Motion Variation Of Objects On Temporal Segmentation

S.K Mandal
Asstisant Professor
E-mail: sudhirmta@gmail.com
Department of Electronics & Telecommunication.
Trident Academy of Technology.
BBSR,Odisha,India

S.K Mohapatra
Asstisant Professor.
E-mail: sumsusmeera@gmail.com
Department of Electronics & Telecommunication.
Trident Academy of Technology.
BBSR,Odisha,India

Abstract- **We can take different sets of images with different environments and with different speeds. We know that temporal segmentation fails in slow movements or where there is very small movement like news readers. We will also test by taking some fast and moderately moving video object where there are sufficient movements between consecutive frames. Here the reference frame is available so we can detect the objects easily even if there is slight movement. But if the reference frames are not available it will very difficult to detect the moving parts, but by adding history we can detect the moving parts easily. History implementation is discussed in this chapter Here we have taken some video sequences having different movements and its effect on temporal segmentation is shown .**

Keywords-Video,Image,Segmentation,VOP,CDM Foreground,BackgroundThresholding,

I. Introduction

Reliable detection of moving objects is an important requirement for many computer vision systems. In video surveillance applications, motion detection can be used to determine the presence of people, cars or other unexpected objects and then start up more complex activity recognition steps. Additionally, the segmentation of moving objects in the observed scenes is an important problem to solve for traffic flow measurements, or behavior detection during sport activities. In the literature, the problem of moving object segmentations discussed, identifying three different kinds of approaches: optical flow, temporal differencing, and background subtraction. In particular, back grounding methods, using an opportune thresholding [1] procedure on the difference between each image of the sequence and a model image of the background, are recognized by the scientific community as those that provide the best compromise between performance and reliability. In addition, they produce the most complete feature data and allow the recovery of the most reliable shapes of the segmented moving objects. Any motion detection system based on background subtraction needs to handle a number of critical situations.

II. Segmentation

Segmentation is an important process in automated image analysis [1]. It is during segmentation that regions of interest are extracted from an image for subsequent processing such as surface description and object recognition. It is the low level operation concerned with partitioning images by determining disjoint and homogeneous regions, or, equivalently, by finding edges or boundaries. The homogeneous regions or the edges are supposed to correspond to actual objects or parts of them within the images. Thus, in a large number of applications in image processing and computer vision, segmentation plays a fundamental role as the first step before applying to images for higher level operations such as recognition, semantic interpretation and representation. Segmentation can be defined as follows.

Let I denote an image and H define a certain homogeneity predicate, then the segmentation of I is a partition P of I into a set of N regions R_n, n = 1, 2, ...N ,such that:

1. $U_{n=1}^{N} R_n = I$ with $R_n \cap R_m \neq 0, m \neq n$

2. $H(R_n) = true \quad \forall_n$

3. $H(R_n \cup R_m) = false \forall R_n and R_m adjacent$

Condition 1) states that partition has to cover the whole image; condition 2) states that each region has to be homogeneous with respect to predicate H; condition 3) states that no two adjacent region cannot be merged into a single region that satisfies the predicate H. Regions of image segmentation should be uniform and homogeneous with respect to some characteristics such as gray tone, texture or color. Region interiors should be simple and without many small holes.

Adjacent regions of segmentation should have significantly different values with respect to the characteristic on which they are uniform. Boundaries of each segment should be simple, not ragged and must be spatially accurate.

III. Video Segmentation

Video segmentation refers to the identification of regions in a frame of video that are homogeneous in some sense. Different features and homogeneity criteria generally lead to different segmentation of same data; for example, color segmentation, texture segmentation, and motion segmentation usually result in segmentation maps. Furthermore, there are no guarantees that any of the resulting segmentation will semantically meaningful, since semantically meaningful region may have multiple colors, multiple textures, or multiple motions. Generally motion segmentation is closely related to two other problems, motion (change) detection and motion estimation [3]. Change detection is a special case of motion segmentation with only two regions, namely changed and unchanged regions (in the case of static cameras) or global and local motion regions(in the case of moving cameras). An important distinction between the change detection and motion segmentation is that the former can achieved without motion estimation if the scene is recorded with a static camera. Change detection in the case of a moving camera and general motion segmentation, in contrast, require some sort of global or local motion estimation, either explicitly or implicitly. It should not come as a surprise that motion/object segmentation is an integral part of many video analysis problems, including (i) improved motion (optical flow) estimation, (ii) three-dimensional (3-D) motion and structure estimation in the presence of multiple moving objects, and (iii) description of the temporal variation or content of video.

IV. VOP Generation.

The Video Object Plane (VOP) is obtained by the combination of temporal segmentation result and the original video image frame [3]. In a given scene we consider objects as one class and background as the other thus having a two class problem of foreground and background [4]. Therefore, the temporal segmentation results yield two classes. We denote *FMt* and *BMt* as the foreground and background part of the *CDMt* respectively. The region forming foreground part in the temporal segmentation identified as object and is obtained by the intersection of temporal segmentation and original frame

as

$$VOP = num\,(FM_t \cap y_t)$$

Where the num (.) is the function counting the number of pixel forming the region of interest.

V. Segmentation of slow, fast and moderately moving objects.

We can take different sets of images with different environments and with different speeds. We know that temporal segmentation fails in slow movements or where there is very small movent like news readers. We will also test by taking some fast and moderately moving video object where there is sufficient movements between consecutive frames. Here the reference frame is available so we can detect the objects easily even if there is slight movent. But if the reference frames are not available it will very difficult to detect the moving parts, which are discussed in the next chapter. Here we have taken three video sequences, Andrew video sequence for slow moving, dance sequence for fast moving and hall monitor sequence ,Santu video sequence and Stop Enter Sequence for moderately moving objects.

VI .Temporal Segmentation Without Reference Frame with History.

We know temporal segmentation fails when the reference frame is not available. So it is very difficult to detect the objects. By simple subtracting the background method [3] is not sufficient, here we will get only small portions and the object can't be identified. So to get good results of the moving we use historical information.

VII. The Concept Of history.

Temporal segmentation produces a change detection mask (CDM) that dictates the foreground and background. CDM is firstly obtained by adaptively thresholding the intensity difference between two consecutive frames. And then the results are verified and compensated by historical information, to enhance the coherent segmentation results of moving objects. The historical information of a pixel means whether or not the pixel belongs to the moving object parts in the previous frame [2]. That is represented as a matrix of a frame size $H = \{h_s | 0 \leq S \leq (M_1 - 1)(M_2 - 1)$. If a pixel with hs = 1, it belonged to part of a moving object in the previous frame; otherwise it belonged to the background in the previous frame. Based on this information, CDM is modified: if it belongs to a moving object part in the previous frame and its label is the same as one of the corresponding pixel in the previous frame, the pixel is marked as the foreground area in the current frame.

(a) Original Hall Monitoring Video Sequence Frame No.2

(b) Original Hall Monitoring Video Sequence Frame No.40,41,42

(c) CDM of Frame No.40,41,42 using Frame No. 2 as reference

(d) VOP of Frame No.40,41,42

Figure 1: VOP Generation of moderately moving Hall Monitoring Sequence using Temporal Segmentation.

(a) Original Hall Monitoring Video Sequence Frame No.2

(b) Original Hall Monitoring Video Sequence Frame No.247,248,249

(c) CDM of Frame No.247,248,249 using Frame No. 2 as reference

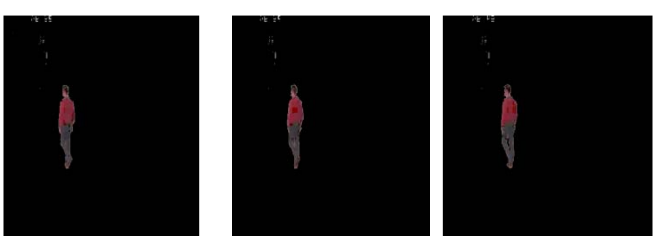
(d) VOP of Frame No.247,248,249

Figure 2: VOP Generation of of moderately moving Hall Monitoring Sequence using Temporal Segmentation

(a) Original stop enter Video Sequence Frame No.515

(b) Original stop enter Video Sequence Frame No.281,282,283

(c) CDM of Frame No.281,282,283 using Frame No. 515 as reference.



(d) VOP of Frame No.281,282,283

Figure 3: VOP Generation of moderately Stop Enter Sequence using Temporal Segmentation.

(a) Original Hall Monitoring video sequence Frame No. 40,41,42

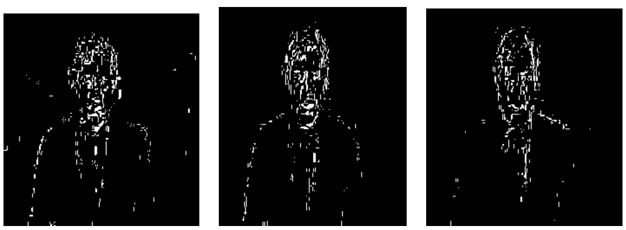
(b) CDM of frame No.40 and 41 (c) CDM of frame No.42 and 41 (d) CDM of frame No.43 and 42

(e) VOP of Frame. No.41 (f) VOP of Frame. No.4 (g) VOP of Frame. No.43

Figure 4: VOP Generation of moderately moving Original Hall Monitoring video using Temporal Segmentation without reference frame.

(a)Original andre video sequence Frame No.2464,2465,2466

(b) Original andre video sequence Frame No.2467,2468,2469

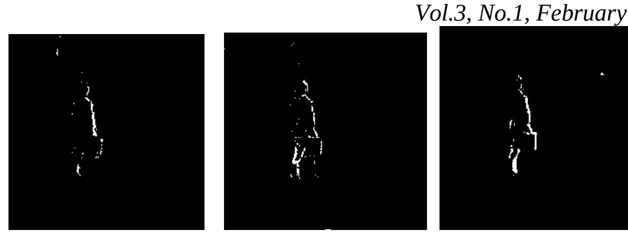


(c) CDM of obtained by subtraction of consecutive frames

(d) VOP of Frame No. 2465 (e) VOP of Frame No. 2466

Figure. 5 CDM & VOP Generation of of slow moving Original Andre video Sequence using Temporal Segmentation without reference frame.

(a) Original Hall Monitoring Video Sequence Frame No.40,41,42

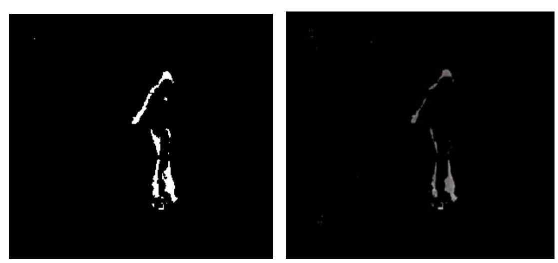


(b) CDM of Frame No. 40&41,41&42.42&43 with history.

(c) VOP of Frame No. 41,42,43 with History.

Figure.6: VOP Generation of Original Hall Monitoring video using Temporal Segmentation without reference frame with history.

(a) Original Dance Video Sequence Frame No. 55025,550



(b) CDM With History. (c) VOP With History .

Figure.7: VOP Generation of Original Fast Dance video Sequence using Temporal Segmentation without reference frame.

VIII. Results And Discussion

In simulation, two types of situations are considered. The first one is when reference frame is available, while the second one is in the absence of reference frames.

Fig.1 shows the Hall monitoring video sequence. The original Hall monitoring sequence which is considered as the reference frame is shown in Fig. 1(a). The movement in the hall Frame No.40,41,42 is shown in different video sequences as shown in Fig. 1(b). The change detection masks are shown in Fig. 1(c). It is observed from the CDMs that there are many other objects i.e parts of the background presence in the CDMs. Temporal Segmentation is carried out and the corresponding VOPs of different frames are shown in Fig. 1(d). It can be observed from Fig. 1(d) that the video objects could be detected but there are few other background patches. However, ignoring the minor background patches in the VOPs it can be concluded that with the availability of reference frames, the objects could be detected accurately. The second example considered is also from hall monitoring sequence as shown in Fig.2 In this case, the reference frame is shown in Fig. 2(a). With the activity in the video, frames 247, 248, 149. are shown in Fig. 2(b) where the moving object is thee human activity. The CDMs obtained with the use of reference frames contains lots of background information besides foreground information. VOPs are generated using Temporal Segmentation and it is observed that the moving object could be detected with less error. Hence in this case also with reference frame, temporal segmentation could produce better results. The second example is the Stop Enter video sequence. The reference frame is shown in Fig. 3(a). The movement in the hall of the man in Frame No.281,282,283 is shown in different video sequences as shown in Fig. 3(b). The change detection masks are shown in Fig. 3(c). It is observed from the CDMs that there are many other objects i.e parts of the background presence in the CDMs. Temporal Segmentation is carried out and the corresponding VOPs of different frames are shown in Fig. 3(d). It can be observed from Fig. 3(d) that the video objects could be detected more accurately because the illumination is better and uniform. The second case considered is when no reference frame is available and without history. The first example considered is the Hall Monitoring Video sequence as shown in Fig. 4.1(a). The VOPs generated are shown in Fig. 4.1(e) where it can be observed that some parts of the moving object could be detected but in a dithered way. Hence, it can be concluded that without availability of reference frames temporal segmentation segmentation method fails to detect the objects. This observation is also corroborated with third example considered as shown in Fig.5 of Andre video sequence of frameNo .2464, 2465,2 466, 2467,2468, also the hall monitoring sequences without reference is shown in the fig.6 where reference frame is not available and hence the VOPs are very much distorted as shown in Fig. 5. It is observed that only some effect of the silhouette is present in the sequence. In the fourth example a fast moving dance video is taken where the object moves fastly in consecutive frames. So the VOPs obtained in Fig.7(c) is better than the results of hall monitoring sequence as shown

in Fig.5,where the movement is very less. Thus it can be concluded that temporal segmentation is not suitable for object detection when reference frame is not available.

The limitation of the existing temporal segmentation methods are as follows

1. It does not give good result in presence of noise and illumination variation 2. It can not able to give good result with poor resolution 3. case will be more critical in absence of reference frame 4. It may not give any result if there are slow movements in the sequences. 5. Substantial amount of object movement is required in order to generate reference frame. 6. If Object size is large it may also fails to generate reference frame.In this case we can add prior knowledge to provide better results. As we know when the reference frame is not available we are getting only few portions of the objects, so it is very difficult to detect the objects. Here in the first example the Hall Monitoring video sequence Frame No.40,41,42 is shown in Fig.1(a) .the CDMs obtained by simply subtracting the frames. Frames are shown in the fig 1(b).this CDM is obtained by adding history i.e an initial ground thuth of the current frame is assumed to be known. The historical information is added accordingly explained above. From the results it is conclude that we are getting still less portion. By adding history we are only detecting only the moving portions in the current frame and unmoved part considered as background. The corresponding VOPs are shown in Fig.1(c).In the second example we have taken the same Hall Monitoring video sequence but the frames are far apart i.e movement is large. The Original Hall Monitoring sequence Frame No. 40,67 and 231,245 is shown in Fig.1(a).The CDMs are shown by subtraction in Fig.1(b). Here we already got some portion of previous and moving parts of current frame in chapter 7 in Fig.4(b) When history is added is added the portions of previous frame is removed and we are getting only the moving part Here the result is better as the movement is large. A dance video sequence is shown in in Fig 6. here the movement is large between two consecutive frames. The VOP obtained is better than shown in fig.6 . Another fast video dance sequence is shown and their results shown in the fig.7. here we will get better results as the movements between the frames is large.

IX. Future work

Temporal segmentation fails due to many reasons like 1. It does not give good result in presence of noise and illumination variation 2. It can not able to give good result with poor resolution 3. case will be more critical in absence of reference frame 4. It may not give any result if there is slow movements in the sequences. 5. Substantial amount of object movement is required in order to generate reference frame. 6. If Object size is large it may also fails to generate reference frame. So to overcome this other methods like spatial segmentation and fusion of temporal and spatial segmentation leads spatio-temporal segmentation can be more effective. Different stochastic models like Markov Random Field Model (MRF), Compound MRF, Double MRF etc are used to model the random behavior of the pixels of the image sequence.

REFERENCES

[1] R .F. Gonzalez, and R. E. Wood,*Digital Image Processing*, Singapore, Pearson Education,2001.

[2] E. Y. Kim and S. H. Park, "Automatic video Segmentation using genetic algorithms,
"*Pattern Recognition Letters*, vol.27, No. 11, pp.1252-1265, 2006.

[3] M. Kim, J. choi, D. Kim and H. Lee, "A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences based on Spatio-Temporal information,"*IEEE Transaction on circuits and Systems for Video Technology*, Vol. 9, No. 8, pp. 1216-1226, Dec 1999.

[4] P.Spagnolo, T.D' Orazio ,M.Leo,A. Distance " Moving Object Segmentation by Background Subtraction And Temporal Analysis" image and vision computing 24 (2006) 411-423.

[5] G. Ananger, T.D.C. Little, A survey of technologies for parsing and indexing digital video,
Journal of Visual Communication and Image Representation 7(1) (1996) 28-43.

[6] Irena Koprinska!, Sergio Carrato Temporal video segmentation: A survey ,Signal Processing: Image Communication, Volume 16, Issue 5, January 2001, Pages 477-500

[7] Zhi Liu, Jie Yang, Ning Song Peng.An improved procedure for the extraction of temporal motion strength signals from video recordings of neonatal seizures *Pattern* Recognition Letters, Volume 26, Issue 5, April 2005, Pages 653-662.

AUTHORS PROFILE

Sudhir Kumar Mandal was born in bhubaneswar odisha in 1982.He received his B.Tech. degree from the Biju pataaik university of technology,bhubaneswar odisha in 2005 & M.Tech in 2009.Currently he is working as assistant professor in Trident Academy og technology,bhubaneswar.His research interests includes image processing & video processing,optical communication & signal processing.

Sumant Kumar Mohapatra was born in Bhubaneswar in 1978.He received his B.Tech (LE) from Biju patanaik university of technology Bhubaneswar ,odisha in 2002 & M.Tech in 2011. Currently he is working as assistant professor in Trident Academy og technology,bhubaneswar.His research interests includes PWM Technology,image & video processing optical communication & signal processing.