International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 1, Issue 2, December 2011)

Human Tracking in Video Surveillance

Helly Patel[1], Mahesh P. Wankhade[2]

[1] *Department of Computer Engineering, Zeal Education Society's Dnyanganga College of Engineering and Research,Pune, Maharastra, India*
[2] *Department of Computer Engineering, Sinhgad College of Engineering, Pune, Maharashtra, India*
[1] hellypatel_786@yahoo.com
[2] mwankhade@yahoo.com

***Abstract*— In the last decade, due to the increase in terrorist activities and general social problems, providing security to citizens have become the top most priorities for almost all the nations and for the same, a very close watch is required to be kept in the areas that needs security. Keeping human watch 24x7 is not possible as we all know that humans can easily be distracted and a small distraction in very sensitive and highly secure area can lead to big loses. To overcome this human flaw in the area of monitoring, the concept of making monitoring automatic came into existence. Since, video surveillance has came in the market, researches have been taking place in order to make to more easy, accurate, fast and intelligent. The goal of visual surveillance is not only to put cameras in place of human eyes, but also to accomplish the entire surveillance task as automatically as possible. In one statement we can say that video surveillance is nothing but taking the video, identifying unwanted entities, tracking their actions, understanding their actions and raising an alarm. In this paper, we will be study the phases of the video surveillance system. We will see the 3 main methods of human detection. Further, we will see most salient region method for tracking and in this paper we propose a method of handling occlusion using velocity and direction information.**

***Keywords*—Image Separation, Video Surveillance System, Noise Removal, Human Detection, Handling occlusion.**

I. Introduction

Video surveillance has been with us since a long time. In the traditional surveillance system, the video captured by the camera would be simply displayed on a monitor in a control room. Many human resources will be sitting who will be continuously monitoring the video for any abnormal activities. But, due to increase in thefts, terrorist activities and other criminal activities the demand for the video surveillance has increased. Also, with the increase in video surveillance, its sensitivity and accuracy has also increased and any small human error may lead to drastic damage.

We all know that humans have their limitations and we cannot afford any abnormal activity in areas under surveillance. Due to this reason, an effort was made to automate the video surveillance system which will help in accurate monitoring of the sensitive areas with less human resources occupied in the whole process.

Automatic video surveillance is becoming increasingly important in many applications, including traffic control, urban surveillance, home security and healthcare. In this paper we will study the process of automatic video surveillance. Also, we will see a proposed method for human tracking.

There are basically three conventional approaches to moving object detection: temporal differencing, optical flow and background estimation methods. Temporal differencing [2] is very adaptive to dynamic environments, but generally does a poor job of extracting all relevant feature pixels [5]. Optical flow[2], [3], [10] can be used to detect independently moving targets in the presence of camera motion, however most optical flow computation methods are very complex and are inapplicable to real-time algorithms without specialized hardware. Background subtraction is a particularly popular method for human detection especially under those situations with a relatively static background. It attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion. However, it is extremely sensitive to changes of dynamic scenes due to lighting and extraneous events.

Human detection is one of the critical phases of video surveillance system as it is not only responsible for the extraction of moving objects but all the remaining phases process based on the output of this phase. A large number of people detection and tracking algorithms rely on the process of background subtracting, a technique which detects changes from a model of the background scene. Let study different techniques that are used to detect human and let us see why its output is importance for the next phase (human tracking).

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 1, Issue 2, December 2011)

II. Video Surveillance System

Video surveillance process that takes video as an input, processes the video and performs actions accordingly. The process of video surveillance consists of many phases as shown in the figure1.

In section III we will discuss about the image generation phase. Section IV discusses about human detection. Following which sections V, VI and VII discuss about Noise removal, human tracking along with occlusion handling and the activity analysis and triggering alarm respectively.

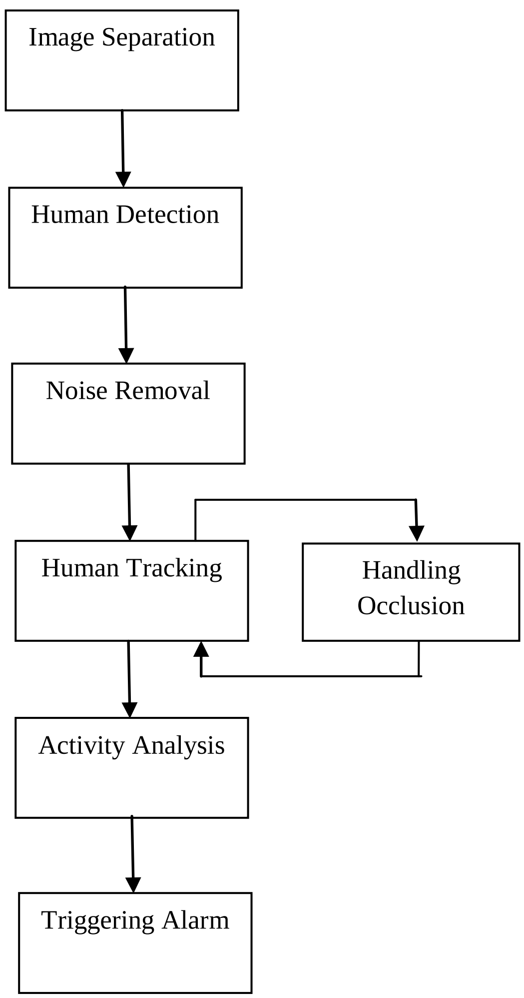

Fig 1. Phases of Video Surveillance system

III. Image Separation

When we talk about processing a video, it is actually not a video that we are dealing with but the video must be divided into sequence of images which are further processed.

The speed at which a video must be divided into images depends on the implementation of individuals. From [1] we can say that, mostly 20-30 images (or frames) are taken per second which are sent to the next phases for further processing.

IV. Human Detection

In this phase, an attempt is made to detect the human is detected in the area under surveillance. There are mainly three ways in which this task can be accomplished 1) Simple Background Subtraction 2) Optical Flow and 3) Temporal Difference.

A. Simple Background Subtraction

The Background subtraction approach is mostly used when the background is static [4]. The principle of this method is to use a model of the background and compare the current image with a reference. In this way the foreground objects present in the scene are detected. It attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion as shown by equation below:

$$| P_c - P_{bk} | > T \qquad (1)$$

where, P_c – Current image pixel,
P_{bk} – Background image pixel,
T – Threshold value.

This method is very useful when there is a static background. Here the threshold is fixed such that the foreground pixels are extracted from the background image. When the pixel difference is above the threshold value, it is considered as foreground image.

B. Optical Flow

Optical flow can be used to segment a moving object from its background provided the velocity of the object is distinguishable from that of the background, and has expected characteristics. Optical flow is the amount of image movement within a given time period [2], [3], [10].

Let us study an optical flow method entitled Lucas-Kanade Method [2], [10]. This method calculated the motion between two image frames which are taken at time t and $t + \delta t$ for every pixel position.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 1, Issue 2, December 2011)

As a pixel at location (*x,y,t*) with intensity *I* (*x,y,t*) will have moved by δ*x*, δ*y* and δ*t* between the two frames, the following image constraint equation can be given:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{2}$$

Assuming that the movement is small enough, the image constraint at *I*(*x, y, t*) with Taylor series can be derived to give:

$$\begin{aligned} &I(x + dx, y + dy, t + dt) = I(x, y, t) \\ &+ \partial I / \partial x * dx + \partial I / \partial y * dy + \partial I / \partial t * dt \end{aligned} \tag{3}$$

From (2) and (3) we get,

$$\partial I / \partial x * dx + \partial I / \partial y * dy + \partial I / \partial t * dt = 0 \tag{4}$$

or

$$\begin{aligned} &\partial I / \partial x * dx / dt + \partial I / \partial y * dy / dt \\ &+ \partial I / \partial t * dt / dt = 0 \end{aligned} \tag{5}$$

Equation (5) can be further written as:-

$$\partial I / \partial x * Vx + \partial I / \partial y * Vy + \partial I / \partial t = 0 \tag{6}$$

Where, *Vx* and *Vy* are the *x* and *y* components of the velocity or optical flow of *I*(*x, y, t*) and ***∂I/∂t*** at a given pixel is just how fast the intensity is changing with time.

C. Temporal Differencing

Unlike, the background subtraction method where the base image was the background image, here in temporal differencing [5] the reference image is the previous images. Hence the previous frame is subtracted from the current image and the subtraction value must be greater than a threshold value in order to give a difference image.

$$\begin{aligned} I_{threshold}(x, y) &= 1, I_{threshold}(x, y) > threshold \\ &= 0, I_{threshold}(x, y) \leq threshold \end{aligned} \tag{7}$$

From the above three human detection methods, the background subtraction is the simplest and mostly used when the background is stationary. In our implementation we will be keeping static background and hence will use the simple background subtraction method of human detection.

V. Noise Removal

The image is expected to contain noises. These noises might be included in the image due to environmental factors (for example, humidity or fog in the area under surveillance), due to illumination changes, during transmission of video from the camera to the processing unit, and similar factors. Noise in a video means there might be a portion that is either added or erased in the obtained images. Hence, due to noise in the acquired image we might get any extra or some lesser image portion.

Such noise has to be handled before the subtracted image is sent for further processing. This can be done by performing morphological operations [6] 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 [15].

VI. Human Tracking and Handling Occlusion

Human tracking means deriving a correspondence of the object detected in one frame with the object detected in the next frame. If a correspondence is found than we can say that the object found in the previous and the current frame is the same and it can be marked with the same color rectangle as the previous object was marked with. In order to find the correspondence a simple strategy is followed. Few features [11] of the object detected in the previous frame are stored and the features are then matched with the object detected in the current frame. If they match, then the object detected in the current and the previous frame are said to be the same. Features can be color, orientation, speed, posture, speed, intensity or any other information that can be obtained from a pixel. Hence, selection of the features plays an important role in tracking an object.

The tracking methodologies [14] can be mainly divided into 2 types:-

1) Region Based Tracking: Here, the features of the blob, detected in one image frame are matched to the blob detected in the other frame. If there is a match then the detected image is linked with the image in the previous frame.

2) Contour Based Tracking[7]: Here, the energy of the boundary/contour of the blob detected in the previous frame is matched with the energy of the boundary of the blob detected in the current image.

As we are taking stationary background, region tracking is very efficient with stationary background. Hence we are taking region tracking. Now region tracking stores the features of whole object and matches the features with the features of the object in the next frame.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 1, Issue 2, December 2011)

This wastes time, instead most salient region method only matches the most salient region of the previous frame with the most salient region of the current frame. Hence, reducing the amount of time required to match the whole image. The most salient region tracking [8], [12] works as follows:

- The initial features for color, orientation and intensity are fetched from the image [12].
- From the above fetched features, the feature vector is calculated for color, orientation and intensity using center-surround method [13].
- Once the feature map is obtained, the 3 features are weighted in order to find out which feature more uniquely identifies the object. Once weight is obtained, using their weight we get the saliency map for the detected object.
- Finally feature weight vector is calculated for this most salient region. This feature weight vector is matched with the subsequent frame's feature weight vector, if the match is above the threshold value then there is a match otherwise, the search area is doubled and the above procedure is repeated again.

Even after doubling the search area, a match is not found then object is expected to be occluded by some other object or any stationary background object.

Now, occlusion handling is not a separate process as seen in the figure1, it is a part of human tracking process. When an object is detected its centroid can easily be obtained and when a match is found between the object in the previous and the current frame object, we will get the centroid of the current image. From the centroids of these two images we will obtain following:

- The speed (velocity) at which object is moving using the simple distance upon time formula as shown below:

$$\begin{aligned} V_{cx1} + V_{cy1} &= S \\ V_{cx2} + V_{cy2} &= S \end{aligned} \tag{8}$$

- The direction of the motion by the tanθ formula.

In case, the object is not found even after doubling the search area, the object can be said to be occluded and its next position can be predicated by the previously calculated velocity and direction of motion. In order to calculate the velocity and the direction of motion, minimum 5 frames should be processed.

VII. Activity Analysis And Alarm Triggering

Activity analysis is like adding intelligence to this whole process. The information about the detected human is sent to a program which will study its position, pose, motion and analyse them. After analysis, in case of any abnormal activity is found then an alarm is triggered.

Abnormal activity can be any action like moving into any highly secure area, moving with speed more than a limit in a secure place, any typical pose that is not normal, and many other actions can trigger the alarm. The list of abnormal activities varies from customer to customer.

Also, alarm triggering [9] may include actually ringing any alarm, sending notification to any department through e-mail or SMS, generating a report, make an entry in database, etc. Thus, this phase totally deals with how you want to utilize the information that has been obtained from the previous phases and notify on abnormal activities.

VIII. Conclusion

Our proposed method of finding the speed and direction while performing most salient region tracking will help in fast and efficient occlusion handling. Also, as discussed above, in case the most salient region method is not able to find the human in initial search window, we just double the search window and if this fails too then the position of the human is predicted. This reduces the processing time in case the object is hidden as we are not searching whole frame instead just doubling the search area.

IX. Future Enhancement

In this paper, we have considered static background; in future it can be enhanced for changing/non-static background.

References

[1] Murat EKINCI, Eyup GEDIKLI, "Silhouette Based Human Motion Detection and Analysis for Real-Time Automated Video Surveillance", VOL.13, NO.2 2005, c T¨UBITAK, pp. 199-229

[2] Nan Lu, Jihong Wang, Q.H. Wu and Li Yang, "An improved Motion Detection Method for Real-Time Surveillance", IAENG International Journal of Computer Science, 35:1, IJCS_35_1_16

[3] Denman, Simon P. and Chandran, Vinod and Sridharan, Sridha (2007) An adaptive optical flow technique for person tracking systems. *Pattern Recognition Letters* 28(10):pp. 1232-1239

[4] K. Srinivasan, K. Porkumaran and G. Sainarayanan, "Intelligent human body tracking, modeling, and activity analysis of video surveillance system: A survey" Journal of Convergence in Engineering, Technology and Science Vol.1 pp 1-8, December 2009

[5] Peng Zhang, Tie-Yong Cao, Tao Zhu, "A Novel Hybrid Motion Detection Algorithm Based On Dynamic Thresholding Segmentation", Communication Technology (ICCT), 2010 12th IEEE International Conference, pg: 853-856

[6] LI Li, XU Jining, "Moving Human Detection Algorithm based on Gaussian Mixture Model", Proceeding of 29th Chinese Conference, 2010, pp. 2853-2856

[7] Ling Cai, Lei He, Yamasita Takayoshi, Yiren Xu, Yuming Zhao, Xin Yang, "Robust Contour Tracking by Combining Region and Boundary Information", IEEE Transactions on circuit and systems for video technology 2011, pp. 1-10

[8] Simone Frintrop, Markus Kessel, "Most Salient Region Tracking", ICRA'09 Proceedings of the 2009 , IEEE 2009, pp. 1869 – 1874

[9] Suzaimah Bt Ramli, Kamarul Hawari B Ghazali, Muhammad Faiz Bin Mohd Ali, Zakaria Hifzan B Hisahuddin, "Human Motion Detection Framework", IEEE 2011, pp. 158-161.

[10] Yokoyama, M.; Poggio, T, "A contour-based moving object detection and tracking", IEEE 2005, pp. 271-276

[11] Robert T. Collins, Yanxi Liu, Marius Leordeanu, "Online Selection of Discriminative Tracking Features", IEEE Transactions on pattern analysis and machine intelligence vol. 27, No.10, October 2005, pp. 1631-1643

[12] Stefan May, Maria Klodt, Erich Rome and Ralph Breithaupt, "GPU-accelerated Affordance Cueing based on Visual Attention", Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 2007, pp. 3385-3390.

[13] Derrick Parkhurst, Klinton Law, Ernst Niebur, "Modeling the role of salience in the allocation of overt visual attention", Vision Research 42 (2002), pp. 107-123

[14] Alpher Yilmaz, Omar Javed, Mubarak Shah, "Object Tracking : A Survey", ACM Computing Surveys, Vol. 38, No.4, Article 13, December 2006, pp. 1-4

[15] Hironobu FUJIYOSHI, Alan J. LIPTON, Takeo KANADE, "Real-time Human Motion Analysis by Image Skeletonization", IEICE TRANS. INF. & SYST., Vol E87-D, No. 1 January 2004 pp 113-120