

Motion Detection and Tracking of Multiple Objects for Intelligent Surveillance

K.Saranya¹, M.Kalaiselvi Geetha², J.Arunnehr³
^{1,2,3}Dept.of Computer Science and Engineering, Annamalai University.

ABSTRACT:- In this paper proposes new strategies for object tracking initialization using automatic moving object detection based background subtraction. The new strategies are integrated into the real-time object tracking system. The proposed background model updating technique and adaptive thresholding are used to produce a foreground object mask for object tracking initialization. Traditional background subtraction technique detects moving objects by subtracting the background model from the current image. Compare to various common moving object detection technique, background subtraction segments foreground objects more accurately and detects foreground objects even if they are non moving. However, one drawback of traditional background subtraction is that it is vulnerable environmental changes, for instance, gradual or fast illumination changes. The reason of this disadvantage is that it assumes a static background, and therefore a background model update is needed for dynamic backgrounds. The most important challenges are how to update the background model, and how to find out the threshold for classification of foreground and background pixels. The proposed technique is to determine automatically and dynamically depending on the intensities of the pixels within the current frame and a technique to update the background model with learning rate depending on the variations of the pixels within the background model and also the previous frame. This paper additionally represented a shape tracking technique to track the moving multiple objects in surveillance video.

Keywords:- moving object detection, background model, shape tracking.

I. INTRODUCTION

Visual surveillance has been a very active research topic in the last few years due to the growing importance for security in public places. A typical automated visual surveillance system consists of moving object detection, object classification, object tracking, activity understanding, and semantic description. Moving object detection is not only useful for object tracking initialization for a visual surveillance system, it is always the first step of many different computer vision applications, for example, automated visual surveillance, video indexing, video compression, and human machine interaction. Since subsequent processes are greatly dependent on the performance of this stage, it is important that the classified foregrounds pixels accurately correspond to the moving objects of interests. There are a lot of common difficulties and problems encountered when performing moving object detection. Some examples of these are illumination change and repetitive motion from clutter such as waving tree leaves. Due to these problems with dynamic environmental conditions, moving object detection from the background becomes very challenging. The major challenges for background subtraction are how to update the background model, and how to determine the threshold for classification of foreground and background pixels. The proposed algorithm is to determine the threshold automatically and dynamically depending on the pixel intensities of the current frame, and a method to update the background model with learning rate depending on the pixel differences between the background model and the previous frame. However, the generated background model may not be applicable in some scenes with some specified issues, including but not limited to six terms.

- 1) **Flexibility to illumination change:** The background model should adapt to gradual illumination changes.
- 2) **Dynamic textures change:** The background model should be able to adapt to dynamic background movements, which are not of interest for visual surveillance, such as moving curtains.
- 3) **Noise acceptance:** The background model should demonstrate appropriate noise immunity.
- 4) **Susceptible to clutter motion:** The background model should not be sensitive to repetitive clutter motion.
- 5) **Bootstrapping:** The background model should be properly generated at the beginning of the sequence.
- 6) **Expedient implementation:** The background model should be able to be set up fast and reliably.

Video surveillance systems generally track moving objects from one frame to another in an image sequence. The tracking algorithms usually have most significant intersection with motion detection during

processing. Tracking over time typically involves matching objects in consecutive frames using features such as points, lines or blobs. There are also other methods to detect objects using shape based model, finding centroid of the target. Objects tracking method is reliable to deal with occlusions, irregular object motions, changing appearances by postponing the decision of object trajectories until sufficient information is accumulated over time.

II. RELATED WORK

Albert Torrent et al [1] propose a general framework to simultaneously perform object detection and segmentation on objects of different nature. This approach is based on a boosting procedure which automatically decides – according to the object properties – whether it is better to give more weight to the detection or segmentation process to improve both results. Feng-Yang Hsieha et al[2] propose an effective approach to the detection of small objects by employing watershed-based transformation. The proposed detection system includes two main modules, region of interest (ROI) locating and contour extraction. In the former module, an image differencing technique is first employed on two contiguous image frames to generate rough candidate objects appearing in the images. ChengEn Lu et al[3] propose a novel framework for contour based object detection from cluttered environments. Given a contour model for a class of objects, it is first decomposed into fragments hierarchically. Then, these fragments are grouped into part bundles, where a part bundle can contain overlapping fragments. Xingwei Yang et al[4] propose a novel solution by treating this problem as the problem of finding dominant sets in weighted graphs. The nodes of the graph are pairs composed of model contour parts and image edge fragments, and the weights between nodes are based on shape similarity. Carlos Cuevas , Narciso García[5] propose a novel background modeling that is applicable to any spatio-temporal non-parametric moving object detection strategy. Through an efficient and robust method to dynamically estimate the bandwidth of the kernels used in the modeling, both the usability and the quality of previous approaches are improved. Furthermore, by adding a novel mechanism to selectively update the background model, the number of misdetections is significantly reduced, achieving an additional quality improvement. Alessandro Leone et al[6] presents a new approach for removing shadows from moving objects, starting from a frame-difference method using a grey-level textured adaptive background. The shadow detection scheme uses photometric properties and the notion of shadow as semi-transparent region which retains a reduced-contrast representation of the underlying surface pattern and texture. Faisal Bashir, Fatih Porikli[7] presents a set of metrics and algorithms for performance evaluation of object tracking systems. Junda Zhu et al [8] present a novel tracking method for effectively tracking objects in structured environments. The tracking method finds applications in security surveillance, traffic monitoring, etc. Jong Sun kim [9] propose a red-green-blue color background modeling with a sensitive parameter, which is used to extract the moving objects. Kinjal A Joshi et al [10] are proposing various object detection and object tracking techniques.

III. MOVING OBJECT DETECTION

Moving object detection is always the first step of a typical surveillance system. Moving object detection aims at extracting moving objects that are interesting out of a background which can be static or 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. The three most popular approaches to moving object detection are background subtraction, frame differencing, and optical flow. Background subtraction is widely used for moving object detection especially for cases where the background is relatively static because of its low computational cost. The name “background subtraction” comes from the simple technique of subtracting the background model from the current frame to obtain the difference image, and by thresholding the difference frame, a mask of the moving object in the current frame is obtained. The major drawback of background subtraction is that it only works for static background, and hence background model update is required for dynamic background scenes.

3.1 Background Model Initialization

To detect moving objects using background subtraction, the first step is to obtain the background model at the beginning of a sequence. This work assumes that the sequence starts with a background in the absence of moving objects. The background model is defined as:

$$B_t(x, y) = \frac{\sum_{n=1}^{K-1} I_n(x, y)}{K} \quad (1)$$

Where $B_t(x, y)$ is the intensity of pixel (x, y) of the background model, $I_n(x, y)$ is the intensity of pixel (x, y) , t represents the frame number, and K is the number of frame used to construct the background model.

3.2 Background Model Update

After generating the initial background model, the proposed background matching mechanism can be used to modify the background image at each frame.

3.2.1 Reference frame generation: To update the background image without the moving object, the reference frame should be calculated at each frame. For each pixel, the reference frame can be calculated as follows:

$$M_t(x, y) = M_{t-1}(x, y) + \text{sgn}(I_t(x, y) - M_{t-1}(x, y)) \quad (2)$$

Where $RF_t(x, y)$ represents each pixel value of the current reference frame, $M_{t-1}(x, y)$ represents each pixel value of the past reference frame, and $I_t(x, y)$ represents each pixel value of the current input frame. Note that M_t is initialized by the initial background model at the frame with index $K - 1$.

3.2.2 Temporal match: Based on reference frame generation, It can easily decide a current set of background candidates to update the background model. For each pixel, $I_t(x, y)$ is regarded as a background pixel if $M_t(x, y)$ is equal to $I_t(x, y)$. Otherwise, the input pixel is eliminated from the background candidates.

3.2.3 Background modification: After determining the background candidates at each frame, the current background model $B_t(x, y)$ can be updated by the following equation:

$$B_t(x, y) = B_{t-1}(x, y) + \text{sgn}(M_t(x, y) - B_{t-1}(x, y)) \quad (3)$$

Where $B_t(x, y)$ represents the current background model, $B_{t-1}(x, y)$ represents the previous background model, and $M_t(x, y)$ represents a current set of background candidates, which is equivalent to $I_t(x, y)$.

3.2.4 Reference frame modification: Based on temporal match and background modification, each modified pixel of the background model can be further used to update the reference frame $M_t(x, y)$. Since each pixel of the input video frame is strictly matched in terms of determining the background candidates, the background model can be updated without the noise pixel of the input image at each frame. Moreover, the reference frame is also updated by the modified background pixel.

IV. MORPHOLOGICAL OPERATIONS

The motion mask is usually contaminated by a large number of erroneous foreground pixels due to image noise, background motion, and camera jitter. Morphology filters are used to remove the noises and holes (removing holes is the same as filling small gaps) of the motion mask and segment the objects. The fundamental morphological operations are erosion and dilation.

4.1 Dilation: Dilation is the dual of erosion. Each background pixel that is touching an object pixel is changed into an object pixel when it is dilated. Dilation adds pixels to the boundary of the object and closes isolated background pixel (fills up the holes of the object). An example is shown in Fig.1 (a) and Fig.1 (b). The definition of dilation is:

$$X \oplus S = \bigcup_{y \in S} X_y \quad (4)$$

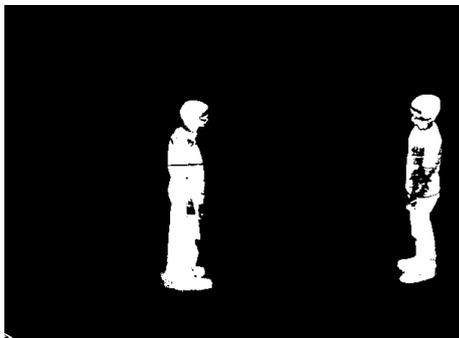


Fig.1 (a) Before dilation



Fig.1 (b) After dilation

4.2 Erosion: To erode an image, each object pixel that is touching a background pixel is changed into a background pixel, erosion removes isolated background pixel. An example is shown in Fig.2 (a) and Fig.2 (b). The definition of erosion is

$$X \ominus S = \bigcap_{y \in S} X_y \quad (5)$$

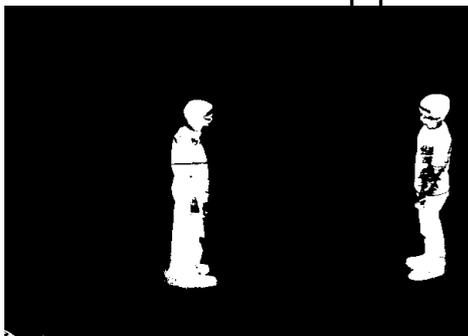


Fig.2 (a) Before erosion



Fig.2 (b) After erosion

In equations (4) and (5), X and S to denote sets in n-dimensional space and y denote position vector in that space.

4.3 Opening: Opening is erosion followed by a dilation operation. Opening removes small islands and thin filaments of object pixels. It is used to remove fine object, separate a target from fine points, and smooth a boundary without changing the area or shape. The definition of opening is

$$\text{Opening}(M) = \text{Dilation}(\text{Erosion}(M)) \quad (6)$$

4.4 Closing: Closing is a dilation operation followed by an erosion operation. Closing removes islands and thin filaments of background pixels. It fills small holes within objects, and joins the neighboring objects without changing the area or shape.

$$\text{Closing}(M) = \text{Erosion}(\text{Dilation}(M)) \quad (7)$$

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 determines the shape of the neighborhood in the image.

V. OBJECT TRACKING

Object tracking is the important issue in human motion analysis. It is higher level computer vision problem. Tracking involves matching detected foreground objects between consecutive frames using different feature of object like motion, velocity, color, texture. Object tracking is the process to track the object over the

time by locating its position in every frame of the video in surveillance system. It may also complete region in the image that is occupied by the object at every time instant. In tracking approach, the objects are represented using the shape or appearance models. The model selected to represent object shape limits the type of motion. 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 objects when they are merged. A new object is formed when a blob has no match with current hypotheses. A track ends when it has no blob similar 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 quite small size are classified as pedestrians; the others to be vehicles. Fig.3 illustrate tracking multiple objects.

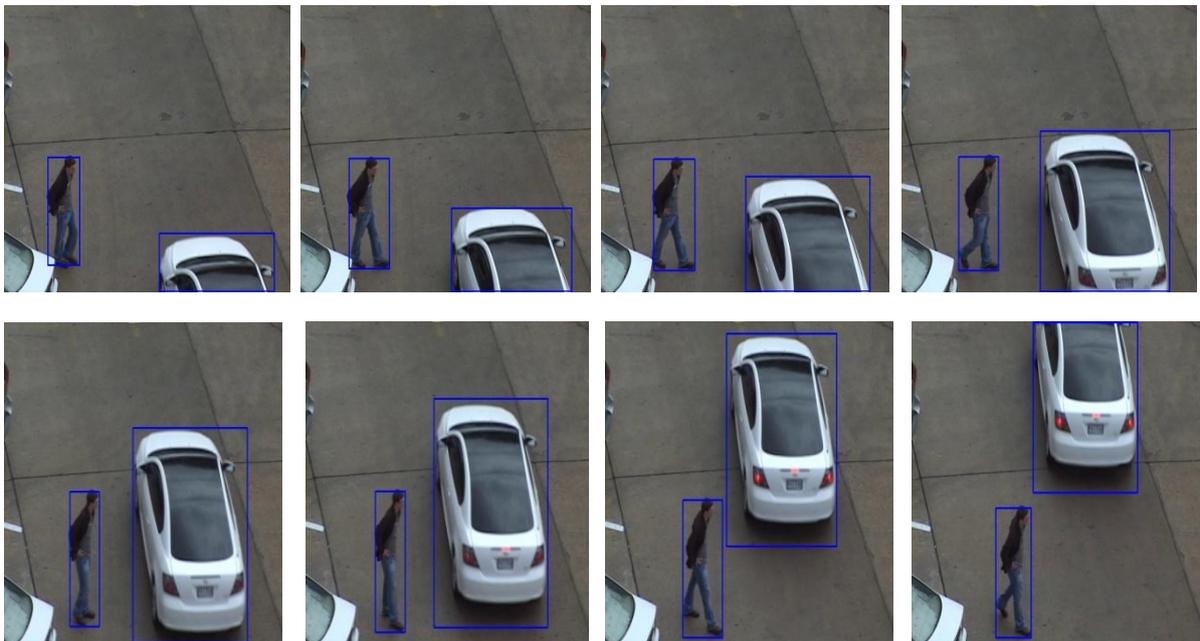


Fig. 3 Tracking different objects

VI. PERFORMANCE MEASURE

In this paper, to propose a set of metrics that compares the output of motion tracking systems to a Ground Truth in order to evaluate the performance of the systems. Before the evaluation metrics are introduced, it is important to define the concepts of spatial and temporal overlap between tracks, which are required to quantify the level of matching between ROI of Ground Truth (GT) tracks and ROI of Tracked Truth, both in space and time. The spatial overlap is defined as the overlapping level between ROI_T and ROI_G . The region overlapping between the ground truth value and the tracked truth value is shown in shaded region. The quantitative evaluation is used to find the rate of success and location error with the reference of object centre. The rate of success and score is defined as given in equation (8).

$$Score = \frac{Area(ROI_T \cap ROI_G)}{Area(ROI_T \cup ROI_G)} \quad (8)$$

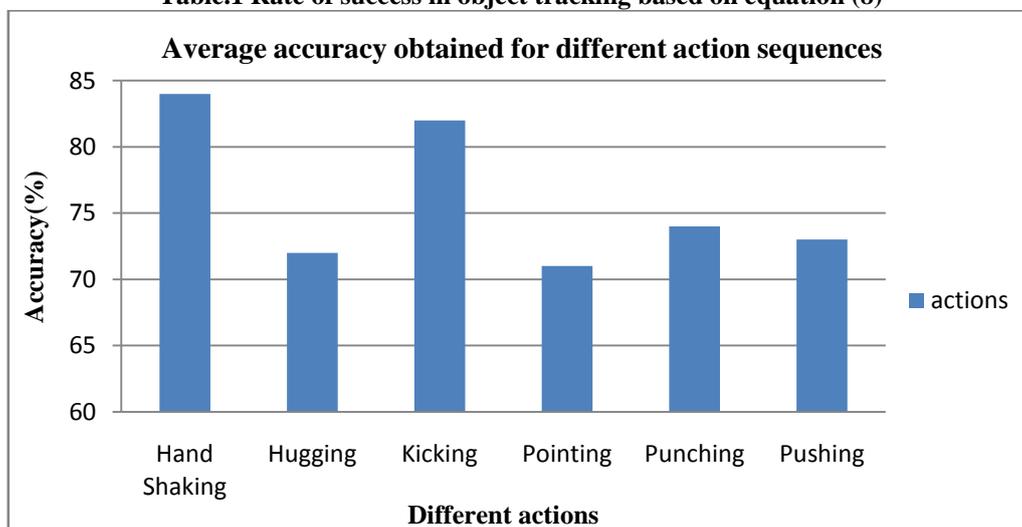
Where, ROI_T is the tracked truth bounding box and ROI_G is the ground truth bounding box. In computer vision field, object's edge is usually identified as the area of pixels that contrasts with its neighbor pixels or area of pixels motion. After getting the edge of the object, the height of the object can be measured by subtracting the top left pixel position of the rectangle and bottom left pixel position of the rectangle. The width of the object can be measured by subtracting the top left pixel position of the rectangle and top right pixel position of the rectangle. The performance is measured by comparing the Tracked Truth (TT) with the Ground Truth (GT), with their Region of Interest (ROI) values. The Region of Interest is found by multiplying the height(x) and width(y) of the bounding box. The score is found by finding area of intersection and union for both Ground truth

value and Tracked truth value, and the area of intersection is divided by area of union, to find the total score. The rate of success is computed in all frames, where this score > 0.5 in each frame is considered as a success.

Table illustrate the success rate.

Sequence Number	Hand shaking	Hugging	Kicking	Pointing	Punching	Pushing
1	0.6	0.5	0.5	0.3	0.7	0.5
2	0.3	0.3	0.2	0.3	0.3	0.2
3	0.5	0.6	0.3	0.3	0.5	0.5
4	0.4	0.3	0.7	0.2	0.3	0.2
5	0.4	0.5	0.4	0.2	0.6	0.4

Table.1 Rate of success in object tracking based on equation (8)



VII. 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 work will be handling the occlusion problem based on assigning unique ID for each object.

REFERENCE

- [1]. Albert Torrent, Xavier Lladó, Jordi Freixenet, Antonio Torralba. A boosting approach for the simultaneous detection and segmentation of generic objects. *Pattern Recognition Letters* 34 (2013) 1490–1498.
- [2]. Feng-Yang Hsieha, Chin-Chuan Hanb, Nai-Shen Wua, Thomas C. Chuangc, Kuo-Chin Fana. A novel approach to the detection of small objects with low contrast. *Signal Processing* 86 (2006) 71–83.
- [3]. ChengEn Lu, Nagesh Adluru, Haibin Ling, Guangxi Zhu, Longin Jan Latecki. Contour based object detection using part bundles. *Computer Vision and Image Understanding* 114 (2010) 827–834.
- [4]. Xingwei Yang, HairongLiu, LonginJanLatecki. Contour-based object detection as dominant set computation. *Pattern Recognition* 45 (2012) 1927–1936.
- [5]. Carlos Cuevas, Narciso García. Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies. *Image and Vision Computing* 31 (2013) 616–630.
- [6]. Alessandro Leone, Cosimo Distante, Francesco Buccolieri. A shadow elimination approach in video-surveillance context. *Pattern Recognition Letters* 27 (2006) 345–355.
- [7]. Faisal Bashir, Fatih Porikli. Performance Evaluation of Object Detection and Tracking Systems, June 2006.

- [8]. Junda Zhu, Yuanwei Lao, Yuan F. Zheng. Object tracking in structured environments for video surveillance applications. *IEEE Transactions on circuits and systems for video technology*, Vol.20, No.2, February 2010.
- [9]. Jong Sun Kim, Hae Yeom, and Young Hoon Joo. Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems. *IEEE Transactions on consumer electronics*, Vol.57, No.3, August 2011.
- [10]. Kinjal A Joshi, Darshak G. Thakore. A survey on moving object detection and tracking in video surveillance system. *International Journal of Soft Computing and Engineering (IJSCE)* ISSN:2231-2307, Volume-2, Issue-3, July 2012.
- [11]. Carlos R. del-Blanco, Fernando Jaureguizar, and Narciso Garcia. An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications. *IEEE Transactions on Consumer Electronics*, Vol.58, No.3, August 2012.
- [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 Multimedia*, VOL.14, No.3, June 2012.
- [13]. J. Arunnehr, M. Kalaiselvi Geetha. A Quantitative Real-Time Analysis Of Object Tracking Algorithm For Surveillance Applications. *International Journal of Emerging Technology and Advanced Engineering*, Volume 3, Special Issue 1, January 2013.
- [14]. X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo. Robust visual tracking based on incremental tensor subspace learning. *Proceedings of the IEEE International Conference on Computer Vision*, 2007.
- [15]. Muharrem Mercimek, Kayhan Gulez and Tarik Veli Mumcu. Real Object Recognition Using Moment Invariants, December 2005.
- [16]. B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 983–990, 2009.
- [17]. J. Kwon and K. Lee. Visual tracking decomposition. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1269–1276, 2010.
- [18]. M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision* 88(2), pp. 303–338, 2010.
- [19]. Fan-Chieh Cheng and Shanq-Jang Ruan. Accurate Motion Detection Using a Self-Adaptive Background Matching Framework. *Proceeding of IEEE Transactions On Intelligent Transportation Systems*, Vol. 13, No. 2, June 2012.
- [20]. J. Arunnehr, M. Kalaiselvi Geetha. "A Quantitative Real-Time Analysis Of Object Tracking Algorithm For Surveillance Applications". *International Journal of Emerging Technology and Advanced Engineering*, Volume 3, Special Issue 1, January 2013.
- [21]. Evan Tan and Chun Tung Chou. A Frame Rate Optimization Framework for Improving Continuity in Video Streaming. *Proceeding of IEEE Transactions On Multimedia*, Vol. 14, No. 3, June 2012.
- [22]. Xianguo Zhang, Tiejun Huang, Yonghong Tian, Wen Gao. Background-Modeling-Based Adaptive Prediction for Surveillance Video Coding. *Proceeding of IEEE Transactions On Image Processing*, Vol. 23, No. 2, February 2014.
- [23]. Chaur-Heh Hsieh, Ping S. Huang, and Ming-Da Tang, Human Action Recognition Using Silhouette Histogram. *Conferences in Research and Practice in Information Technology (CRPIT)*, Vol. 113, 2011.