

Moving Object Detection for Video Surveillance

Abhilash K.Sonara¹, Pinky J. Brahmbhatt²

¹Student (ME-CSE), Electronics and Communication, L. D. College of Engineering, Ahmedabad, India

²Associate Professor, Electronics and Communication, L. D. College of Engineering, Ahmedabad, India

ABSTRACT: Video surveillance has long been in use to monitor security sensitive areas such as banks, department stores, highways, crowded public places and borders. The advance in computing power, availability of large-capacity storage devices and high speed network infrastructure paved the way for cheaper, multi sensor video surveillance systems. Traditionally, the video outputs are processed online by human operators and are usually saved to tapes for later use only after a forensic event. The increase in the number of cameras in ordinary surveillance systems overloaded both the human operators and the storage devices with high volumes of data and made it infeasible to ensure proper monitoring of sensitive areas for long times. In order to filter out redundant information generated by an array of cameras, and increase the response time to forensic events, assisting the human operators with identification of important events in video by the use of “smart” video surveillance systems has become a critical requirement. The making of video surveillance systems “smart” requires fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis.

Keywords: Video-Based Smart Surveillance, Moving Object Detection, Background Subtraction, Object Tracking.

I. Introduction

Video surveillance systems have long been in use to monitor security sensitive areas. The history of video surveillance consists of three generations of systems which are called 1GSS, 2GSS and 3GSS. The first generation surveillance systems (1GSS, 1960-1980) were based on analog sub systems for image acquisition, transmission and processing. They extended human eye in spatial sense by transmitting the outputs of several cameras monitoring a set of sites to the displays in a central control room. They had the major drawbacks like requiring high bandwidth, difficult archiving and retrieval of events due to large number of video tape requirements and difficult online event detection which only depended on human operators with limited attention span. The next generation surveillance systems (2GSS, 1980-2000) were hybrids in the sense that they used both analog and digital sub systems to resolve some drawbacks of its predecessors. They made use of the early advances in digital video processing methods that provide assistance to the human operators by filtering out spurious events. Most of the work during 2GSS is focused on real-time event detection. Third generation surveillance systems (3GSS, 2000-) provide end-to-end digital systems. Image acquisition and processing at the sensor level, communication through mobile and fixed heterogeneous broadband networks and image storage at the central servers benefit from low cost digital infrastructure. Unlike previous generations, in 3GSS some part of the image processing is distributed towards the sensor level by the use of intelligent cameras that are able to digitize and compress acquired analog image signals and perform image analysis algorithms like motion and face detection with the help of their attached digital computing components. The ultimate goal of 3GSS is to allow video data to be used for online alarm generation to assist human operators and for offline inspection effectively. In order to achieve this goal, 3GSS will provide smart systems that are able to generate real-time alarms defined on complex events and handle distributed storage and content-based retrieval of video data. The making of video surveillance systems “smart” requires fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis. Starting from the 2GSS, a considerable amount of research has been devoted for the development of these intelligent algorithms. Moving object detection is the basic step for further analysis of video. It handles segmentation of moving objects from stationary background objects. This not only creates a focus of attention for higher level processing but also decreases computation time considerably.

Commonly used techniques for object detection are background subtraction, statistical models, temporal differencing and optical flow. Due to dynamic environmental conditions such as illumination changes, shadows and waving tree branches in the wind object segmentation is a difficult and significant problem that needs to be handled well for a robust visual surveillance system [1].

II. Moving Object Detection

Each application that benefit from smart video processing has different needs, thus requires different treatment. However, they have something in common: moving objects. Thus, detecting regions that correspond to moving objects such as people and vehicles in video is the first basic step of almost every vision system since it provides a focus of attention and simplifies the processing on subsequent analysis steps. Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving in blowing wind), motion detection is a difficult problem to process reliably. Frequently used techniques for moving object detection are background subtraction, statistical methods, temporal differencing and optical flow whose descriptions are given below [2].

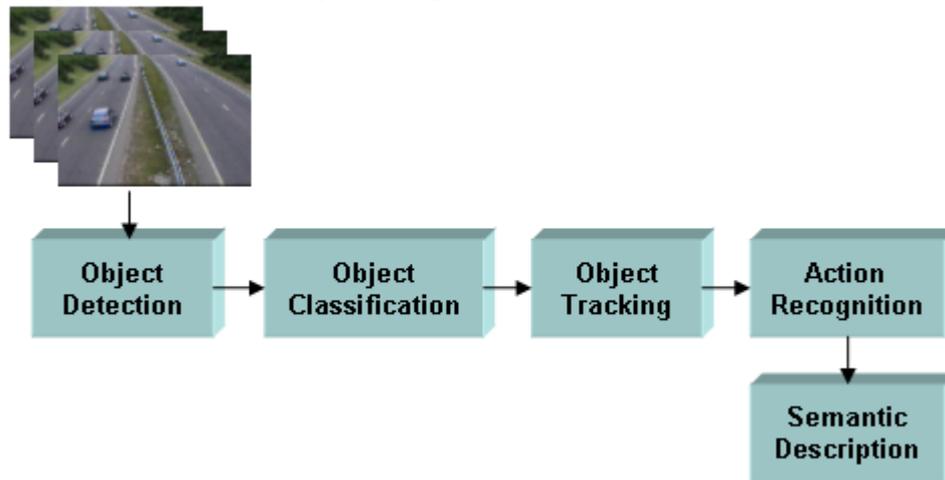


Fig 1: A generic framework for smart video processing algorithms

A. Background Subtraction

Background subtraction is particularly a commonly used technique for motion segmentation in static scenes. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image *that is created by averaging images over time in an initialization period. The pixels where the difference is above a threshold are classified as foreground. After creating a foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt to dynamic scene changes. There are different approaches to this basic scheme of background subtraction in terms of foreground region detection, background maintenance and post processing. In Heikkila and Silven uses the simple version of this scheme where a pixel at location (x, y) in the current image I_t is marked as foreground if*

$$|I_t(x, y) - B_t(x, y)| > \tau$$

Is satisfied where τ a predefined threshold is. The background image B_t is updated by the use of an Infinite Impulse Response (IIR) filter as follows:

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t$$

The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions. Although background subtraction techniques perform well at extracting most of the relevant pixels of moving regions even they stop, they are usually sensitive to dynamic changes when, for instance, stationary objects uncover the background (e.g. a parked car moves out of the parking lot) or sudden illumination changes occur [3].

B. Statistical Methods

More advanced methods that make use of the statistical characteristics of individual pixels have been developed to overcome the shortcomings of basic background subtraction methods. These statistical methods are mainly inspired by the background subtraction methods in terms of keeping and dynamically updating statistics of the pixels that belong to the background image process. Foreground pixels are identified by comparing each pixel's statistics with that of the background model. This approach is becoming more popular due to its reliability in scenes that contain noise, illumination changes and shadow. The W4 system uses a statistical background model where each pixel is represented with its minimum (M) and maximum (N) intensity values and

maximum intensity difference (D) between any consecutive frames observed during initial training period where the scene contains no moving objects. A pixel in the current image I_t is classified as foreground if it satisfies:

$$|M(x, y) - I_t(x, y)| > D(x, y) \text{ or } |N(x, y) - I_t(x, y)| > D(x, y)$$

After thresholding, a single iteration of morphological erosion is applied to the detected foreground pixels to remove one-pixel thick noise. In order to grow the eroded regions to their original sizes, a sequence of erosion and dilation is performed on the foreground pixel map. Also, small-sized regions are eliminated after applying connected component labelling to find the regions. The statistics of the background pixels that belong to the non-moving regions of current image are updated with new image data. As another example of statistical methods, Stauffer and Grimson described an adaptive background mixture model for real-time tracking. In their work, every pixel is separately modeled by a mixture of Gaussians which are updated online by incoming image data. In order to detect whether a pixel belongs to a foreground or background process, the Gaussian distributions of the mixture model for that pixel are evaluated [4].

III. Object Detection and Tracking

The overview of our real time video object detection, classification and tracking system is shown in Figure 3.1. The proposed system is able to distinguish transitory and stopped foreground objects from static background objects in dynamic scenes; detect and distinguish left and removed objects; classify detected objects into different groups such as human, human group and vehicle; track objects and generate trajectory information even in multi-occlusion cases and detect fire in video imagery. In this and following chapters we describe the computational models employed in our approach to reach the goals specified above. Our system is assumed to work real time as a part of a video-based surveillance system. The computational complexity and even the constant factors of the algorithms we use are important for real time performance. Hence, our decisions on selecting the computer vision algorithms for various problems are affected by their computational run time performance as well as quality. Furthermore, our system's use is limited only to stationary cameras and video inputs from Pan/Tilt/Zoom cameras where the view frustum may change arbitrarily are not supported. The system is initialized by feeding video imagery from a static camera monitoring a site. Most of the methods are able to work on both color and monochrome video imagery. The first step of our approach is distinguishing foreground objects from stationary background. To achieve this, we use a combination of adaptive background subtraction and low-level image post-processing methods to create a foreground pixel map at every frame. We then group the connected regions in the foreground map to extract individual object features such as bounding box, area, center of mass and colour histogram [5].

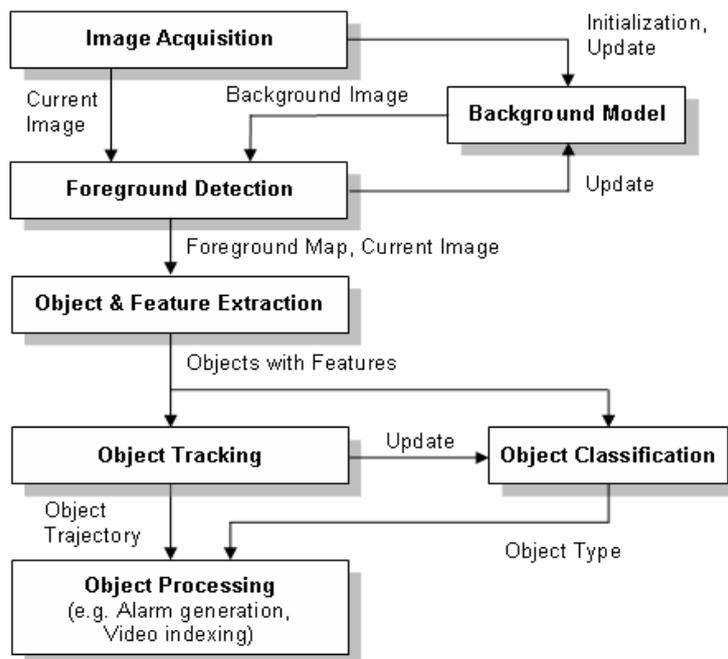


Fig 2: The system block diagram.

Our novel object classification algorithm makes use of the foreground pixel map belonging to each individual connected region to create a silhouette for the object. The silhouette and center of mass of an object are used to generate a distance signal. This signal is scaled, normalized and compared with pre-labeled signals in a template database to decide on the type of the object. The output of the tracking step is used to attain temporal consistency in the classification step. The object tracking algorithm utilizes extracted object features together with a correspondence matching scheme to track objects from frame to frame. The color histogram of an object produced in previous step is used to match the correspondences of objects after an occlusion event. The output of the tracking step is object trajectory information which is used to calculate direction and speed of the objects in the scene. After gathering information on objects' features such as type, trajectory, size and speed various high level processing can be applied on these data. A possible use is real-time alarm generation by pre-defining event predicates such as "A human moving in direction d at speed more than s causes alarm a_1 ." or "A vehicle staying at location l more than t seconds causes alarm a_2 .". Another opportunity we may make use of the produced video object data is to create an index on stored video data for offline smart search. Both alarm generation and video indexing are critical requirements of a visual surveillance system to increase response time to forensic events. The remainder of this chapter presents the computational models and methods we adopted for object detection and tracking [6].

A. Object Detection

Distinguishing foreground objects from the stationary background is both a significant and difficult research problem. Almost the visual surveillance systems' entire first step is detecting foreground objects. This both creates a focus of attention for higher processing levels such as tracking, classification and behaviour understanding and reduces computation time considerably since only pixels belonging to foreground objects need to be dealt with. Short and long term dynamic scene changes such as repetitive motions (e. g. waiving tree leaves), light reflectance, shadows, camera noise and sudden illumination variations make reliable and fast object detection difficult. Hence, it is important to pay necessary attention to object detection step to have reliable, robust and fast visual surveillance system. Our method depends on a six stage process to extract objects

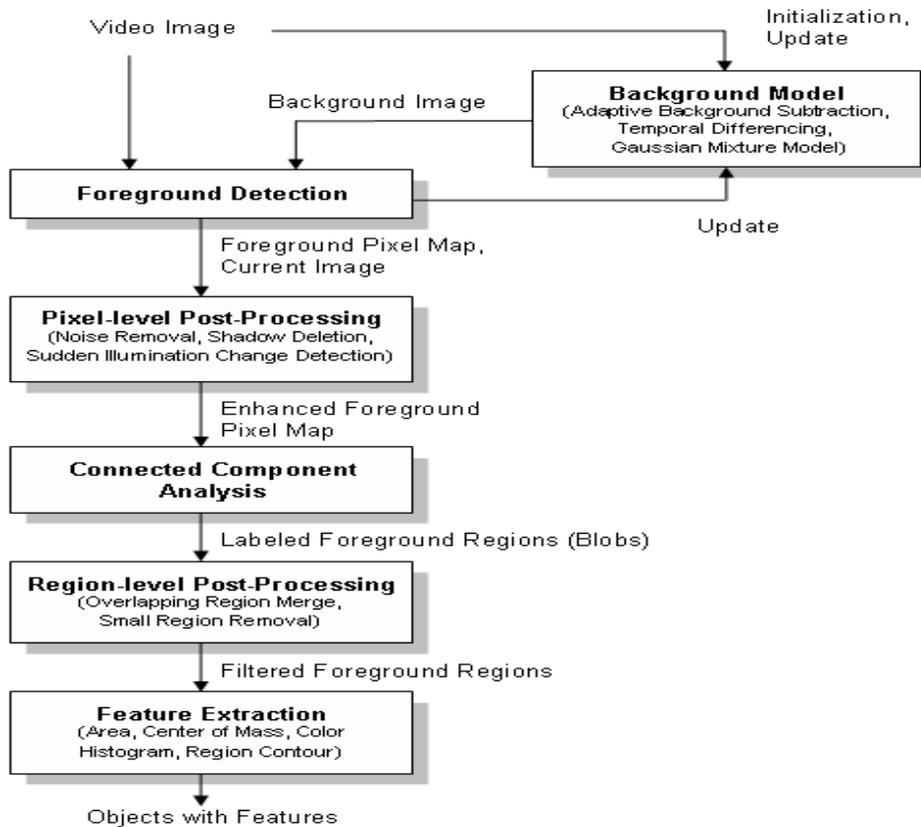


Fig 3: The object detection system diagram.

With these features in video imagery, first step is the background scene initialization. There are various techniques used to model the background scene in the literature. In order to evaluate the quality of different background scene models for object detection and to compare run-time performance, we implemented three of these models which are adaptive background subtraction, temporal frame differencing and adaptive online Gaussian mixture model. The background scene related parts of the system is isolated and its coupling with other modules is kept minimum to let the whole detection system to work flexibly with any one of the background models. Next step in the detection method is detecting the foreground pixels by using the background model and the current image from video. This pixel-level detection process is dependent on the background model in use and it is used to update the background model to adapt to dynamic scene changes. Also, due to camera noise or environmental effects the detected foreground pixel map contains noise. Pixel-level post-processing operations are performed to remove noise in the foreground pixels [7].

B. Foreground Detection

We use a combination of a background model and low-level image post-processing methods to create a foreground pixel map and extract object features at every video frame. Background models generally have two distinct stages in their process: initialization and update. Following sections describe the initialization and update mechanisms together with foreground region detection methods used in the three background models we tested in our system [8].

IV. Results And Analysis

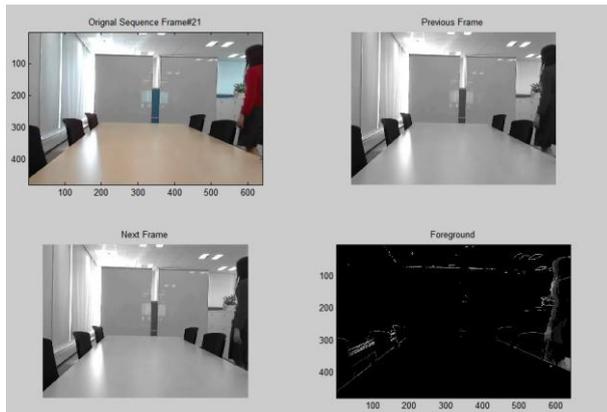


Fig. 4(a)

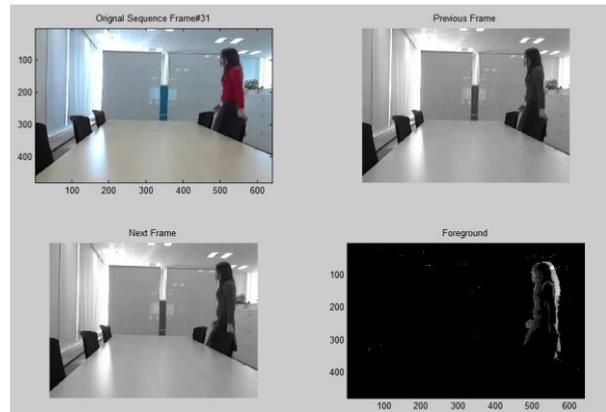


Fig. 4(b)

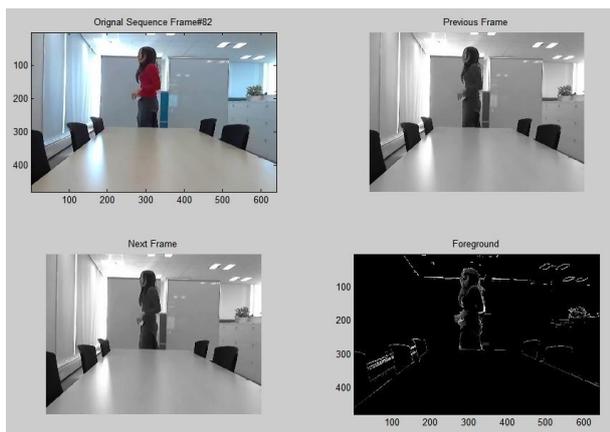


Fig. 4(c)

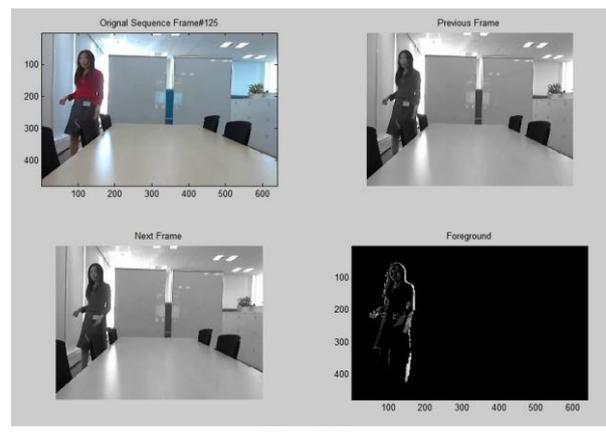


Fig. 4(d)

Fig. 4 Position wise detection of moving object along with their corresponding reference frames

V. Conclusions and Future Work

In this work we have described a unique moving object detection technique based on separation of background and foreground. The approximate position of moving object is captured by comparing the reference frame with consecutive frames.

In this work we have focussed mainly on the detection of a single object from a video sequence. As a part of future work look forward to incorporate methods to enable our algorithm to detect multiple objects present in the video sequence. Also we propose to work on video sequences having complex background.

REFERENCES

- [1] Komagal, E.; Vinodhini, A.; Srinivasan, A.; Ekava, B.” Real time Background Subtraction techniques for detection of moving objects in video surveillancesystem” Computing, Communication and Applications (ICCCA), 2012.
- [2] Soumya. T “A Moving Object Segmentation Method for Low Illumination Night Videos” Proceedings of the World Congress on Engineering and Computer Science 2008.
- [3] A. Amer. Voting-based simultaneous tracking of multiple video objects. In Proc. SPIE Int. Symposium on Electronic Imaging, pages 500–511, Santa Clara, USA, January 2003.
- [4] R. Bodor, B. Jackson, and N. Papanikolopoulos. Vision-based human tracking and activity recognition. In Proc. of the 11th Mediterranean Conf. on Control and Automation, June 2003.
- [5] H.T. Chen, H.H. Lin, and T.L. Liu. Multi-object tracking using dynamical graph matching. In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 210–217, 2001
- [6] Lian Xiaofeng; Zhang Tao; Liu Zaiwen “A Novel Method on Moving-Objects Detection Based on Background Subtraction and Three Frames Differencing Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on, Issue March 2010.
- [7] Jin Wang; Lanfang Dong “Moving objects detection method based on a fast convergence Gaussian mixture model Computer Research and Development (ICCRD), 2011 3rd International Conference on, Issue March 2011.
- [8] C. B. Liu and N. Ahuja. Vision based fire detection. In IEEE International Conference on Pattern Recognition, Cambridge, UK, August 2004.