional space and y denote position vector in that space. A morphological operation analyzes and manipulates the structure of an image by marking the locations where the structuring element fits. In mathematical morphology, neighborhoods are, therefore, defined by the structuring element, i.e., the shape of the structuring element defines the shape of the neighborhood in the image. For dilation, if any pixel of the rectangle fits at or under the image intensity profile, the center pixel of the rectangle is given the maximum intensity of the pixel and its two neighbors in the original image; otherwise the pixel is set to zero intensity. For erosion, if the whole rectangle fits at or under the image intensity profile, the center pixel is given the minimum intensity of the pixel and its two neighbors in the original image; otherwise the pixel is set to zero intensity. Hence by using this method the motion of an object can be obtained accurately.

4.2 Motion detection:

An effective shape descriptor is a key component of multimedia content description, since shape is a fundamental property of an object. There are two types of shape descriptors: contour-based shape descriptors and region-based shape descriptors (Kim & Sung 2000). Regular moment invariants are one of the most popular and widely used contour-based shape descriptors is a set of derived by Hu (1962) [15]. Two-dimensional moments of a digitally sampled $M \times M$ image that has gray function f (x, y), $(x, y = 0, 1, 2 \ldots M - 1)$ is given as,

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

$$m_{pq} = \sum_{x=0}^{x=M-1} \sum_{y=0}^{y=M-1} x^p \,.\, y^q \, f(x,y) \qquad p,q = 0,1,2,3 \ldots (4)$$

The moments *f (x, y)* translated by an amount *(a, b)*, are defined as,

$$\mu_{pq=} \sum_x \sum_y (x+a)^p . (y+b)^q \; f(x,y) \qquad (5)$$

Thus the central moments μ_{pq} can be computed from (3) on substituting $a = -\bar{x}$ and $b = -\bar{y}$ as,

$$\bar{x} = \frac{m_{10}}{m_{00}} \; and \; \bar{y} = \frac{m_{01}}{m_{00}},$$

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p . (y-\bar{y})^q f(x,y) \qquad (6)$$

When a scaling normalization is applied the central moments change as,

$$\eta_{pq} = \mu_{pq}/\mu_{00}^{\gamma}\,, \; \gamma = [(p+q)/2]+1 \qquad (7)$$

In particular, Hu (1962), defines seven values, computed by normalizing central moments through order three, that are invariant to object scale, position, and orientation. In terms of the central moments, the seven moments are given as,

$$M_1 = (\eta_{20} + \eta_{02}),$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2,$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2,$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2,$$

$$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2],$$

$$M_6 = (\eta_{20} + \eta_{02})[(\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}),$$

$$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] - (\eta_{30} + 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \qquad (8)$$

5. OBJECT TRACKING

Tracking of multiple objects is one of the most complicated issues in video surveillance systems since the behaviors of these objects are unpredictable and cannot be assumed. All the tracking method usually follow a certain trajectory, which means that studying all the objects positions in the previous frame in order to predicate their positions in the next frame. This approach considers the objects to be tracked as blobs. The goal is to determine when an object enters the field of view, compute the correspondence matching between previous foreground regions and regions currently being tracked and estimate the spatial position of each object. The shape based detector does not work well on images of resolution less than 24 pixels wide, as in the case when the object are far from the camera. Taking the motion detection results as input, the approach applies morphological operations like opening and closing to connect foreground pixels to generate motion blobs.

Due to environment factors the image is expected to contain noises. To get proper boundary lines without any noise portion morphological operations like opening and closing on the subtracted image. Opening is a combination of erosion and dilation operations with erosion followed by dilation whereas closing is dilation followed by erosion. For simplicity, this work model moving objects as rectangles. Each object is associated with an appearance model and a dynamic model. At each new frame, the approach predicts the objects position with its dynamic model. Appearance model is used to differentiate the objects when they are merged. A new object is formed when a blob has no match with current hypotheses. A track end up when it has no blob matches for more than a set number of frames. However, if multiple objects are combined in one blob from the beginning to the end, the blob tracker cannot segment them. The moving objects with somewhat small size are classified as pedestrians; the others to be vehicles.

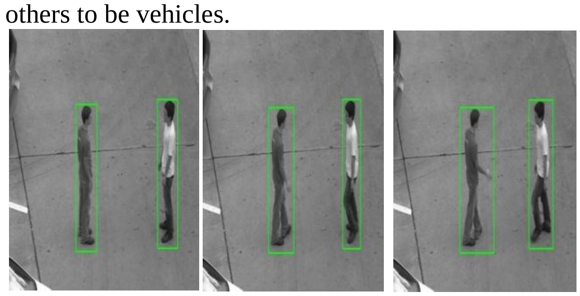


Figure 4. Tracking multiple objects

5.1 Occlusion handling:

Occlusions of moving objects are one of the major problems in any surveillance system. During the moving object detection process, occlusions cause moving objects to be either segmented in an erroneous shape or became completely lost. In this paper occlusion handling is done with the tracking of object. When motion area of the moving object has been detected then its centroid of an object can easily obtained from both images. Each object in our system is assigned a unique ID. The unique ID is dynamically generated based on contour. Contour of each object is calculated based on the height and width of each object. Successful occlusion identification occurs when this ID is correctly maintained before and after occlusion.

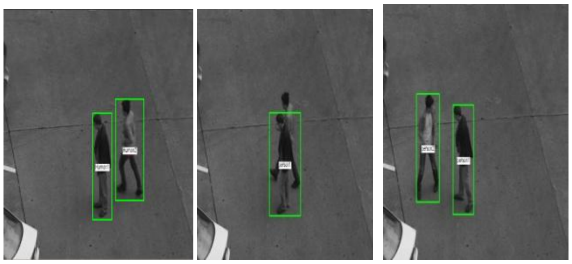


Figure 5. Occlusion Handling

6. PERFORMANCE MEASURE

This paper proposes two novel metrics for the evaluation of multiple objects tracking system. The proposed metrics - Multiple Object Tracking Precision (MOTP) and the Multiple Object Tracking Accuracy (MOTA) are general and intuitive, and allow for objective comparison of the main characteristics of tracking systems, such as their precision in localizing objects, their accuracy in recognizing object configurations and their ability to consistently track objects over time.

1. The Multiple Object Tracking Precision (MOTP).

$$MOTP = \frac{\sum_{i,t} d_{i,t}}{\sum_t e_t} \qquad (9)$$

Where $d_{i,t,}$ e_t are the distance between the object and number of matches found for time t. It is the total position error for matched object-hypothesis pairs over all frames, mean by the total number of matches made. It shows the capability of the tracker to estimate precise object positions, independent of its skill at recognizing object configurations, maintaining consistent trajectories, etc.

2. The Multiple Object Tracking Accuracy (MOTA).

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t n_t} \qquad (10)$$

Where m_t,fp_t,and mme_t are the number of misses, number of false positives and number of mismatches respectively for time t. n_t represents number of objects presents at time t.The MOTA can be seen as composed of 3 error ratios:

$$Number\ of\ misses(m) = \frac{\sum_t m_t}{\sum_t n_t} \qquad (11)$$

the ratio of misses in the sequence, computed over the total number of objects present in all frames,

$$Number\ of\ false\ positive(fp) = \frac{\sum_t fp_t}{\sum_t n_t} \qquad (12)$$

the ratio of false positives, and

$$Number\ of\ missmatches(mm) = \frac{\sum_t mme_t}{\sum_t n_t} \qquad (13)$$

the ratio of mismatches.

Summing up over the different error ratios gives the total error rate E_{tot}, and 1-E_{tot} is the resulting tracking accuracy. The MOTA accounts for all object configuration errors made by the tracker, false positives, misses, mismatches, over all frames. It is similar to metrics widely used in other domains and gives a very intuitive measure of the tracker's performance at keeping accurate trajectories, independent of its precision in estimating object positions. In this implementation evaluates four different types of video sequences which consist of two to four persons in the outdoor, with a total length of 2 minutes.

Table 1. Results for multiple object tracking

Video sequences	Miss rate	False positive	Miss matches	MOTA
Sequence 1	3.3%	2.3%	6.8%	88%
Sequence 2	5.2%	6.4%	7.2%	81.1%
Sequence 3	6.1%	7.1%	11.8%	74.9%
Sequence 4	3.4%	6.1%	10.5%	79.9%

7. CONCLUSION AND FUTURE WORK

In this paper, it describes the framework of the video surveillance system and provides the algorithms and implementation results of the current work on multi-person tracking. It is done by doing background subtraction and extracting the foreground object, using the extracted foreground object the object containing motion alone is detected and tracked. This system works well in real-time situations. It can be used in large number of applications particularly in anti crime systems. It can track multiple persons in the camera's field of view accurately and the performance is higher. In future this paper will be implemented in Particle Filter, Kalman Filter for addressing motion correspondence between objects.

REFERENCES

[1] Chandrasekhar D. Badgujar, Dipali P. Sapkal, "A survey on object detect, track and identify

using video surveillance", IOSRJEN, Vol. 2, Issue 10, PP 71-76, Oct 2012.

[2] Shilpee A. Dave, Dr. M. S. Nagmode, Aditi Jahagirdar, "Statistical survey on object detection and tracking methodologies", International Journal of Scientific & Engineering Research, Vol. 4, Issue 3, Mar 2013.

[3] Sreepathi B, Halaharvi Keerthi, "Person detection and tracking system for visual surveillance", International Journal of Research in Computer Science and Information Technology (IJRCSIT), Feb 2013.

[4] M. Vinod, T. Saravanthi, Brahma Reddy, "An adaptive algorithm for object tracking and counting". International Journal of Engineering and Innovative Technology (IJEIT), Vol. 2, Issue 4, Oct 2012.

[5] R. O'Malley, E. Jones, M. Glavin, "Rear-lamp vehicle detection andtracking in low-exposure color video for night conditions", IEEE Trans. Intell. Transp. Syst., Vol. 11, pp. 453–462, Jun 2010.

[6] L. Oliveira, U. Nunes, P. Peyote, "On exploration of classifier ensemble synergism in pedestrian detection", IEEE Trans. Intell. Transp. Syst., Vol. 11, pp.16–27, Mar 2010.

[7] Kaiqi Huang, Liangsheng Wang, Tieniu Tan, Steve Maybank, "A real time object detection and tracking system for outdoor night surveillance", Elsevier Ltd on behalf of Pattern Recognition society, 2008.

[8] Mahbub Murshed, MD. Hasanul Kabir and Oksam Chae, "Moving object tracking-an edge segment based approach", International Journal of Innovative Computing, Information and Control, Vol. 7, Jul 2011.

[9] Brajesh patel, Neelam Patel, "Motion detection based on multi frame video surveillance system", International Journal of Computer Science and Network Security (IJCSNS), Vol.12, Mar 2012.

[10] G. Raja Raman, M. Sarath Chandran, S. R. Vinotha, "Motion based security alarming system for video surveillance", International Conference on Computational Techniques and Artificial Intelligence(ICCTAI), 2011.

[11] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking", IEEE Trans. Intell. Transp. Syst., vol. 11, pp. 267–276, Jun 2010.

[12] Houari Sabirin and Munchurl Kim, "Moving object detection and tracking using a spatio-temporal graph in h.264/avc bitstreams for video surveillance", IEEE Transactions of Mutimedia, Vol. 14, Jun 2012.

[13] W. L. Khong, W. Y. Kow, H. T. Tan, H. P. Yoong, K. T. K. Teo, "Kalman filtering based object tracking in surveillance video system", Proceeding of the 3rd CUTSE International Conference, Nov 2011.

[14] Carlos R. del-Blanco, Fernando Jaureguizar, Narciso Garcia, "An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications", IEEE Transactions on Consumer Electronics, Vol. 58, Aug 2012.

[15] Muharrem Mercimek, Kayhan Gulez and Tarik Veli Mumcu. Real Object Recognition Using Moment Invariants, December 2005.

[16] Aishwarya P, "Object Recognition in Smart Environment", International Journal for Advance Research in Engineering and Technology, Vol. 1, Issue IV, May 2013.

[17] Tarun Dhar Diwan, Deepa Gupta , Praveen Sahu, "Application of Gesture Recognition in Virtual Touch in Virtual Environments", International Journal for Advance Research in Engineering and Technology, Vol. 1, Issue VII, Aug 2013.