MOTION BASED OBJECT DETECTION AND CLASSIFICATION FOR NIGHT SURVEILLANCE

Usham Dias[1] and Milind Rane[2]

Department of Electronics and Telecommunications, Vishwakarma Institute of Technology, India
E-mail: [1]ushamdias@gmail.com and [2]me_rane@yahoo.com

Abstract
This paper describes a simple technique for object detection and temporal data association of thermal image sequences. Night surveillance system using thermal imaging involves object detection, temporal data association and tracking of object. Object detection could be motion based or feature based. The temporal data association in multi-object classification involves finding the minimum distances between an object in current frame to the objects in previous frame. A performance comparison is made between two techniques for object detection based on timing constraints and qualitative analysis. The second method proposed clearly outperforms the first in terms of timing. Target classification using neural network dynamically identifies the moving object. Implementation is done using MATLAB software.

Keywords:
Object Detection, Thermal Imaging, Night Surveillance, Neural Network

1. INTRODUCTION

The different types of object detection techniques include feature-based, template-based and motion based. In feature-based object detection, standardization of image features and registration (alignment) of reference points are important. The images may need to be transformed to another space for handling changes in illumination, size and orientation. One or more features are extracted and the objects of interest are modelled in terms of these features. Object detection and recognition then can be transformed into a graph matching problem e.g. Shape-based approach, Colour-based approach. In template-based technique, if a template describing a specific object is available, object detection becomes a process of matching features between the template and the image sequence under analysis. Object detection with an exact match is generally computationally expensive and the quality of matching depends on the details and the degree of precision provided by the object template. There are two types of object template matching, fixed and deformable template matching. Detecting moving objects, or motion detection, obviously has very important significance in video object detection and tracking. A large proportion of research efforts of object detection and tracking focused on this problem in last decade. Compared with object detection without motion, on one hand, motion detection complicates the object detection problem by adding object's temporal change requirements; on the other hand, it also provides another information source for detection and tracking [1].

New Night Vision technologies are rapidly being developed and an ever-expanding pool of users is looking to night vision to solve challenging security and surveillance problems. Night vision technologies include image intensifiers, thermal imaging, low-light CCD cameras and new infrared illuminations technologies.

The different night vision imaging techniques involve image intensifiers, infra red imaging and thermal imaging. Human body temperature is around 37°C therefore the thermal camera range should be adjusted to 35-40°C. Practically the range is set to 30-40°C because the person wears jacket, etc. Care should be taken while selecting range because a large range e.g. 20-40°C (Room Temperature included) will be more sensitive to noise due to sunlight, fluorescent tube light, etc[2].

Earlier, the availability of infrared cameras was only for military applications. In applications like outdoor surveillance, where the background temperature is largely different from human beings, thermal imaging can play vital role in identifying and tracking persons. Further thermal images cannot track shadow or light illumination. On the flip side, clutters like cool body, variation in temperature across same subjects, blowing winds with different temperature gradients, person overlap while crossing each other, put challenges in thermal imaging and will have to be handled intelligently to make the motion tracking system using only thermal imaging efficient [5].

2. MOVING OBJECT DETECTION BASED ON W.K WONG'S METHOD[2]

The method proposed in [2] divides the image into $m \times n$ regions and detects changes in the intensity level for RGB images. This algorithm has been modified for gray scale images. The detailed steps are mentioned below,

Step 1: Input the video frames

Step 2: Divide each image into $m \times n$ equal regions

Step 3: Define a matrix, M with size of $(m \times n)$ to represent the characteristic of each corresponding region.

Step 4: Define a threshold Q. If the difference between current frame and previous frame is greater than Q, it is considered a valid motion pixel.

Step 5: Define a variable, h for counting the number of pixels exceeding Q in each region. Initially, h is set to 0.

Step 6: Define the value H, as the minimum number of pixels exceeding Q in a region which can be considered as a potential object movement.

Compare current taken image with previous taken image. For each corresponding region, find out the difference between a particular current image and previous image pixels value. If the difference value for a particular current image pixel to previous image pixel $\geq Q$, then, $h = h + 1$. If $h \geq$ H, mark a "1" into the corresponding element of M, else if $h < H$, mark a "0" into the corresponding element of M. This M indicates if a human is

detected in the region or not. Human have more height than width therefore care should be taken to select $m \times n$ such that number of rows in region is more than the number of columns.

3. PROPOSED METHOD FOR MOVING OBJECT DETECTION

The proposed algorithm tries to find the centroid so as to pin point the location of the object. The detailed steps are mentioned below,

Step 1: Input the video frames.

Step 2: Define a threshold Q. If the difference between current frame and previous frame is greater than Q, it is considered a valid motion pixel.

Step 3: If difference image pixel $> Q$, then thresholded image pixel = 1 else 0.

Step 4: Remove small noisy pixel from thresholded image.

Step 5: Use morphological close operator to combine disjoint parts of the same object.

Step 6: Find boundaries of each object.

Step 7: For boundary coordinates of each object calculate the centroid.

4. TEMPORAL DATA ASSOCIATION

If an object from the previous frame with distance less than some threshold is found, then its label is assigned to the object under test in the current frame indicating that the object has now moved to this new position. The temporal buffer is updated with the new object parameters. In case nearest object has distance greater than threshold, an object with new label is created indicating the entry of new person [5].

A simple algorithm for temporal data association is mentioned below,

Step 1: Calculate the moving object centroids based on the proposed algorithm mentioned in section 2.

Step 2: Select the object that should be tracked.

Step 3: Calculate the distance of this object (in current frame) from the other objects in the next frame.

Step 4: Associate the object to the one which has minimum distance. Repeat the process.

5. TARGET CLASSIFICATION

The moving object detection algorithm based on section 3, gives the centroid of the white blobs as shown in Fig.2. Based on the dimensions of these blobs, the objects are segmented. Paper [6], finds edges of the objects and then feeds it to a classifier. But in our case edges fail because the human structure varies continuously. Hence the intensity values are directly fed to the classifier. Here artificial neural network is used as a classifier. It is trained to classify humans, vehicles and armed men. The training and testing results are given in section 6.

6. EXPERIMENTAL RESULTS

IEEE OTCBVS WS Series Bench is an OTCBVS Benchmark Dataset Collection, has been used for experimentation in this paper. The codes were tested for Dataset 01: OSU Thermal Pedestrian Database which had 10 different sequences and few sequences from Dataset 05: Terravic Motion Infrared Database.

6.1 RESULTS FOR W.K WONG'S METHOD

The results for Dataset01 are shown below in Fig.1 for one of the 10 sequences present. Here the image is divided into 20×40 regions. The threshold value Q to find motion pixel is taken as 50. The minimum value H that exceeds threshold that can be considered as a potential object movement is taken as 15.

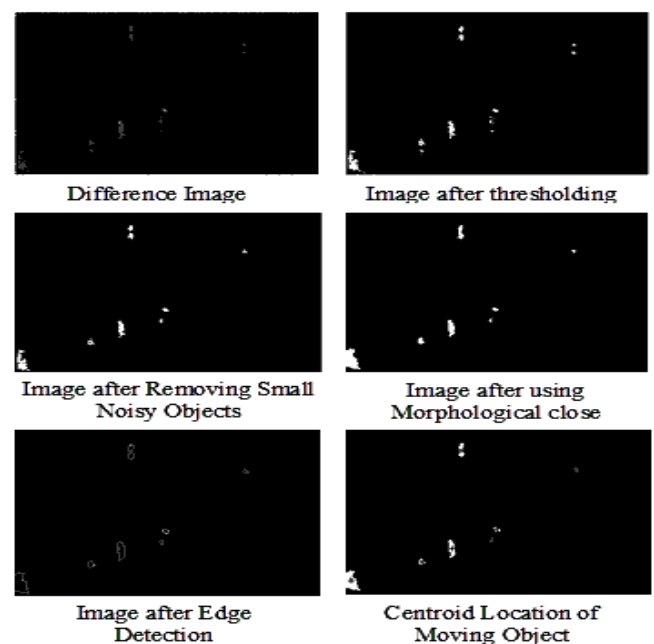Fig.1. Result using Method1 on Dataset01

Fig.2. Result using Method2 on Dataset01

Table.1. Timing results on Dataset01 for object detection

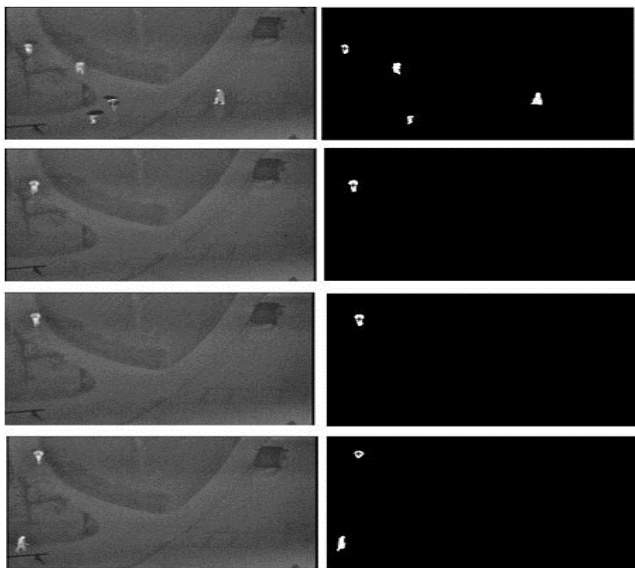| Sequence/Video	Time Taken (sec)	
	W.K. Wong's Method	Proposed Method
01	328.831562	7.325645
02	296.424570	6.278622
03	240.631474	4.731066
04	187.099693	3.562597
05	241.122828	4.815423
06	186.017957	3.675107
07	228.939521	4.692295
08	253.029345	5.264392
09	790.857740	28.126269
10	252.696583	5.328195

Fig.3. Temporal data association

6.2 RESULTS FOR PROPOSED METHOD

The results for Dataset01 are shown in Fig.2 for one of the 10 sequences present. The threshold value Q to find motion pixel is taken as 50.

The time taken for moving object detection using the two methods is given in Table.1. It is important to note that each sequence within the dataset has different number of frames and hence the table cannot be used for comparison within a particular method, but can be used to compare between methods.

6.3 RESULTS FOR TEMPORAL DATA ASSOCIATION

The results for temporal data association are shown in Fig.3. The top left object has been tracked successfully with its centroid marked in successive frames using temporal data association.

6.4 RESULTS FOR TARGET CLASSIFICATION

The training data and testing data are shown in Fig.4. There are a total of 22 segmented images for training, where $h1$ through $h10$, $v1$ through $v4$ and $a1$ through $a8$ indicate 10 humans, 4 vehicles and 8 armed men for training. There are a total of 33 segmented images for testing, where $h1$ through $h21$, $v1$ through $v4$ and $a1$ through $a8$ indicate 21 humans, 4 vehicles and 8 armed men for testing. All the segmented images are resized into 32×32 matrix. The results with and without histogram equalization of the segmented images are shown in Table.2, along with neural network details.

The trained network was then used to classify the segmented objects as explained in section 5, after contrast enhancement. Since all the moving objects in the video database are humans, the network had to give output 1, indicating it's a human. The results on DATABASES\OTCBVS Benchmark Dataset Collection\Dataset 01 OSU Thermal Pedestrian Database\00001\00001 and DATABASES\OTCBVS Benchmark Dataset Collection\Dataset 01 OSU Thermal Pedestrian Database\00002\00002, gave one and two misclassification of detected objects respectively.
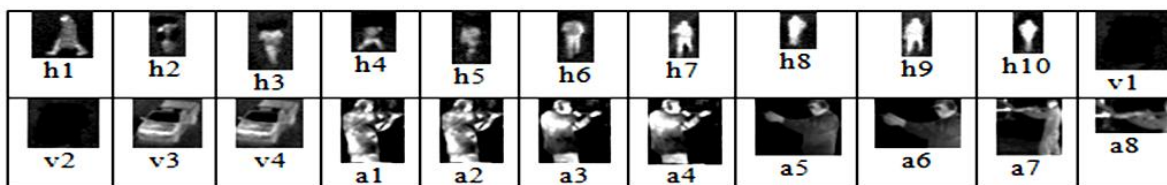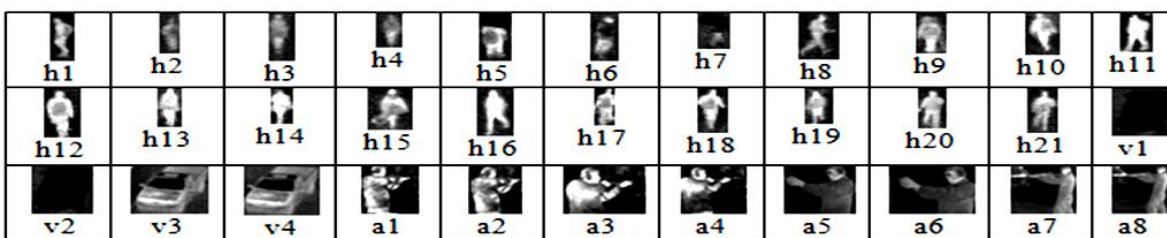

Fig.4. Training and testing data set

Table.2. Target classification results based on Gradient Descent Back propagation algorithm with adaptive learning rate

Case	No. of hidden nodes	Transfer function for hidden & output layer	Epoch	Performance (mse)	Training time	No. of Misclassification on testing data
Without histogram equalization	513	Tansig, Tansig	10,000	73.6e-05	31 min 22 sec	1
With histogram equalization	513	Tansig, Tansig	7000	4.19e-05	3 min. 9 sec	1

7. CONCLUSION

In this paper we discus two methods for object detection. The first method divides the image into regions. By selecting longer vertical length of the region we can take into account height of the human body, thus eliminating the need for merging regions. But this method still fails to give exact location of the human/ moving body. The second proposed method proves to be more precise since it finds the centroid of the object, thus locating nearby objects more easily and precisely. The proposed method clearly outperforms the first in terms of timing. The performance of the neural network for classification improved considerably after contrast enhancement of the segmented images.

The Temporal data association technique was implemented successfully but only for tracking single objects. In future this could be extended for tracking multiple objects.

REFERENCES

[1] Zhong Guo, "Object Detection and Tracking in Video", *Advances in Internet based systems and Applications*, 2001.

[2] Wai Kit Wong *et al.*, "An Effective Surveillance System Using Thermal Camera", *International Conference on Signal Acquisition and Processing*, pp. 13-17, 2009.

[3] Nathan Funk, "A Study of the Kalman Filter applied to Visual Tracking", *Report- University of Alberta, Project for CMPUT 652*, pp. 1-26, 2003.

[4] Leow Wee Kheng, "Motion Tracking", *CS6240 Multimedia Analysis*, Department of Computer Science, School of Computing, National University of Singapore.

[5] Chandrashekhar N. Padole and Luis A. Alexandre, "Wigner Distribution based motion tracking of Human Beings using Thermal Imaging", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, pp. 9-14, 2010.

[6] Alan E. Putman, Gene A. Tagliarhi and Edward W. Page, "Using Adaptive Resonance Theory Networks and Fuzzy Matching to Recognize Target Features in Thermal Images", *IEEE International Joint Conference on Neural Networks,* Vol. 6, pp. 4284-4288, 1999.