Automatic Video Segmentation Using Modified HEM Algorithm

Chandan Kumar
M.Tech Scholar
PCST, Bhopal, India
E-mail-mckv.chandan@gmail.com

Sumit Utareja
Asst. Professor
PCST, Bhopal, India
E-mail-sumit.utareja@gmail.com

Deepak Rathore
Asst. Professor
PCST, Bhopal, India
E-mail-rathore.rath@gmail.com

Abstract
Video segmentation plays an important role in computer vision. The process of video segmentation uses in various fields such as road transportation, security surveillance and medical diagnosis of critical stage of human body damage detection. Various authors proposed various methods for the segmentation of video such as color based segmentation, texture based segmentation and shape based segmentation. In this paper modified the HEM clustering technique for video segmentation using genetic algorithm. Genetic algorithm is population based optimization technique. The genetic algorithm used for the selection of M estimation parameter. The value of M decides the better clustering process. For the empirical evaluation used MATLAB software and Google texture video data and YouTube video data. The proposed method improved the value of F-measure and accuracy in compression of HEM method. The proposed clustering algorithm is capable of both clustering Dynamic texture and searching novel Dynamic texture cluster centers that are representative of the cluster members in a manner that is consistent with the underlying generative probabilistic model of the Dynamic texture. We demonstrate the efficacy of the proposed clustering algorithm for dynamic textures on several computer visions problems. First, we perform hierarchical clustering of video textures, showing that proposed groups perceptually similar motion together.

Keywords: - Video segmentation, HEM, GPCA.

.

INTRODUCTION

Recently, there is a rapid increase in the amount of digital video in multimedia applications due to improvement in the technology. Digital videos are widely used in the areas such as TV domains, geographic information systems, monitoring systems, education domains, mobile phones. For these types of applications large video databases are created and stored. The ability to browse the stored video data or to retrieve the content of interest is becomes a fundamental requirement of any video archiving system. For example, a large amount of audiovisual material has been archived in television and film databases. If these data can be properly segmented and indexed, it will be easier to retrieve the desired video segments for the editing of a video clip. The idea of applying dynamic texture representations to the segmentation of video has previously appeared in the video literature. In fact, some of the inspiration for our work was the promise shown for temporal texture segmentation (e.g. smoke and fire) by the dynamic texture model. For example, segments video by clustering patches of dynamic texture using the level-sets framework and the Martin distance. More recently, clusters pixel-intensities (or local texture features) using auto-regressive processes (AR) and level-sets, and segments video by clustering pixels with similar trajectories in time, using generalized PCA (GPCA). The researches on the audio information shows that audio can also be used for video segmentation. By being aware of the audio information importance for the video segmentation process. The initial problem is segmentation of the audio into acoustically similar regions such as “silence”, “speech”, “music”, and “crowd” .For the silence segment detection a simple threshold comparison method on the short time energy of the signal is proposed. In the process for segmentation “speech”, “music”, “crowd”, “unclassified” regions, a pattern classification scheme has been adopted. This pattern classification scheme is composed of two sections: a feature extraction and selection stage, training and a classification stage. Object Tracking stations Receive analogue or IP camera images directly and perform the most time consuming video content analysis and image processing tasks. OTS Stations implement the camera level and the scene level intelligent functions, like object tracking based on multiple camera views, shape classification, and detection of crossing perimeters of virtual zones. Results of the OTS calculations are collected and processed in a central Site Wide Object Tracking Server, which implements the site level intelligent functions, like identity tracking, evaluation of complex rules based on the identity and security clearance of moving persons or vehicles, and moreover this module is capable of recognizing the suspicious activities. Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, clutter, etc. It is necessary to distinguish objects from each other in order to track and analyze their actions reliably. Currently, there are two major approaches towards moving object classification, which are shape-based and motion-based methods. The objects' 2D spatial information whereas motion-based methods use temporal tracked features of objects for the classification solution. Visual tracking is one of the most important fields in video processing and computer vision has been widely applied to traffic surveillance system, suspicious person monitoring system etc. In practical application, since the camera moves and rotates, it needs to track objects in a dynamical background. How to select the initial target objects automatically and establish objects motion model, and how to update object and background models at each frame are the key in real-time Visual tracking with an active camera. Aimed at solving the problem of initial selection of object, Reference used optic flow field and internal limitation constrains of motion to detect motion object. The current system is able to distinguish transitory and stopped foreground objects from static background objects in dynamic scenes, detect and distinguish left and removed objects, classify detected objects into different groups such as human, human group and vehicle, track objects and generate trajectory information even in multi-occlusion cases and detect fire in video imagery. A video can be thought of as being composed of a series of objects, each of these objects having edges. This edge information can be captured for a sequence of frames, and used to determine whether edges remain in the shot from one frame to the next, or whether they disappear and are replaced by different edges, compared pixel differences, statistical differences and several different histogram methods and found that the histogram methods were a good trade-off between accuracy and speed. Clustering is the task of partitioning a data set into subsets so that the data points in each subset are more similar to each other, according to some distance measure, than those from different subsets. It is a fundamental technique in data analysis and has many applications in computer vision, including image and video segmentation. Video clustering is an important problem in various areas of computer vision. For example, it can be used to uncover high-level patterns of structure in a video stream (e.g. recurring events, events of high and low probability, outlying events, etc.) and has, therefore, application to problems such as surveillance, novelty detection, video summarization (by shot clustering), or remote monitoring of various types of environments. It can also be applied to the entries of a video database in order to automatically create taxonomy of video classes that can then be used for database organization or video retrieval. The rest of paper is organized as follows. In Section II state the problem. The Section III Related work IV discusses proposed methodology. Section V discusses Implementation and results analysis followed by a conclusion in Section VI.

II PROBLEM STATEMENT

The increasing rate of multimedia data and transmission facility induces some problem of data loss and delay of delivery. Now in the process of video segmentation is important factor for analysis. For the background updating used segmentation process and segmentation used clustering technique. Now in our paper used POS model for segmentation process and reduces the loss of frame and video data during object tracking process.

In the process of review we found that some performance affected problem related to the video object detection. These problem are affected the performance and accuracy of video object tracking and result overcome in fact of loss of frame. The segmentation region increase, decrease the accuracy and performance of object tracking. Some problems are mentioned here :

1. Segmentation errors
2. Change of lighting conditions
3. Shadows
4. Occlusion
5. Automatic updating of background
6. False frame hit

III RELATED WORK

In this section discuss the related work of video segmentation and dynamic texture analysis. Some techniques discussed here.

[1] In this paper author surveys existing multi target tracking performance scores and, after discussing their limitations, they propose three parameter-independent measures for evaluating multi target video tracking. The measures consider target-size variations, combine accuracy and cardinality errors, quantify long-term tracking accuracy at different accuracy levels, and evaluate ID changes relative to the duration of the track in which they occur. They proposed three measures (METE, MELT, and NIDC) that quantify key factors in extended multi-target tracking: accuracy, cardinality and ID changes. These measures are parameter independent, numerically bounded and account for target-size changes. METE provides a holistic error assessment using an effective trade-off between accuracy and cardinality errors. MELT enables the analysis of tracking performance at varying accuracy levels that can facilitate the selection of trackers for specific applications. NIDC penalizes ID changes as a function of the length of the track in which they occur.

[2] In this paper author proposed a short video that will be a synopsis of an endless video streams to overcome the problems, and this video synopsis is an effective way to solve this problem to provide a compact video representation, while preserving the essential activities of the original video. Firstly, using Visual Background

Extractor (ViBe) motion detector algorithm extracted interest motion blob and frame segmented from the video, and that information compose a spatio-temporal object tube element; Secondly, we propose Multiple Object Tracking (MOT) technology match to interest blobs with effective frame build up the object motion tube in the camera surveillance view; Thirdly, using this paper propose the tracker sort rule, that according to the interest object maintain the view time, from long to short array to the tracker vector; Finally, the selected different array vector and pick up objects are stitched to the background image. Some experiments are performed using the proposed algorithm and the results are acceptable.

[3] In this paper, a new fusion state estimation method by fusing extended Kalman filter with particle filter is proposed to realize efficient and robust video target tracking. Extended Kalman filter has the time performance close to the Kalman filter and is more suitable for nonlinear video target tracking. Particle filter is based on non-parameter estimation and outperforms in robustness in video tracking. Fusion state estimation can obtain more accurate and reliable motion state of video target by optimizing the state estimation and prediction of video target. To further boost the efficiency of video tracking, this paper also presents an adaptive frames sampling method which utilizes the motion state of video target to skip some frames and then avoid frame by frame sampling. In addition, an efficient video target state observation method is introduced.

[4] In this work Author provided the first solid comparison of state-of-the-art general object trackers on hand tracking with a primary focus on grey-scale high-speed videos. Novel annotated high-speed video data were collected and made publicly available for evaluation purposes. The algorithms were tested with both finger and hand targets, and with grey-scale and color videos. In addition to tracking accuracies, the stability, sensitivity, and the processing speeds of the algorithms were evaluated. They presented the first solid comparison of the state-of-the-art general object tracking algorithms in hand and finger tracking focusing on high-speed videos. The experiments clearly showed that hand tracking is still a challenging task, although considerable advances have been made during the recent years.

[5] In this paper author addressed the problem of detecting and tracking multiple moving people based on color interest points. The proposed method uses the statistical Gaussian Mixture Model (GMM) for the segmentation, extraction of moving people and background area. After that, from the detected foreground they determine the rules that define skin regions for good people detection. Color Interest Points are identified in the detected regions of skin using Harris algorithm. The use of an interest points set allows us to track people by matching these ones from image to image based on ZNCC correlation approach (Zero mean Normalized Cross Correlation). Finally, by calculating Euclidean-distance between the best matches and other interest points detected on each consecutive images of video sequence, they can observe the motion of people tracked in the scene.

[6] In this paper author described a real-time highway surveillance system (RHSS), which operates autonomously to collect statistics (speed and volume) and generates incident alerts (e.g., stopped vehicles). The system is designed to optimize long-term real-time performance accuracy. It also provides convenient integration to an existing surveillance infrastructure with different levels of service. Innovations include a novel 3-D Hungarian algorithm which is utilized for object tracking and a practical, hands-off mechanism for camera calibration. Speed is estimated based on trajectories after mapping/alignment with respect to dominant paths learned based on an evolutionary dynamics model. The system, RHSS, is intensively evaluated under different scenarios such as rain, low-contrast and high-contrast lightings. Performance is presented in comparison to a current commercial product.

[7] In this paper author correlated automated activity measures through video analysis to ethological scores of pig activity, using off-line video recordings off our pens with grower pigs. Human observations (HO) of different behavioral activities were carried out by 2-mins can sampling during four 30-min sessions on 6 observation days. HO of pig activity was expressed as a mean proportion per session. Automated observations (AO) of pig activity were calculated by their lative number of moving pixels between two consecutive image frames (1frame/s) and expressed as a mean image activity index per session. The overall correlation between pig activity data from AO and HO was strong and positive (Rs¼0.92, Po0.0001).When comparing AO and HO data at session level, the correlation coefficients for the two afternoon sessions were lower.

[10] In this paper author describes a novel approach for semantic music automation and the details are to automatic music annotation and retrieval that captures temporal aspects as well as timbral content. The proposed approach leverages a recently proposed song model that is based on a generative time series model of the musical content the dynamic texture mixture (DTM) model that treats fragments of audio as the output of a linear dynamical system. To model characteristic temporal dynamics and timbral content at the tag level, a novel, efficient, and hierarchical expectation–maximization (EM) algorithm for DTM (HEM-DTM) is used to summarize the common information shared by DTMs modeling individual songs associated with a tag.

IV PROPOSED METHODOLOGY

In this section discuss the modified algorithm of HEM for clustering technique using genetic algorithm. The HEM of cluster used Random swap in terms of data iteration and reduction of iteration in processing of cluster. In the process of modification set the auto level center selection using genetic algorithm. The genetic algorithm processes the data in fashion in random manner. The auto swapping process of clustering technique assigned the variable of center point. Here steps are given below

1: Auto = (X,C) ←empty

2: C_list ← K-means (Ci_list, K_{auto})

3: Input C_list X , the clustering number pn , population scale XN , probability auto P stop conditions cS ;

4: Code the data in real number and initialize population S(i),i = 0 at random;

5: Evaluate the fitness of all individual in the current instant D(s);

6: HEM clustering requires optimization of cluster center, which way thrashing of data of waiting cluster. Hence the fitness function of algorithm is determined by f(x).

7: $$G(s)=\frac{N(s)}{D(s)}=\frac{\sum_{i=0}^{n-1}A_i s^i}{\sum_{i=0}^{n}a_i s^i}$$ Umpire the termination conditions. If the termination situation are satisfied, then turn to step 9, if not, turn to step 10;

8: Crack to find and compute the optimal clustering centers.

9: find final population of GA

10: Take the CR optimization on population P (i) and generate the next generation A (i +1). Then turn to step

11: for h ∈ A(i+1) do

12: h.nn ← CR (A(i+1)- {h})

13: h.sc ← Compute-SC (h, h.nn)

14: AUTO←AUTO ∪{h}

15: AUTO←AUTO ∪{h.nn}

16: if h.sc < th_{sc} then

17: E←E ∪ {(h,h.nn)}

18: End if

19: end for

20: count ← Matrix

for each pair of components (g1,g2) ∈ G do

21: μ_1←mean-dist (g1), μ_2←mean-dist (g2)

22: if $\frac{\mu_1+\mu_2}{2*centroid\_dist(g1,g2)}$ > 1 then g1←Merge (g1, g2)

23: end for

24: N_type ← empty

25: for x ∈ N list do

26: h←PseudopointOf(x)

27: HEM

28: end for

Algorithm step:

Step1

In first step the video is loaded. And apply feature extraction technique GMM model. The GMM model gives the texture features of given video. The extracted texture feature input the process of segmentation.

Step 2

After the process of population set the selection of center point find. The center point finds the position location of Centered value.

Step 3

The variable of auto level chose by the fitness constraints function, the fitness constraints decide the selection of parameter value for centric ratio.

Step4

The selection of centric process done and after that the processing of data are from and iteration process are done.

Step5

The process of iteration generates the number of maximum cluster and merging process is done.

Step 6 finally video are segmented.

Step 7 measures the value of accuracy and f-measure

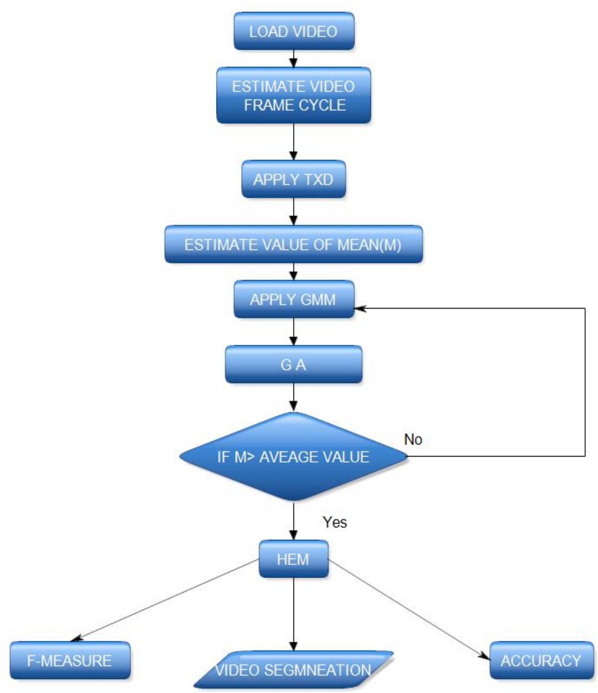

Figure 1: Proposed Model of video segmentation based on HEM and Genetic Algorithm.

V IMPLEMENTATION AND RESULT ANALYSIS

YouTube is a video sharing website, created by three former PayPal employees in February 2005 and owned by Google since late 2006, on which users can upload, view and share videos. The company uses Adobe Flash Video and HTML 5 technology to display a wide variety of user-generated video content, including video clips, TV clips, and music videos, and amateur content such as video blogging, short original videos, and educational videos. Here we using video clip of animal, sports athletes etc. all videos are divided into five different video such as V1, V2, V3, V4, and V5.

Name of video	Length	Frame Width	Frame Height	Data Rate	Total Bit Rate	Frame rate(in frame per Second)
V1	46	320	240	200 Kbps	264 Kbps	25
V2	43 S	320	240	200 Kbps	264 Kbps	25
V3	39	320	240	200 Kbps	264 Kbps	25
V4	37	320	180	200 Kbps	264 Kbps	29
V5	48	320	240	200 Kbps	264 Kbps	25

Table 1: Shows that the Different video description used in experimental process.

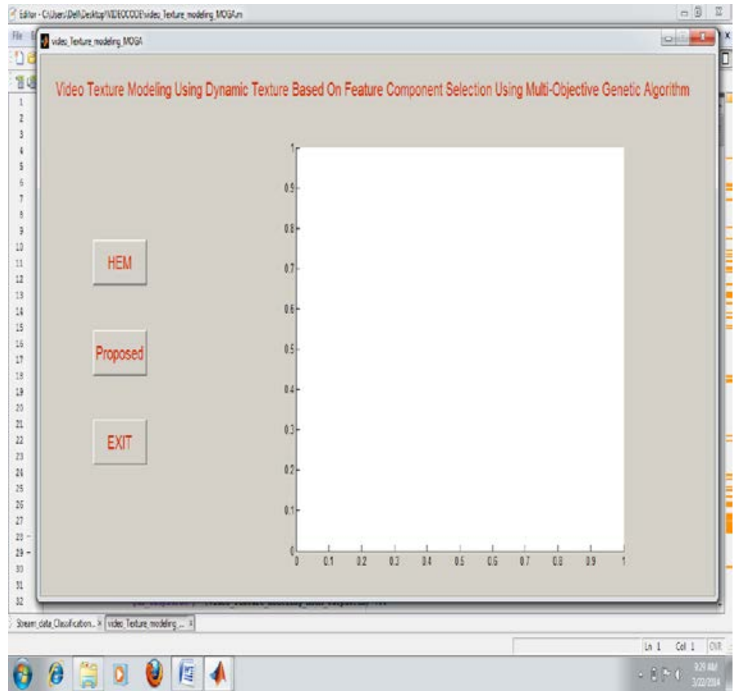

Figure 2: Shows that the main implementation window for result.

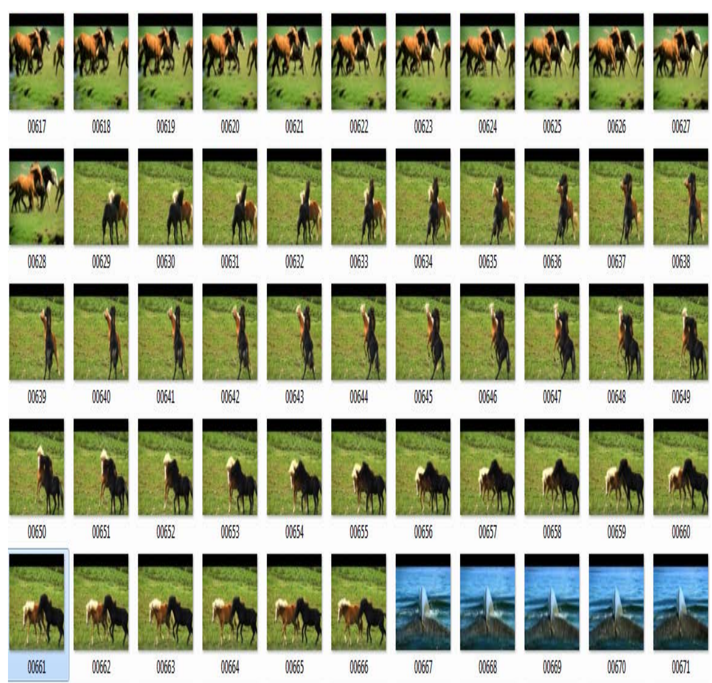

Figure 3: Shows that the Video track from animal video V1 with using HEM method.

Figure 4: Shows that the Video track from Animal video V1 with using proposed method.

Figure 5: Shows that the Video track from Sports video V4 with using HEM method.

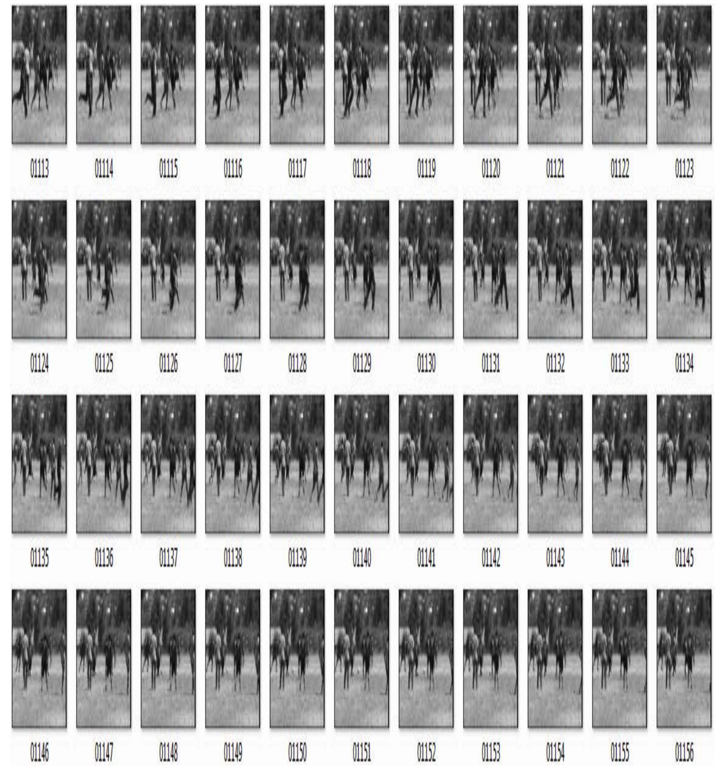

Figure 6: Shows that the Video track from Sports video V5 with using proposed method.

Method name	Average Precision	Average recall	Average F-Measure
HEM	93.81	90.81	86.81
Proposed Method	94.26	91.47	89.75

Table 2: Shows that the Average precision, Average recall and Average F- Measure for Video V1 using both HEM and proposed method.

Method name	Average Precision	Average recall	Average F-Measure
HEM	94.76	91.52	87.88
Proposed Method	95.46	92.32	90.26

Table 3: Shows that the Average precision, Average recall and Average F- Measure for Video V2 using both HEM and proposed method.

Method name	Average Precision	Average recall	Average F-Measure
HEM	95.16	92.67	88.93
Proposed Method	96.41	93.12	91.48

Table 4: Shows that the Average precision, Average recall and Average F- Measure for Video V3 using both HEM and proposed method.

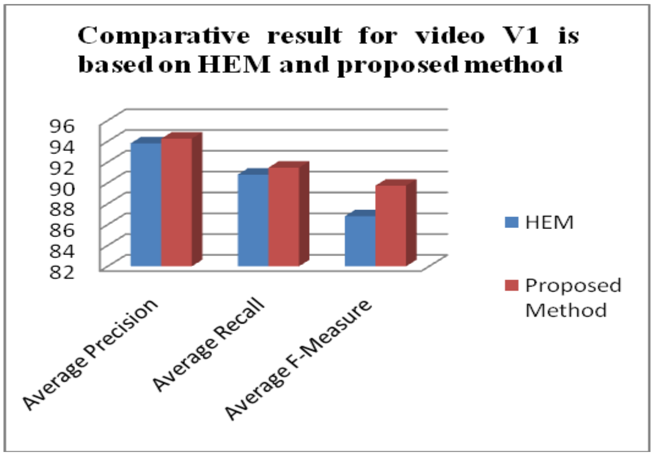

Figure 7: Shows that the Average precision, Average recall and Average F- Measure for Video V3 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

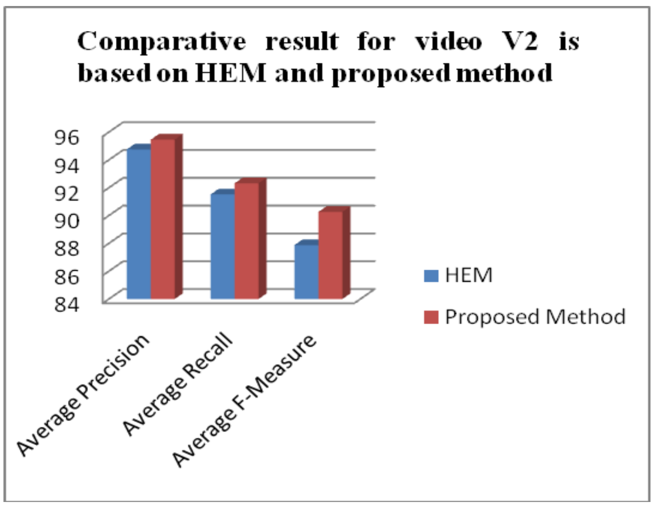

Figure 8: Shows that the Average precision, Average recall and Average F- Measure for Video V2 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

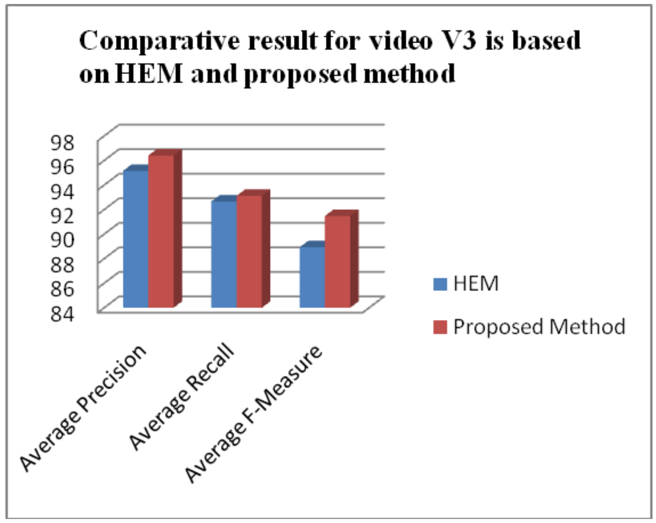

Figure 9: Shows that the Average precision, Average recall and Average F- Measure for Video V3 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

VI CONCLUSION AND FUTURE WORK

In this paper, we proposed a dynamic texture based video segmentation. The dynamic texture based video segmentation play an important role in object tracking and human detection. The video frame divided into number of frames, the number of frames changes the texture position during motion. The changing of texture position is raised a problem for video segmentation and clustering. In Future work will be directed at extending proposed algorithm to general graphical models, allowing a wide variety of generative models to be clustered or used as codeword in a bag-of-X representation. Finally, in this work we have not addressed the model selection problem, i.e., selecting the number of reduced mixture components. Since proposed is based on maximum likelihood principles, it is possible to apply standard statistical model selection approaches, such as Akaike information criterion (AIC) and Bayesian information criterion (BIC).

REFERENCES

[1] Tahir Nawaz, Fabio Poiesi, Andrea Cavallaro “Measures of Effective Video Tracking” IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, 2014. Pp 376-388.

[2] Bin Zhu,WenLiu,GangWeiandLin Yuan “A Method for Video Synopsis Based on
Multiple Object Tracking” IEEE 2014. Pp 414-418.

[3] Howard Wang, Sing Kiong Nguang “Video Target Tracking Based on Fusion State

Estimation" International Symposium on Technology Management and Emerging Technologies (ISTMET), 2014. Pp 337-343.

[4] Ville Hiltunen, Tuomas Eerola, Lasse Lensu, Heikki Kalviainen "Comparison of general object trackers for hand tracking in high-speed videos" 22nd International Conference on Pattern Recognition,2014. Pp 2215-2220.

[5] Khadija Laaroussi, Abderrahim Saaidi, ,Mohamed Masrar, Khalid Satori "Video-Surveillance System for Tracking Moving People Using Color Interest Points" World Applied Sciences Journal 32 (2), 2014. Pp 289-301.

[6] Yiwen Wan, Yan Huang, Bill Buckles "Camera calibration and vehicle tracking: Highway traffic video analytics" Transportation Research Part C 44, 2014.Pp 202-213.

[7] S. Ott, C.P.H.Moons , M.A.Kashiha , C.Bahr , F.A.M.Tuyttens , D. Berckmans, T.A.Niewold "Automated video analysis of pig activity at pen level highly correlates to human observations of behavioural activities" Livestock Science 160, 2014. Pp 132-137.

[8] Antoni B. Chan ,Nuno Vasconcelos "Mixtures of Dynamic Textures" Appears in IEEE International Conference on Computer Vision, Beijing, 2005. Pp 1-7.

[9] Rui Li, Tai-Peng Tian, Stan Sclaroff "Simultaneous Learning of Nonlinear Manifold and Dynamical Models for High-dimensional Time Series" To appear in Proc. International Conference on Computer Vision, 2007. Pp 241-249.

[10] Emanuele Coviello, Antoni B. Chan, Gert Lanckriet "Time Series Models for Semantic Music Annotation" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, 2011. Pp 1343-1359.

[11] Hayit Greenspan, Jacob Goldberger, Arnaldo Mayer "Probabilistic Space-Time Video Modeling via Piecewise GMM" 2000. Pp 1-29.

[12] Antoni B. Chan, Emanuele Coviello, Gert. R. G. Lanckriet "Clustering Dynamic Textures with the Hierarchical EM Algorithm" 2011. Pp 1-8.

[13] Xiaofei Di, Hong Chang, Xilin Chen "Multi-layer Spectral Clustering for Video Segmentation" 2011. Pp 1-12.

[14] Lin, F, Cohen, W.W. "Power iteration clustering" ICML 2010.

[15] Comaniciu, D, Meer, P. "Mean shift: a robust approach toward feature space analysis" PAMI 2002.

[16] Kohli, P Ladicky, L, P.Torr"Robust higher order potentials for enforcing label consistency" 2008.

[17] Hu, G, Gao, Q "A non-parametric statistics based method for generic curve partition and classification" ICIP. 2010.

[18] Vazquez-Reina, A., Avidan, S., P_ster, H., Miller "Multiple hypothesis video segmentation from superpixel ows" 2010.

[19] Arbelaez, P., Maire, M., Fowlkes, C., Malik "From contours to regions:an empirical evaluation. In: CVPR. 2009.

[20] Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri "Action as space-time Shapes" IEEE TPAMI 2007.