A New Hybrid Algorithm for Video Segmentation

K. Mahesh, K. Kuppusamy

Associate Professor, Department of Computer Sci & Engg, Alagappa University, Karaikudi Tamilnadu, Indi
Email: mahesh.alagappa@gmail.com, kkdiksamy@yahoo.com

Abstract - Video segmentation became popular and most important in the digital storage media. Effective segmentation is a challenging problem in digital storage media. In this video segmentation technique, initially similar shots are segmented, and subsequently the track frames in every shots are assorted using the extracted objects of every frame which highly reduces the processing time. In this hybrid video segmentation technique, it yields the effective video segmentation results by performing intersection on the segmented results provided by both the frame difference method as well as consecutive frame intersection method. The frame difference method considers the key frame as background and it segments the dynamic objects whereas the frame difference method segments the static and dynamic objects by intersection of objects in consecutive frames. The new hybrid 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, k-means clustering, frame difference algorithm, Euclidean distance

I. INTRODUCTION

Due to the rapid increase in the number of digital video documents produced, the real challenge for computer vision is the development of robust tools for their utilization. It entails tasks such as video indexing, browsing or searching. Video Segmentation is one of the most important processes performed to achieve these tasks [14]. The primary aspect of video indexing is it has the potential to segment the video into consequential or meaningful segments [1]. Video applications such as surveillance, communications and entertainment are very crucial for industries. Among these applications, segmentation of moving object is one of the basic operations for several automated video analysis tasks [15]. One of the main challenges in computer vision is automatic comprehension of complex dynamic content of videos, such as detection, localization, and segmentation of objects and people, and understanding their interactions [9]. Image and video segmentation is very beneficial in several applications for finding the regions of interest in a panorama or annotating the data [11]. MPEG-4 is a promising standard for multimedia communications. MPEG-4 provides standardized ways to encode the video and audio objects, and the scene description, which represents how the objects are structured in a scene [4].

The video segmentation is an imperative technique used for the improvement of video quality on the basis of segmentation [8]. The function of video segmentation is to segment the moving objects in video sequences [19]. Video segmentation is an imperative component of numerous video analysis and coding problems, such as 1) video summarization, indexing, and retrieval, 2) advanced video coding, 3) video authoring and editing, 4) enhanced motion (optical flow) estimation, 5) 3D motion and structure assessment with several moving objects [16]. 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.

II. RELATED WORK

YasiraBeevi P and S. Natarajan [1] have proposed a video segmentation algorithm for MPEG-4 camera system by means of change detection, background registration methods and real time adaptive threshold techniques

PanagiotisSidiropoulos*et al.*[2] have proposed a technique, where the low-level and high-level features extracted from the visual and the aural channel have been used jointly.

KuoLiangchungi*et al.*[3] have developed a promising predictive watershed-based video segmentation algorithm using motion vectors.

EnginMendi, *et al.*[4] have presented a medical video segmentation and retrieval research initiative.

OnurKucüktunc [5] has proposed a fuzzy color histogram-based shot-boundary detection algorithm devoted for content based copy detection (CBCD) applications.

III. DYNAMIC AND STATIC FOREGROUND SEGMENTATION USING HYBRID TECHNIQUE

The proposed technique segments both the dynamic and static foreground objects without considering the global motion constraints. 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.

A. Shot Segmentation

The term video commonly refers to several storage formats for storing moving pictures. The video consists of consequent sequence of frames which is a single picture or still shot run in succession to produce what appears to be seamless piece of film or video tape. Let be the video to be segmented where $'L'$ is the total no of frames present in the video. A shot is defined as a sequence of frames which are captured from a single camera operation. Prior to video segmentation, the shot segmentation is 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

$$T_{ik} = D_{ik} \times f_{ik} \times D'$$

Where

$$\eta_a = \begin{cases} \left(\frac{1}{2}\right)^{1/2} & if \quad a = 0 \\ 1 & if \quad 1 \le c \le m-1 \end{cases}$$

$$\alpha_c = \begin{cases} \left(\frac{1}{2}\right)^{1/2} & if \quad b = 0 \\ 1 & if \quad 1 \le c \le n-1 \end{cases}$$

$1 \le k \le P_{(i)}$ where P_i is the number of patches present in the ith frame. 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.

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

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.

B. Object Extraction

Initially the objects in every frame are identified for segmentation. Let $\delta = \{\delta_a \mid 1 < a \le \mathbf{A}\}$ be the result of shot segmentation where 'A' is total no of shots and

$\delta_a = \{f_{aj} \mid 1 < a \le \mathbf{A}; 1 < j \le |\delta_a|\}$ be the set of similar shots where '$|\delta_a|$' are the total no of frames in athshot 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 $f_{key(1)}$. Likewise each shot having its own key frames. Initially all the frames which are in RGB color format are converted to grey scale format. A RGB color is another format for color images and it represents an image with three matrices of sizes matching the image format while each matrix corresponds to one of the colors red, green and blue. [26] When we convert it into a grey scale (or "intensity") image, it depends on the sensitivity response curve of detector to light as a function of wavelength [27]. This conversion process yields grey scale image.

The objects are identified by the clustering process which is carried out using fuzzy k-means clustering. The value of 'k' can be identified easily for static single image but since the video images are dynamic in nature, the 3d color histogram is calculated for every frame to find the no of objects in a frame. Before calculating 3D histogram value the frames in 'RGB' format is converted into LAB color space. The number of peaks values in the 3D color histogram determines the total no of objects in a frame which is given as the value of 'k' in fuzzy k-means clustering. Finally the clustering process yields the no of objects in a frame. The overlapping objects are also identified using fuzzy k-means clustering.

C. Track frame 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 kth index of shot then the index is stored in T_1^1. Likewise all the track frames of the every shot are identified.

D. Dynamic foreground Segmentation based on frame difference algorithm

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

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

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.

E. Static and Dynamic foreground Segmentation using intersection of 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.

F. Binarization:

Binarization is a technique by which the gray scale images are converted to binary images. Binarization separates the foreground (text) and background information. The most common method for binarization is to select a proper threshold for the intensity of the image and then convert all the intensity values above the threshold to one intensity value ("white"), and all intensity values below the threshold to the other chosen intensity ("black"). $\hat{\hat{f}}_{aj} = \begin{cases} 1 & ; \; if \; \hat{f}_{aj} > \Gamma \\ 0 & ; \; otherwise \end{cases}$ Where 'Γ' is a

global threshold value for binarization. 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

$$G2_{aj} = \hat{\hat{f}}_{11} - f_{aj}^{'}$$ Where as the static foreground are segmented as follows $f_{aj}^{i} = \hat{\hat{f}}_{11} \cap \hat{\hat{f}}_{12}$

Like wise all the consecutive frames are intersected to achieve the static and dynamic object segmentation.

G. Hybridization of Segmentation methods

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.

$$G = G1 \cap G2$$

IV. EXPERIMENTAL ANALYSIS

Our proposed video segmentation approach has been validated by experimenting with variety of video sequences. 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. It represents the dynamic object segmentation results by using frame difference algorithm which considers the key frame of every shot as background. It illustrates the segmented results of intersection methods which contains both the segmented results of static and dynamic objects. It represents the final video segmented results of the proposed hybrid segmentation technique. In the green color represents the dynamic objects and the other color represents the static objects.

A. 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

Sensitivity value represents the percentage of recognition of actual values. Specificity value represents the percentage of recognition of actual negatives. Accuracy is degree of closeness of measurements of a quantity to its actual (true) value. The performance is also analysis by the kappa coefficient which is as below. It represents the statistical measures of the proposed system for the different frames in a video sequence-I

B. 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.

V.   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).
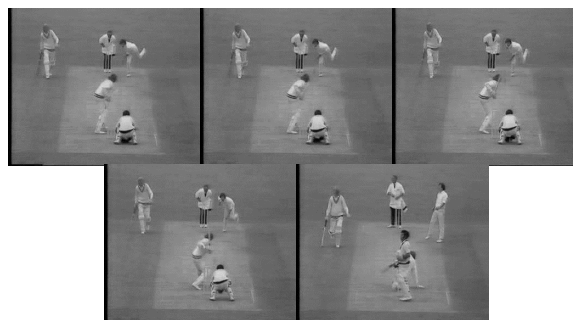

(a)

(b)

(c)

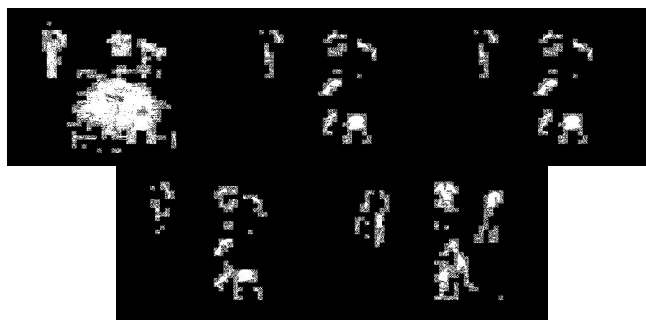Figure 3: Sample similar shot segmented results

Initially the similar shots are grouped together by patch processing method. The fig. 3(a), (b), (c) represents the sample results of shot segmentation process.

(a)

(b)

(c)

Figure 4: (a) Sample frames, (b) Grayscale images, (c) Extracted Objects
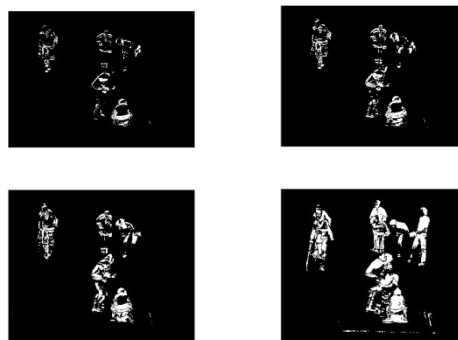

Figure 5: Segmented dynamic objects using frame difference algorithm

The fig. 4(a) represents the sample frames(Frame :2,Frame :3, Frame :4, Frame :5, Frame :44) in a tested video and the fig. 4(b),4(c) represents their corresponding gray scale converted image .Considering the initial frame as key frame all the objects in the every frames are extracted. The fig. 4(c) represents the extracted objects of the frames in fig. 4(a).

Figure 6: Segmented static and dynamic objects using intersection method

Figure 7: Segmented static and dynamic objects using proposed hybrid method

VI.   CONCLUSION

In this paper, we have proposed a new hybrid video segmentation technique to segment both the static and dynamic objects. This is intended to overcome the existing frame difference based segmentation techniques. The segmentation technique based on frame difference algorithm segmented the objects by considering the key frame as background which only produced the motion difference from key frame with remaining frames but the proposed technique also considered the consecutive frame differences by using the consecutive frame intersection method and hence provided better results. The comparison shows the effective accuracy of the proposed technique and also the proposed technique achieves 88% accuracy. The accuracy is evaluated by different video sequences.

REFERENCES

[1]   "A Hidden Markov Model Framework for Video Segmentation Using Audio and Image Features" citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20...rep.

[2]   Panagiotis Sidiropoulos1, VasileiosMezaris, IoannisKompatsiaris, Hugo Meinedo, Miguel Bugalho, Isabel Trancoso,"Video scene segmentation system using audio visual features", Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 2010.

[3]   AlirezaFathi, Maria FlorinaBalcan ,XiaofengRen,James M. Rehg,"Combining Self Training and Active Learning for Video Segmentation", British Machine Vision Conference,2011.

[4]   ÇigdemErogluErdem and BülentSankur, "Performance evaluation metrics for object-based video segmentation" , Proceeding of IEEE International Conference on Image Processing, 2001.

[5]   Sohaib Khan and Mubarak Shah," Object Based Segmentation of Video Using Color, Motion and Spatial Information", Proceedings of the Conference on IEEE Computer Society, vol.2, no.1, 2003.

[6]   KashinathMurmuVineet Kumar," Wavelet Based Video Segmentation and Indexing",EE678 Wavelets Application Assignment, April 2005.

[7]   Michal Haindl, PavelZid and RadekHolub,"Range video segmentation", Proceedings of the IEEE 10th International Conference on Information Science, Signal Processing and their Applications, 2010.

[8]   Ram Kumar Yadav, Sanjeev Sharma and Jitendra Singh Verma,"Deformation and Improvement of Video Segmentation Based

[9]   on morphology Using SSD Technique" IJCTA, Vol. 2, no. 5, pp.1322-1327, 2011.

[10] Jose Lezama, KarteekAlahari, Josef Sivic and Ivan Laptev, "Track to the Future: Spatio-temporal Video Segmentation with Long-range Motion Cues", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[11] "Exploiting Nonlocal Spatiotemporal Structure for Video Segmentation",https://wiki.engr.illinois.edu/.../report_video_segmentation.pdf?.

[12] Yining Deng and B. S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Video", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.8, 2002.

[13] YasiraBeevi C and S. Natarajan, "An efficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, no.4, December 2009.

[14] Si Wu, Yu-Fei Ma, Hong-Jiang Zhang, "Video quality classification based home video segmentation" IEEE, 2005.

[15] N. Peyrard and P. Bouthemy, "Content-Based Video Segmentation using Statistical Motion Models", In Proceeding of British Machine Vision conference, 2002