

# A Level Set Approach on Human Action Prediction

M. Sushma Sri

Faculty in ECE, Govt. Institute of Electronics, Secunderabad, India

**Abstract:** Human activity prediction is a challenging task. The aim of this paper is to track the human motion and probabilistically predict the future action of the target. This is done in three ways detect the target, estimate the direction and then predict the future action. We employ K-means algorithm which results were not very encouraged hence we shifted to level sets for better target localization. A probabilistic approach is used to estimate the direction of motion. MAP estimator is implemented to estimate the direction of the object motion. The results of estimated directions are used to predict the position of the object in future frames.

**Keywords:** Direction estimation, K-mean clustering, Level sets, Prediction, Target localization.

## I. INTRODUCTION

Today Security is given very much importance and lot of electronic equipment is being used in security applications. Monitoring continuously the movements of persons or vehicles and reporting when predefined events take place is very common security applications. A human observation based system for implementing this has several disadvantages. In the olden days persons use to be employed for doing such observations. Decades before electronic cameras could solve the problem of man being physically present at such place. Instead man has to observe the camera's output on a TV and can detect when any expected events occur. The present day technology allows automatic detection based on predefined measures. In this foreground detection based moving object detection algorithm is implemented targeting a wide class of applications. Various operations are carried out and the moving objects are detected. Clustering and level sets at various phases are used to decide the possibility of identifying the moving object of certain sizes. Moving objects also tracked in it. However, in a video surveillance, some information or frames are missing due to technical and software problems otherwise that information stolen or skipped by others. This information can be obtained from

previous frames of video by predicting the ongoing frames. Human activity prediction is an inference of the ongoing activity in incomplete videos. Activity prediction is used to find the unfinished activities in a video. Prediction method computes the posterior probability of which activity has progress to which point of the activity.

## II. BACK GROUND

### A. PREDICTION

The human activity prediction is defined as the inference of unfinished activity from incomplete video. The estimated direction and speed, obtained in the previous phase, is used to predict the position of the object in future frames, for example, in the immediately following frame or N frames after the last frame available in the video.

### A. TARGET LOCALIZATION

The moving objects in a video can be easily detected through sensing the changes in the adjacent frames of the video. The difference between adjacent frames provides positions of the pixels that change their values. A moving object may produce many such pixels. The first phase in our work is to localize the target i.e. producing one point per moving object. This point should mark the same position on the moving object so that the object can be tracked to estimate its motion in our next phase. The video may contain one or more moving objects. With a single object, the localization can be done by locating the centric of all the pixels that change their value from frame to frame. But with multiple objects, the change points need to be clustered so that each cluster represents an object [1]. In this paper, the clustering is done through K-means algorithm [2]. The results were not very encouraged and hence the clustering action is carried out through a deformable model called level set.

### **1. K-mean clustering**

In this method, randomly selected  $k$  points from the difference data are defined as the position of  $k$  objects in the frame. Then the other points are added to the cluster whose mean is nearest to it. The process is repeated iteratively till the means of these clusters become stationary. At steady state, the number of clusters will equal the number of moving objects in the video. The mean of each cluster represents the location of each object. The main drawback of this method is that the number of difference points and their positions relative to the object keep varying from frame to frame and hence the mean of the cluster do not point to the same part of the object. This drawback forces in erroneous estimation of the objects motion.

### **2. Level Sets**

Level set is a deformable model that is a hardy tool in segmenting an image into object [7]. In our work, level set technique defined by Chan & vese is used [5], [7]. In this model, a closed curve is evolved iteratively so as to make the inner and outer regions of the curve uniform. The non-uniformity of the region is measured through energy and the curve is evolved in such a way that both these energies, related to inner and outer region, minimize at each iteration. If the curve evolution is defined to tolerate some non-uniformity in the inside region, the curve at saturation will almost surround the whole object even through all the object points are not available in difference data [8]. This allows the centroid of these points to represent relatively the same position on the moving object in all the frames. The centroid of each object moving in the video using level set is located for each frame of the video. This information is further used to estimate the direction of moving each object.

### **C. DIRECTION ESTIMATION**

In this phase, a probabilistic approach is used to estimate the direction of motion. The interframe displacement of centroid point will vote for computing the likelihood of the motion direction [8], [10]. A posteriori for each direction is computed using Baye's theorem [9]. A MAP estimator is implemented to estimate the direction and speed of the object motion [10].

## **III. EXPERIMENT**

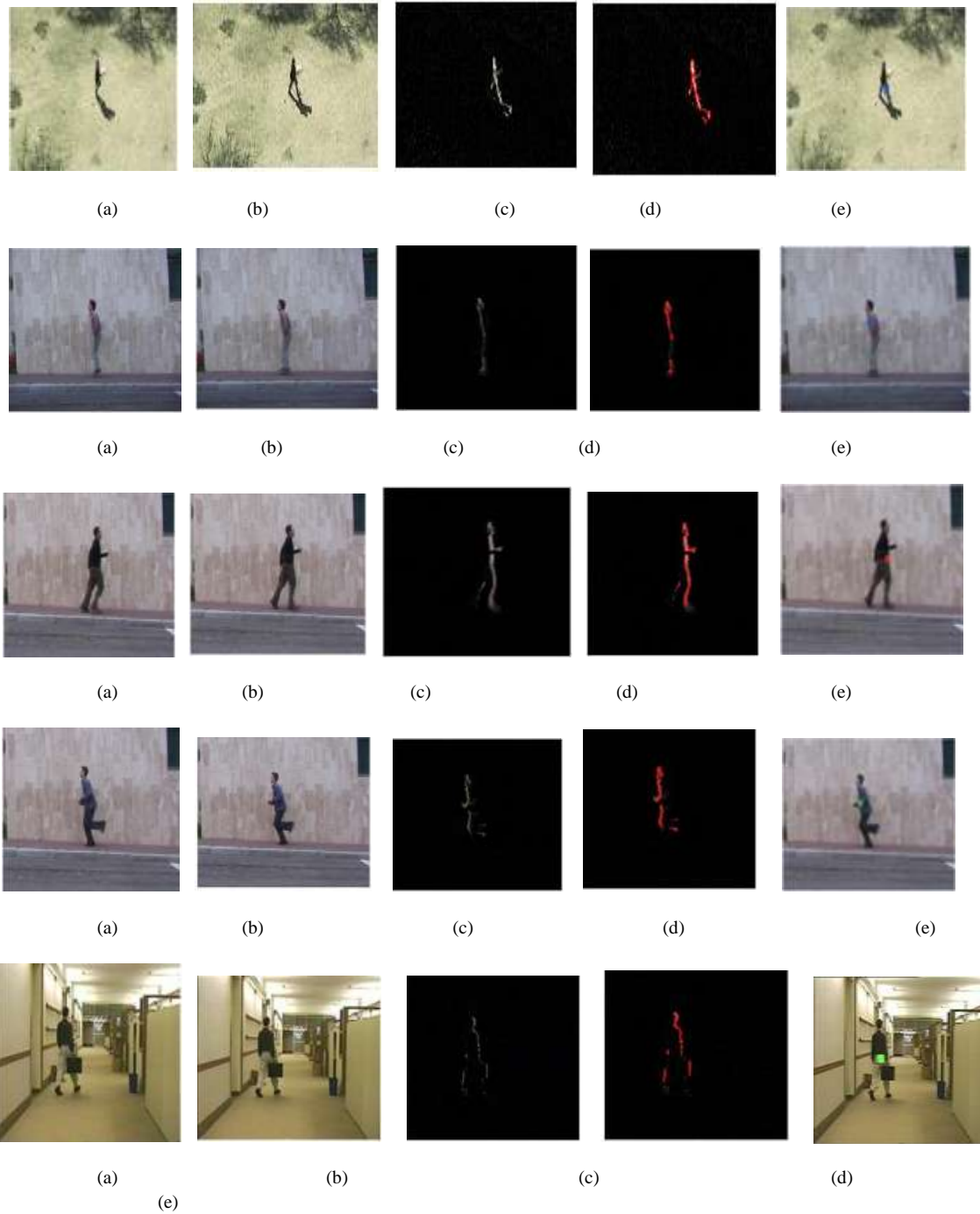
For our Experiment, we used two datasets: Weizmann and UT interaction dataset [5], [12]. Each dataset contains stationary background and multiple actions. We present experiments and results in this phase.

We implemented one and multiple human activity recognition based on the two methods presented in this paper: K-mean clustering and level sets. Human activity prediction is done by likelihood probability. Dataset contains videos of stationary background. Target localization is computed by comparing any two consecutive frames at a given time. To do first the frames of a video should be converted into images. The complete operation is performed on the images itself. Image subtraction method is performed on the successive images in order to know the changes in the image. To track the direction of a single person, find centroid to all moving objects, which is done by labeling method. If more than one person is present in a video, then if we wish to extract the direction of different persons in that video then we go for k-mean clustering algorithm. The results were not very encouraged and hence the clustering action is carried out through a deformable model called level set. We represented direction points in any one of RGB colours. A MAP estimator is implemented to estimate the direction of motion which is used to predict the position of moving objects in future frames.

## **IV. RESULTS**

### **A. RECOGNITION AND LOCALIZATION OF SINGLE ACTION**

Weizmann dataset contains videos of seven types of human actions: walking, running, hand waving, bending, skipping, carrying and jumping. We observed 5 videos. In video 1, a person is walking towards. In video 2, a person is jumping towards west. In video 3, a person is running towards east. In video 4, a person is hopping towards east. In video 5, a person is carrying a briefcase towards north east as shown in below figure 1.



**Figure 1:** Example clips of recognition results. (a), (b) are two successive frames of videos. (c) Subtracted image, (d) the shape of moving person enclosed with red curve by level set, (e) tracked the location of person with a square box one of RGB colours.

**Table I:** Directions for single person videos

Video	Actual Direction	Most probable Direction
1	East	East
2	West	West
3	East	North east
4	West	South west
5	North east	North east

From Table I, these videos have one person. In video 1, person is moving towards east, we predicted direction is east. For video 3, Actual direction is east, but predicted direction is northeast. Predicted directions of single persons are exactly matched to actual directions for video 1, video 2 and video 5. But for remaining videos, predicted directions are not exactly matched to actual directions.

### V. CONCLUSION

Chan-Vese model of level set provides more stable localization of moving objects and it is robust, efficient and simple to implement. The direction is estimated by voting among the 8 directions. The position of the object at entry will make us to predict among 3 directions only and hence simplifying the computation. The work predicts the position of the object in the next frame and after N frames. The work is not efficient when the objects are overlapped. The object tracking becomes difficult when a moving object ceases to move. This can be overcome by applying some cumulative method

### REFERENCES

[1] Adnan khashman, “ Automatic detection, extraction and recognition of moving objects”. International journal of systems applications, engineering & development . Issue 1, volume 2, 2008.

[2] Anil Z Chitade, “colour based image segmentation using k-means Clustering”. International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5319-5325.

[3] Chan Tony F, Vese Lumanita A, “Active contours without edge”, IEEE Trans. “On Image processing”, 10(2): 266-277, 2001.

[4] D.Waltisberg, A. Yao, J. Gall, and L.V.Gool, “Variations of a Hough- voting action recognition system”. In ICPR Contest on Semantic Description of Human Activities (SDHA), in Proceedings of ICPR Contests, 2010.

[5] F. Yuan, V. Prinet, and J. Yuan, “Middle-level representation for human activities recognition: the role of spatio-temporal relationships”. In ECCV Workshop on Human Motion: Understanding, Modeling, Capture and Animation, 2010.

[6] Mario F. Triola, Bayes' Theorem.

[7] M. Airouche, L. Bentabet and M. Zelmat, “Image Segmentation Using Active Contour Model and Level Set Method Applied to Detect Oil Spills”, Proceedings of the World Congress on Engineering 2009 Vol I WCE 2009, July 1 - 3, 2009, London, U.K.

[8] M.S. Ryoo and J.K. Aggarwal, “UT-Interaction Dataset”, ICPR Contest on semantic Description of Human Activities (SDHA), 2010.

[9] M. S. Ryoo, C.C. Chen, J. K. Aggarwal, and A. Roy-Chowdhury, “An overview of contest on semantic description of human activities (SDHA)”. in Proceedings of ICPR Contests, 2010.

[10] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, “An Efficient k-Means Clustering Algorithm: Analysis and Implementation”. IEEE transactions on pattern analysis and machine intelligence, vol. 24, no. 7, July 2002.

[11] Thomas Brox and Joachim Weickert, “Level Set Based Image Segmentation with Multiple Regions”. In Pattern Recognition, LNCS 3175, C.E. Rasmussen, H. Bulthoff, M. Giese, and B.Scholkopf (Eds.), pp. 415-423, Tubingen, Germany, Aug. 2004.

[12] Yonggang Shi, W.Clem Karl, “Real-time Tracking Using Level Sets”. CVPR'05.