A New Hybrid Video Segmentation Algorithm using Fuzzy C Means Clustering

K.Mahesh[1] and Dr.K.Kuppusamy[2]

[1] Department of Comp.sci & Engg, Alagappa University, Karaikudi
630003, Tamilnadu, India

[2] Department of Comp.sci & Engg, Alagappa University, Karaikudi
630003, Tamilnadu, India

Abstract

This paper is concerned with the development of video segmentation technique for segmenting both the static and dynamic objects in the video sequence. The effective extraction of objects from the frame is an important process of video segmentation. The proposed technique uses the fuzzy C means approach for object extraction. The proposed approach uses the joint use of frame difference algorithm with the background subtraction method as well as consecutive frame difference for segmentation of static and dynamic objects. The prior research used the FKM based object extraction method whereas the proposed approach uses the improved FCM based object extraction method. The proposed technique is evaluated by varying video sequences and the efficiency is analyzed by calculating the statistical measures and kappa coefficient.

Keywords: - Video segmentation, Discrete cosine transform, Fuzzy C Mean Clustering, Fuzzy K Mean clustering , backround subtraction method

1.Introduction

Video, a rich information source, is commonly used for capturing and sharing knowledge in learning systems[17].With the rapid progress in video technology, video has quickly become an essential component of today's multimedia applications [15] [13], including consisting of VCR (Video Cassette Recording), Video-on-Demand, virtual walkthrough and more with the speedy development in video technology [1]. A huge amount of video data is now extensively available owing to the technology's progress in multimedia, digital TV and information systems [2]. Due to the sudden growth in digital video content, an efficient way to access and manipulate the information in a huge video database has become a difficult and timely issue [3]. Therefore, the need for developing tools and systems that can effectively search and retrieve the desired video content has gained enormous popularity among researchers.

In modern times, video object segmentation has emerged as one of the most imperative and challenging area of research. Segmentation is highly dependent on the model and criteria for grouping pixels into regions [10].The principal objective of video object segmentation is to facilitate content-based representation by extracting objects of interest from a series of consecutive video frames[16] [4]. Video segmentation aims to segment moving objects in video sequences. [5]. several video segmentation algorithms have been proposed. They can be classified into types shape based video segmentation, edge information based video segmentation, image based video segmentation, texture based video segmentation and color based segmentation. Video segmentation can be divided into temporal segmentation and spatio-temporal segmentation [11]. Both are used in coding and analysis. All video coding standards include some type of temporal segmentation in the encoder [12].Various approaches have been proposed for video or spatio-temporal segmentation. The video object extraction could be based on change detection and moving object localization, or on motion field segmentation, particularly when the camera is moving [7].

The structure of the paper is organized as follows: A brief review of the researches related to the video segmentation is given in Section 2. The proposed video segmentation technique is given in section 3. The experimental results of the proposed approach are presented in Section 4. Finally, the conclusions are given in Section 5.

2. Related work

Chasanis *et al.* [1] have proposed herein, local invariant descriptors were used to represent the key-frames of video shots and a visual vocabulary was created from these descriptors resulting to a visual words histogram representation (bag of visual words) for each shot.

Kinoshenko *et al.* [2] have discussed with unstructured information traditionally relies on shot boundary detection and key frames extraction. For content interpretation and for similarity matching between shots, video segmentation, i.e. detection of similarity-based events, were closely related with multidimensional time series representing video in a feature space.

Yadav *et al*. [3] have analyzed the segmentation of video as deformation Video quality improvement and blurring of video. Various authors improved the quality of video using sliding window technique and using different noise filter but all these method were not accurate and suffering for object prediction in noise video.

Boltz *et al.* [4] have proposed to define the energy as a function of (an estimation of) the MCE distribution. This function was chosen to be a continuous version of the Ahmad-Lin entropy approximation, the purposed being to be more to outliers inherently presented in the MCE.

3. Dynamic and static objects segmentation by means of efficient object extraction

The Extraction of all the individual objects from the frame is the most critical issue in the video segmentation. The proposed video segmentation technique concentrates more on extraction of objects from each frame. The global motion constraints are not considered in the segmentation of dynamic and static objects. The motion segmentation process is carried out by both the frame difference algorithm and intersection method subsequently the most common and accurate segmented objects are retrieved from both the segmented results whereas the static foreground are segmented using the intersection of consecutive frames.

3.1 Zoning of shots

Video decomposition into temporal unit is an essential pre-processing task for a wide range of video manipulation applications. Shot zoning is one of the video decomposition techniques which aim to partition a video sequence into shots.

Let

$$V_{M\times N} = \{ f_{(i)}(x,y) \mid j=1,2,....L; x=0,1,2,..M-1; y=0,1,2,...N-1 \}$$

be the video to be segmented where L is the total no of frames present in the video. Shots are elementary structural segments that are defined as sequences if images taken without interruption by a single camera. Prior to video segmentation, the shot segmentation are necessary for grouping the similar shots. In this shot segmentation similar shots are grouped together for improving the performance of the segmentation. To accomplish this task initially all the frames are partitioned into $m\times n$ patches and every patch are converted to its equivalent frequency coefficients by means of Discrete Cosine Transform (DCT) developed by Ahmed, Natarajan and Rao (1974) (i.e) DCT is applied to every patch in the frames as follows

$$T_{ik} = D_{ik} \times f_{ik} \times D' \qquad (1)$$

Thus all the patches are transformed to transform domain subsequently the Euclidean distance of every patches of consequent frames and their total mean are calculated as follows.

$$D_j = \frac{\sum_{i=1;k=1}^{L;P_i} \sqrt{(T_{i,k} - T_{i+1,k})^2}}{P_{(i)}} \qquad (2)$$

Where $1 < i, j \le L$ and $1 \le k \le P_{(i)}$. The frames belongs to the similar shots are identified based on the mean distance. If $D_j \le E$ f_i and f_{i+1} belong to the same shots otherwise they are belong to the different shots. The fig. 1 illustrates the process of shot segmentation. In fig.1 f_{11} is firs patch in the frame1 like wise f_{21} is the first patch in the second frame. Euclidean distance of every patch in the consequent frames is calculated as sample shown in fig.1

3.2. FCM Based Object Extraction

Partitioning of the video sequence by detecting the scene changes is essential for characterization and categorization of video. Basically, there are two types of scene changes, abrupt change and gradual change. Abrupt change or camera changes are apparently easy to detect because the difference in image properties between two consecutive frames is expected to be large. But the detection of changes in gradual transition is very difficult because the changes occurred over period of frames. To segment the static and dynamic objects in the environment of abrupt and gradual changes, the object detection is a crucial process. The proposed segmentation technique uses fuzzy c means algorithm for detect the objects in every frame. Initially the objects in every frame are identified for segmentation. Let $\delta = \{\delta_a \mid 1 < a \leq \mathbf{A}\}$ be the result of shot segmentation where 'A' is total no of shots and $\delta_a = \{ f_{aj} \mid 1 < a \leq \mathbf{A}; 1 < j \leq |\delta_a| \}$ be the set of similar shots where '$|\delta_a|$' are the total no of frames in ath shot in the segmented results. The initial frames in every shot are taken as key frame for object extraction for example the f_{11} is key frame for shot δ_1 which is known as $f_{key(1)}$. Like wise each shot having its own key frames. Let $F_{key} = \{ f_{key(i)} \; 1 < i < A\}$ be the key frames set of every ith shot. The objects in every frame are identified using fuzzy c means algorithm. The FCM is developed by Dunn in 1973 and improved by Bezdek in 1981. The value of 'C' is calculated using the 3D histogram value. The 3D color histogram is calculated for every frame to find the no of objects in the frame. The number of peak values in the 3D color histogram determines the total number of objects in a frame which is given as the value of c in FCM. Finally the clustering process yields the no of objects in a frame. The overlapping objects are also identified using fuzzy c means clustering.

3.2.1. Fuzzy K-means

The prior research [] of this author used the Fuzzy K-means clustering for object extraction. The main idea behind the fuzzy K-means is the minimization of objective function, which is normally chosen to be the total distance between all patterns from their respective cluster centers. The distribution of objects among clusters and the updating of cluster centers are the two main steps of the K-means algorithm. The algorithm alternates between these two steps until the value of the objective function cannot be reduced anymore. Given the frame

$$\hat{f}_{aj} \in \delta_a \quad . \qquad (4)$$

K-means determine the c cluster Vi=1.C centers by minimizing the objective function

$$Min_{\ j}(\mu, V) = \sum_{i=1}^{C} \sum_{k=1}^{n} \mu_{ik} \left\| f_{aj_{ik}} - V_i \right\|^2 \qquad (5)$$

Where $\mu_{ik} \in [0,1] \forall_i, k$ and $\|.\|$ is the Euclidean distance measure .

3.2.2. Fuzzy C-Means Clustering

The fuzzy C-means clustering do the fuzzy partition rather than the hard partition, by using the objective function given in eq (7). This algorithm is proposed as an improvement to the fuzzy k-means clustering. The FCM partition the collection of 'n' vector into C groups and finds a cluster center in each group such that a cost function of dissimilarity measure is minimized. The FCM algorithm assigns pixels to each category using fuzzy membership functions. Let $\hat{f}_{aj}$

be the frame which is to be clustered into 'C' clusters. The algorithm is an iterative optimization that minimizes the cost function defined as follows.

$$M' = \sum_{k=1}^{m \times n} \sum_{i=1}^{C} \mu_{ik}^{m} \left\| f_{aj_{ik}} - V_i \right\| \quad 1 \leq m < \infty \qquad (6)$$

Where 'm 'is the any real number greater than '1', μ_{ik} is the degree of membership of $f_{aj_{ik}}$ in cluster 'I, V_i is the i[th] cluster center , $\|.\|$ is a norm metric. The parameter 'm' controls the fuzziness of the resulting partition. The cost function is minimized when pixels close to the centroid of their clusters are assigned high membership values and low membership values are assigned to pixels far from the centroid. The membership function represents the probability that a pixel belong to a specific cluster. In the FCM algorithm, the probability is dependent on the distance between the pixel and each individual cluster center in the feature domain.

3.3. Frame Sequence Assortment

After performing shot segmentation, the track frames of the every shot are identified using the objects of their key frame. The objects that appear simultaneously in at least two consecutive frames can be compared directly in terms of their motion so the assortment of the track frames is a required preprocessing step for segmentation this track frame selection process reduces the computational time of segmentation. The objects of the key frame are compared with the other frames of the shot for their presence in the frame. If the object is present in any of the frame then its frame index is stored in $T_j^{\,i}$. For example if the object of key frame $f_{key(1)}$ is presented in the k[th] index of shot then the index is stored in $T_1^{\,1}$. Like wise all the track frames of the every shot are identified.

3.4. Dynamic and Static object Segmentation based on hybrid methods

The proposed method uses frame difference algorithm as well as the consecutive frame intersection method to detect the static and dynamic objects in the current frame. In the background subtraction method the key frame of every shot is consider as background. At each $\hat{f}_{aj}$ frame, the $\hat{f}_{aj}(p,q)$ pixel's value can be classified as foreground pixel if the following inequality

$$\hat{f}_{aj}(p,q) - \hat{f}_{a1}(p,q) > \lambda \qquad (7)$$

Holds; otherwise $\hat{f}_{aj}(p,q)$ will be classified as background pixel value. Where $\hat{f}_{aj}(p,q)$ is the current frame pixel value, $\hat{f}_{a1}(p,q)$ is the key frame value and 'λ' is the threshold pixel value in foreground. The frame difference algorithm provides the only the difference between the current frame and the key frame while the frame intersection method provides the difference between consecutive frames. The motion analysis and segmentation of dynamic objects is performed by intersection process of track frames. Initially the frames in every shot are converted to binary form. After performing the binarization the consecutive frames are intersected to segment the dynamic and static objects. Let $\hat{\hat{f}}_{11}$ and $\hat{\hat{f}}_{12}$ be the binarized form of frame1 and frame2 in shot1 respectively. The dynamic motion objects are found as follows.

$$G2_{aj} = \hat{\hat{f}}_{11} - f_{aj}^{'} \qquad (8)$$

Where as the static foreground are segmented as follows

$$f_{aj}^{\,i} = \hat{\hat{f}}_{11} \cap \hat{\hat{f}}_{12} \qquad (9)$$

Like wise all the consecutive frames are intersected to achieve the static and dynamic object segmentation. Let $G1 = \{S_i \mid 1 < i \le n\}$ and $G2 = \{S_i^{'} \mid 1 < i \le n\}$ be the segmented results of dynamic objects using frame difference algorithm and frame intersection method respectively. The prior said technique yields the segmented motion objects by subtracting the background and the later segmentation technique yields the motion object by intersection method. The ' G 'consists of final segmented motion objects.

$$G = G1 \cap G2 \qquad (10)$$

3.5. Morphological Processing

The segmented motion objects are subjected to dual morphological operations for getting the improved segmentation results. Morphological processing is constructed with operations on sets of pixels. The four most basic operations in mathematical operations are dilation, erosion, opening and closing.

4. Experimental Results

Our proposed video segmentation approach has been validated by experimenting with variety of video sequences. The proposed system has been implemented in Matlab (Matlab7.10). The performance of the proposed approach is compared with the results of the prior research of the author.

(a)

(b)

Figure 1: (a) Sample frames, (b) Grayscale images,

The fig. 1(a) represents the sample frames(Frame :2,Frame :3, Frame :4, Frame :5, Frame :44) in a tested video and the fig. 1(b) represents their corresponding gray scale converted image .Considering the initial frame as key frame all the objects in the every frames are extracted. The proposed hybrid segmentation technique yields the dynamic object segmentation results by intersection of segmented results of both the frame difference algorithm and intersection methods and hence produces the better enhanced segmented results. Also the proposed system segments the static objects in every frame.

4.1. Performance Evaluation

The performance of the proposed system is evaluated by the statistical measures like sensitivity and specificity. The output of the proposed system may be positive (Segmenting the objects) or negative (non-segmenting the objects). The output of the proposed system may or may not be match with the original status of the image. Consider the following setting for the statistical measures.

True Positive (TP) : Valid objects correctly segmented.

False Positive (FP) : Invalid objects incorrectly segmented.

True Negative (TN) : Invalid objects correctly non-segmented.

False Negative (FN) : Valid objects incorrectly non-segmented

4.2 Comparative Analysis

The performance of the proposed hybrid segmentation technique is also evaluated by comparing its segmented results with that of the traditional video segmentation technique which uses background substraction method.

Figure 2: Segmented static and dynamic objects(prior research)

Figure 3: Segmented static and dynamic objects (proposed method)

The fig. 2 represents the video segmentation results of the frames (Frame: 3, Frame: 4, Frame: 5 and Frame: 7) using the prior research (using FKM clustering) and fig.3 represents the segmentation results using the proposed research (using FCM clustering) of the author. The table-1 and table-2 represents the comparison statistical measures of the segmentation of video-I using the proposed technique as well as the prior method.

Table1: Statistical Measures of the segmented results of Video-I using the prior research

Measures	TP	TN	FP	FN	Accuracy	Positive Predictive value	Negative Predictive value(NPV)	False discovery rate	Mathews Correlation Coefficient	Kappa Coefficient
Frame1	5270	92966	3140	1756	95.25	62.66	98.15	37.34	0.66	0.94149
Frame2	5814	93355	2207	2841	95.16	72.48	97.05	27.52	0.67	0.957513

Table2: Statistical Measures of the proposed System for Video-I using the proposed segmentation approach

Measures	TP	TN	FP	FN	Accuracy	Positive Predictive value	Negative Predictive value(NPV)	False discovery rate	Mathews Correlation Coefficient	Kappa Coefficient
Frame4	5370	94858	1125	1158	97.77	82.51	98.79	17.49	0.81	0.946997
Frame5	7569	97385	1266	1266	97.64	85.67	98.72	14.33	0.84	0.975712

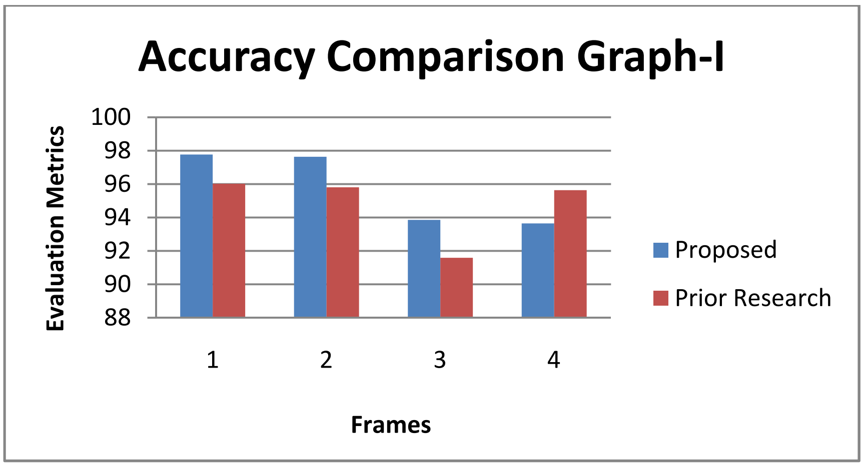

Figure 4: Accuracy Comparison graph-I

The fig. 4 represents the efficiency of the proposed approach than the

The fig. 5 and fig. 6 represents the video segmentation results of the frames (Frame: 23, Frame: 24, Frame: 25 and Frame: 65) using the prior video segmentation research of the author and by using the proposed segmentation technique respectively. The table-3 and table-4 represents the comparison statistical measures of the segmentation of video-II using the proposed technique as well as the prior method and the fig.7 illustrates the corresponding accuracy comparison graph.

Figure 5: Segmented static and dynamic objects(prior research)

Figure 6: Segmented static and dynamic objects (proposed method)

5. Conclusion

In this paper, we have proposed a hybrid video segmentation technique with the FCM based object detection technique to segment both the static and dynamic objects. This has intended to overcome the prior video segmentation techniques of the author. The FCM based object detection and the dual morphological operation process had efficiently increased the segmentation. The proposed approach shows 92% accuracy and this is evaluated by different video sequences

References

[1]Chih-Wen Su, Hong-Yuan Mark Liao, Hsiao-Rong Tyan, "Motion Flow-Based Video Retrieval", IEEE Transactions On Multimedia, Vol. 9, No. 6, October 2007.

[2]S.Padmakala and G.S.Anandha Mala, "A Novel Video Object Segmentation Approach for Noisy Video Sequences towards Effective Video Retrieval", International Journal of Computer Theory and Engineering, Vol.2, No.6, pp.1793-8201, December, 2010.

[3]Kuo-Liang Chung, Yah-Syun Lai and Pei-Ling Huang, "An Efficient Predictive Watershed Video Segmentation Algorithm Using Motion Vectors", Journal of Information Science and Engineering, Vol.26, pp. 699-711, 2010.

[4]Alessandro Parolin, Guilherme P. Fickel, Claudio R. Jung, Tom Malzbender, Ramin Samadani, "Bilayer Video Segmentation For Videoconferencing Applications", Proceeding of the IEEE International Conference on Multimedia and Expo , pp.1-6, 2011.

[5]ChunHui Cui, Qian Zhang, and KingNgi Ngan, "Multi-view Video Based Object Segmentation - A Tutorial", ECTI Transactions on Electrical Eng., Electronics and Communications, Vol.7, No.2 August 2009.

[6]Ramya G, Vidhyalakshmi J, A.Umamakeswari, "Character Identification In Videos Using Iris Recognition," Journal of Theoretical and Applied Information Technology, Vol.28, No.2, 2011.

[7]Vasileios Chasanis, Argyris Kalogeratos, Aristidis Likas, "Movie Segmentation into Scenes and Chapters Using Locally Weighted Bag of Visual Words", ACM International Conference on Image and Video Retrieval, 2009.

[8]Dmitry Kinoshenko, Sergey Mashtalir, Andreas Stephan, Vladimir Vinarski, "Neural Network Segmentation Of Video Via Time Series Analysis", International Journal of Information Theories and Applications", Vol. 18, No. 3, 2011.

[9]Ram Kumar Yadav, Sanjeev Sharma, Jitendra Singh Verma, "Deformation and Improvement of Video Segmentation Based on morphology Using SSD Technique", International Journal of Computer Technology and Applications, Vol. 2, No. 5, pp.1322-1327, 2009.

[10]Gerald Friedland and Raul Rojas. "Anthropocentric Video Segmentation for Lecture Webcasts", EURASIP Journal on Image and Video Processing, Vol. 2008, 2008.

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012
ISSN (Online): 1694-0814
www.IJCSI.org

BIOGRAPHY

K.MAHESH

Mr. K. Mahesh is working as an Associate professor in the Department of Computer Science and Engineering, Alagappa University, Karaikudi, Tamilnadu, India. He has received his M.Phil in Computer Science from Bharathidasan University, Tiruchirapalli, Tamilnadu in the year 2006. He has 21 years of teaching experience at PG level in the field of Computer Science. He has published many papers in international journals and presented his work in national and international conferences. His areas of research interests include Video Segmentation, Video Processing, Clustering techniques, Image processing.

Dr.K. KUPPUSAMY

Prof. Dr K.KUPPUSAMY is working as an Associate Professor in the Department of Computer Science and Engineering, Alagappa University, Karaikukdi, Tamilnadu, India. He has received his Ph.D in Computer Science and Engineering from Alagappa University, Karaikudi, Tamilnadu in the year 2007. He has 22 years of teaching experience at PG level in the field of Computer Science. He has published many papers in International Journals and presented in the National and International conferences. His areas of research interests include Information/Network Security, Algorithms, Neural Networks, Fault Tolerant Computing, Software Engineering, Software Testing and Optimization Techniques.