Segmentation of Moving Object In Video Using Background Registration and GMEC

Shilpa Shegaonkar[1], Shubhangi Vaikole[2] , Dr.Sudhir Sawarkar[3]

[1](ME Student, Department of C.E., Datta Meghe College of Engineering, Airoli, Navi Mumbai, India)
[2](Research Scholar, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India)
[3](Prncipal, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India)

Abstract: *Emerging multimedia applications demand content-based video processing. Content based video retrieval or concept detection systems require video to be segmented in to objects. A large number of video object segmentation algorithms have been developed such as semiautomatic and automatic. Semiautomatic methods requires the human intervention and also not suitable for some applications. Many applications require automatic segmentation but still there is lot of scope for the improvement. The proposed work aims at contributing to identify the lacunas that are present in the existing segmentation system and also to give the possible solutions to overcome those lacunas so that the precise and effective video segmentation system can be developed. The proposed system deals with the issue of automatic video segmentation system such as uncovered background, Temporary poses and Global motion of background.*
Keywords: *Moving Object, Region based Segmentation, Block Matching, Global Motion Estimation and Compensation*

I. Introduction

VIDEO segmentation, which extracts the shape information of moving object from the video sequence, is a key operation for content-based video coding. In personal databases and Web repositories use of video data is continuously increasing. To retrieve the objects from video, a process is required called as segmentation. In many computer vision and video processing applications object segmentation and background subtraction are important tasks. They are mainly used as the first step in applications like concept detection, automatic surveillance systems [1].

Two methods are mostly used in video object segmentation (VOS), one is semiautomatic, in which some kind of user interference is needed to define the semantic object and the second one is automatic, where object is separated from the background without any help from the user. Most of the applications require segmentation of video objects automatically, especially those with real time requirements [4].

A large number of video object Segmentation methods have been proposed, most aiming to specific applications, and trying to fulfill specific constraints. Good results have been obtained so far in semiautomatic methods, since there is also human interference in the segmentation process. However, the human assistance involved in these methods is not required because it unnecessary adds work of users and also it is not suitable for some applications. On the other hand, fully automatic video segmentation systems are still a summons, although they are required by many applications.

Many automatic segmentation systems are developed for particular problems and with simplified assumptions like videos with fixed background. So it is very important to have flexible automatic segmentation system for different types of videos. Most of the existing automatic segmentation systems involve complicated techniques. Also each stage of the segmentation process involves computationally acute operations to obtain good results. Thus, reducing the complications of the techniques involved is required while keeping good performance of segmentation results. This can be done by selecting efficient algorithms with reduced complicated methods in each step of the segmentation process. Accuracy of segmentation can be improved by applying post-processing.

II. Review Of Literature

A lot of work is done on video segmentation algorithms. This section provides a critical review of the various approaches available for video segmentation.

Primary object regions are extracted for video object segmentation. In this method the author has extracted the object proposal and used DAG approach which gives good segmentation result. To find out which image regions are objects vs. background, it makes these methods very slow [2].

A Causal Graph based video segmentation is proposed by Camille Courier. This method uses the graph based matching method. It is gives good results to large camera displacements but spanning trees method takes more computation time [3]

MacFarlane N.J.B. proposed a method for segmentation and tracking of piglets in images. This paper is based on approximate median method which applies frame difference with constantly updated background model. Storage requirements of median filtering are eliminated by this technique but it requires continuous updating of background model [4].

Mixture of Gaussian model which uses Background model is proposed by Ricardo. In this model is parametric but results of this method varies as modelling of background changes. [5].

Background registration method for segmentation of moving object is proposed by Shao-Yi Chien. Computational complexity of this method is low but slow movements or temporary movements are not identified and works only for fixed camera [6].

Change detection method separates foreground regions by applying a higher order statistics (HOS) significance test to inter-frame differences is proposed by Neri which separates potential. The earliest methods were comparing successive frames by relying pixels. Comparison could be performed on a global level, so methods based on histograms were also proposed [7].

A. Summarized findings of literature review

Change detection based methods proposed till date has applied frame difference information of two successive frames (the current and the previous frame) only. One of the problems that confuse the conventional change detector is that of the temporary poses or slow movements. In these cases, the motion information disappears if we use the difference in the frames only. But, if we use background difference information, we can see very clearly that these pixels belong to the object region and should be included in the object mask.

Most algorithms fail in segmenting the foreground with slow movements and temporary poses. Optical flow methods based on gradient have shown good results but generally come with increased computational overhead. Block-based algorithm gives satisfactory results for slow movements and small object motion from frame to frame. There are various methods of segmentation of video object, but the faster video object segmentation techniques are based on change detection approach.

If videos are captured using a fixed camera the segmentation will be easier and accurate results may be obtained. However, when the videos are captured through moving camera, and when no initial background reference frame is present, the segmentation will be difficult, and the results of segmentation may not be that good. This shows that there is still a lot to be done to obtain better segmentation system.

III. Proposed Work

The present study aims at developing a system to segment objects from the video automatically. In the proposed work we will try to resolve the issue of moving camera and uncovered background. For moving camera we will use motion estimation and compensation. The proposed work focuses on to resolve the issue of temporary poses by integrating the region based segmentation with the system.

Proposed Method
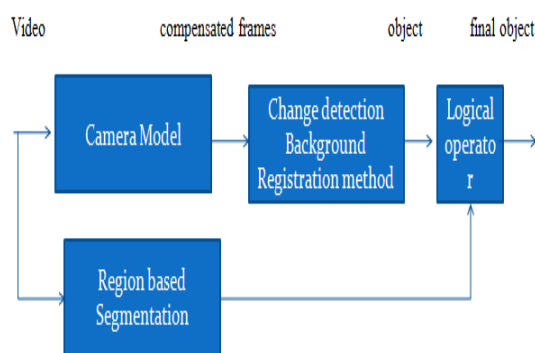
Fig 1: Block diagram of proposed method

Global motion of background

Movement of the camera creates the unnecessary disturbance in the motion of the object. This undesired motion should be removed at the initial step before segmenting the moving object. This is done in three steps as motion vector estimation, motion vectors of background elimination and finally calculation of new frame [18]. Frame is divided in to blocks of n*n (8*8/16*16) to calculate the motion vectors. Then motion vectors are calculated by searching for the best match in the reference or previous frame. Best match is found between current block and blocks in previous frame which are having minimum measure of matching error.

$MAD (m, n) = 1/n*m (\Sigma\ gl\ (I)-gl\ (I-1))$………………….. *(1)*

(u,v)=min (MAD (m, n)) …………………………………. *(2)*
Where MAD represents the mean absolute difference and gl represents the grey level and (u, v) is motion vector.

After calculating motion vectors, motion vectors that totally different from their neighbourhood are rejected. The mean of 3*3 group of motion vector is calculated and compared with reference motion vector. Then frame warping is used to associate the previous or next frame to current frame. New frame is calculated from previous frame by transforming the co-ordinates of previous frame into new position defined as

X'=a1*X + a2* Y+ a3 and Y'= a4*X + a5* Y+ a6

Where a1 to a6 represents camera parameters.

Temporary Poses/Slow Movements

Block Matching algorithm gives good results for slow movements so we have used this method to calculate the motion vectors. To resolve the issue of temporary poses we have combined region based segmentation with our system. Region based segmentation partitions the frame in to regions which are same in some characteristics such as colour and intensity. Result of region based segmentation is added with the object detection result to give final output.

Background Registration

In the background registration the difference between current frame and previous frame is used to generate the frame difference mask and background difference mask. Based on the contents of these two masks the object is detected [6].
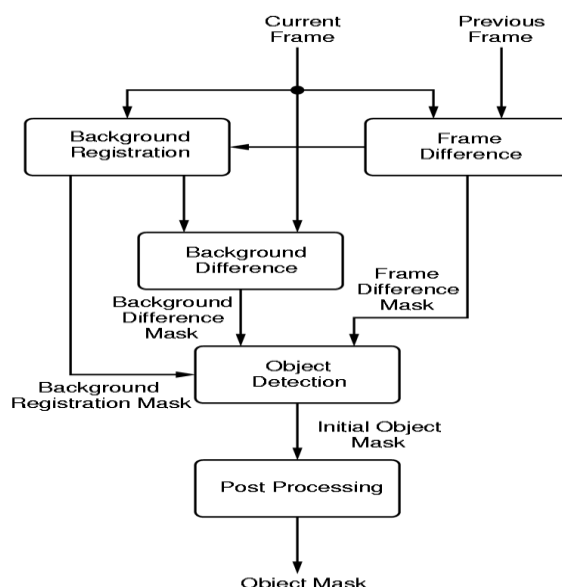

Fig 2: Block diagram of Background registration method

In this system uses three consecutive frames as past, current and future. The past frame is normalized with respect to current frame and future frame is normalized with respect to current frame .This two are combined by a logical AND operator. The operator removes all areas except the foreground object detected which is the region that overlaps in two masks.

A. Frame Difference

In *Frame Difference*, the frame difference between current frame and previous frame, which is stored in *Frame Buffer*, is calculated and thresholded.
It can be presented as

$$FD(x,y,t) = |I(x,y,t) - I(x,y,t-1)|$$

$$FDM(x,y,t) = \begin{cases} 1 & \text{if } FD \geq Th \\ 0 & \text{if } FD < Th \end{cases}$$

where FD is frame data, is frame difference, and FDM is *Frame Difference Mask*. Note that there is a parameter TH needed to be set in advance. The method to decide the optimal
is shown in Section VI. Pixels belonging to are viewed as "moving pixels."

B. Background Registration

Background Registration can extract background information from video sequences. According to , pixels not moving for a long time are considered as reliable background pixels. The procedure of *Background Registration* can be shown as

$$SI(x,y,t) = \begin{cases} SI(x,y,t-1)+1, & \text{if } FDM = 0 \\ 0, & \text{if } FDM = 1 \end{cases}$$

$$BG(x,y,t) = \begin{cases} I(x,y,t), & \text{if } SI(x,y,t) = Fth \\ BG(x,y,t-1), & \text{else} \end{cases}$$

Where SI is Stationary *Index* and BG is the background information.

C. Background Difference

The procedure of *Background Difference* is similar to that of *Frame difference*. What is different is that the previous frame is substituted by background frame BG . After *Background Difference*, another change detection mask named *Background Difference Mask(BDM)* is generated. The operations of *Background Difference* can be shown by

$$BD(x,y,t) = |I(x,y,t) - BG(x,y,t-1)|$$

$$BDM(x,y,t) = \begin{cases} 1, & \text{if } BD \geq Th \\ 0, & \text{if } BD < Th, \end{cases}$$

Where BD is background difference, BFD is background frame and BDM is *Background Difference Mask* respectively.

D. Object Detection

Both of FDM and BDM are input into *Object Detection* to produce *Initial Object Mask (IOM)* . The procedure of *Object Detection* can be presented as the following equation.

$$IOM(x,y,t) = \begin{cases} BDM(x,y,t), & \text{if } BI(x,y,t) = 1 \\ FDM(x,y,t), & \text{else.} \end{cases}$$

E. Post processing

The *Initial Object Mask* generated by *Object Detection* has some noise regions because of irregular object motion and camera noise. Also, the boundary may not be very smooth. Therefore, there are two parts in Postprocessing: noise region elimination and boundary smoothing.
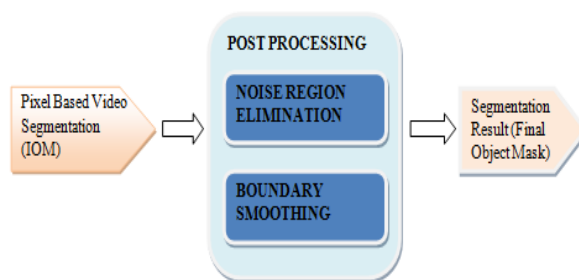
Fig 3: Post processing block diagram

IV. Result And Discussion

Result of Proposed Method

The system is tested on a Segtrack standard dataset. It consists of 14 videos some of which are having interacting objects, slow movements, deformation, motion blur, and occlusion.

Proposed method is evaluated on Segtrack dataset [20]. There are 6 videos in these dataset, and also a pixel-level segmentation ground-truth for each video is available. We follow the setup in the literature ([13, 14]), and use 5 (birdfall,cheetah, girl, monkeydog and parachute) of the videos for evaluation (since the ground-truth for the other one (penguin) is not useable).

Proposed method is also evaluated on Segtrack v2 dataset [20]. There are 14 videos in this dataset, and also a pixel-level segmentation ground-truth for each video is available. It consists of 14 videos with camera in motion some of which are having interacting objects, slow movements, deformation, motion blur, and occlusion. In SegTrack dataset there are 6

Videos in addition to that 8 more videos are added in SegTrack v2 dataset. Also there are changes in the ground truth for videos which contains multiple objects. In SegTrack dataset only one object is detected whereas in SegTrack v2 dataset multiple objects are detected.
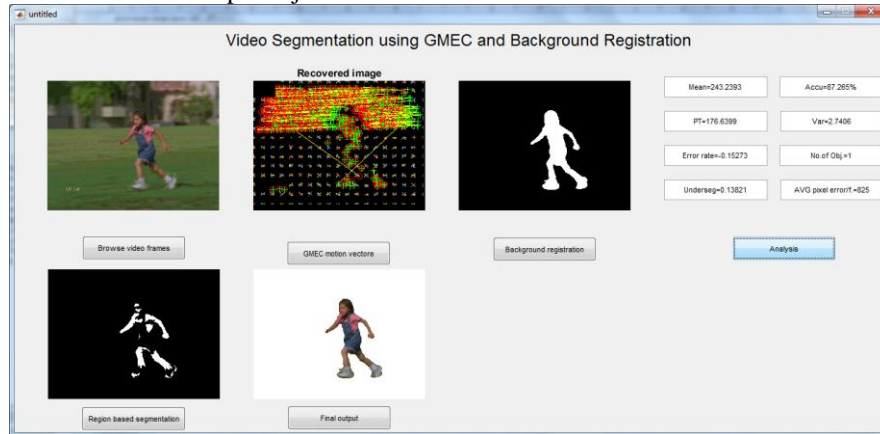

Fig 4: Snapshots of proposed method

TABLE I Evaluation Parameters

Mean	243.2393
Variance	2.0476
Processing time	176.63
Under/over segmentation	0.13821
Error rate	0.15273
Accuracy	87.265
Avg.per frame pixel error rate	825

TABLE II Average per frame pixel error

Video	ours	[2]	[4]	[5]	[8]	[10]
Birdfall	**106**	155	189	288	252	454
Cheetah	**540**	633	806	905	1142	1217
girl	**825**	1488	1698	1785	1304	1755
monkeydog	**386**	365	472	521	563	683
parachute	**179**	220	221	201	235	502

For each video average per frame pixel error is calculated by dividing the XOR result of ground truth and detected object by number of frames of that video .we have compared our result with the other methods and found that pixel error is reduced by our proposed method.

From Table II we can say that average per frame pixel error rate is improved as compared to the existing methods.

Table III shows segmentation accuracy of the proposed methods and the state of the art methods including tracking and graph based approaches [22, 23, 24, 25, 26, and 27]. The proposed method's segmentation accuracy is improved as compared to other methods.

TABLE III segmentation Accuracy

Video	VS-BRM	[22]	[23]	[24]	[25]	[26]	[27]	[28]
Girl	87.26	83.7	**89.2**	87.7	31.9	53.6	52.4	87.9
Birdfall	**88.42**	77.5	62.5	49.0	57.4	56.0	32.5	57.4
Parachute	**94.32**	94.9	93.4	96.3	69.1	85.6	69.9	94.5
Cheetah-Deer	**65.30**	63.1	37.3	44.5	18.8	46.1	33.1	33.8
Cheetah-Cheetah	63.82	35.3	40.9	11.7	24.4	47.4	14.0	**70.4**
Monkeydog-Monkey	**93.94**	82.2	71.3	74.3	68.3	61.0	22.1	54.4
Monkeydog-Dog	**56.26**	21.1	18.9	4.9	18.8	18.9	10.2	53.3
Penguin-#1	71.72	92.7	51.5	12.6	72.0	54.5	20.8	**93.9**
Penguin-#2	43.40	**91.8**	76.5	11.3	80.7	67.0	20.8	87.1
Penguin-#3	69.57	**91.9**	75.2	11.3	75.2	7.6	10.3	89.3
Penguin-#4	44.29	**90.3**	57.8	7.7	80.6	54.3	13.0	88.6
Penguin-#5	51.08	76.3	66.7	4.2	62.7	29.6	18.9	**80.9**
Penguin-#6	72.84	**88.7**	50.2	8.5	75.5	2.1	32.3	85.6
Drifting#1	44.93	67.3	74.8	63.7	55.2	62.6	43.5	**84.3**
Drifting#2	54.74	**63.7**	60.6	30.1	27.2	21.8	11.6	39.0
Hummingbird#1	**73.98**	58.3	54.4	46.3	13.7	11.8	28.8	69.0
Hummingbird#2	**74.01**	50.7	72.3	74.0	25.2	-	45.9	72.9
BMX-Person	**96.74**	88.9	85.4	87.4	39.2	2.0	27.9	88.0
BMX-Bike	**76.72**	5.7	24.9	38.6	32.5	-	6.0	7.0
Frog	**90.67**	61.9	72.3	0.0	67.1	14.5	45.2	81.4
Worm	76.52	76.5	82.8	84.4	34.7	36.8	27.4	**89.6**
Soldier	**93.88**	81.1	83.6	66.6	66.5	70.7	43.0	86.4
Monkey	79.63	86.0	84.8	79.0	61.9	73.1	61.7	**88.6**
Bird of Paradise	92.75	93.0	94.0	92.2	86.8	5.1	44.3	**95.2**

V. Conclusion

The proposed work solved the issue of Moving camera which adds the unwanted disturbance in the video and also solved the problem of temporary poses or slow movements. The performance analysis show that the Accuracy is improved as well as average per frame pixel error is also improved.

References

[1]. Shao-Yi Chien, Yu-Wen Huang, Bing-Yu Hsieh, Shyh-Yih Ma, and Liang-Gee Chen,"Fast Video Segmentation algorithm with Shadow Cancellation, Global Motion compensation, and Adaptive Threshold Techniques," IEEE Trans. on Circuits and System for Video Technology., vol. 6, pp. 732- 748, no. 5, Oct. 2004.

[2]. Dong Zhang1, Omar Javed2, Mubarak Shah1," Video Object Segmentation through Spatially Accurate and Temporally Dense Extraction of Primary Object Regions," 2013 IEEE Conference on Computer Vision and Pattern Recognition

[3]. Camille Couprie,"Causal Graph based video segmentation"(2012)

[4]. McFralane, N. J. B. and Schofield, C.P. "Segmentation and tracking of piglets in images". Machine Vision and Applications, Vol. 8, No. 3, 187-193. 2005.

[5]. Ricardo Augusto Castellanos Jimenez "Event Detection In Surveillance Video" Florida Atlantic University Boca Raton, Florida May 2010

[6]. Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen, "Efficient Moving Object Segmentation Algorithm Using Background Registration Technique," IEEE Trans. on Circuits Syst. Video Technol., vol. 12, no. 7, pp. 577-586, 2002.

[7]. Tung-Chien Chen "Video Segmentation Based on Image Change Detection for Surveillance Systems".

[8]. D. Tsai, M. Flagg, and J. Rehg. Motion coherent tracking with multi-label mrf optimization. In *BMVC*, page 1, 2010.

[9]. T. Ma and L. Latecki. Maximum weight cliques with mutex constraints for video object segmentation. In CVPR, pages 670–677, 2012.

[10]. Y. Lee, J. Kim, and K. Grauman. Key-segments for video object segmentation. In ICCV, pages 1995–2002, 2011.

[11]. D. Tsai, M. Flagg, and J. Rehg. Motion coherent tracking with multi-label mrf optimization. In BMVC, page 1, 2010.

[12]. H. Jiang, A.S. Helal, A.K. Elmagarmid, and A. Joshi. "Scene change detection techniques for video database systems". Multimedia Systems, 6(3):186–195.

[13]. A. Dailianas, R.B. Allen, and P. England. Comparison of automatic video segmentation algorithms". In SPIE Conference on Integration Issues in Large Commercial Media Delivery Systems, volume 2615, pages 2–16, Philadelphia, PA

[14]. M.K. Mandal, F. Idris, and S. Panchanathan" A critical evaluation of image and video indexing techniques in the compressed domain". Image and Vision Computing, 17(7):513–529.

[15]. S. Y. Chien, Y. W. Huang, and L. G. Chen, "Predictive watershed: a fast watershed algorithm for video segmentation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, May 2003, Page(s):453-461
[16]. R. Zabih, J. Miler, K. Mai, "A feature-based algorithm for detecting and classifying production ejects", Multimedia Systems 7 (1999) 119}128.
[17]. K. Zhang and J. Kittler, "Using background memory for efficient video coding," in Proc. IEEE Int. Conf. Image Processing, 1998, pp. 944–947.
[18]. R. M. Haralick and L. G. Shapiro, Computer and Robot Vision. Reading, MA: Addison-Wesley, 1992, pp. 28–48.
[19]. Chen T-H, Liau H-S and Chiou Y-C. (2005) "An Efficient Video Object Segmentation Algorithm Based on Change Detection and Background Updating." Kun Shan University, National Computer Symposium, MIA1-2 (MI14).
[20]. Hanqing Jiang, Guofeng Zhang, Huiyan Wang, Hujan Bao," Spatio-Temporal Video Segmentation of Static Scenes and its Applications." IEEE Transaction on Multimedia ,VOL.17,2015
[21]. Stephen Liwicki,Stefanos P. Zafeiriou,Maja Pantic, "Online Kernel Slow Feature Analysis for Temporal Video Segmentation and Tracking", IEEE Transaction on Image Processing ,VOL.24,2015
[22]. L. Wen, D. Du, Z. Lei, S. Z. Li, and M.-H. Yang. Jots: Joint online tracking and segmentation. In CVPR, 2015. 1, 2, 6, 7,
[23]. F. Li, T. Kim, A. Humayun, D. Tsai, and J. M. Rehg. Video segmentation by tracking many figure-ground segments. In ICCV, 2013. 1, 2, 5, 6
[24]. Y. J. Lee, J. Kim, and K. Grauman. Key-segments for video object segmentation. In ICCV, 2011.
[25]. M. Grundmann, V. Kwatra, M. Han, and I. Essa. Effi- cient hierarchical graph-based video segmentation. In CVPR, 2010.
[26]. M. Godec, P. M. Roth, and H. Bischof. Hough-based tracking of non-rigid objects. In ICCV, 2011.
[27]. S. Wang, H. Lu, F. Yang, and M.-H. Yang. Superpixel tracking. In ICCV, 2011.
[28]. Yi-Hsuan Tsai,"Video segmentation via Object Flow",CVPR2016