Object Tracking and Suspicious Activity Identification during Occlusion

Ravi Teja Yakkali
NXP Semiconductors
Noida
India

Raunaq Nayar
Nanyang Technological
University
Singapore

S. Indu
Professor
Delhi Technological University
Delhi, India

ABSTRACT

Rising criminal activities and demand of robust security solutions, detection and tracking of every minute detail of suspicious activity or object has become a topic of interest for researchers all around the world. In this paper, we propose an approach based on Digital Image and Video processing to detect and track the motion of multiple objects during the phenomenon of occlusion and activate an alert if an object is dropped for a long period of time in the region of concentration of camera. The proposed method can be utilized in video surveillance system and the method has been verified through extensive experimentation for multiple video.

General Terms

Image Processing, Video Processing, Computer Vision

Keywords

Occlusion, Digital Image Processing, Suspicious Object, Object Detection, Object Tracking, Video Processing

1. INTRODUCTION

Digital Image and Video Processing has been a hot research topic for the past few years due to increased processing and reduced hardware requirement particularly for security surveillance in the crowded areas where the chances of terrorist attacks and other criminal activities are growing day by day. This involves the concept of framing of a video into a sequence of images and then performing image processing as well as establishing relationship between all image frames to detect motion of multiple objects in the video. In this method of motion detection using background subtraction, a reference image is taken from the still video camera without including any objects which act as the background image. The background image is subtracted from the subsequent image frames in the video to detect objects in motion. For multiple object detection and tracking, difference in object features based on the area and grayscale level is usually taken into account. The main objective of the paper is to develop an algorithm which is used to identify a suspicious activity and track the multiple objects as well as retaining their identity which does not get affected during occlusion of multiple objects. The remainder of this paper is organized in the following manner. Section 2 describes related work in the corresponding field. Section 3 presents the problem statement. Section 4 presents the proposed algorithm of multiple object detection and tracking during occlusion. Suspicious activity identification method is proposed in Section 5. The corresponding algorithm of suspicious activity identification is explained in Section 6. The simulation results are illustrated in Section 7. Finally, we conclude with Section 8.

2. RELATED WORK

There have been many motion detection algorithms which have been in use like temporal difference technique [1], optical flow [2], [3] and background subtraction [4]. Motion detection is followed by median filter smoothening [5]. Identification of moving objects using pixel wise difference between consecutive image frames followed by application of classification matrix to identify objects as human, vehicle or background clutter [10] has also been carried out. Various other techniques of object tracking in video sequences through different phases using image processing [15] have also been discussed in detail. However, the detection methods do not take the effect of occlusion into account. Occlusion is the phenomenon in which, when multiple objects cross each other, at the point when one object overlaps other or when they touch each other on screen they behave like single object and after the occlusion, usually it is very difficult to determine the exact identity of each object. A method for detection of occlusion areas [8] and motion discontinuities using a stochastic approach to the estimation of 2D motion vector fields from time-varying images [9] is used to check whether the occlusion has occurred and to determine portion of the images where occlusion has occurred. This was further improved to accommodate [12] different methods to determine occlusion areas in high resolution images. Occlusion detection has also been carried out by judging the geometric relation between the projection of virtual points and projection of real points according to the theory of stereoscopic geometry in recent past [7]. Some more methods which have also been used are disparity maps [11], random forest method [14] to detect occlusion. Later, a method to detect and track multiple objects which do not get affected by occlusion was formulated based on difference in the area and the threshold level of the various frames in a video taken from a still camera. This was experimentally carried out on detecting and tracking multiple fish swimming in shallow water. [6]. Theoretical model of real fish occlusion tracking, imaginary fish occlusion tracking, and real and imaginary fish concurrent occlusion tracing aiming to achieve 3D fish tracking without occlusion, fulfilling the requirements beforehand for fish target behavior analysis [13] was also proposed recently.

3. PROBLEM STATEMENT

However, the existing algorithms do not detect any suspicious activity happening in the scenario and there is a need to extend the existing algorithms for detecting any suspicious activity in the scenario. So, an area as well as a grayscale level based multiple object detection technique for tracking of multiple suspicious objects along with the concept of raising a red alert if a person drops an apparently suspicious object or if theft occurs in the area under surveillance is proposed in the existing paper. Apart from the identification of suspicious

activity and raising red alert, the proposed method also identifies the exact location in the scenario where the suspicious activity has occurred. The proposed algorithm ensures stringent security check in the area under video surveillance.

4. PROPOSED ALGORITHM

Initially a reference image is taken from the still video camera which is assumed as default scenario. While taking the reference, the image must include all those objects whose position is expected to remain still throughout the video. It may happen that in some cases useful objects are there in the scenario which remain for long time after initial reference frame was taken, in that case background image needs to be taken again for proper operation. For detection of multiple objects, first the video is captured for a fixed amount of time and then for another fixed interval which is called processing time, the video is sampled into image-frames to ensure that the proposed DIP algorithm can be implemented. The acquired color image is converted to grayscale image format for further processing. An intermediate frame is created by subtracting the sampled frame from the initial reference frame and taking its absolute value

$$Inter\ Frame = mod(Sampled\ Frame - Reference\ Frame) \quad (1)$$

A Binary Image is then created using threshold operation to form the difference image frame and to eliminate low intensity objects and noise [4]. Finally, all the difference Image Frames are passed through a median filter [5] to remove background noise during image acquisition. An area filter is then applied to the image which removes all the objects below a certain area threshold value called clutters which are of low significance. Parameters for the area filter can be adjusted. In the final difference image, all the white pixels indicate motion and presence of objects. Thus, each group of white pixels separated by black pixels throughout in all the directions is considered as a single object. In this manner, the algorithm is used to detect multiple objects. The positions of all the objects are indicated by the locations of their corresponding centroids calculated based on the mean values of their x and y coordinates. Each object is identified by its mean grayscale value taken over the entire pixel range corresponding to that object. After detection of multiple objects, tracking of the detected multiple objects is performed. Multiple object tracking involves tracing the path followed by each object separately in a video. Tracking of multiple objects while considering the effect of occlusion between multiple objects is done by considering the various parameters such as difference in the mean grayscale levels of various objects being traced (when areas of objects under consideration is comparable).Therefore the object with maximum area (or grayscale value) is tracked by highlighting its centroid with one color, while the one with second largest area (or grayscale value)is tracked with highlighting its centroid with other color and so on. [6] Such peculiar characteristics of people are being tracked for every person even after occlusion.

5. PROPOSED SUSPICIOUS ACTIVITY IDENTIFICATION ALGORITHM

For identification of a suspicious activity and raising a corresponding alert the processing on centroid of each object is performed. If the centroid corresponding to each object in the scenario in different images does not change for a long number of consecutive frames, it implies that either an additional object has been added during the video surveillance

which was not initially present in reference frame (indicating the possibility of a suspicious object being dropped which has not changes its position for a long time) or an object initially present in reference frame is missing (indicating the possibility of a theft).The time for which the object remains still with respect to the reference frame before raising the alarm can be adjusted. Once the time limit is exceeded an alarm is raised (in the form of red mark at the position corresponding to the centroid and its neighboring pixels) indicating an emergency.

6. STEPWISE EXPLANATION OF SUSPICIOUS ACTIVITY IDENTIFICATION

Step 1 - Initialize a 'Warning' matrix of the same size as that of the reference image with all elements set to zero

Step 2 - Frame=1

Step 3 - Calculate the difference image of the current image with respect to reference image.

Step 4 - Detect the position corresponding to the centroids of multiple objects in the difference image.

Step 5 - For the centroid positions corresponding to the current frame, increment the corresponding Warning pixel by one. For a non - centroid position in the current frame, Warning pixel makes zero.

Step 6 - If Value of any Warning pixel>Max Count for which object should remain, Mark the position corresponding to the warning pixel and neighbouring pixels 'red' indicating an alert.

Step 7 - Frame= Frame+1

Step 8 - If Frame < = total number of frames, Go to step 3 else exit

Algorithm carried out during suspicious activity identification (Algorithm 1) is explained in this section-

Algorithm.1- Proposed suspicious activity identification system during video surveillance.

The flowchart of the proposed algorithm is shown in Figure 6.

7. SIMULATION RESULTS

The simulation results corresponding to the proposed algorithm are shown in this section. A Real time video is taken from a Webcam and processed using MATLAB. Initially we have simulated the detection of the location of two people before (fig.1), during (fig.2) and after (fig.3) the occurrence of occlusion. From the simulation results, it is clearly evident that the mean grayscale level for two persons is different because of different feature which is colour of objects. Fig. 1(a) shows the actual video frame of the movement of both the people before the occurrence of occlusion. Fig.1(b) displays the difference video frame with respect to the reference background frame. The pixels which are white are the corresponding difference pixels. Figure 1 (c) shows the position of the two objects indicated by their

centroid (mean grayscale level). The blue mark indicates the position of the person with white shirt along with green stripes and blue jeans while the green marker indicates the person with white shirt and black jeans. Fig.1(d) checks if any activity could be termed as suspicious and since the objects have been there for short period of time less than max permissible value for scenario so there is no suspicious activity. Fig.2(a) shows the occurrence of occlusion between both the people. Fig.2(b) shows the difference image. Fig.2(c) shows the position where the occlusion has occurred. Since both the people have been passing each other during occlusion they appear as a single object, the position of which is shown by a white mark. Fig.2(d) ensures that no suspicious activity has occurred.Fig.3(a) shows the frame of the video after the occurrence of occlusion. Fig.3(b) shows the difference image.Fig.3(c) shows the position of the two objects indicated by their mean grayscale level. The position of the two objects can still be predicted accurately even after the occurrence of occlusion. The blue mark still indicates the position of the person with white shirt along with green stripes and blue jeans while the green mark continues to indicate the person with white shirt and black jeans. This shows that our object tracking works even after the occurrence of occlusion. Fig.3 (d) ensures that no suspicious activity has occurred. (Since there is no suspicious activity there are no red marks in the figure). For performing the simulation of suspicious object detection in our video we drop a chair assuming it to be an introduction of a new object into the range of view which could be a suspicious if kept still for a certain number of frames. Fig.4(a) shows the video frame where the suspicious object has been introduced into the scenario and has not been detected. Fig.4(b) shows the difference image Fig.4 (c) shows the person dropping the suspicious object with green mark and the dropped suspicious object (for demonstration purpose, we used a blue coloured chair) has been indicated with blue mark. In Fig.4(d), the suspicious activity has not been indicated yet (No red mark). However, after some time the dropped object remains still for a certain threshold number of frames. Fig.5 (b) shows the difference image. Fig.5 (c) shows the person who had dropped the suspicious object with green mark and the dropped suspicious object has been indicated with blue mark. Fig. 5 (d), after the suspicious object remains still for a threshold number of frames, the suspicious activity is indicated with a red mark (alert signal). (This algorithm even works well for theft detection where red mark appears after a certain threshold number of frames indicating the missing object). All the figures (Fig.1 to Fig.5) demonstrate the proposed video surveillance algorithm is robust which is capable of detecting and tracking the motion of multiple objects undergoing multiple occlusions while raising a RED alert when anything apparently suspicious activity occurs in the scenario under video surveillance.
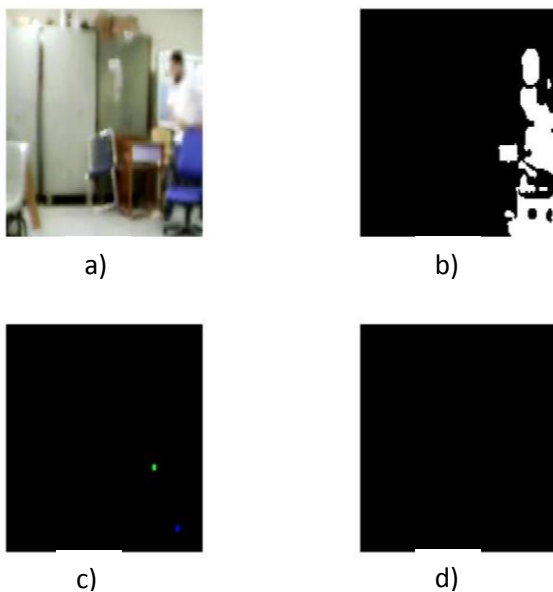

a)

b)

c)

d)
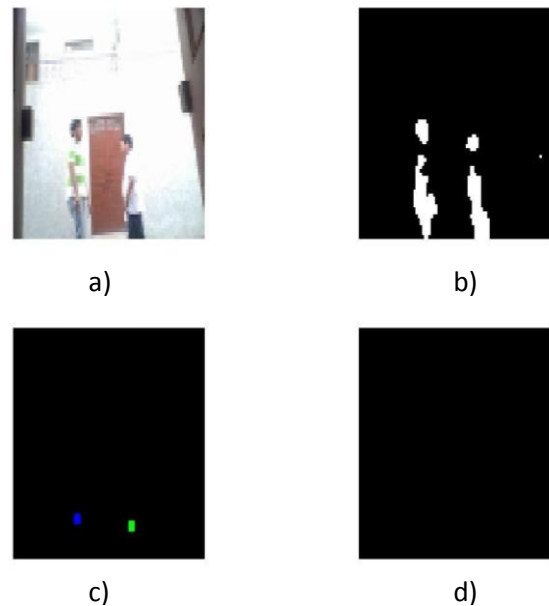
**Fig.1-Two moving people before undergoing occlusion –
(a) Original Frame (top left)
(b) Difference Image corresponding to the frame (top right)
(c) Multiple Object Tracking (bottom left) based on difference in mean grayscale level (mean grayscale level corresponding to person with shirt with green and white stripes and blue jeans (indicated with blue mark) is different from the mean grayscale level of the person with white shirt and black jeans (indicated with green mark).
(d) Suspicious Object detection (No red spot implies no suspicious or missing object detected)**

a)

b)

c)

d)
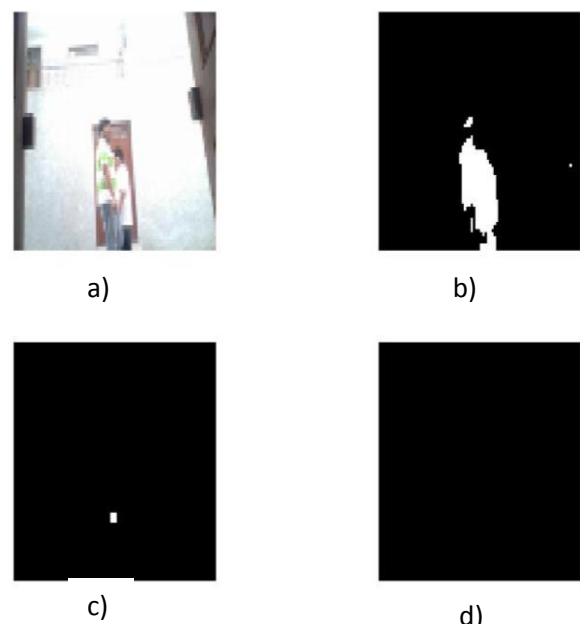
**Fig.2-Two moving people during occlusion-(a)Original Frame(top left)
(b)Difference Image corresponding to the frame(top right)
(c)Multiple Object Tracking(bottom left)based on difference in mean greyscale level(During occlusion the two objects merge and appear to be in form of a single object)
(d)Suspicious Object Detection**

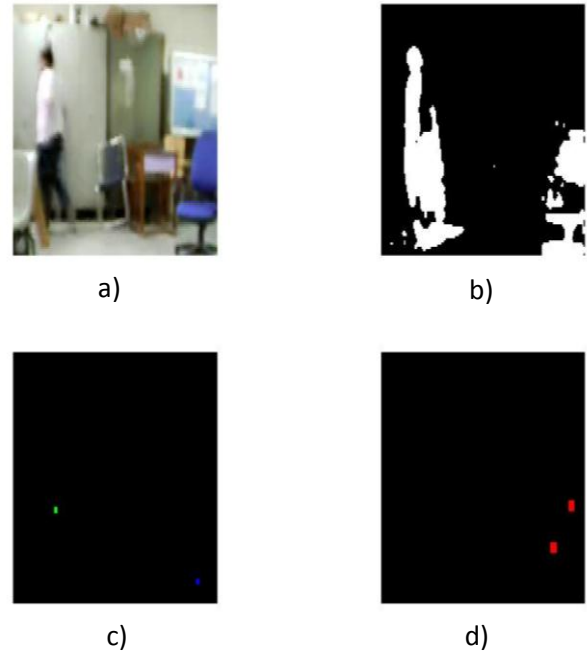a)   b)

c)   d)
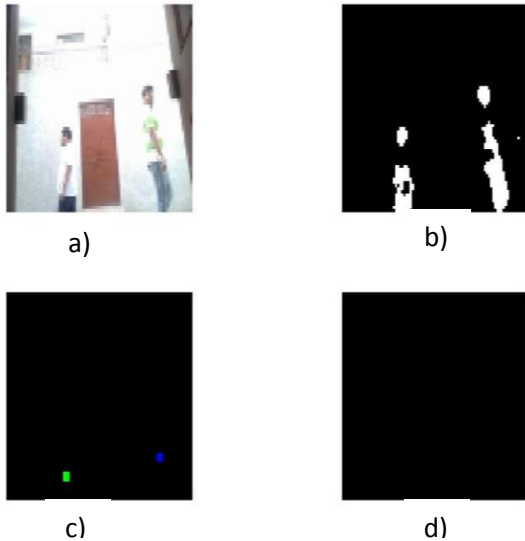
Fig.3 Two moving people after occlusion-(a)Original Frame(topleft)
(b)Difference Image corresponding to the frame(top right)
(c)Multiple Object Tracking(bottom left) based on difference in mean grayscale level(The track of two objects is not lost after occlusion)
(d)Suspicious Object Detection.(No Red spot implies no suspicious or missing object detected)

a)   b)

c)   d)

Fig.4-Before suspicious object detection(a)Original Frame(topleft)
(b)Difference Image Corresponding to the frame(top right)
(c)Multiple Object Tracking(bottom left) based on difference in mean greyscale level(Person dropping the object(green) and the object(blue)
(d)Suspicious Object Detection

a)   b)

c)   d)

Fig.5-After suspicious object detection-
(a) Original Frame(top left).
(b)Difference Image corresponding to the frame(top left).
(c)Multiple Object Tracking (bottom left)based on difference in mean greyscale level(Person who dropped the object(green)and the object(blue))
(d)Suspicious Object Detection(Occurrence of Red spot implies detection of suspicious or missing Object).

8. CONCLUSION

In this paper, multiple object tracking during occlusion and suspicious activity detection has been described using proposed DIP Algorithm. We have used Real Time Data and it has been shown through results that multiple occlusions of objects do not affect the behaviour of the suspicious object/activity identification. The whole scenario was implemented on MATLAB 2012b. Extensive experiments were carried out to check the correctness of the proposed algorithm. The proposed algorithm can be employed for video surveillance, especially in the crowded areas like railway and bus station and can be used on CCTV footage for detecting suspicious objects on FLY as computation can be performed in very less time if multiple short videos are processed instead of a long video which consumes processing as well as capturing time.

In future, we would like to use optimization algorithms which use Machine Learning and Data Mining techniques to extract important features from the Video and Compressing it for faster processing with less overall size while not losing significant information.
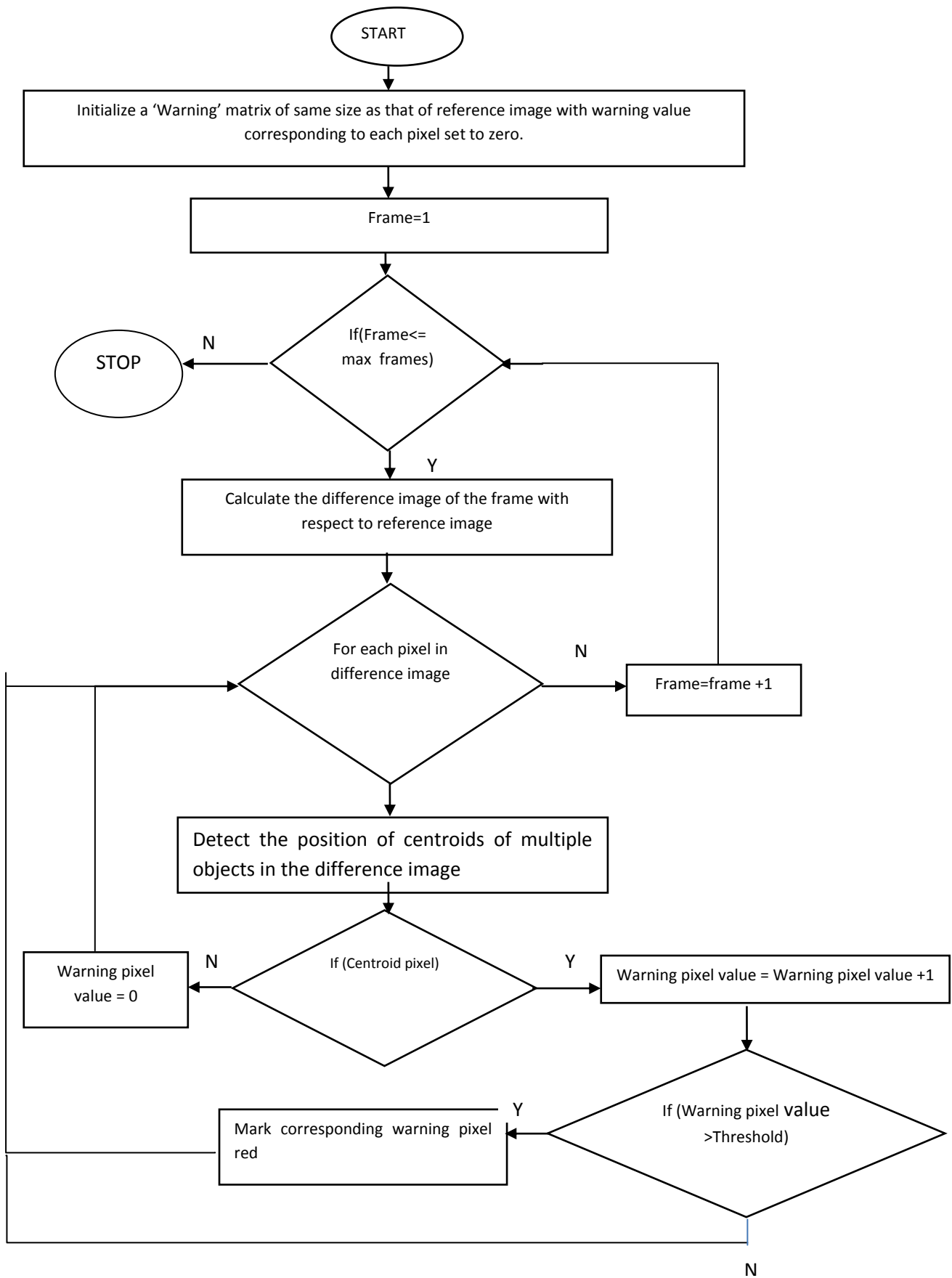
START

Initialize a 'Warning' matrix of same size as that of reference image with warning value corresponding to each pixel set to zero.

Frame=1

If(Frame<= max frames)

STOP

N

Y

Calculate the difference image of the frame with respect to reference image

For each pixel in difference image

N

Frame=frame +1

Detect the position of centroids of multiple objects in the difference image

Warning pixel value = 0

N

If (Centroid pixel)

Y

Warning pixel value = Warning pixel value +1

Mark corresponding warning pixel red

Y

If (Warning pixel value >Threshold)

N

Fig: 6- Flowchart of the proposed algorithm

9. REFERENCES

[1] M. Kalpana Chowdary , S. SuparshyaBabu, S. Susrutha Babu, Dr. Habibulla Khan. "FPGA Implementation of Moving Object Detection in Frames by Using Background Subtraction Algorithm", International conference on Communication and Signal Processing, April 3-5, 2013, pp 1032-1036.

[2] K.Kinoshita, M.Enokidani, M. Izumida and K.Murakami. "Tracking of a Moving Object Using One-Dimensional Optical Flow with a Rotating Observer," 9th International Conference on Control, Automation, Robotics and Vision, 2006. ICARCV'06. 5-8 Dec. 2006, pp.1 – 6.

[3] J. L. Barron, D. J. Fleet and S. S. Beauchemin. "Performance of optical flow technique" IJCV 12:1, 1994, pp 43-77.

[4] Lijing Zhang and Yingli Liang. "Motion human detection based on background Subtraction", 2010 Second International Workshop on Education Technology and Computer Science, pp.284-287.

[5] Massimo Piccardi. "Background subtraction techniques: a review" 2004 IEEE International Conference on Systems, Man and Cybernetics, pp.3099-3104

[6] Mr.Zhi Ming Qian, Xi En Cheng, YanQiu Chen.(2014)"Automatically Detect and Track Multiple Fish Swimming in shallow Water with Frequent Occlusion",Plos ONE 9(9):e106506,doi:10.1371/journal.pone.0106506.

[7] Zhang Ruling, Sun Hanxu, JiaQingxuan, Yao Fusheng. "Research on fast and accurate occlusion detection technology of augmented reality system", Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on 13-16 July 2008, pp.111-116.

[8] R. Depommier and E. Dubois. "Motion Estimation with Detection of Occlusion Areas", 1992 IEEE, pp.269-272.

[9] J. Konrad and E. Dubois, "Bayesian estimation of discontinuous motion in images using simulated annealing", INRS-Telecommunications, 3 Place du Commerce,Verdum, Quebec, Canada, H3E 1H6,pp.51-60.

[10] Alan J. Lipton, Hironobu Fujiyoshi, Raju S. Patil, "Moving Target Classification and Tracking from Real-time Video", Fourth IEEE Workshop on Applications of Computer Vision,1998, WACV'98 Proceedings, pp.8-14.

[11] C. Lawrence Zitnick and Takeo Kanade, "A Cooperative Algorithm for Stereo Matching and Occlusion Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 7, July 2000, pp.675-684.

[12] Karlheinz Gutjahr, HannesRaggam. "Determination of Occlusion Areas in HighResolution Remote Sensing Data", 2nd GRSS/ISPRS Joint Workshop on "Data Fusion and Remote Sensing Over Urban Areas", pp.191-194.

[13] Mao Jia-Fa, Xiao Gang, Sheng Wei-Guo, Xiaohui Liu. "A 3D occlusion tracking Model of the underwater fish", 2015 IEEE International Conference on Electro/Information Technology (EIT), 21-23 May 2015, pp.82-86.

[14] Yingkun Xu, Lei Qin, Guorong Li, Qingming Huang. "An efficient occlusion detection method to improve object trackers", 2013 20th IEEE International Conference on Image Processing (ICIP), 15-18 Sept. 2013, pp.2445-2449.

[15] S. V. Kothiya, K. B. Mistree. "A review on real time object tracking in video sequences", 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), 24-25 Jan. 2015.