Volume-11, Issue-07, July 2022
JOURNAL OF COMPUTING TECHNOLOGIES (JCT)
International Journal
Page Number: 01-09

Video Texture Modeling Using Dynamic Textures Based On Feature Component Selection Using POS Optimizations

Komal Sharma[1], Prof . Shivraj Singh[2],
PG Scholar(M.Tech)[1], Assistant Professor[2],
Department of Electronics and Communication[1,2],
[1,2,]Technocrats Institute of Technology, Bhopal(M.P), India

Kom.elex@gmail.com[1], 86.shivraj@gmail.com[2]

***Abstract*— Dynamic texture video segmentation plays an important role in computer vision and gait recognition. The dynamic texture video segmentation used various clustering and classification technique for improvement of accuracy value. Now in current research trend hierarchical estimation clustering technique (HEM), the HEM method is unsupervised learning technique and work as markov model estimation. Segmentation is the art of automatically separating video into different regions in a fashion that mimics the human visual system. It is therefore a broad term that is highly dependent on the application at hand. one might want to segment each object individually, groups of objects, parts of objects. In order to segment a particular image, one must first identify the intended result before a set of rules can be chosen to target this goal. The human eye uses low-level information such as the presence of boundaries, regions of different intensity or colors, brightness and texture, etc., but also mid-level and high-level cognitive information, for example, to identify objects or to group individual objects together. In this reseach paper modified the HEM clustering technique for video segmentation using particle of swarm optimization. Particle of swarm optimization is population based optimization technique. The particle of swarm optimization 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.**

Keywords - HEM clustering, MATLAB software, Dynamic texture, Video Segmentation, Markov Model Estimation and empirical evaluation.

I. 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[1]. 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. As the volume of the database becomes huge, manual segmentation and indexing became very hard to apply [2]. Automatic segmentation and indexing through computer will be very useful. Thus, segmentation of a video into its constituent short pieces is fundamental functionality for video retrieval and management tasks. The basic idea is to embed on information into the signal of the media (audio, video, or photo). Today all DVD movies, video games, audio CDs, etc. have fingerprints that prove the ownership of the material [7]. As a disadvantage, watermarks are generally fragile to visual transformations (e.g., re-encoding, change of the resolution/bit rate). For example, hidden data embedded on a movie will probably be lost when the clip is compressed and uploaded to a video sharing web site.

Besides, temporal information of the video segments (e.g., frame number, time-code) are also important in some applications. Watermarking technique is not designed to be used for video retrieval by querying with a sample video clip [10].

1.2MULTIMEDIA SYSTEM

In the multimedia indexing and retrieval a lot of system has been developed. However, no system or technology has yet become widely pervasive. By the improvement in the audio-visual analysis techniques, better results are getting place. Most of the developed multimedia indexing and retrieval systems used visual features for indexing. Recently, the audio features are used in combination with the visual features in multimedia indexing and retrieval systems[11].

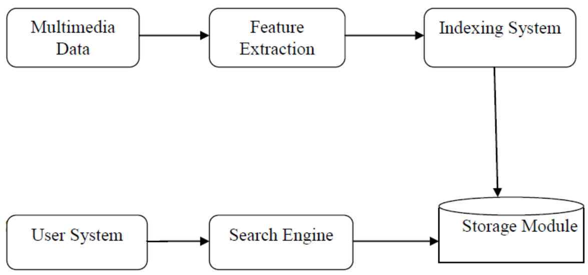

Fig. 1.1: A multimedia indexing and retrieval system

1.3 VIDEO SEGMENTATION

The idea of applying dynamic texture representations to the segmentation of videohas 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 autoregressive processes (AR) and level-sets, and segments video by clustering pixels with similar trajectories in time, using generalized PCA (GPCA). While these methods have shown promise, they do not exploit the probabilistic nature of the dynamic texture representation for the segmentation itself. On the other hand, the segmentation algorithms proposed in the following section are statistical procedures that leverage on the mixture of dynamic texture to perform optimal inference [12].

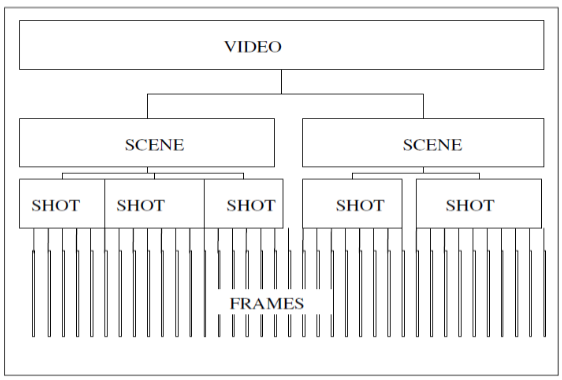

Fig. 1.2: A Video Structure

II. VIDEO SEGMENTATION TECHNIQUES

2.1 INTRODUCTION

Now describes video as a medium, and gives a thorough explanation of the process of video segmentation. Some known methods and algorithms regarding this matter will be introduced. Since segmentation algorithms are a very important part of this thesis, this section is quite comprehensive. I think it is important to illustrate a wide selection of earlier results and conclusions, and use this knowledge to explore new possibilities instead of doing work that already has been done and evaluated. Further, this chapter suggests how to select the best pictures from a video file to build a complete picture storyline, which from a user's point of view will represent a compact summation of the video file content. The final section presents methods for speeding up the entire segmentation process, with the goal of keeping the output quality as unaffected as possible [13].

2.2 VIDEO AS A MEDIUM To design efficient video shot detection algorithms, it is important to examine video as a medium to fully understand this domain. Video is a continuous series of pictures displayed sequentially at a fixed rate. Because all the pictures in a video file have equal size, the pictures are called frames. Digital video information consists of a series of 25 frames per second. These frames can be grouped together in related collections, to make the handling of the video file easier. The collections are defined as follows, and will be used through the entire research work [14].

- Clip: A clip is a digital video document. It can last from a few seconds to several hours, and consists of a sequence of contiguous video frames. A video segment is any contiguous portion of a clip.
- Scene: A scene is a sequential collection of shots unified by a common event or locale. A clip can have one scene or several scenes.
- Shot: A shot is captured between a record and a stop camera operation. A scene can have one shot or several shots.

Frame: A frame is the atomic part related to digital video, and is one picture from the picture sequence. This can be organized hierarchical. A shot is a collection of frames, a scene is a collection of shots and a clip is a collection of scenes.

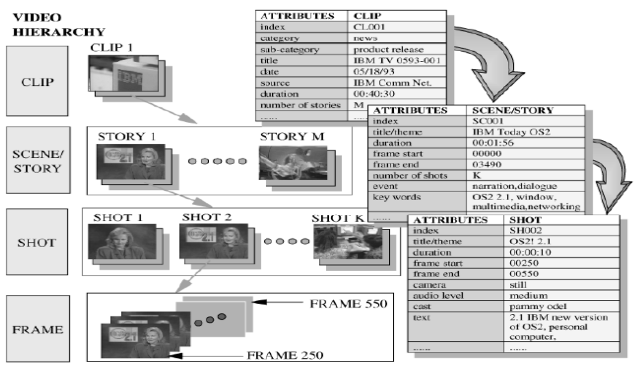

Figure 2.1: A hierarchy of digital video.

A typical digital video codec design starts with conversion of camera-input video from RGB color format to YCbCr color format. The conversion to YCbCr provides some benefits. First, it improves compressibility by providing de-correlation of the color signals. Second, it separates the luma signal, which is perceptually much more important, from the chroma signal, which is less perceptually important and which can be represented at lower resolution. To reduce the raw data rate before the basic encoding process, spatial and temporal down sampling may be used. The most popular transform is the discrete cosine transform (DCT), which is similar to the discrete Fourier transform (DFT), but using only real numbers. DCT is used both in digital video and image compression. A two-dimensional DCT of N × N blocks is computed, and the results are quantized and entropy coded. Quantization is the process of approximating a continuous range of values by a relatively small set of discrete integer values, while entropy coding involves the actual compression process. N is typically 8, and the following formula is applied to each row and column of the block [15]:

III LITERATURE SURVEY

***Adee Mumtaz, et.al*, " Clustering Dynamic Textures With The Hierarchical EM Algorithm For Modeling Video"** Those author proposed a clustering algorithm with hierarchical EM algorithm for dynamic texture and video, the details are is capable of both clustering DTs and learning novel DT cluster centers. To derive an efficient recursive algorithm for sensitivity analysis of the discrete-time Kalman smoothing filter, which is used as the basis for computing expectations in the E-step of the HEM algorithm. Finally, they demonstrate the efficacy of the clustering algorithm on several applications in motion analysis, including hierarchical motion clustering, semantic motion annotation, and learning bag-of-systems (BoS) codebooks for dynamic texture recognition. They address the problem of clustering dynamic texture models, i.e., clustering linear dynamical systems. Given a set of DTs (e.g., each learned from a small video cube extracted from a large set of videos), the goal is to group similar DTs into K clusters, while also learning a representative DT “center” that can sufficiently summarize each group. This is analogous to standard K-means clustering, except that the data points are dynamic textures instead of real vectors[1].

***JorgStuckler et.al.* "Hierarchical Object Discovery And Dense Modeling From Motion Cues In Rgb-D Video",** Those author proposed an approach for simultaneously segments rigid-body motion within key views, and discovers objects and hierarchical relations between object parts. The poses of the key views are optimized in a graph of spatial relations to recover the rigid-body motion trajectories of the camera with respect to the objects. In experiments, they demonstrate that our approach finds moving objects, aligns partial views on the objects, and retrieves hierarchical relations between the objects. They introduced a novel method for learning object maps with hierarchical part relations from motion cues. Motion segmentation between RGB-D key views finds the rigid parts in images and estimates their motion. It is based on an efficient expectation-maximization algorithm and employs a compact local multi-resolution 3D representation of RGB-D images to process images efficiently. They integrate our motion segmentation method with SLAM into a framework for simultaneous motion segmentation, localization, and mapping. Our mapping approach extracts moving objects from key views and aligns the parts by optimizing a graph of spatial relations. From the overlap of motion segments, we deduce a hierarchy of object parts. In experiments, They demonstrated that our approach is capable of extracting motion segments and aligning multiple views on objects. In each of the sequences, there approach deduces hierarchical object relations [2].

***Katherine Ellis et.al.*,"A Bag Of Systems Representation For Music Auto tagging",** Those author proposed a content-based automatic tagging system and the details are, it is use for music that relies on a high-level, concise “Bag of Systems” (BoS) representation of the characteristics of a musical piece. Compared to estimating a single generative model to directly capture the musical characteristics of songs associated with a tag, the BoS approach offers the flexibility to combine different generative models at various time resolutions through the selection of the BoS code words. Additionally, decoupling the modeling of audio characteristics from the modeling of tag-specific patterns makes BoS a more robust and rich representation of music They have presented a semantic auto-tagging system for music that leverages a rich “bag of systems” representation based on generative modeling. The Bo representation allows for the integration of the descriptive qualities of various generative models of musical content with different time resolutions into a single histogram descriptor of a song [3].

***Katherine Ellis, et.al.* " Semantic Annotation And Retrieval Of Music Using A Bag Of Systems Representation",** Those author presented a semantic auto-tagger that leverages a rich “bag of systems” representation of music. The latter can be learned from any representative set of songs, which need not be annotated, and allows integrating the descriptive quality of various generative models of musical content, with different time resolutions. This approach improves performance over directly modeling tags with a single type of generative model. It also proves significantly more robust for tags with few training examples [4].

Antoni B. Chan et.al., "Modeling, Clustering, And Segmenting Video With Mixtures Of Dynamic Textures"Those author study about the dynamic textures and the details are, a statistical model for an ensemble of video sequences that is sampled from a finite collection of visual processes, each of which is a dynamic texture. An expectation maximization (EM) algorithm is derived for learning the parameters of the model, and the model is related to previous works in linear systems, machine learning, time-series clustering, control theory, and computer vision. Through experimentation, it is shown that the mixture of dynamic textures is a suitable representation for both the appearance and dynamics of a variety of visual processes that have traditionally been challenging for computer vision (e.g. fire, steam, water, vehicle and pedestrian traffic, etc.). When compared with state-ofthe-art methods in motion segmentation, including temporal texture methods and traditional representations (e.g. optical flow or other localized motion representations), the mixture of dynamic textures achieves superior performance in the problems of clustering and segmenting video of such processes They have studied the mixture of dynamic textures, a principled probabilistic extension of the dynamic texture model. They derived an exact EM algorithm for learning the parameters of the model from a set of training video, and explored the connections between the model and other linear system models, such as factor analysis, mixtures of factor analyzers, and switching linear systems. Through extensive video clustering and segmentation experiments [5].

Emanuele Coviello et.al. " Automatic Music Tagging With Time Series Models " Those author describes about the music tagging and the details are, a novel Approach to automatic music annotation and retrieval that captures temporal aspects. The proposed approach shows here at that 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 hierarchical EM algorithm for DTM (HEM-DTM) is used to summarize the common information shared by DTMs modeling individual songs associated with a tag. Experiments show learning the semantics of music benefits from modeling temporal dynamics. They have presented the dynamic texture mixture model; a principled approach for capturing the temporal, as well as timbral qualities of music. They derived a hierarchical algorithm for efficiently learning DTM models from large training sets, enabling its usage as a tag model for semantic annotation and retrieval. Experimental results demonstrate that the new model improves accuracy over current bag-of-feature approaches [6].

IV.PROPOSED METHODOLOGY AND ARCHITECTURE

Video segmentation and dynamic texture analysis play an important role in computer vision. Segmentation and dynamic texture used for human tracking and object detection. The better prediction and detection of human and object precede the segmentation process. The process of segmentation performs by different method such as clustering technique and neural network. The process of clustering technique is unsupervised learning and process is done by iteration. The process of clustering applied in form of diversity such as k-means clustering hierarchal clustering and estimated clustering technique for segmentation. In this desertion apply the video segmentation process improved clustering technique. The process of improved clustering technique applies on dynamic texture of video. For the texture analysis of video used Gaussian mixture model, GMM model well knows feature extraction method for video segmentation and object tracking. for the improvement of hierarchal clustering technique used particle of swarm optimization. Particle of swarm optimization is population based optimization technique used a concept of bird fork. In this chapter also discuss GMM model particle of swarm optimization, hierarchal clustering technique and finally proposed algorithm and model.

4.2 GAUSSIAN MIXTURE MODEL (GMM) Gaussian mixture model is originally designed by Fisher for taxonomic segmentation. GMM searches for those vectors in the underlying space that best discriminate among classes (rather than those that best describe the data). More formally, given a number of independent features relative to which the data is described, GMM creates a linear combination of these which yields the largest mean differences between the desired classes. It tries to find an optimal reducing dimensionality linear projection that maximizes the scatter of all projected samples. However, for segmentation, the between class scatter should be maximized, while the within-class scatter should be minimized. If a data set is categorized, it makes sense to use the class information to build a more desirable projection space to improve discrimination while reducing the dimensionality of the feature space. GMM is an example of a class specific method, in the sense that it tries to "shape" the scatter in order to make it more favorable for segmentation. This method seeks the projections that maximize the ratio of the between-class scatter to the within-class scatter in the projection space.

4.3 PARTICLE OF SWARM OPTIMIZATION In Particle Swarm Optimization optimizes an objective function by undertaking a population based search. The population comprise of possible solutions, named particles, which are metaphor of birds in flocks. These particles are at random initialized and freely fly across the multi dimensional seek space. During flight, each particle updates its own velocity and position based on the best experience of its own and the entire population.

4.4 HEM ALGORITHM

The HEM algorithm is a general technique for maximum likelihood estimation. In practice HEM has been applied almost exclusively to unsupervised learning problems. This is true of the neural network literature and machine learning literature, in which HEM has appeared in the context of clustering and density estimation as well as the statistics literature, in which applications include missing data problems, mixture density estimation and factor analysis. Another unsupervised learning application is the learning problem for Hidden Markov Models, for which the Baum-Welch estimation formulas are a special case of HEM. There is nothing in the HEM framework that precludes its application to regression or classification. HEM is an iterative approach to maximum likelihood estimation. Each iteration of an HEM algorithm is composed of two steps: an Estimation (E) step and a Maximization (M) step.

PROPOSED MODEL

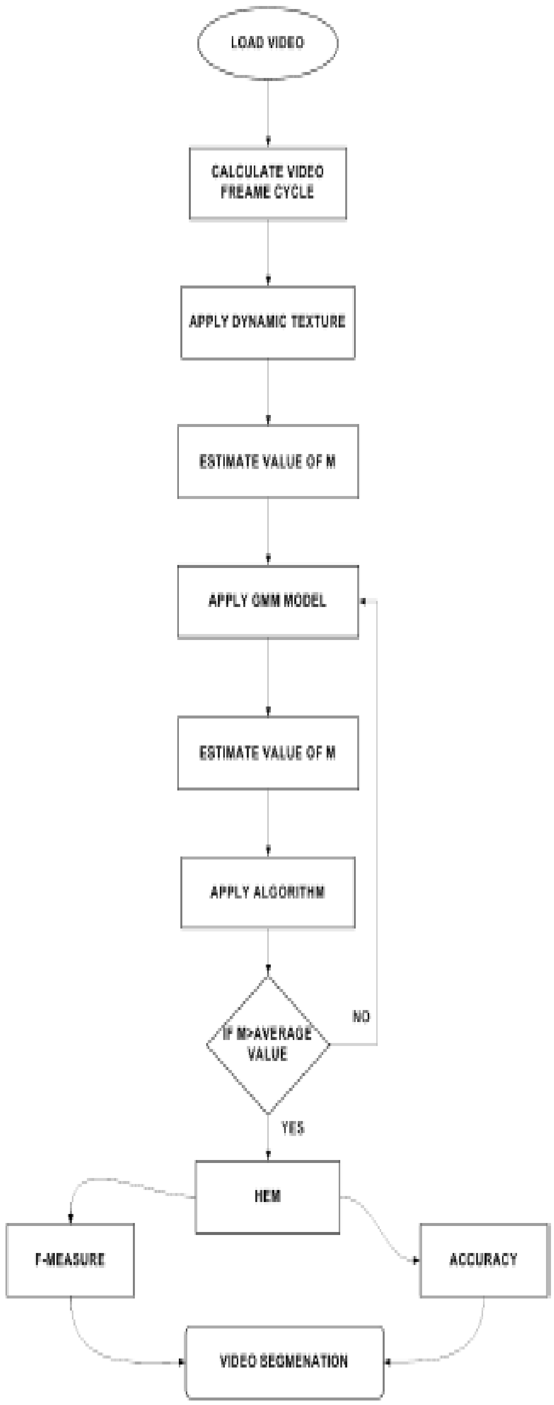

Figure 4.1: proposed model of video segmentation

V. RESULT AND DISSCUSUION

DESCRIPTION OF MATLAB

For the performance evaluation of ensemble classifier technique and our cascaded model used MATLAB software package. MATLAB is a software package for high- performance numerical computation and visualization. It provides an interactive environment with hundreds of built-in function for technical computation, graphics and animation. Best of all, it also provides easy extensibility with its own high- level programming language. The MATLAB stands for matrix laboratory. There are also several optional "toolboxes" available from the developers of MATLAB. These toolboxes are collections of functions written for special applications such as symbolic computation, image processing, statistics, control system design, neural networks etc. the list of toolboxes keeps growing with time. There are now more than 50 such toolboxes. One of the best features of MATLAB is its platform independence. Once you are in MATLAB, for the most part, it does not matter which computer you are on. Almost all commands work the same way.

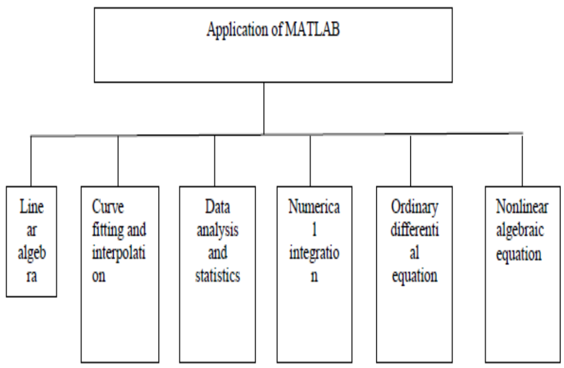

Figure 5.1: Application of MATLAB.

The efficiency of programs that are used often and by several different people can be improved by simplifying the input and output data management. The use of Graphic User Interfaces (GUI), which provides facilities such as menus, pushbuttons, sliders etc, allows Programs to be used without any knowledge of MATLAB. They also provide means for efficient data management. A graphic user interface is a MATLAB script file customized for repeated analysis of a specific type of problem. There are two ways to design a graphic user interface**.** It is a user interface built with graphical objects, such as buttons.

5.2 DESCRIPTION OF DATASET

YouTube is a video sharing website, created by three former PayPal employees inFebruary 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 5technology 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.

Video Description

Name of video	Length	Frame Width	Frame Height	Data Rate	Total Bit Rate	Frame rate
V1	46 seconds	320	240	200 Kbps	264 Kbps	25 Frames/second
V2	43 Seconds	320	240	200 Kbps	264 Kbps	25 Frames/second
V3	39 Seconds	320	240	200 Kbps	264 Kbps	25 Frames/second
V4	37 seconds	320	180	200 Kbps	264 Kbps	29Frames/ second
V5	48 seconds	320	240	200 Kbps	264 Kbps	25 Frames/second

Table 5.1: Shows that the Different video description used in experimental process.

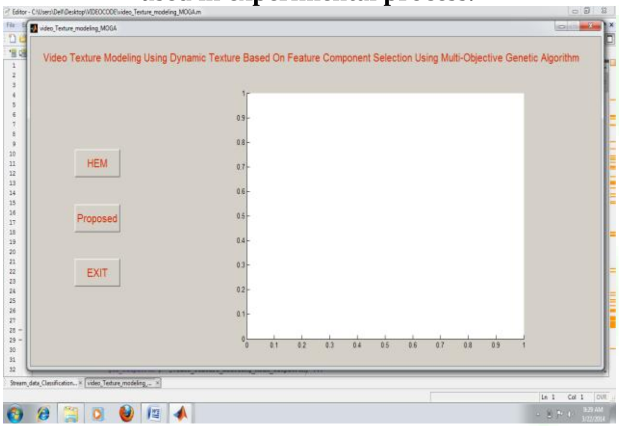

Figure 5.1: Shows that the main implementation window for result.

Figure 5.3.1: Shows that the Video track from animal video V1 with using HEM method.

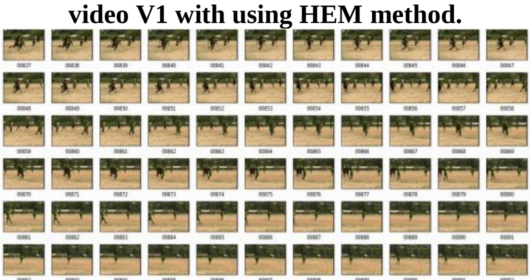

Figure 5.3.2: Shows that the Video track from Sports video V2 with using HEM method. Figure

Figure 5.3.3: Shows that the Video track from Sports video V3 with using HEM method.

Figure 5.3.4: Shows that the Video track from Sports video V4 with using HEM method.

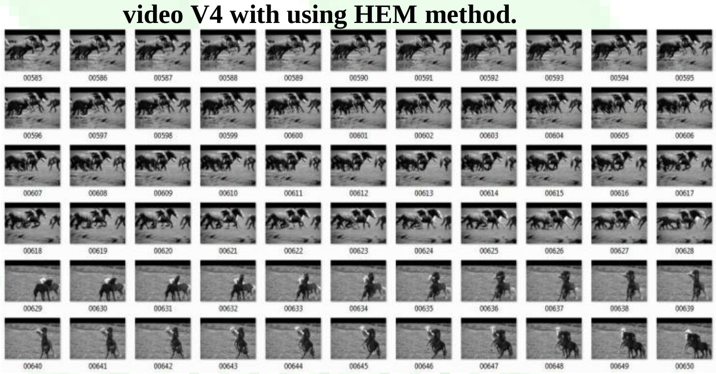

Figure 5.3.5: Shows that the Video track from Animal video V1 with using proposed method.

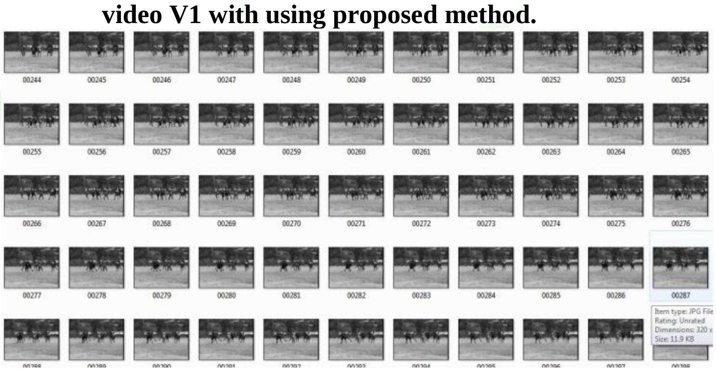

Figure 5.3.6: Shows that the Video track from Sports video V2 with using proposed method.

Figure 5.3.7: Shows that the Video track from Sports video V3 with using proposed method.

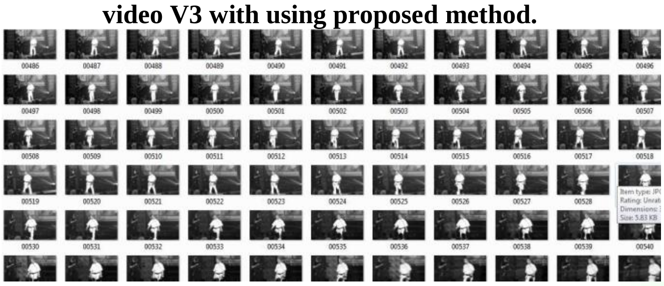

Figure 5.3.8: Shows that the Video track from Sports video V4 with using proposed method.

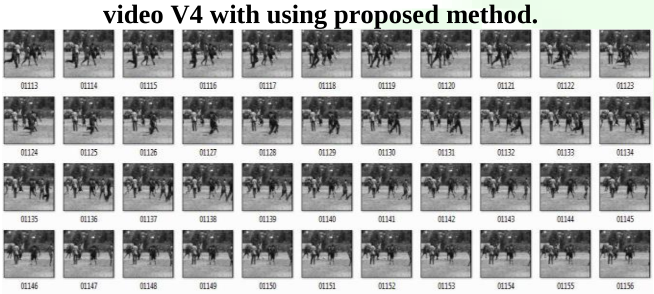

Figure 5.3.9: Shows that the Video track from Sports video V5 with using proposed method.

5.4 COMPARATIVE RESULT ANALYSIS :-

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 5.2: Shows that the Average precision, Average recall and Average FMeasurefor 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 5.3: 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 5.4: Shows that the Average precision, Average recall and Average F-Measure for Video V3 using both HEM and proposed method.

Method name	Average Precision	Average recall	Average F-Measure
HEM	95.83	93.46	89.33
Proposed Method	97.60	94.50	92.81

Table 5.5: Shows that the Average precision, Average recall and Average F-Measure for Video V4 using both HEM and proposed method.

Method name	Average Precision	Average recall	Average F-Measure
HEM	96.17	94.40	90.57
Proposed Method	98.81	95.86	93.40

Table 5.6: Shows that the Average precision, Average recall and Average F-Measure for Video V5 using both HEM and proposed method.

5.5 COMPARATIVE RESULT GRAPH

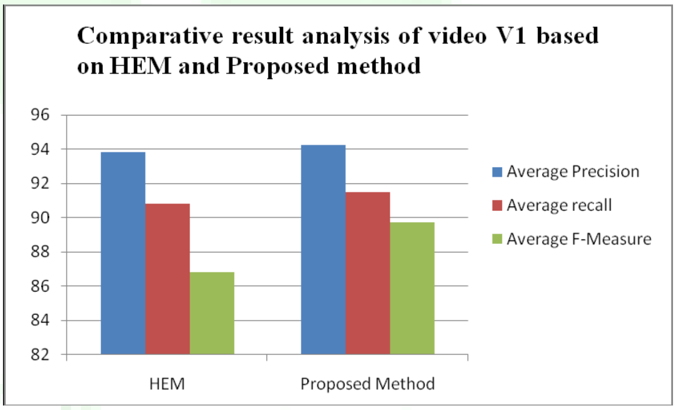

Figure 5.5.1: Shows that the Average precision, Average recall and Average F- Measure for Video v1 using both HEM and proposed method, and here ourproposed method shows the better results than HEM method.

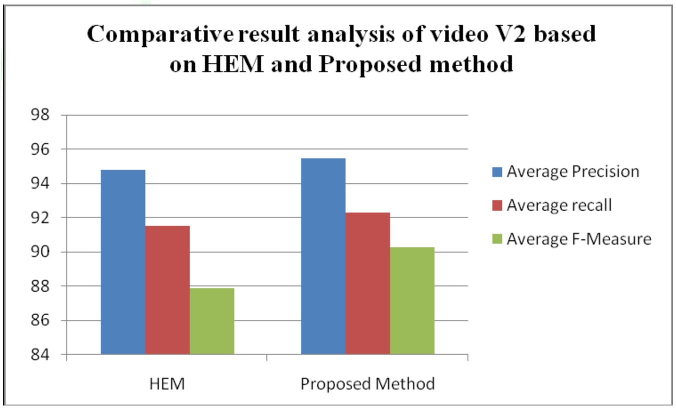

Figure 5.5.2: Shows that the Average precision, Average recall and Average F- Measure for Video V2 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

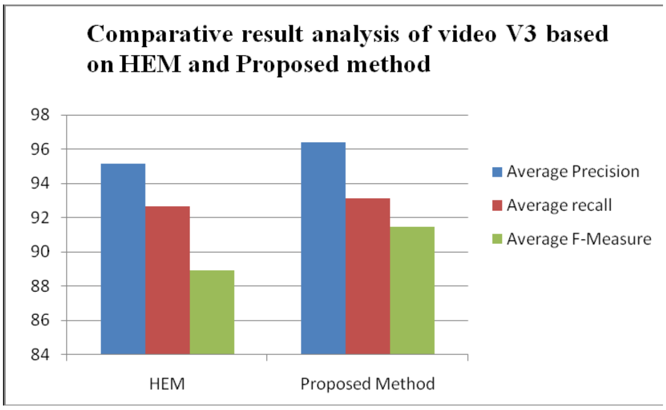

Figure 5.5.3: Shows that the Average precision, Average recall and Average F- Measure for Video V3 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

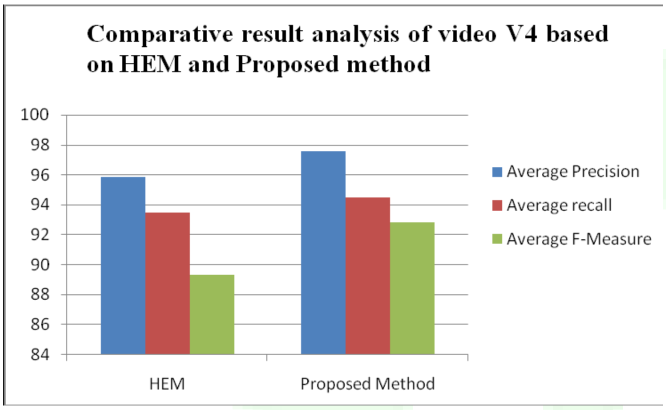

Figure 5.5.4: Shows that the Average precision, Average recall and Average F- Measure for Video V4 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

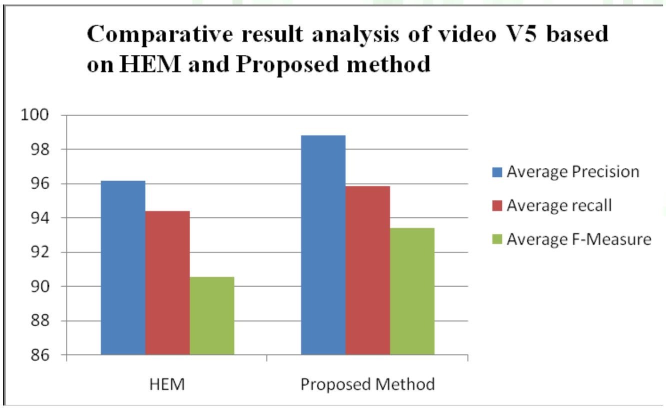

Figure 5.5.5: Shows that the Average precision, Average recall and AverageF- Measure for Video V5 using both HEM and proposed method, and here our proposed method shows the better results than HEM method.

V.CONCLUSION

In this dissertation 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. One method for video segmentation is HEM very promising in the case of dynamic texture. But this method suffered from a problem of selection of M value. For the selection of M value used optimization technique for better prediction of clustering. For the optimization of M value used particle of swarm optimization. Particle of swarm optimization is population based optimization technique. The processes ofoptimization find local and global value of optimal selection of M. For the validation and testing of proposed algorithm used MATLAB software implementation process and used Google texture video and YouTube video. For the measurement of performance used some standard parameter such as average precision, average recall and accuracy. The proposed method gives better result in compassion of HEM method. The process of texture feature extraction process is done by dynamic feature descriptor is called TXD. The texture feature descriptor representation is a descriptor of motion in a video, where dynamic texture codeword's represents the typical motion patterns in spatiotemporal patches extracted from the video. The TXD representation of videos is analogous to the bag-of-visual-words representation of images, where images are represented by counting the occurrences of visual code word in the image. Specifically, in the TXD framework the codebook is formed by generative time-series models instead of words, each of them compactly characterizing typical textures and dynamics patterns of pixels in a spatiotemporal patch. Hence, each video is represented by a TXD histogram with respect to the codebook by assigning individual spatiotemporal patches to the most likely code word and then counting the frequency with which each codeword is selected.

SUGGESTIONS FOR FUTURE WORK

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 Akanke information criterion (AIC) and Bayesian information criterion (BIC).

REFERENCES

[1] AdeelMumtaz, Emanuele Coviello, Gert R.G. Lanckriet, Antoni B. Chan"Clustering Dynamic Textures with the Hierarchical EM Algorithm for Modeling Video" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 35, 2013. Pp 1606-1621.

[2] JorgStuckler, Sven Behnke "Hierarchical Object Discovery and Dense Modelling From Motion Cues in RGB-D Video" IJCAI 2013. Pp 1-8.

[3] Katherine Ellis, Emanuele Coviello, Antoni B. Chan, GertLanckriet "A Bag of Systems Representation for Music Auto-Tagging" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, 2013. Pp 2554-2570.

[4] Katherine Ellis, Emanuele Coviello, Gert R.G. Lanckriet "SEMANTIC ANNOTATION AND RETRIEVAL OF MUSIC USING A BAG OF SYSTEMS REPRESENTATION" 2012. Pp 1-6.

[5] Antoni B. Chan, NunoVasconcelos "Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2007. Pp 1-18.

[6] Emanuele Coviello, Luke Barrington, Antoni B. Chan, Gert. R. G. Lanckriet"Automatic Music Tagging With Time Series Models" 2010.Pp 1-6.

[7] GwenaellePiriou, Patrick Bouthemy, Jian-Feng Yao "Recognition of Dynamic Video Contents With Global Probabilistic Models of Visual Motion" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, 2006. Pp 3418-3431.

[8] Antoni B. Chan ,NunoVasconcelos "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, Oct. 2007. Pp 1-9.

[10] Emanuele Coviello, Antoni B. Chan, GertLanckriet "Time Series Models for Semantic Music Annotation" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, 2011. Pp 1343-1360.

[11] Hayit Greenspan, Jacob Goldberger, ArnaldoMayer "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 superpixelows" 2010.

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

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

