A

MAJOR PROJECT REPORT

On

FALL DETECTION OF HUMAN BEING AND NOTIFICATION SYSTEM FROM VIDEO FILES

Submitted in partial fulfilment of the

Award of Degree

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by:

Ashutosh Kumar Mishra(156320027)

Pratyush Ratna(156320051)

Under the guidance of:

Mr. Vivek Arya

(Asst. Professor, ECE Department)

Department of Electronics and Communication Engineering

Faculty of Engineering and Technology, Gurukula Kangri University
Haridwar-249404,Uttarakhand

ACKNOWLEDGEMENT

We would like to acknowledge the contribution of all those people who have been associated with us in this project. We would like to thank our guide **Mr. Vivek Arya**, Asst. Professor, Department of Electronics and Communication engineering, Faculty of Engineering and Technology, Gurukula Kangri Vishwavidyalaya, Haridwar. For his supervision, knowledge, support and persistent encouragement during our project work. He steered us through this journey with his invaluable advice, positive discussions and consistent encouragement.

We also express our deep sense of gratitude to other staff members of the department who have given us help and valuable advice time to time.

Date :

Ashutosh Kumar Mishra

Pratyush Ratna

CANDIDATE'S DECLARATION

We hereby declare that the work, which is being presented in the minor project, and titled **"FALL DETECTION OF HUMAN BEING AND NOTIFICATION SYSTEM FROM VIDEO FILES"** in partial fulfilment of requirement for the award of degree of Bachelor of Technology in Electronics and Communication Engineering submitted in the department of Electronics and Communication Engineering, Faculty of Engineering and Technology, Gurukula Kangri University, Haridwar is an authentic record of our own work carried out during 15/03/2019 to 25/04/2019 under the supervision of **Mr. Vivek Arya,Asst. Prof. ,Dept. of ECE.**

The matter embodied in this record has not been submitted by us for the award of any other degree or diploma.

Date: **Signature of candidate**:

Ashutosh Kumar Mishra

Pratyush Ratna

CERTIFICATE

This is to certify that **Ashutosh Kumar Mishra** and **Pratyush Ratna** of Electronics and Communication Engineering batch 2015-2019 has satisfactorily completed the major project **FALL DETECTION OF HUMAN BEING AND NOTIFICATION SYSTEM FROM VIDEO FILES** in fulfilment of the requirement for the award of Bachelor of Technology in Electronics and Communication Engineering in Faculty of Engineering and Technology,Haridwar during the session 2015-2019 .

Signature

Internal Examiner

Signature

External Examiner

INDEX

Table of contents **Page no.**

CHAPTER-1
INTRODUCTION

Elderly people have a higher risk of death or injury resulting from falls . More than a third of the over-aged population fall each year which can result into fear and loss of independence. The increase in population of elderly people also increases the demand for healthcare systems . Fall detection systems are categorized into the following three groups: ambience device, camera-based systems and wearable devices. Ambience devices are attached around an area which can detect falls using the following sensors: pressure, PIR, Doppler, microphone and accelerometer sensors. The advantages of this method is that it is cheap and non-intrusive however, disadvantages include the range and environmental factors which can result in low accuracy. Computer vision makes use of cameras to track the user movements. A fall may be detected when the user is inactive for a long time . Advantages of the system is that it can detect multiple events simultaneously and are less intrusive since they are not worn . The disadvantages of this system are that it is limited to a specific area and does not guarantee privacy and is expensive. Wearable devices which are attached on the user include the following sensors: an accelerometer and a gyroscope. Advantages of these devices are that they are portable, cheap and easy to use. Disadvantages of these worn devices are that some users forget or refuse to put it on, are intrusive and can produce a lot of false alarms .

There are three types of methods mainly used in moving object detection. These methods are the frame subtraction method, the background subtraction method and the optical flow method . In the Frame subtraction method the difference between two consecutive images is taken to determine the presence of moving objects. The calculation in this method is very simple and easy to develop. But in this method it is difficult to obtain a complete outline of moving object; therefore the detection of moving object is not accurate. In the Optical flow method , calculation of the image optical flow field is done. The clustering processing is done according to the optical flow distribution characteristics of image. From this, the complete movement information of moving body is found and it detects the moving object from the quantity of calculation, poor antinoise performance makes it unsuitable for real-time applications. The background subtraction method is the method in which the difference between the current image and background image is taken for the detection moving objects by using simple algorithm. But it is very sensitive to the changes which occur in the external environment and it also has poor anti interference ability. One advantage of this method is, it can provide the most complete object information in the case of the background is known . In the background subtraction method, in a single static camera condition, the dynamic background modeling is combined with dynamic threshold selection method which depends on the background subtraction. The background is updated on the basis of accurate detection of object.

1.1 Frame Separation

Frame processing is the first step in the background subtraction algorithm, the purpose of this step is to prepare the modified video frames by removing noise and unwanted objects in the frame in order to increase the amount of information gained from the frame. Preprocessing of a image is a process of collecting simple image processing tasks that change the raw input video info into a format. This can be processed by subsequent steps. Preprocessing of the video is necessary to improve the detection of moving object's, For example; by spatial and temporal smoothing, snow as moving leaves on a tree, can be removed by morphological processing of the frames after the identification of the moving object.

1.2 Moving Object Detection

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 are classified as foreground where the difference is above a threshold. 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 the basic scheme of background subtraction in terms of foreground region detection, background maintenance and post processing. In [5] Heikkila and Silven uses the simple version of this scheme where a pixel at location (x, y) in the current image, it is marked as foreground if is satisfied.

$$|It(x, y) - Bt(x, y)| > \tau \text{-----------------------}1$$

Where, τ is a predefined threshold. The background image Bt is up- dated by the use of an Median filter as follows:

$$Bt+1 = \alpha It + (1 - \alpha) Bt \text{----------------------}2$$

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.

1.2.1 Background Modeling

In the background modeling process, the reference background image and some parameters associated with normalization are computed over a number of static background frames. The background is modeled statistically on a pixel by pixel basis. A pixel is modeled by a 4-finite sequence of pixels Ei; si; ai; bi where Ei is the expected color value, si is the standard deviation of color value which is defined in ai is the variation of the brightness distortion, and bi is the variation of the chromaticity distortion of the ith pixel. The expected color value of pixel i is given by Where $\mu R(i)$, $\mu G(i)$ and $\mu B(i)$ are the arithmetic means of the ith pixel's red, green and blue values

$$Ei = [\mu R(i), \mu G(i), \mu B(i)] \text{----------------} 3$$

computed over N background frames. So far, we have defined Ei and si.

1.2.2 Background Update

For accurately extracting the moving object the background needs to be updated in real time and the background model can better adapt to light changes. In the proposed method, the update algorithm is as follows: In the moving object detection, the pixels judged as belonging to the moving object maintain the original background gray values, not be updated. We update the background model according to following rule for the pixels which are judged to be the background

$$Bk+1(x,y) = \beta Bk(x,y) + (1-\beta)\, Fk(x,y) \text{--------------} 4$$

Where B(x,y) is background image, Fk(x,y) is current image and F3 E (0,1) is update coefficient, in this paper f3 = 0.004. Fk(x,y) is the pixel gray value in the current frame. Bk(x,y) and Bk+1(x,y) are respectively the Background value of the current frame and the next frame. As the camera is fixed, the background model can remain relatively stable at one position for very long period of time. Using this method we can avoid the unexpected phenomenon of the Background, such as the sudden appearance of something in the background which is not included in the original background. Moreover, the impact brought by light, weather and other changes in the external environment can be effectively adapted by the updating of pixel gray value of the background.

1.2.3 Moving Object Extraction

When the background image B(x, y) is obtained, subtract the background Image B(x,y) from the current frame Fk (x, y). Set threshold as T. If the pixel difference is greater than threshold T, then determines that the pixels appear in the moving object, otherwise, as the background pixels. The moving object can be detected after threshold operation . Its expression is as follows:

$$Dk\,(x, y) = \{1 \mid Fk(x,y) - Bk\text{-}1(x,y)\mid T \text{---------}(5)$$
$$=\{0 \text{ others Where, } Dk\,(x, y)$$

is the binary image of differential results. T is gray-scale threshold. Its size determines the accuracy of object identification. As in the algorithm T is a fixed value, only for an ideal situation, is not suitable for complex environment with lighting changes. Therefore, this paper proposes the dynamic threshold method, we dynamically changes the threshold value according to the lighting changes of the two images obtained. On this basis, add a dynamic threshold T to the above algorithm. d) Extraction of Moving Human Body Some accurate edge regions are got after doing the median filtering and morphological operations. But the moving human body regions could not be determined. By observation, we can find out that when moving object appears, shadow will appear in some regions of the scene. The presence of shadow it is difficult to extract the moving object accurately. By analyzing the characteristics of motion detection, we combine the projection operator with the previous methods [6]. Based on the results of the methods above, adopting the method of combining vertical with horizontal projection to detect the height of the motion region. This can remove the impact of the shadow to a certain degree. Then we analyze the vertical projection value and set the threshold value (determined by experience) to remove the pseudo-local maximum value and the pseudo-local minimum value of the vertical projection [7] to determine the number and width of the body in the motion region, we will get the moving human body with precise edge. We are assuming that people in the scene are all in upright-walking state.

1.3 Detection of Moving Objects in the Video

Motion detection has been done using spatio-temporal differencing. For motion detection based on the spatiotemporal filter, the motion is characterized via the entire three-dimensional (3D) spatio-temporal data volume spanned by the moving person in the image sequence. Its advantages are low computational complexity and a simple implementation process. Here, Motion detection is carried out by the background elimination algorithm.

Moving object detection is an important research topic of computer vision and video processing areas. Detection of moving objects In video streams is the first relevant step of information extraction in many computer vision applications.. This paper puts forward an improved background subtraction of moving object detection of fixed camera condition. Then combining the adaptive background subtraction with symmetrical differencing obtains the integrity foreground image. Using chromaticity difference to eliminate the shadow of the moving target, effectively distinguishes moving shadow and moving target. The results show that the algorithm could quickly establish the background model and detect integrity moving target rapidly.

Moving object detection is an important part of digital image processing techniques and it is the base of the many following sophisticated processing task such as target recognition and tracking, target classification, behavior understanding and analysis .Aside from the intrinsic usefulness of being able to segment video streams into moving and background components, detecting moving objects provides a focus of attention for recognition, classification and activity analysis. The technology has a wide application prospect such as smart monitor, autonomous navigation, human computer interaction, virtual reality and so on.

This paper studies the method of obtaining the data of moving object from video images by background extraction. Object detection requires two steps: background extraction and object extraction. Moving object detection needs static background image. Since each frame of video image has moving object then background extraction is necessary. Each frame image subtracting the background image can get the moving object image. This is object extraction. Then the moving object detection can be achieved.

This paper firstly introduces two moving object detection algorithms of fixed scenes — frame difference method and moving edge method and analyzes their advantages and disadvantages, and then presents a new algorithm based on them, lastly gives the experimental results and analysis

1.3.1 Background extraction of moving object

Background extraction means that the background, the static scene, is extracted from the video image. Because the camera is fixed, each pixel of the image has a corresponding background value which is basically fixed over a period of time.

Well known issues in background extraction include

- **Light changes:** background model should adapt to gradual illumination changes.
- **Moving background:** background model should include changing background that is not of interest for visual surveillance such as moving trees
- **Cast shadows:** the background model should include the shadow cast by the moving objects that apparently behaves itself moving in order to have a more accurate detection of moving object shape.
- **Bootstrapping:** the background model should be properly setup even in absence of a complete and static training set at the beginning of the segment
- **Camouflage:** moving objects should be detected even if their chromatic features are similar to those of the background model.

1.3.2 Calculation of consecutive frames subtraction

The method utilizes current two frames or the differences between the current frame and its previous frame to extract a motion region. In this paper, we adopt its improvement methods namely symmetrical differencing, that means image differences of the three current frames. This method can remove effects of unveiling background which is caused by motion, accurately obtain contour of moving targets. In the conventional background subtraction method, a fixed reference background model for the intended surveillance area is constructed in advance. The conventional background subtraction method extracts moving targets based on the difference between the current image and the reference background model. It works well for applications in controlled environments, in which a constant illumination scenario can be achieved artificially.

However, for other visual tracking applications such as traffic monitoring and security/surveillance, the illumination conditions change over time so that a fixed reference background model is not realistic and may eventually lead to a detection failure. Consequently, construction and maintenance of a reliable and accurate reference background model is crucial in background

1.3.3 Typical moving object detection algorithms

- **Frame difference method**

To detect moving object in the surveillance video captured by immobile camera, the simplest method is the frame difference method for the reason that it has great detection speed, can be implemented on hardware easily and has been used widely. While detecting moving object by frame difference method, in the difference image, the unchanged part is eliminated while the changed part remains.

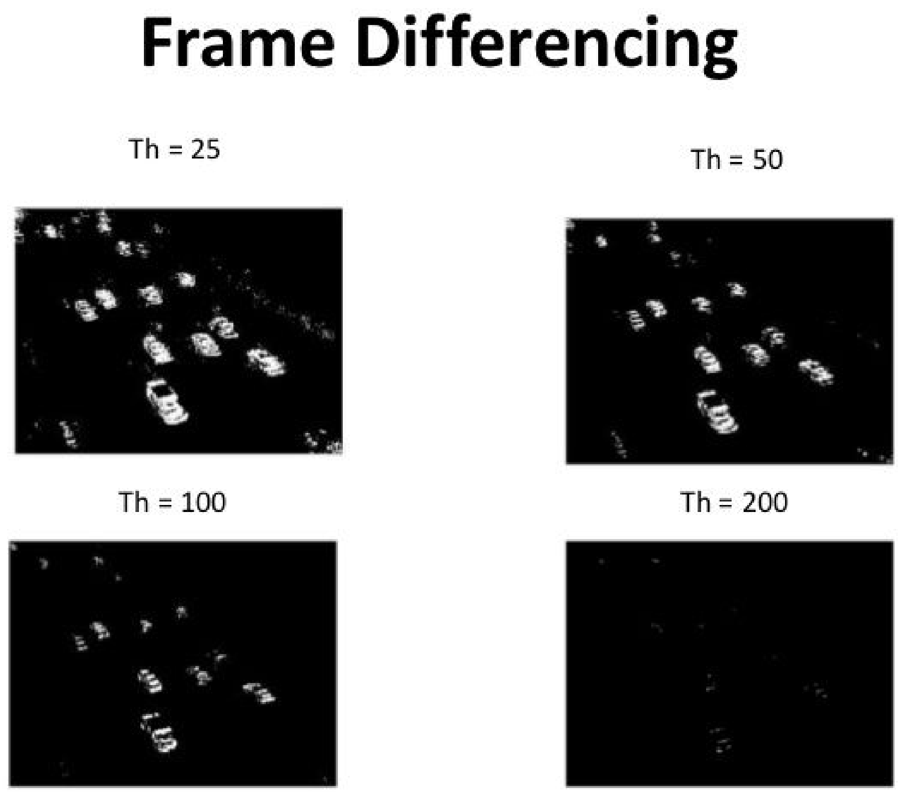

Fig 1 frame detection method

This change is caused by movement or noise, so it calls for a binary process upon the difference image to distinguish the moving objects and noise. Connected component labeling is also needed to acquire the smallest rectangle containing the moving objects. The noise is assumed as Gaussian white noise in calculating the threshold of the binary process. According to the theory of statistics, there is hardly any pixel which has dispersion more than 3 times of standard deviation.
While u is the mean of the difference image is the standard deviation of the difference image.

The flow chart of the detecting process by frame method is shown in fig 2

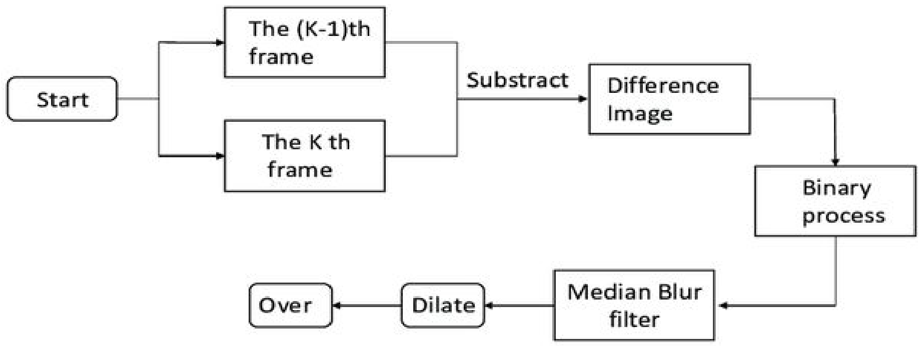

Fig 2 flow chart of the detecting process by frame method

- **Moving edge method**

Difference image can be regarded as time gradient, while edge image is space gradient. Moving edge can be defined by the logic AND operation of difference image and the edge image .

The advantage of frame difference method is its small calculation, and the disadvantage is that it is sensitive to the noise. If the objects do not move but the brightness of the background changes, the results of frame difference methods may be not accurate enough. Since the edge has no relation with the brightness, moving edge method can overcome the disadvantage of frame difference method.

The flow chart of the detecting process by moving edge method is shown in fig 3

Improved Moving object detection algorithm based on frame difference and edge detection

Moving edge method can effectively suppress the noise caused by light, but it still has some misjudgments to some other noise. This paper proposes an improved algorithm based on frame difference and edge detection. Upon analysis, the method has better noise suppression and higher detection accuracy.

The mostly used edge detecting methods are gradient method and Laplacian method. In the gradient method by using maximum and minimum in the first derivative of the image detects the edges. While in case of Laplacian method searches for zero crossings in the second derivative of the image to find edges. Some examples of gradient method are Roberts, Prewitt, and Sobel and it is widely used.Mostly used edge detection method is Gradient edge detection operator. First order derivatives of a digital image are based on various approximations of the 2D gradient.Computation of the gradient of an image is based on obtaining partial derivative of $\partial f/\partial x$ and $\partial f/\partial y$ at every pixel location.For every point in image, applying Sobel operator results in either the corresponding gradient vector or the norm of this vector. Prime drawback of Sobel operators such as over segmentation and sensitivity.

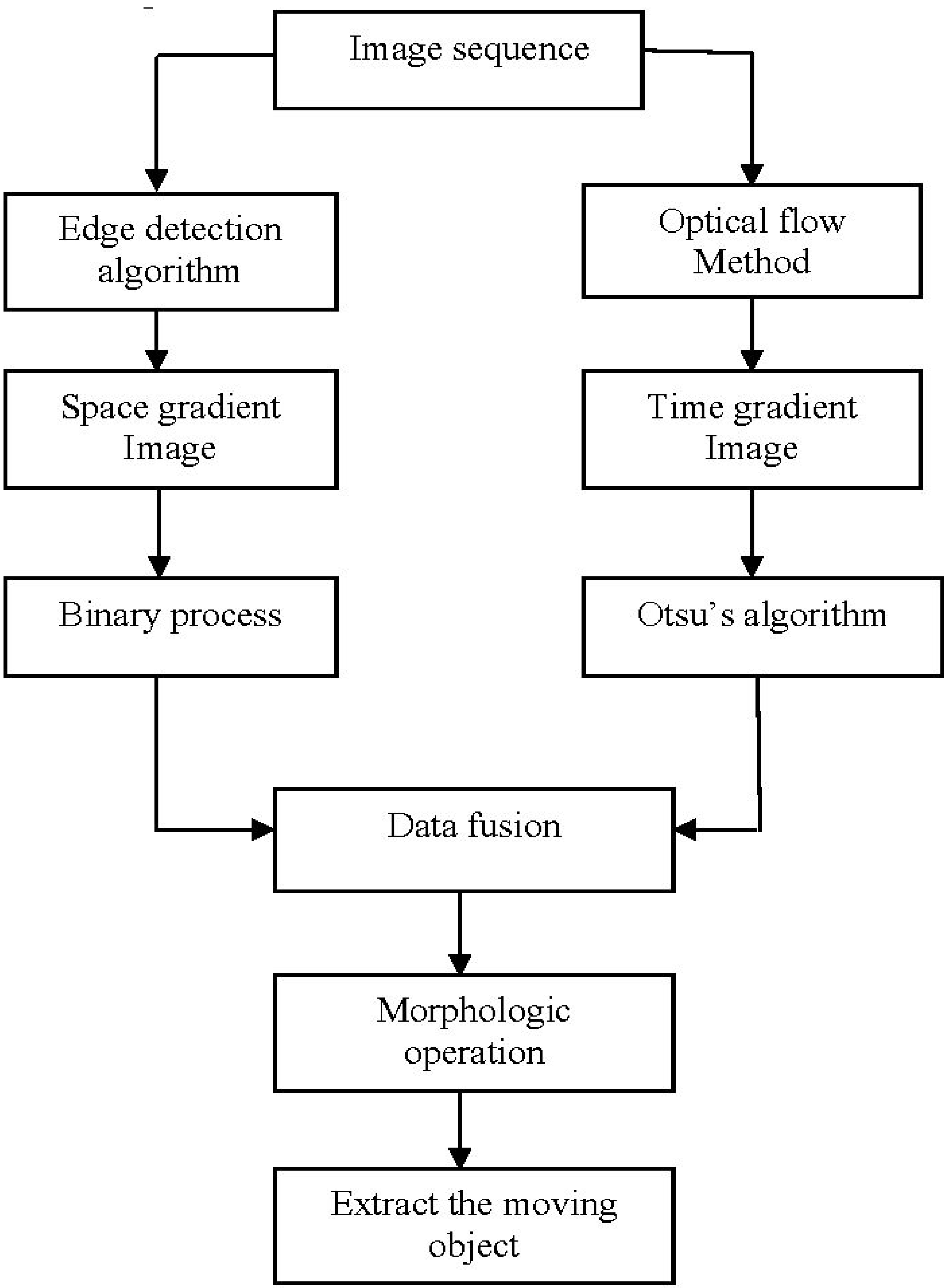

Fig 3.Flow chart of moving edge method

- **Moving cast shadow method**

To prevent the moving shadows being misclassified as moving objects or parts of moving objects, this paper represents an explicit method of detection of moving cast shadows on a dominating scene background. These shadows are generated by objects between a light source and the background. Moving cast shadows cause a frame difference between two succeeding images of a monocular video image sequence. For shadow detection these frame differences are detected and classified into regions covered and regions uncovered by a moving shadow. The detection and classification assume plane background and a non negligible size and intensity of light sources. A cast shadow is detected by temporal integration of the covered background regions while subtracting the covered background regions. The shadow detection method is integrated into an algorithm for 2D shape estimation of moving objects. The extended segmentation algorithm compensates first apparent camera motion. then a spatially adaptive relaxation scheme estimates a change detection mask for two consecutive images. An object mask is derived from the change detection mask by elimination of changes due to background uncovered by moving objects and the elimination of changes due to background covered or uncovered by moving cast shadows.

1.4 Background Elimination

The background subtraction system is used to provide foreground image through the threshold of difference image between the current image and reference image. As the reference image is the previous frame, this method is called temporal differencing. The temporal differencing is very adaptive to dynamic environment. Background elimination was carried out using mean squared error concept.

The first phase's main purpose is to process the video and perform background subtraction on every frame. First a statistical based Gaussian model is made. This model will be used as the reference for every frame when executing background subtraction. The output of the phase will contain sequence of images that requires further filtering process.

1.4.1 Gaussian Model

The first step in developing a foreground extraction software is to build a model of the background. Since there are no preset background images to use, the software will have to generate a model automatically. Using the statistical approach, the software will build a Gaussian Model. A Gaussian Model calculates each pixel-value from all the sample pixels' mean and variance. The model will set a lower bound and an upper bound that will eliminate pixels that are outside of the norm. If a video is to run for an extended period of time, the pixels' average will equal to the background's value unless the foreground object stays static. On the Matlab's official website,I was able to find a downloadable software package that creates a Gaussian Model from a video. The code works with Hue, Saturation, Value (HSV) color space instead of the Red, Green, Blue (RGB) color space since HSV can minimize the effect that shadow have on images. After loading the video into the software package, a background can be calculated. (See Figure 2)

Fig 4 Background Model

1.4.2 Background Subtraction

The background subtraction is very straight forward. This step takes every frame from the video and subtracts it by the Gaussian Model that was calculated in the previous step. The resulting frame shows a general shape of where the foreground objects locate, however, the frame is very noisy due to the imperfection of the subtraction method. See Figure 5.

There was one problem that was faced during this step. The Matlab software occupies too much memory and often the process would have to be stopped short. Instead of saving the pixel data into one large variable, the program now reads each frame individually during each subtraction and deletes the temporary frames immediately after each loop.

Fig 5 Subtracted Frame

CHAPTER-2
LITERATURE REVIEW

This chapter gives an overview, of the various methods and techniques generally used for fall 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. As mentioned in Chapter 1 it is a challenging problem with many potential applications. This chapter reviews the state of the art in automatic fall detection and notification, with particular attention to fall 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 typically geared toward surveillance applications where 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 can be 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.

Most of the cited works on object detection & tracking can be classified according to:

- The motion segmentation techniques that they use
- The image descriptors or feature vectors that they use.
- The object detection framework that is built over these descriptors.

This section reviews the relevant work according to these categories. Section 2.1 provides an overview of the key approaches used to build detectors. Section 2.2 covers the different set of image features that have been used including ones specific to human detection. Section 2.3 and Section 2.4 provides, respectively, states of the art on person detection in videos and on approaches to fuse multiple overlapping detections. Section 2.5 concludes by presenting motivations behind our approach and reviewing its relation to prior work

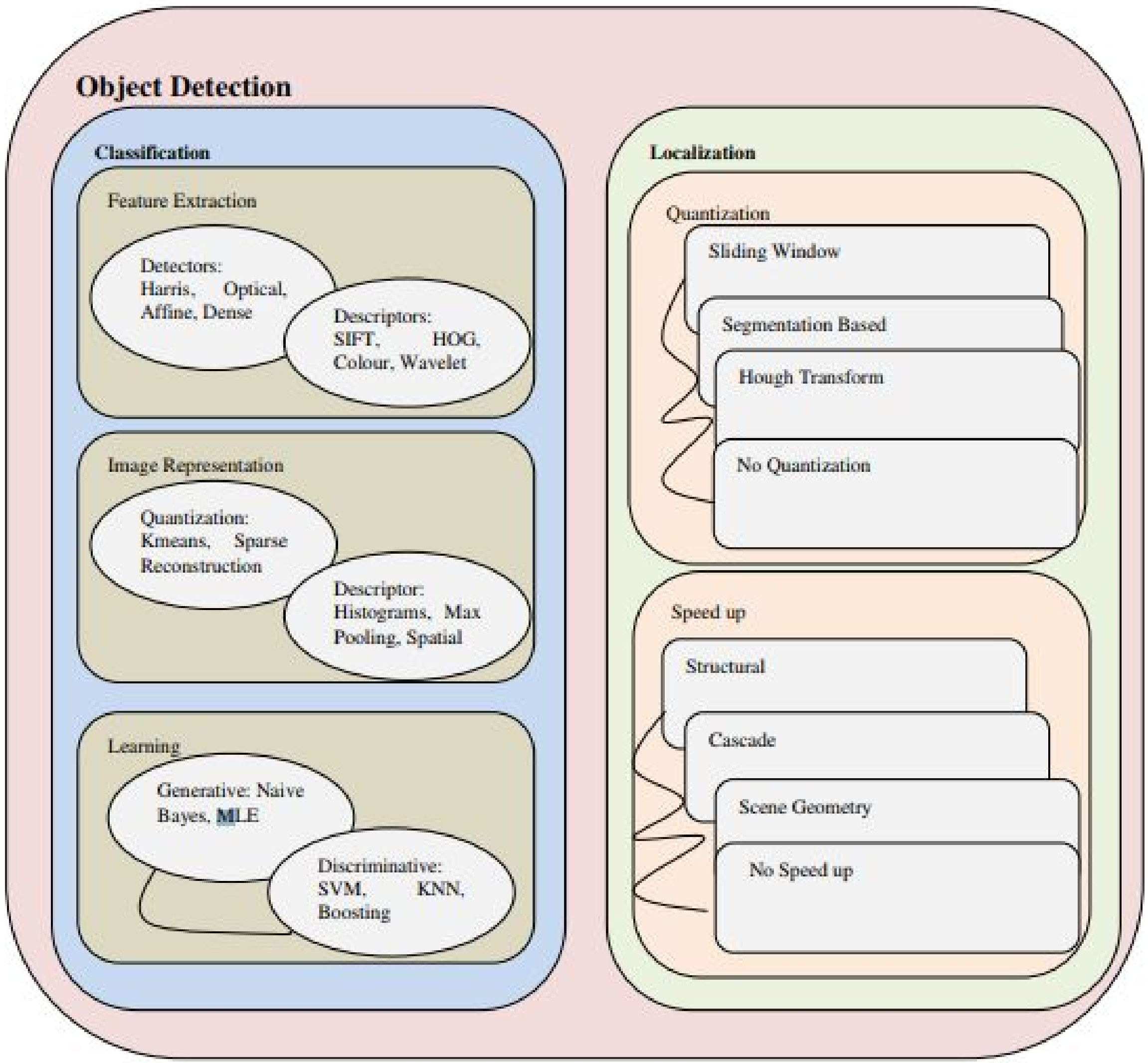

Fig 6 Object detection taxonomy presented by Marco Pedersoli in Hierarchical Multiresolution Models for fast Object Detection.

2.1 Motion segmentation

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. Different conventional approaches for motion segmentation are outlined as follows.

2.1.1 Background subtraction

Background subtraction is a well-known method for motion detection & tracking, particularly under those conditions with a reasonably unmoving background (N. Prabhakar et al. 2012 [13]). It identifies moving regions in an image by computing the disparity between the present image and the reference background image in a pixel-by-pixel manner. Background subtraction is simple, but extremely sensitive to variations in vibrant scenes derived from irrelevant and illumination events etc. Therefore, it is enormously dependent on an excellent background model to lessen the control of these changes (I. Haritaoglu et al. 2000 [22], S. McKenna et al. 2000 [23], & C. Stauffer et al. 1999 [17]) as part of environment modelling. (Toyama et al. 2012 [24]) propose the Wall flower algorithmic technique in which background subtraction and background maintenance are performed at three different levels: the basic pixel level, the middle region level, and the last frame level. Recently, some statistical methods to find modified regions from the background are stimulated by the basic background subtraction methods. The statistical approaches uses the features of an individual pixels or groups of pixels to construct more advanced background models, and the information of the backgrounds can be vigorously updated during processing. Each pixel in the current image can be classified into foreground pixel or background pixel by comparing the statistics of the current background model. This approach is becoming progressively more popular due to its robustness to shadow, changing of lighting conditions etc. (C. Stauffer et al. 2010 [28]). It is also possible that a few elements of background might actually move, such as tress moving in a breeze, Shadow, Rain drops etc. An example of motion mask is shown in figure . Eigen background subtraction [25] is proposed by Oliver et al. 2000. It presents an Eigen space model for moving object segmentation. In this technique, dimensionality of the sample images constructed from space is decreased with the help of Principal Component Analysis. It is proposed that the condensed space following PCA should signify only the static parts of the scene, residual moving targets, if an image is projected on this space.

Notice that the subtle lighting differences at the subject feet where her shadow is cast. Kang et al. 2003 [9] also applied head detection after background segmentation to locate people, who are characterised by bounding boxes. In addition, results from the previous frame's head detection are used in the next frame through a feedback loop to aid the process. So It identifies moving regions in an image by computing the disparity between the present image and the reference background image in a pixel-by-pixel manner. Background subtraction is simple, but extremely sensitive to variations in vibrant scenes derived from irrelevant and illumination events.So we can say Background subtraction is a major preprocessing step in many vision-based applications. For example, consider the case of a visitor counter where a static camera takes the number of visitors entering or leaving the room, or a traffic camera extracting information about the vehicles etc. In all these cases, first you need to extract the person or vehicles alone. Technically, you need to extract the moving foreground from static background.

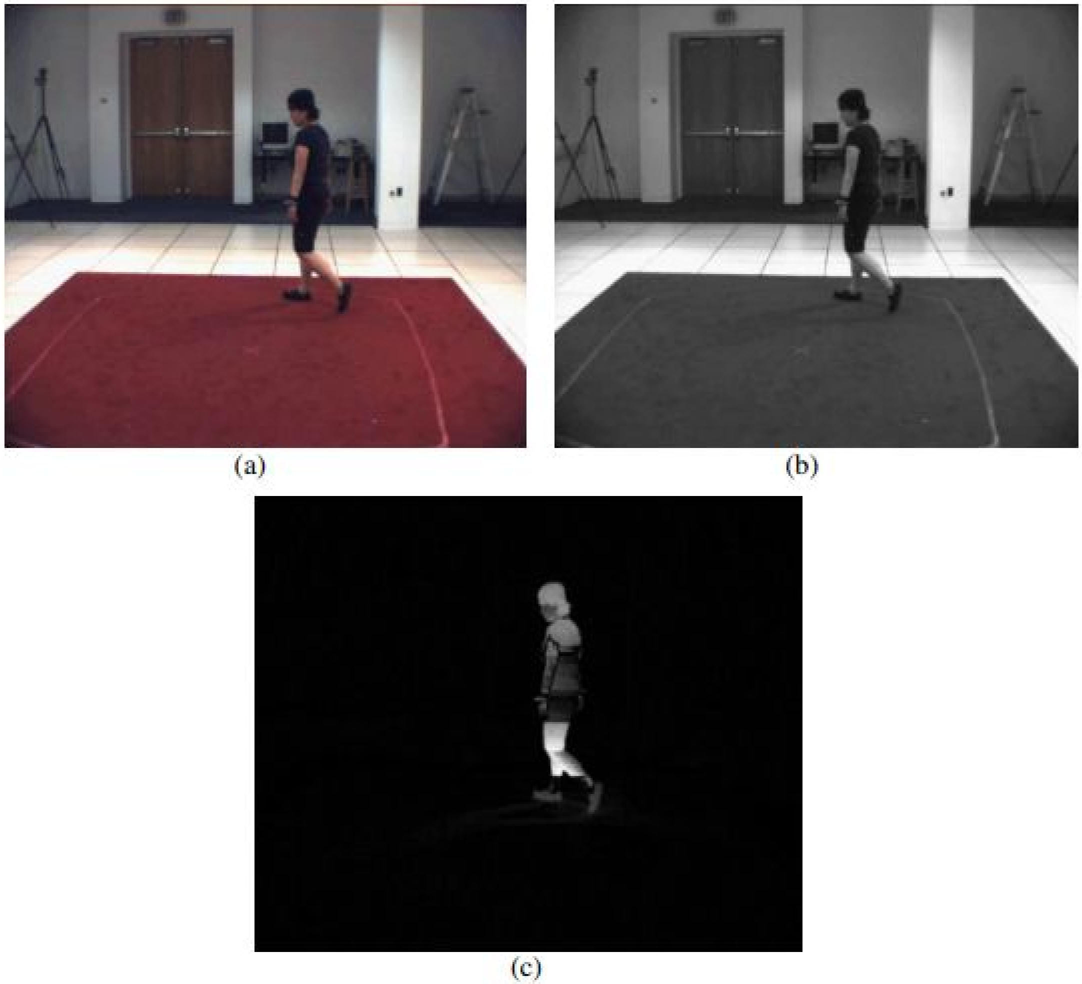

Fig 6 (a) walking video in original colour RGB. (b) Gray scale Frame.
(c) Subtracting Background Frame from Current Frame.

2.1.2 Temporal differencing

Temporal differencing makes the use of pixel-wise differences in between two or three consecutive frames in a video sequence to find motion regions. Temporal differencing is generally adaptive to changing environments, but usually does not work well for extracting all the appropriate pixels, for an example there might be holes present within moving entities. As an example of temporal differencing algorithm, Lipton et al. 1998 [27] detect moving objects in a real video streams using temporal differencing method. After the complete difference between the existing and the preceding image is obtained, a threshold formula is used to find the alteration. By using a connected component analysis, the segmented moving foreground fragments are grouped into the motion regions. An enhanced edition uses three image frames as an alternative to two frame differencing method. This technique is computationally less difficult and adaptive to vibrant modifications in the video frames. In temporal difference method, removal of changing pixel is fast and simple. Temporal difference may induce holes in foreground regions, and is more susceptible to the threshold value when finding the variations in difference of successive video frames. Temporal difference method requires a special supportive algorithm to detect the moving object which suddenly becomes stationary.

2.1.3 Optical flow

Optical-flow-based motion detection uses features of flow vectors of moving targets over time to identify moving pixels in a video sequence. For example, Meyer et al. 1998 [15], [17] calculate the displacement vector field to begin a contour based tracking algorithm, called active rays, for the extraction of articulated targets. The results are used for gait analysis. Optical flow based techniques can be used to recognize separately moving objects even in the presence of camera movement.

However, most flow computation techniques are computationally difficult and very sensitive to errors, and therefore cannot be applied to video sequences in real time simulation without using specialized hardware. Further comprehensive description of optical flow can be found in a work of Barron's 1994 [16]. Of course, besides the fundamental methods described above, there are several other ways for motion detection & segmentation. By means of the extended expectation maximization (EM) algorithm, Friedman et al. 1997 [12] implement a mixed Gaussian classification mixed model for every pixel. This model categorizes the pixel values into separate three fixed distributions corresponding to shadow, foreground and background. It also updates mixed component involuntarily for every class according to the probability of membership.

Thus, slowly moving objects are handled faultlessly, while shadows are eliminated much more effectively. Figure 2.3 shows optical flow method output. R. T. Collins et al. 2000 [28] has successfully developed a fusion system for motion segmentation by combining three-frame difference method with an adaptive background subtraction model. This hybrid algorithm is very fast and unexpectedly efficient for segmenting moving pixels in an image sequences. Also, E. Stringa 2000 [17] proposes a new morphological technique for detecting motion in scenes. This algorithm acquires steady segmentation results yet in altering ecological conditions.so Optical-flow-based motion detection uses features of flow vectors of moving targets over time to identify moving pixels in a video sequence.However, most flow computation techniques are computationally difficult and very sensitive to errors, and therefore cannot be applied to video sequences in real time simulation without using specialized hardware so This model categorizes the pixel values into separate three fixed distributions corresponding to shadow, foreground and background. It also updates mixed component involuntarily for every class according to the probability of membership

(a) (b)

Fig 7 (a) Frame of a Video Sequence. (b) Output of an Optical Flow Algorithm.

2.1.4 Object Tracking

The aim of a target tracker is to create the path of an object over time by locating its pixel position in all images of the video. Object tracker may also give the complete region in the image that is engaged by the object at each time instant. The tasks of finding the target and setting up correspondence between the object instances across the frames can also be performed jointly or separately. In the first case, probable object regions in each frame are obtained by means of an object detection technique, and then the tracker correspond objects across frames. In the latter case, the target region and correspondence is jointly estimated by iteratively updating object region and location information obtained from previous frames. In either tracking approach, the objects are represented using the appearance and shape models. The model selected to characterize object appearance limits the type of motion or deformation changes it can undergo. For an example, if an object is defined as a point, then only a translational model can be used. In the view where a geometric shape representation like an ellipse is used for the object, parametric motion models like projective transformations or affine are appropriate. These representations can approximate the motion of stiff objects in the scene. For a nonrigid object, shape or contour is the most descriptive representation and both nonparametric and parametric models can be used to specify their motion. Figure 9 represents the object detection taxonomy presented by Alper Yilmaz et al. 2006 [8]

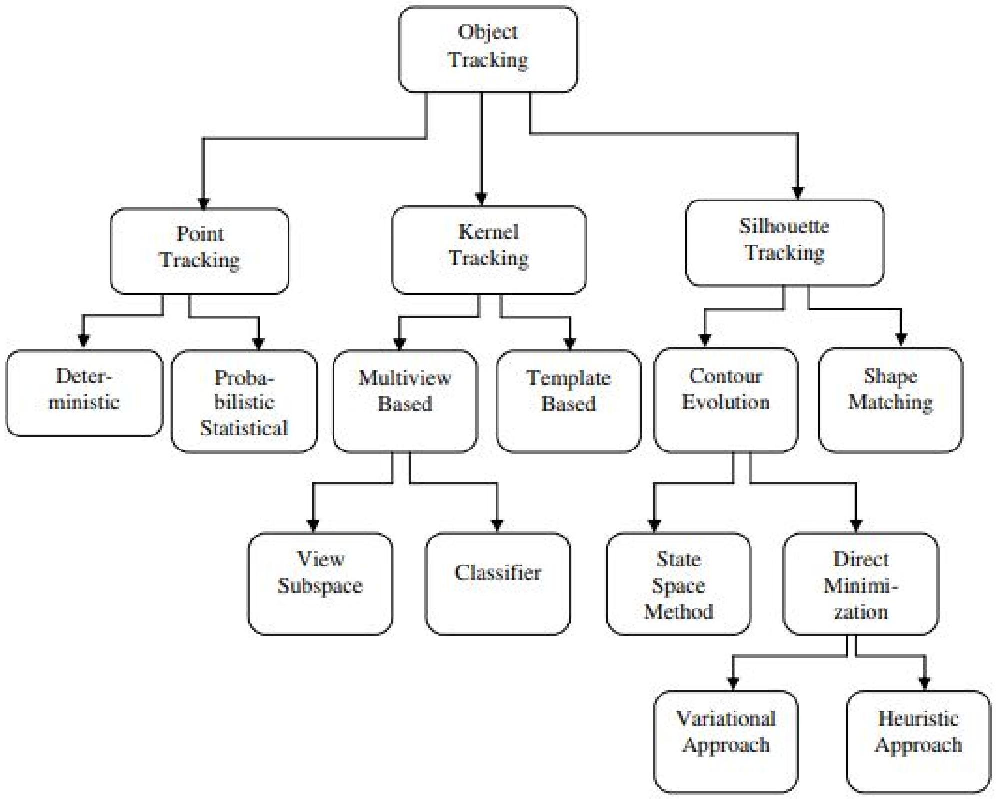

Fig 8 Taxonomy of Object tracking methods

2.2 Image Features

The image feature set needs to extract the most relevant features for object detection or classification while providing illumination changes invariance, differences in viewpoint and shifts in object contours. To attain this, rather than directly using images intensities or gradients, one frequently uses several forms of more advanced local image descriptors. V. R. Satpute, et al. 2014 [4] presented 2D - DWT and variance method as feature extraction for Human detection and tracking in surveillance applications. More such features can be based on points, Harris et al. 1998, Mikolajczyk et al. 2002 [17 & 19], blobs (Laplacian of Gaussian [Lindeberg 1998] or Difference of Gaussian, gradients [Ronfard et al. 2002, Mikolajczyk et al. 2004] [12], texture, colour, or combinations of some or all of these (Martin et al. 2004 [21]). The final descriptors need to symbolize the image adequately well for the detection and classification task in hand. For precise object representation, uniqueness and salience are the most important characteristics. One popular approach is to represent an object or an image by a set of local features, which are unique and distinguish the object from others. The Harris corners [11], the Kanade-Lucas-Tomasi (KLT) features 1991 [14] and the Scale Invariant Feature Transform (SIFT) 2004 [8] are three typical examples. These features can be applied to specific object recognition and tracking, image rectification and matching, 3D structure reconstruction from 2D images for scene or motion analysis, and even Simultaneous Localization and Mapping for robotics. Tracking systems are required to function in a wide variety of conditions. System must be able to function in both indoor and outdoor environments, and need to be able to deal with challenges such as illumination changes and changing weather conditions (i.e. fog, rain, snow). Other challenges such as occlusions are common place in real world scenarios, and system must be able to maintain an objects identity and a reasonable approximation of its position during these occlusions. In order to overcome some of these problems, algorithms that are able to utilise multi-camera setups have been developed. This however introduces additional challenges such as camera calibration, track handover between views, and how to utilise cameras

The various approaches are divided into two broad categories: sparse representations based on points, image fragments or part detectors; and dense representations using image intensities or gradients.

2.2.1 Sparse Local Representations

Sparse matrix value representations are based on local descriptors of relevant local image regions. The regions can be selected using either key point detectors or parts detectors. Point Detectors The use of salient local points or regions for object detection has a long history. Tianzhu Zhang, Si Liu, Narendra Ahuja, et. al, 2014 [5] proposed Consistent low-rank sparse tracker (CLRST) that builds upon particle filter framework for tracking. Wei Zhong, Huchuan Lu, and Ming-Hsuan Yang in 2014 [3] provides an approach for Robust Object Tracking via Sparse Collaborative Appearance Model. Weber et al. 2000 , Schmid and Mohr 1997, Lowe 2001 , Agarwal and Roth 2002 [4], Fergus et al. 2003 [3], Dork´o and Schmid 2003 [18], Leibe et al. 2005 [6], Opelt et al. 2004 [7], Mikolajczyk et al. 2004 [7]. These approaches extract local image features at a sparse set of salient image points usually called points of interest or key points. The final detectors are then based on feature vectors computed from these key point descriptors. The hypothesis is that key point detectors select stable and more reliable image regions, which are especially informative about local image content. The overall detector performance thus depends on the reliability, accuracy and repeatability with which these key points can be found for the given object class and the informativeness of the points chosen. Regarding the computation of feature vectors or descriptors over the local image regions surrounding the key points, many approaches have been tried. Currently the most popular approaches are image gradient based descriptors such as the Scale Invariant Feature Transformation (SIFT) Lowe 2001, 2004 and shape contexts William Robson Schwartz et al. 2013 [9] Belongie et al. 2001, 2002. Both compute local histograms of image gradients or edges. SIFT uses the local scale and dominant orientation given by the key point detector to vote into orientation histograms with weighting based on gradient magnitudes. It therefore computes rotation and scale invariant feature vectors. The scale information is also used to define an appropriate smoothing scale when computing image gradients. SIFT computes histograms over rectangular grids, whereas shape contexts use log-polar grids. The initial shape context method (Belongie et al. 2002 [19]) used edges to vote into 2-D spatial histograms, but this was later extended to generalised shape contexts by Mori and Malik [2003] who use gradient orientations to vote into 3-D spatial and orientation histograms with gradient magnitude weighting similar to SIFT. Part or Limb Detectors Local "parts"-based detectors are also widely used in object and human recognition systems Forsyth and Fleck 1997 , Ioffe and Forsyth 1999 , Schneiderman and Kanade 2004, Ronfard et al. 2002, Ramanan and Forsyth 2003 , Sigal et al. 2003 . For example, Forsyth and Fleck , Ioffe and Forsyth and Ramanan and Forsyth use precise human body segments (upper leg, torso, forearm, upper arm, lower leg, etc.) which are assumed to be well represented by cylinders. Parallel edge detectors are then used to detect the corresponding image segments, and body-geometry based detectors are generated by means of articulation constraints or graphical models to constrain the relative geometry of the limbs. 3D limb detectors have also been used, c. f. Sigal et al. One problem with these approaches is that the assumption that limbs can be represented by parallel lines is rather simplistic and its scalability to real world examples is questionable. This may explain the lack of extensive testing on real world images in these works.

2.2.1.1 Trajectory based Detectors

B. Morris et al. 2008 Trajectories, derived from the location of particular points on an object in time, are very trendy since they are comparatively simple to extract and their interpretation is obvious. The generation of trajectories of motion from a sequence of images typically involves the detection of tokens in each frame and the relevance of such tokens from one frame to another frame. The tokens need to be distinctive enough for easy detection and steady during time so that they can be detected & tracked. Tokens include edges, corners, regions, interest points, and limbs. Several proposed solutions for human actions modeling and recognition using the point-based features approach. In the first step, a random changing number of objects are tracked. From the record of the states of tracked object, temporal trajectories are formed which illustrates the motion paths of these objects. Secondly, typical motion patterns are learned by e.g. clustering these trajectories into the prototype curves. In the concluding step, motion recognition is then tackled by tracking the position of pixels within these prototype curves based on the same method used for the object tracking.

2.2.1.2 Texture

The Texture is measure of intensity variation of a surface which enumerates characteristics such as smoothness and regularity. Compared to colour, texture requires a processing step to create the descriptors. There are different texture descriptors: GrayLevel Cooccurrence Matrices (GLCM's) [18] (a 2D histogram which shows the Cooccurrence of intensities in a specified direction and distance), Law's texture measures [18] (twenty-five 2D filters produced by five 1D filters corresponding to level, edge, spot, wave, and ripple), wavelets [16] (orthogonal bank of filters), and steerable pyramids [19]. Like to edge features, the texture features are less sensitive to illumination changes compared to colour.

2.2.2 Dense Representation of Image Regions

Another approach is to extract image having significantly maximum number of features (often pixel-wise) over an entire image or detection window and to collect them into a high-dimensional descriptor vector that can be used for the discriminative image classification or labelling window as object or non-object.

2.2.2.1 Regions and Fragments Based on Image Intensity

Region-based tracking method tracks objects according to variations of the image regions corresponding to the moving targets. For these algorithms, the background image is maintained vigorously (K. Karmann et al. 1990 [14], M. Kilger et al. 1992 [11]), and motion regions are generally identified by deducting the background from the current image. Wren et al. 1997 [18] quoted the use of small blob features to track a single human in an indoor environment. McKenna et al. 2000 [12] propose an adaptive background subtraction method in which colour and gradient information are combined to cope with shadows and unreliable colour cues in motion segmentation. Tracking task is then performed at three different levels of abstraction: people, regions, and groups. Each region has a bounding box and regions can merge and split. Accordingly, these techniques cannot satisfy the needs for surveillance against a multiple moving objects or with cluttered background.

2.2.2.2 Edge Based Detectors

Image edges and gradient filters have also been used for object detection. Active contour-based tracking algorithms track moving targets by representing their outlines as bounding contours and updating these contours dynamically in successive frames [6], [9], [10]. These algorithms aim at directly extracting shapes of subjects and provide more effective descriptions of the objects than region based algorithms. Paragios [18] detect and track multiple moving objects in an image sequences using a geodesic active contour objective function and a level set formulation scheme. Peterfreund [12] explores a new active contour model based on a Kalman filter for tracking nonrigid moving objects such as people in spatio-velocity space. Isard et al. 1996 [11] adopt stochastic differential equations to describe complex motion models, and merge this approach with deformable templates to cope with people tracking. D. Koller et al. 1994 [16], Malik et al. 1997 [19] have successfully applied active contour-based methods to vehicle tracking.

2.2.2.3 Wavelet Based Detectors

Some well known approach to object detection are described in Yang Songy, Xiaolin Fengy 2000 [13], Papageorgiou and Poggio 2000 [17], Mohan et al. 2001, Viola and Jones 2003 [7]. These approaches use dense encoding of image regions based on operators similar to Haar wavelets. Tan et al. 1994 [16] propose a generalized Hough transformation algorithm based on a single characteristic line segment matching to calculate vehicle pose. Further, Tan et al. 1994 [18] analyze the one-dimensional (1-D) correlation of the image gradients and determine the vehicle pose by voting. As to the refinement of the vehicle pose, research group in the University of Reading have utilized an independent 1- D searching method [16] in their past work. Pece et al. 2000 [12], introduce a statistical Newton method for estimating the vehicle pose. Recently, Kaihua Zhang, et al. 2014 [6] formulates the spatio-temporal relationships between the object of interest and its regionally dense contexts in a Bayesian framework, which models the statistical correlation between the simple low-level features (i.e., image position and intensity) from the target and its surrounding regions.

2.2.2.4 Hough Transform

Hough transform is a popular technique for finding primitives in images like lines and circles R. O. Duda et al. 1972 [8]. The main idea of the method is to collect the number of occurrences the sought primitive over the primitive parameters. In case of a line for instance, the Hough space is composed by line orientation and line intersection with the x axis. A slightly changed version of this algorithm has been applied for object detection. In this case the Hough space is composed by position and scale of the center of the object bounding box (assuming to deal with a single aspect ratio object). During training the correspondence of a visual word to the object center is learned. Then in testing, every new feature is associated with the closest visual word and a vote on the possible object center is added. The locations with more votes are the final detections. The advantage of this method is that is not necessary to really evaluate the score of the detector on each bounding box. However, it has been noticed that the Hough space is quite noisy because a single hypothesis can collect votes from different objects. Thus, generally the Hough procedure is improved with a final refinement often based on standard SVM classification [20, 16 & 14]. Similarly to Hough transform, the so called jumping window (A. Vedaldi et al. 2009 [21]) has revealed a useful strategy to sample image windows for detection. As in Hough transform detection hypotheses are generated from visual words: the words with best discriminative power and stable location in the bounding boxes of the training samples are used to generate detection hypotheses. In [22] it is shown that with a reduced set of hypotheses jumping windows can retrieve more than 90% of the objects in the dataset.

2.2.2.5 Contour, shape & Skeletal representation

- **Object contour and silhouette:** Contour representation defines the boundary of an object Figure 9 (a) & (b). The region inside the contour is called the silhouette of the object. Contour and silhouette representations are suitable for tracking complex non-rigid shapes [Yilmaz et al. 2004] [17].
- **Articulated shape models:** Articulated objects are composed of different body parts that are held together by means of joints. For example, the human body structure is an articulated object with torso, legs, head, hands, and feet connected by joints. The relationships between the bodies parts are governed by kinematic motion models, for example, joint angle, etc. In order to characterize an articulated object, one can model the constituent parts using cylinders or ellipses as shown in Figure 9 (c).

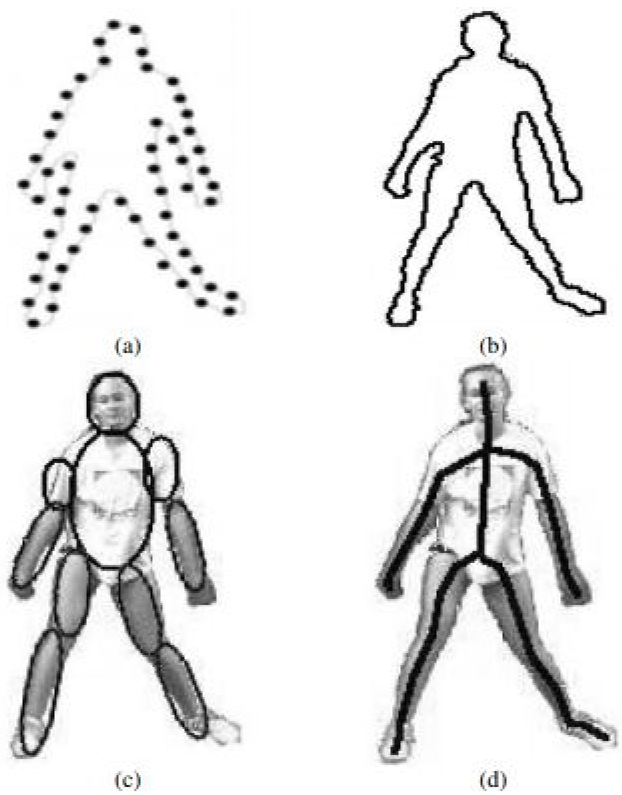

Fig 9 Object representations. (a) Complete object contour, (b) control points on object contour, (c) part-based multiple patches, (d) object skeleton.

- **Skeletal models:** Object skeleton can be extracted by applying medial axis change to the object outline [Ballard and Brown 1982] [17]. This model is commonly used as a silhouette representation for identifying objects [Ali and Aggarwal 2001] [12].

Skeleton representation can be used to model both articulated and rigid objects Figure 9 (d).

2.3 Image Classifications

Classification methods can be divided into discriminative approaches such as shape based, & motion based. Different moving regions may correspond to different moving objects in natural scenes. For instance, the image sequences captured by the surveillance cameras mounted in road traffic scenes possibly include vehicles, humans and other moving targets such as moving clouds and flying birds etc. To further track objects, it is essential to correctly classify moving objects. The target categorization can be considered as a model pattern recognition issue. At present, there are two major categories of approaches for classifying moving objects.

2.3.1 Shape-based classification

Different descriptions of shape information of motion regions such as blobs, silhouettes, points, boxes are available for classifying moving objects. Vikas Reddy, Conrad Sanderson 2013 [10] suggested Block-based Classifier Cascade with Probabilistic Decision Integration which can be formulated as a binary classification problem. VASM 2000 [12] takes image blob area, image blob dispersedness, apparent aspect ratio of the blob bounding box, etc, as key characteristics, and classifies moving-object blobs into four classes: vehicles, single human, human groups, and clutter, using a viewpoint-specific three-layer neural network classifier. Lipton et al.1999 [23] use the dispersedness and area of image blobs as classification metrics to classify all moving-object blobs into vehicles, humans, and clutter. Temporal consistency constraints are considered so as to make classification results more precise. Kuno et al. 1990 [12] use simple shape parameters of human silhouette patterns to separate humans from other moving objects.

2.3.2 Motion-based classification

In general, non-rigid articulated human motion demonstrates a periodic property, so this has been used as a sturdy signal for classification of moving objects. Cutler et al. 2000 [13] describe a similarity-based technique to analyze and detect periodic motion. By tracking an interesting moving target, its own-similarity is approximated as it progress over time. As we know, for periodic motion, its own-similarity measure is also periodic. Hence time-frequency analysis is applied to characterize and detect the periodic motion, and tracking and classification of moving objects are implemented using periodicity. In 1999, Lipton’s work [14], residual flow is used to analyze rigidity and periodicity of moving objects. It is expected that rigid targets present small residual flow; while a nonrigid moving object such as a human being has generally higher standard residual flow and still display a periodic component. Based on this useful cue, individual motion is separated from motion of other objects, such as vehicles. The two common approaches mentioned above, namely motion-based and shapebased classification can also be successfully combined for categorization of moving objects. Furthermore, C. Stauffer 1999 [15] proposes a novel method which relies on a time co-occurrence matrix to hierarchically categorize both objects and behaviours. It is expected that more accurate classification results can be generated by using extra features such as colour and velocity.

2.3.3 Other Methods

Agarwal and Roth 2002 [9] use the Perceptron-like method Winnow as the underlying learning algorithm for car recognition. The images are symbolized as binary feature vectors and classification is done by using a trained linear function over the feature space. Dong Wang 2013 [11] adopts algorithm with sparse prototypes, which uses both standard principal component analysis (PCA) model with new sparse representation schemes for learning effective appearance models.

Tal Ben-Zvi, Jeffrey V. Nickerson 2012 [14] uses sensor network aims to protect a stationary target (e.g., a large aircraft at anchor) and identifies signals from approaching objects (e.g., small ships travelling in a harbour).

William Robson Schwartz 2013 [9] uses Partial Least Squares Analysis method that augments widely used edge-based features with texture and colour information for human motion detection. Leonid Sigal et al. 2011 [17] propose Particle Message Passing (PAMPAS) & Non-parametric Belief Propagation algorithm for Estimating 3D Human Pose and Motion. A Discriminative Model of Cross and Motion Ratio for Invariant View Action Recognition are used by Kaiqi Huang 2011 [18] for video based view invariant action recognition system. In 2010, Golnaz Abdollahian et al. [13] use a location-based saliency map which is produced based on camera motion features. This map is combined with other saliency

maps generated using features such as colour contrast, object motion and face detection to determine the ROI. One of the most advanced systems research efforts in large-scale supervision systems is the united state control program titled Combat Zones That See [16]. This program explores rapidly deployable intelligent camera tracking systems that communicate over improvised wireless networks, transmitting track information to a central station for the uses of activity recognition and long-term movement pattern analysis

The Markov chain Monte Carlo (MCMC) algorithm was employed to recursively approximate the solution of multi-target data association problem by Jianguo Lu, Anni Cai, 2010 [29] for multi target detection. In 2009, Koichiro Goya, Xiaoxue Zhang et al. [8] presented a method that make use of the morphological functional information between detected objects in video sequence to classify the video scenes. However, the system is limited to the longitudinal viewpoint of the object. David A. Ross Jongwoo Lim 2008 [11] describes Incremental Learning for Robust Visual

Tracking method that incrementally learns a low-dimensional subspace representation, powerfully adapting online to changes in the form of the object but, suffers from the limitation of target detection in cluttered background conditions.

Viola and Jones 2001 [104], Viola et al. 2003 [23] use AdaBoost to train cascades of the weak classifiers for pedestrian and face detection, using spatial and temporal difference-of-rectangle based descriptors. Opelt et al. 2004 [24] use a similar AdaBoost framework for their interest point based weak classifiers.

Veenman et al. 2001 [17] extend the work of Sethi and Jain 1987 [18], and Rangarajan and Shah 1991 [19] by introducing the frequent motion restraint for the correspondence. The common motion constraint provides a strong constraint for coherent tracking of the different points that lie on the same object; however, it is not suitable for points lying on isolated objects moving in various directions.

The method is initialized by generating the initial tracks using a two-pass algorithm, and the time function is reduced by Hungarian assignment algorithm in two consecutive frames.

This approach can handle misdetection and occlusion errors; nevertheless, it is presumed that the numbers of objects are identical all through the sequence, that is, no object entries or exits.

2.4 Occlusion Detection

In practice, occlusions and self-occlusion between different moving objects or between moving objects and the background are unavoidable. Multiple camera systems present promising and efficient methods for coping with occlusion. Utsumi et al. 1998 [20] utilize multiple cameras to track people, successfully determining the mutualocclusion and self-occlusion by choosing the finest view. Multiple target tracking becomes a data association problem where detection responses need to be reliably linked to form target trajectories. Nevertheless, this is still a complex and only partially solved problem. Several recent tracking algorithms address the association problem offline, i.e., by optimizing detection assignments over large temporal windows, e.g., K-shortest paths, Hungarian algorithm, and hyper graphs [12, 19 & 20]. Recently, Horst Possegger et al. 2014 [7] proposed an online multi-object tracking-by-detection approach for real-time applications using geometric information to efficiently overcome detection failures

.

Shafique and Shah 2003 [25] propose an approach of multiframe to protect temporal coherency of the position and speed. For misdetected or occluded objects, the path will consist of absent points in consequent frames. The directed graph, which is produced using the points in k frames, is transformed to a bipartite graph by splitting each object into two (- and +) nodes and representing directed boundaries as undirected edges from − to + nodes. The association is then established by a greedy algorithm. They employ a window of frames during point correspondence to deal with occlusions whose durations are shorter than the temporal window used to perform matching.

Qian Zhang et al. 2011 [21] proposes adaptive background penalty with occlusion reasoning to separate the foreground regions from the background in the first frame. Multiple cues are employed to fragment individual objects from the group. To spread the segmentation through video, each object region is independently tracked by uncertainty refinement and motion compensation and the motion occlusion is tackled as layer transition.

Nizar Zarka 2005 [9] describes a real-time system for motion analysis; human detection and tracking. Motion segmentation and tracking are attained through several steps: First, design of a robust, adaptive background model that can deal with objects occlusions, lightning changes and long term changes in the scene. This method is used to extract foreground object pixels by using the background subtraction method. Afterwards,object detection and noise cleaning are applied, followed by human pose modeling to monitor and recognize human activity in the scene such as human walking or running.

However, system suffers from limitation of tracking with camera motion, recognizing various types of human moving activities such as jumping, identifying other moving objects like animals and vehicles, & shadow detection.

To reduce ambiguities due to occlusion, better methods need be developed to cope with the correspondence between body parts and features, and consequently eliminate a correspondence error that arises throughout tracking multiple objects. When objects are occluded by stationary objects such as street lamps, and buildings a few resolution is achievable through motion region analysis and partial matching. Although, when multiple moving objects occlude each other, principally when their directions, shapes and speeds are very near, their motion regions combines, which makes the segmentation and tracking of objects predominantly difficult.

The self-occlusion of a human body is also an important and complex problem. Interesting progress is being made using arithmetical methods to predict object position, pose and so on, from available image information.

2.5 Evolution of object detection

Object detection on static images started in the seventies. Fishler et al. 1973 [21] introduced the pictorial structure model that proposes to detect an object as a set of independently-learned parts that have spring-like geometrical constraints. Currently, most of the state-of-the-art methods are still based on that deformable model [13, 17].

In the year 1994, S. H. Courellis and V. Z. Marmarelis [1] introduced artificial neural network for motion detection with non-linear spatio-temporal features. System structure was fairly simple and straight forward, and the process was carried out in three layers. The input to the network is moving spots and edges with a number of velocity profiles. The network uses the principle of directional selectivity to determine the direction of motion, by selectively activating output nodes.

Further in 1995 [2] Scott A. Nichols and W. Brian Naylor developed Advanced Exterior Sensor (AES) program for highly adaptive video motion detection and tracking.
In the early 2000 object detection using an adaptive background subtraction method was developed. I. Haritaoglu, et al. [7] focused on Scene Modeling and Maintenance for Outdoor Surveillance using fast background subtraction. Later on, Yang Song et al. [4] presented an unsupervised learning model that can derive the probabilistic dependence structure of parts of an object (a moving human body) automatically from unlabeled data. The distinguished part of their work was that it was based on unlabeled data, i.e., the training features set contains both useful foreground parts and background clutter and the correspondence between the parts and detected features were unknown. Unfortunately, the method does not work very well due to the highest degree of variability that a pedestrian can have depending on clothes and pose he was assuming.
In 2005, Pierre-Marc Jodoin et al. [11] introduced accurate and fast motion detection in the existence of camera jitter by applying background subtraction to video scene dynamics as an alternative of scene photometry. In their work, an object is assumed moving if its dynamical behaviour is unlike from the standard dynamics observed in a reference sequence. Subsequently, several improvements on the object tracking have been proposed [18, 12, 11 & 10]
Further in 2005, evolution in the field of object detection is generated by the introduction of the Weizmann challenge [26] that presents an object detection challenge, where the proposed methods are applied on a large dataset of 9 different object categories performing 10 natural actions.
An important step forward for enhancing detection accuracy is the work of Vikas Reddy, Conrad Sanderson et al. 2013 [10], where the authors propose a block-based method capable of dealing with noise, dynamic backgrounds, illumination variations and while still obtaining smooth curve of foreground objects. Specifically, image sequences are analysed on an overlapping segment-by-segment basis.A low-dimensional texture descriptor obtained from each block is passed through an adaptive cascade classifier, where every stage deals a distinct problem. A probabilistic foreground mask generation approach then use blocks overlaps to incorporate interim block-level decisions into final pixel-level foreground segmentation.
Table 2.1 provides the list of methods for object detection in the last few years. Each method is described using the year of publication, features used, classifier & network used.

2.6 Motivations for Proposed Work

Despite the tremendous progress of Computer Vision during the last four decades, the general vision problem is far away from being solved. As a consequence, current vision applications need to be quite heavily constrained to produce meaningful results. People detection and tracking is no exception. 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. However, this area is still in a transition step between image processing and pattern recognition. With respect to motion segmentation, it can be concluded that although remarkable advances have been achieved by presenting a wide set of different approaches; the segmentation task is still an open problem. These techniques must be enhanced to cope successfully with the numerous difficulties expected, specifically in outdoor scene.

Among these difficulties it includes different weather conditions, lighting changes, background in motion like water of river, leaves of trees, shadow of an object etc. Further it is still not clear how to deal with background objects which unexpectedly move at a given moment with foreground objects which stop momentarily. The solution may come from the combination and development of some of the existing approaches, thereby providing the system with redundancy. Taking advantage of context knowledge and making use of high level information may also be a way of solution.

With respect to tracking, numerous approaches have been proposed to perform this task. Data association techniques on their own are not reliable enough, since they completely depend on a proper segmentation. Prediction updating approaches should be flexible and general enough to cope with complex environments. The combination of several of the aforementioned techniques may lead to a way of solution; several segmentation and tracking methods may be combined with prediction updating techniques in order to provide the system with error recovery capabilities.

From the above literature survey, it is concluded that some sort of structured architecture with cooperative levels are needed in order to cope with such a complex problems as the analysis of human motion.

CHAPTER-3
PROPOSED MODEL

In this system we will try to build a system using matlab which can detect humans falling on the ground from a CCTV camera feed and so that we can limit the physical damage to the person by alerting the hospital authorities.The heart of this system is computer vision and edge detection algorithm.The video is processed frame by frame in the computer vision algorithm,it draw outline to the edge of the object,that traces the object movement.The algorithm detects the falling of human body by determining all the factors that include instantaneous speed,average speed,area change ,orientation change,average area change ,average orientation change.

When the object changes its speed ,orientation and area above a threshold value as determined in the algorithm,the body is concluded to be fall down.And sms is then send to the family member of the person who fell down,this is done using php CURL,and the hospital authorities are also,informed.

Firstly we import the video into the project,the video can be in any format including mp4,avi and various other stander format , the video imported in the file is then segmented in and various layers at 10 frame per second ,these frames are then send to the edge detection algorithm which determine edge of the image and then the parameters such as speed , area , orientation is calculated by the respective algorithm,these parameter are calculated for every frame of the image and then compared if the changing parameter is larger than the threshold value if it is then the object is consider to be fallen if the system detect object to be fallen then a sms api is requested by the hook in the system which connects to the backend php engine,this php engine request sms api with the help of curl command,then a sms is sent to the relative and the hospital ,awaring them about the emergency by the person who is falling.

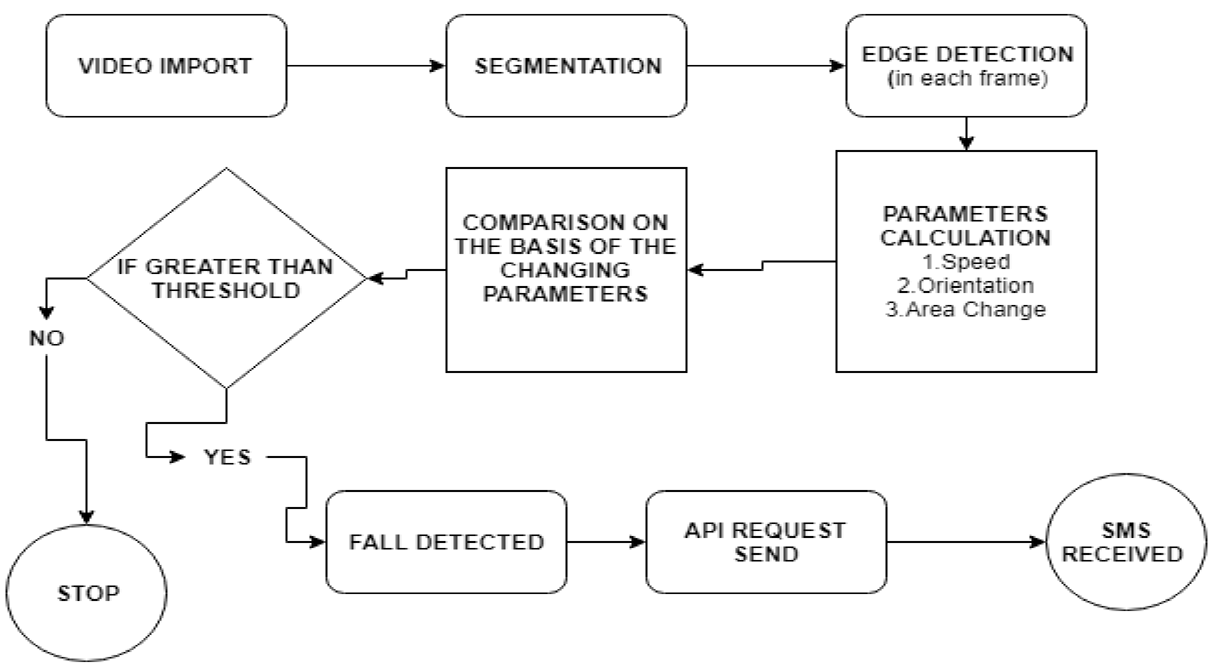

Fig 10 Flow chart of proposed modal

3.1 Video Import

The First block is the video import block ,in this block we import the video file from the desired source.we can import video in various format which include mp4, avi, mov, mkv and other standard format.These stander video format which can be imported are as follows

MP4
The .MP4 container is probably the closest thing to a universal standard that currently exists. It can use all versions of MPEG-4 and H.264 and is compatible with a huge range of players. Videos using the .MP4 container can have relatively small file sizes while retaining high quality. Many of the largest streaming services, including YouTube and Vimeo, prefer .MP4.

AVI
One of the oldest and most universally accepted video file formats is .AVI. It can use an enormous range of codecs, resulting in a large variety of different file settings. While .AVI videos can be played on a wide range of players, file sizes tend to be large making it less ideal for streaming or downloading. It’s a great option for videos you plan to store on a computer.

MOV
Apple developed the .MOV container to use with its Quicktime player. Videos using .MOV generally have very high quality but also fairly large file sizes. Quicktime videos don’t have as much compatibility with non-Quicktime players, though there are third party players that will read them.

FLV
Made for Adobe’s Flash player, .FLV videos were extremely common for a number of years thanks to their very small file size and a wide range of browser plugins and third party Flash video players. There has been a significant decline in Flash videos recently.

WMV
Windows Media videos tend to have the smallest file size, which makes them a good option if you need to send through email or other methods with file size limits. However, this comes with the tradeoff of having a significant drop in quality. A common use for .WMV is emailing video previews to clients.

3.2 Segmentation

Image segmentation is the process of dividing an image into multiple parts. This is typically used to identify objects or other relevant information in digital images. There are many different ways to perform image segmentation, but we are performing it by using Gaussian mixture models.

The ForegroundDetector compares a color or grayscale video frame to a background model to determine whether individual pixels are part of the background or the foreground. It then computes a foreground mask. By using background subtraction, you can detect foreground objects in an image taken from a stationary camera.

To detect foreground in an image :

- Create the vision.ForegroundDetector object and set its properties.
- Call the object with arguments, as if it were a function.

3.2.1 Algorithm for background Elimination

1. Read video using mmreader command in MATLAB.
2. Extract number of frames, height and width of frames.
3. Pre-allocate output video structure with all its elements assigned value 0.
4. Divide each frame into blocks of size 16x16 each.
5. Compare blocks of first frame with the corresponding blocks of each frame using mean squared error(MSE) concept.
6. If MSE is less than 5 percent the blocks are considered to be matched.
7. If more than half the corresponding blocks match, the blocks are considered to be a partof background and are made white in the output video structure.
8. If none or less than half the number of corresponding blocks does not match, they are considered to be moving object and original video values are assigned to those blocks in the output video structure.
9. The output video now contains only the moving object and background is eliminated.

3.3 Edge Detection

Edge detection is an image processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation and data extraction in areas such as image processing, computer vision, and machine vision.This technique is carried out to detect the following

3.3.1 Human Detection from background eliminated video

A shape-based approach for classification of objects is used following background subtraction based on frame differencing. The goal is to detect the humans for threat assessment. The target intruder is classified as human or animal or vehicle based on the height to width ratio (H/W) of the moving object detected during background subtraction.

3.3.2 Algorithm for human detection

1. Read the original and background eliminated videos using mmreader function in MATLAB.
2. Extract the number of frames and frame size.
3. Pre-allocate output video structure assigning zeros to all its elements.
4. Make the pixels corresponding to moving object white and the rest black.
5. For each frame number of blocks containing moving object are checked to satisfy H/W ratio depending on area covered by camera.
6. Draw top, bottom, left and right lines in red color to highlight the detected human.

3.4 Parameter Calculation

We calculate parameter such as speed ,area change and orientation change to define if the body is falling or not these parameter are calculated on the basis of the edge detected in the last block. Speed is calculated by determining the change of the edge from the past value,where the first frame is determined as the beginning frame or the reference frame, all the speed is calculated on the basis of this reference frame.Then the orientation is calculated and similar to the calculation of the speed it is calculated on the basis of the edge detected and determining its change by setting the first frame as the reference frame.The area is also calculated in the similar way ,the area is determined by calculating the area of the closed edge figure generated by the edge detection algorithm.The area change is calculated on the basis of the last frame as well as the reference frame.

3.5 Fall Detection

The parameters are calculated and the average change in the parameter are also calculated , if the changing parameter exceed the value of the threshold value as defined in the algorithm the algorithm will consider the object to be fallen down,or is falling.

3.6 SMS is Send

The final step of this system is sms sending system, if the fall is detected ,this system does not require any kind of mobile device attached to it.The sms is sent with the help of the desired api integration,in this project we are request the php api with parameter which request the way2sms api with the same parameter,which in turn send the sms to the required authorities and thus we get the alert sms ,which make us aware about the injury happened with the patient.

CHAPTER-4
RESULTS AND DISCUSSION

This study developed a fall detection system based on an matlab. Identifying the video patterns and extracting parameters that can be used to distinguish falls from daily activities and sending the notification sms to the hospitals and the relatives of the falling person.This solution will help many old age people in emergency.And can incorporate in many automation system to detect the falling,so that injury of the person can be reduced.

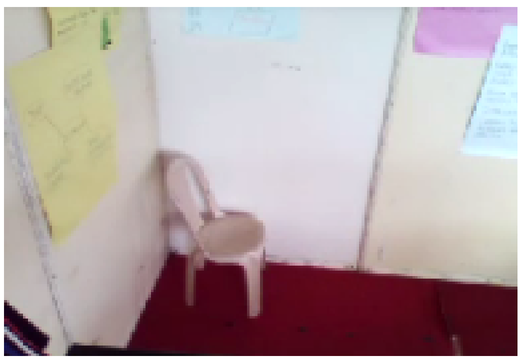

Fig 11 Initial Frame

Fig 12 Video Current Frame

Fig 13 Human Detection Current Frame

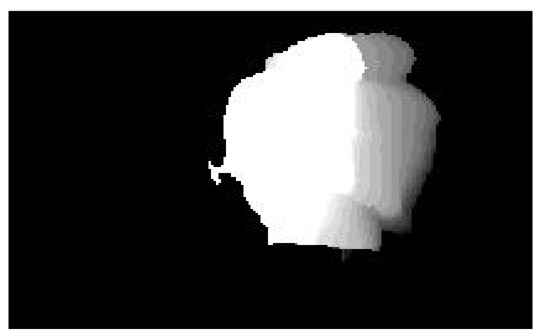

Fig 14 Motion history image
to calculate speed

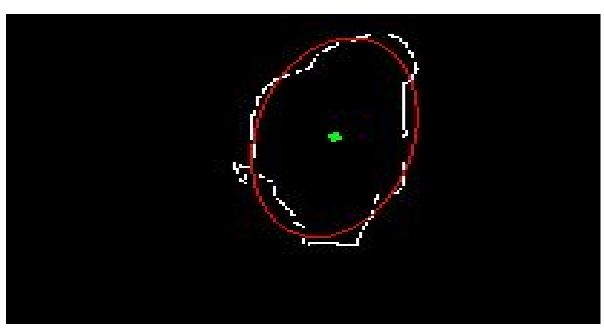

Fig 15 Shape of body to calculate orientation

REFERENCES

[1] CHUN-MING LI, YU-SHAN LI, SHU-HAI WANG, XIU-QING ZHANG: MovingHuman Body Detection In Video Sequences. 2007: 19-22.

[2] Ismail Haritaoglu, David Harwood, Larry S. Davis. W4: Real-Time Surveillance of People and Their Activities. IEEE Transactions on Pattern Analysis and MachineIntelligence. 2000, 8, 22(8): 809-830.

[3] C. Wren, A. Azerbayejani, T. Darrell, et al. Pfinder: A Real-Time Tracking of Hu man Body. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997, 7, 19(7): 780-785.

[4] 19Robert T. Collins, Alan J. Lipton, Takeo kanade, et al. A System for Video Surveillance and Monitoring: VSAM Final Report. Carnegie Mellon University. Technical Report: CMU-RI-TR-00-12. 2000.

[5] Oren M, Papageorgious C, Sinha P, et al. Pedestrian Detection using Wavelet Tem- plates. Proceedings of the IEEE Conference Computer Vision and Pattern Recognition. Puterto Rico: IEEE Press. 1997: 193-199.

[6] Lipton A, Fujiyoshi H, Patil R. Moving Target Classification and Tracking from Real- time Video. In: Proc IEEE Workshop on Applications of Computer Vision, Princeton, NJ. 1998: 8-14.

[7] Ross Cutler, Larry S.Davis. Robust Real-Time Periodic Motion Detection, Analysis, and Applications. IEEE Transactions on PAMI. 2000, 8, 22(8): 781-796.

[8] F.Liu, R.Picard. Finding Periodicity in Space and Time. Proc. Intl Conf. Computer Vision. 1998, 1: 376-383.

[9] R. Polana, R. Nelson. Detection and Recognition of Periodic, Non-Rigid Motion. Intl J. Computer Vision. 1997, 6/7, 23(3): 261-282

[10]David A.Forsyth, Jean Ponce. Computer Vision: A Modern Approach [M]. Beijing: Tsinghua University Press.2003.

[11]Clement Fredembach, Michael Schroder, Sabine Su sstrunk. Eigenregions for Image Classification. IEEE transctions on Pattern Analysis and Machine Intelligence. 2004, 12, 26(12): 1222-1229.

[12]Dubuisson M P, Jain A K Contour Extraction of Moving Objects in Complex Out- door Scenes. [J] International Journal of Computer Vision, 1995, 1, 14 (1): 83105.

[13]R. T. Collins, A. J. Lipton, and T. Kanade, Introduction to the special section on video surveillance, IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 745746, Aug. 2000.

[14]C. Regazzoni, V. Ramesh, and G. L. Foresti, Special issue on video communications, processing and understanding for third generation surveillance system, Proc. IEEE,vol. 89, no. 10, pp. 13551367, Oct. 2001.

[15]Pavlidis, V. Morellas, P. Tsiamyrtzis, and S. Harp, Urban surveillance systems: From the laboratory to the commercial world, Proc. IEEE, vol. 89, no. 10, pp. 14781497, Oct. 2001.

[16]R. T. Collins, A. J. Lipton, H. Fujiyoshi, and T. Kanade, Algorithms for cooperativemultisensor surveillance, Proc. IEEE, vol. 89, no. 10, pp. 14561477, Oct. 2001.

[17]E. F. Lyon, The application of automatic surface lights to improve airport safety, IEEE Aerosp. Electron. Syst. Mag., vol. 8, no. 3, pp. 1420, Mar. 1993.

[18]G. Donohue, Vision on aviation surveillance systems, in Proc. IEEE Int. Conf. Radar, 1995, pp. 14.

[19]M. J. Cattle, The use of digital CCTV in an airport carpark application, in Proc. IEEE Int. Carnahan Conf. Security Technology, 1995, pp. 180185.

[20]Partial Least Squares (PLS) Analysis, Edge-Based Features With Texture & Color Information and Human Detection Method Using Richer Descriptor Set Including Edge-Based Features , in IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009
[21]Dense Spatio-Temporal Context For Visual Tracking Bayesian Framework Two Local Context Models Spatial Context & SpatioTemporal Context Models ,Conference: European Conference on Computer Vision, September 2014
[22]Spatio-Temporal Evolution, Horst Possegger, Thomas Mauthner, Peter M. Roth, Horst Bischof; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1306-1313
[23]Dense Optical Flow & Filter, Robust Fingertip Detection in a Complex Environment; Article in IEEE Transactions on Multimedia ,June 2016
[24]Hybrid Approach,Gaussian Mixture Model with Optical Flow ,Combination of two methods accurate results but first 50 frames are required to initialize gaussian model;2010 International Conference on Electrical and Control Engineering,25-27 June 2010
[25]C.Rougier,J.MeunierFall,A.Arnaud and J.Rousseau. Detection from Human Shape and Motion History using Video Surveillance;Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops,2007
[26]Olfa Ben Ahmed;Mahmoud Mej doub;Chokri Ben Amar;New approach for moving point detection: Application to video space-time description,2013 International Conference on Computer Applications Technology (ICCAT)
[27]Hironobu Fujiyoshi,Moving target classification and tracking from real-time video;Conference: Applications of Computer Vision, 1998. WACV '98. Proceedings., Fourth IEEE Workshop
[28]Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., Zhang, Z.: Spatialtemporal convolutional neural networks for anomaly detection and localization in crowded scenes. Sig. Process. Image Commun. 47, 358–368 (2016)
[29]Yen, S.H., Wang, C.H.: Abnormal event detection using HOSF. In: 2013 International Conference on IT Convergence and Security, ICITCS 2013 (2013)