Real Time Motion Detection and Tracking

Saurabh Gade[1], Jay Shiurkar[2], Milind Rane[3]
[1]*Vishwakarma Institute of Technology*
[2]*Vishwakarma Institute of Technology*
[3]*Vishwakarma Institute of Technology*
(E-mail: saurabh7998gmail.com)

Abstract— This paper describes a real-time system and a constant framework for human detection, tracking /following and motion analysis. This system depicts automated video surveillance for detection and monitoring of people in both indoor & outdoor environments. Detection and tracking can be achieved through a few stages. First, we design a robust, and a strong foundation for adaptive background model that can deal with lightning changes, long term changes in the scene and objects occlusions. This technique is used to get foreground pixels by taking the background subtraction method. Afterwards, noise cleaning and object detection along with techniques like thresholding, dilation, erosion are applied, followed by human modeling to recognize and monitor human activity in the scene such as human walking, running.

Keywords— motion analysis, detecting and monitoring, object occlusions.

Introduction

Human detection and tracking in a complex environment can be a very hard task, because people always interact with each other, form groups and may make motion in unexpected ways. This therefore requires a robust method, that can cope with the different motions, without being affected at all by occlusions and changes of environmental features. In order to overcome changes in the environment monitored by the system, we need to design a robust background model that can deal with very slow illumination changes like light switches between day and night, fast lighting changes like clouds blocking the sun.

There are indeed various successful vision systems for human motion detection and tracking like Pfinder[1], W4[2]. These systems apply and make use of human features like head or body shape, leg symmetry analysis and statistical models which restrict them to human figures. They also may need large number of pixels on target due to the shape based nature of the model. This often leads to misidentification of small objects. Initially a foundation is kept up, and after that a forefront picture is obtained, at that point the background and the highlights are separated. Along these lines moving items are then followed and distinguished.

I. LITERATURE SURVEY

This paper gives an overview, of the various methods and techniques generally used for object detection & tracking. In spite of being a relatively new research area, a massive number of contributions related to surveillance system using motion analysis have been published in the last few years. This paper reviews the state of the art in automatic object detection and localisation, with particular attention to object detection & tracking. In particular, it is still beyond the current state-of-the-art to expect a very general tracker, which would be able to follow people accurately in any situation, regardless of the environment, light, people density and activity, etc. Tracking is the process of following an object of interest within a series of frames, from its initial appearance to its last. The type of object and its description within the system depends upon the application. During the time that it is present in the scene, it may be occluded (either partially or fully) by other objects of interest or fixed obstacles within the scene. A tracking system should be able to predict the position of any occluded objects through the occlusion, ensuring that the object is not temporarily lost and only detected again when the object appears after the occlusion. The process can be extended to design an algorithm which can help to overcome occlusions.

Object tracking systems are generally geared toward surveillance applications wherein it is desired to monitor people and or vehicles moving about an area. Systems such as these need to perform in real time, and be able to deal with real world environments and effect such as changes in lighting and spurious movement in the background (such as tress moving in the wind). Other surveillance applications include data mining applications, where the aim is to annotate video after the event. Target representation is categorized into two major classes. One is for a collection of general objects, such as human bodies or faces, computer monitors, motorcycles, and so on. The other is for one precise target including a specific person, car, toy, building and so on. The targets can be images, concrete objects or even abstract feature points. Motion detection & segmentation in video sequences aims at detecting regions that corresponds to mobile objects such as humans and vehicles. Detecting moving regions gives a centre of attention for later processes such as tracking and behaviour analysis since only these particular regions need be considered in the later processes. At present, many segmentation methods use either spatial or temporal information in the image sequence.

II. BACKGROUND MODELLING

A background is an image which contains all the non-moving objects in a video footage. Obtaining a background model can be done in mainly two steps: first, the background initialization, wherein we obtain the background image from a specific time of the video sequence or footage. Secondly, the background maintenance, where the background can be updated due to the changes that may occur in the real.[3]

Model Initialisation:
The median algorithmic program is wide utilized in background model low-level formatting [4]. It is based on a principle that the background at each and every pixel has to be visible more than fifty percent of the time at the time of the training sequence. The median algorithm results in wrong background intensity value, particularly when a moving article or an object stops for in excess of 50% of the preparation succession time. More efficient algorithm is the highest redundancy ratio (HRR) algorithm [5]. HRR considers that a pixel intensity belongs to the background image only if it is having the highest redundancy ratio among intensity values of the pixel taken from a training sequence. This assumption is a lot close to the actual meaning of the background. Thus, the HRR algorithm is definitely more flexible and applicable for real events than median algorithm.

III. IMAGE SEGMENTATION

The main motive behind performing the task of image segmentation is to separate foreground regions from background region in order to detect any moving objects. This process is categorically divided into foreground region extraction, object detection and contour detection.

A. *Foreground Region Extraction:-*

Foreground region extraction will be detected exploitation background subtraction technique [6]. With time the newly updated (latest) background image is subtracted from current acquired frame. The resulting image pixels are then flagged as foreground pixels if they have large intensity values, and are considered as background pixels if they have a close to zero value. A mathematical analysis of background pixel intensity shows that its value can be modeled using Gaussian distributions. Thus if a difference pixel is having a value greater than the standard deviation (3σ) for this pixel value from the background pixel intensity, then it is marked as foreground pixel. Figure 1 shows the result of varying these thresholds. A threshold at σ ends up in additional background pixels known as foreground (noise), and a threshold at 5*σ ends up in additional foreground pixels lost and classified as background pixels, therefore a worth of 3*σ is taken into account a decent trade-off between the **2** extremes. A 3X3 median filter is then accustomed scale back error classification.

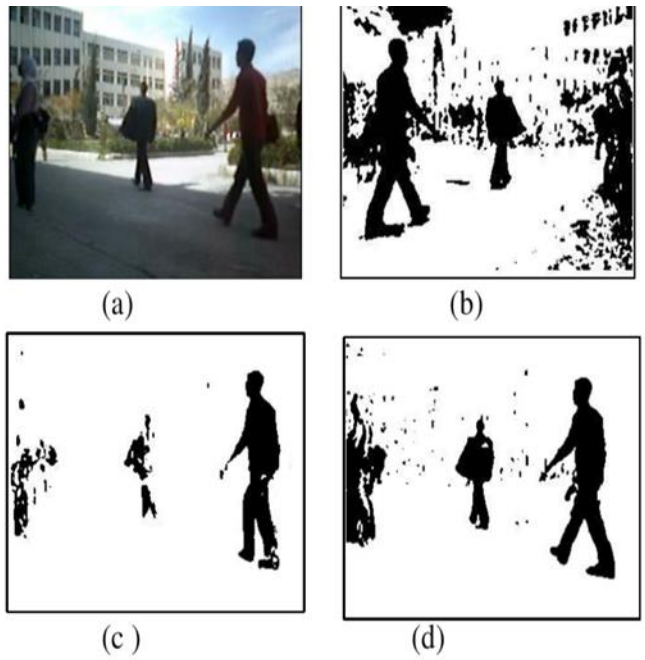

FIGURE 1. FOREGROUND EXTRACTION

B. *Object Detection:-*

Determining an accurate position of objects in the foreground pictures is a basic task of any tracking system; in order to achieve this aim efficiently we present the 2D image of foreground by its 1D projections histogram over the vertical and the horizontal axis [5]. These projections help in finding the limit box wrapping the protest as it appears in figure 2. With the end goal to distinguish objects that cover in their projections, we have actualized a similar calculation for each bouncing box in an iterative way so we can part them separated and find them precisely. After getting the bounding boxes for the objects, we implement a size filtering to eliminate any small objects, considering them as irrelevant or wrongly classified objects.

C. *Contours:-*

Foreground region extraction and object detection may contain prattles and gabs in the object which is detected, and it may cause objects to be divided or splitted into more than one connected region, that could make the object to be detected as multiple objects. Therefore we must restore the objects to their original state and size by applying a sequence of two erosions and one dilation. The contours are therefore later detected using Duda & Hart algorithm [7]. Figure 3 shows the previous steps to get the moving object contour.

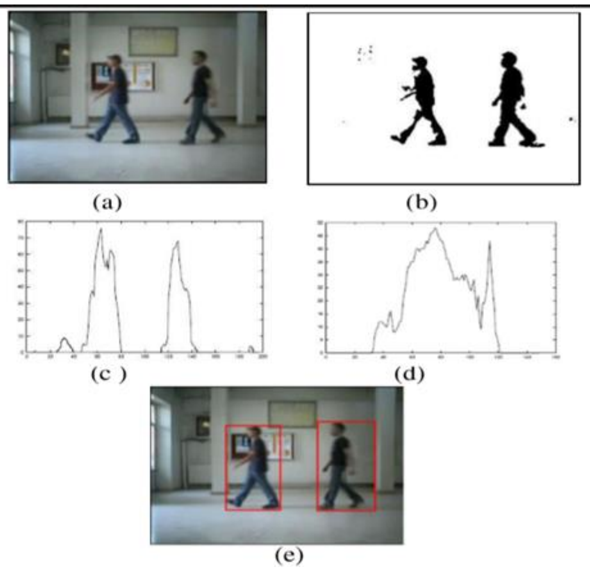

FIGURE 2. IDENTIFYING ENCLOSING BOXES FOR EACH OBJECT. (A) ORIGINAL IMAGE,(B)FOREGROUND IMAGE, (C) THE FOREGROUND HISTOGRAM PROJECTION OVER THE HORIZONTAL AXIS, (D) THE FOREGROUND HISTOGRAM PROJECTION OVER THE VERTICAL AXIS, (E) THE RESULTING BOUNDING BOXES.

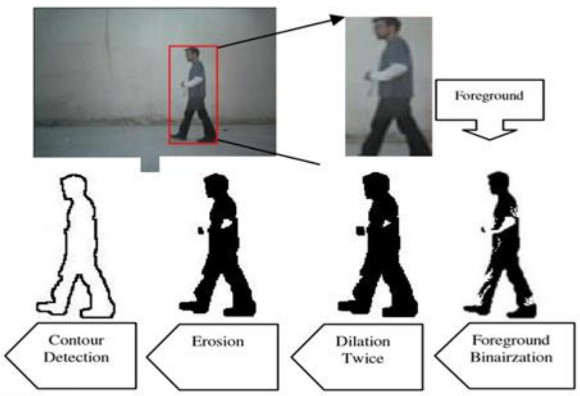

FIGURE 3. MORPHOLOGICAL OPERATORS AND CONTOUR EXTRACTION APPLIED ON THE MOVING OBJECTS LOCATED BY THEIR BOUNDING BOX.

Morphological Dilation and Erosion:-

Morphology is a vast set of image processing operations and operands/operators that help in processing of images based on their shapes. These operations mostly apply a structuring element to an given input image, which creates an output image of the same size. In a given morphological operation, the value associated with each pixel of the o/p image is based on a true comparison of the respective pixel in the input image with its neighbours. By selecting the size and shape of the neighbourhood, you are able to construct a morphological operation that may be sensitive to already existing shapes in the input image. The simplest morphological operations are dilation and erosion. Dilation is a technique adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries.

Dilation: The value associated with the output pixel is the *maximum* value of all the pixels in the vicinity of input pixel. For a binary image, if the value of any of the pixels is set to 1, then the output pixel is set to 1.

Erosion: The value associated with the output pixel is the *minimum* value of all the pixels in the vicinity of input pixel. For a binary image, if the value of any of the pixels is set to 0, then output pixel is set to 0.

IV. OBJECT TRACKING

Basically, mapping of an object in successive frames or progressive casings of a video is termed as tracking. This definition may sound very straight forward but in computer vision and machine learning areas, tracking is a very broad term that incorporates conceptually similar but technically different ideas. The basic motive of the object tracking are [8]: determining when a new object enters the scene, then calculating the correspondind relationship between the foreground regions detected by background subtraction and the elements being tracked presently, and finally, employing tracking algorithms to estimate the position of each object.

RESULTS: The motion is therefore detected in the real time as can be seen in the figure 4 below. The window on the left shows the original video footage. The window on the right shows the video in which the features are identified and the motion is detected as well as tracked with the help of contours. Various techniques like Thresholding, Dilation, Erosion etc. help in detecting and applying the contours very well. The system is tested using a video camera of resolution 600x800. The output contains an indoor scene of a group of students in a classroom. The percentage of accurate tracking is 90%.

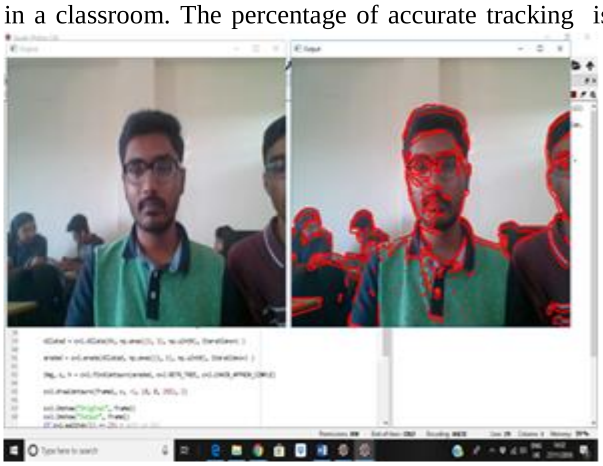

Figure 4. Output showing the motion tracked with the help of contours.

REFERENCES:

[1] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland "Pfinder: Real-Time Tracking of the Human Body" IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.

[2] Isail Haritaoglu, David Harwood and Larry S. Davis "W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People" Computer Vision Laboratory, University of Maryland College Park, 1998.

[3] Hironobu Fujiyoshi, Alan J. Lipton and Takeo Kanade "Realtime Human Motion Analysis by Image Skeletonization" IEICE TRANS, 2004.

[4] D. Gutchessy, M. Trajkovi´cz, E. Cohen-Solalz, D. Lyonsz, A. K. Jainy "A Background Model Initialization Algorithm for Video Surveillance "Computer Science and