

Moving Object Segmentation with camera in motion Using GMEC and Change Detection Method

Shubhangi L Vaikole¹, S.D.Sawarkar²

¹Datta Meghe College of Engineering, Airoli, India

²Datta Meghe College of Engineering
Airoli, India

shubhangiv@rediffmail.com



ABSTRACT: Emerging multimedia applications demands content-based video processing. Video has to be segmented into objects for content-based processing. A number of video object segmentation algorithms have been proposed such as semiautomatic and automatic. Semiautomatic methods adds burden to users and also not suitable for some applications. Automatic segmentation systems are still a challenge, although they are required by many applications. The proposed work aims at contributing to identify the gaps that are present in the current segmentation system and also to give the possible solutions to overcome those gaps so that the accurate and efficient video segmentation system can be developed.

Keywords: Content based Application, Semiautomatic Segmentation, Change Detection, Block Level Matching

Received: 10 February 2015, Revised 23 March 2015, Accepted 28 March 2015

© 2015 DLINE. All Rights Reserved

1. Introduction

In digital video processing technology video segmentation generated by objects is an important application domain. Segmentation of foreground objects from background has a lot of applications in human-computer interaction, video compression, multimedia content editing and manipulation. The extraction of the moving foreground from a stationary background from a general video sequence has various applications such as compression of videos and also in the cinematographic effects. One of its important applications is digital composition, in which the object of interest is extracted from a video clip and pasted to a new background.

Most video effects in movies involve this task. In video object segmentation(VOS) two methods are mostly used, one is semi-automatic, in which some kind of user intervention is required to define the semantic object and other one is automatic, where segmentation is performed without user intervention, but usually with some a priori information[1]. Most of the applications require automatic segmentation of video objects, especially those with real time requirements. A number of Video object Segmentation algorithm s have been proposed, most aiming to specific applications, and trying to fulfill specific requirements. Promising results have been obtained so far in semiautomatic methods, since there is also human assistance in the segmentation process. However, the human assistance involved in these methods is not required because it adds burden to users and also it is not suitable for some applications. On the other hand, fully Automatic Segmentation (AS) systems are still a challenge, although they are required by many applications.

Many automatic segmentation systems are designed for specific problems and with simplified assumptions like videos with fixed background. So it is necessary to have flexible automatic segmentation system for different types of videos. Most of the existing AS systems involves complex techniques. Also each stage of the segmentation process involves computationally intense operations to obtain good segmentation results. Thus, reducing the complexity of the techniques involved is required while keeping accuracy of segmentation results. This can be done by selecting efficient algorithms with reduced computational intensity in each step of the segmentation process. Accuracy of segmentation can be improved by applying post-processing.

2. Review of Literature

A number of video segmentation algorithms have been proposed. Rough classification of Automatic video object segmentation algorithms includes the edge-feature based segmentation, spatial-temporal based segmentation, and change detection based segmentation. This section provides a critical review of the various approaches available for video segmentation.

2.1. Video Segmentation based on Edge-Feature based Method

Canny edge detector is applied on each frame to find edge information and it can obtain correct segmentation object for stable moving-object, but this approach will require an absolute background from video sequence and involve a computation-intensive processing [4]. Various segmentation algorithms have been proposed in previous works for different applications, such as Mixture of Gaussians model and Bayesian background model. In these methods, pixel-wise variation of visual features is represented by using statistical model along the temporal domain. Performance of these algorithms greatly depends on the modeling of the background and its update. Dailianas compared several segmentation methods and introduce a filtering algorithm in order to limit the false detections. Mandal focused on methods working in compressed domain. [10]

2.2. Video Segmentation based on Spatio-Temporal Method

Spatial and temporal features in video are exploited and integrated in Spatio-temporal methods. Accurate segmentation results are obtained by global motion of background segmentation, but with an increased complexity. Moving objects(MO) have distinct motion patterns from the background hence they can easily identify Moving object in the Temporal segmentation. If underlying objects have a different visual appearance (color, luminance, etc) from background, Spatial segmentation can determine object boundaries accurately.

The problem of automatic VOS is formulated as graph labeling approach over a region adjacency graph based on MI, that is, the approach is to classify regions obtained in an initial partition as foreground or background, based on MI. Watershed algorithm takes initial spatial partition of each frame[18]. Hierarchical region matching estimates the motion of each region. A classification stage labels regions as foreground or background. It begins with an initial classification based on statistical significance test which marks regions as foreground candidates. The GMOB process increases the complexity of the segmentation system, since it is a computationally intensive process, and involves a number of steps, this is a problem in GMOB approaches. GMOB usually produces over-segmentation of frames, and region merging is required to reduce the number of regions. In classification motion estimation (ME) is applied for each region to classify into foreground and background. It is a computation intensive process, which is another limitation of these methods. Watershed transform is used to separate a frame into many homogeneous regions. It can obtain a good result of segmentation with high accurate boundaries, but it will suffer over-segmentation due to noise. Though this problem can be solved by smoothing the image previously, it will reduce the performance of the algorithm[5].

2.3. Video Segmentation based on Change Detection Method

Among the motion detection algorithms change detection is widely used method due to its simplicity and efficiency. At two different times the movement of objects between two frames able the detection of change. Hence, it is used for detecting MOs efficiently and also is ideal for automatic object detection and segmentation. Comparison of current frame with the background frame is used for video sequences with static backgrounds in change detection. Change detection compares the current and the previous frame when the background frame is difficult to obtain. In this situation however, one problem arises, which is detection of uncovered background as change. Automatic VOS based on change detection is reported to be more efficient than the GMOB approaches because it is only the motion feature that is used to distinguish the (moving) object from the background. Algorithms that perform GMOB at first will waste much of the computing power in segmenting the background also without knowing the motion information (MI).

In Background subtraction scheme, the basic step is the selection of GMOB (RF) or background to be subtracted. In some

approaches, accumulated frame difference pictures are analyzed to reconstruct stationary scene component to compare with frames to detect change. In these approaches, there is a strong assumption of stationary background. Other approaches align consecutive frames to construct a reference background image before applying change detection. Whenever there is a big deviation of the background, it is updated as necessary. Changes between two frames can also be identified by using two consecutive frames instead of using a reference background frame.

An algorithm based on change detection is proposed by Neri which separates potential foreground regions employing a higher order statistics (HOS) significance test to inter-frame differences. 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[6]. The problem of moving camera is represented, which is often encountered in real life for target tracking surveillance[17]. To solve the problem of uncovered background the object detection algorithm uses three consecutive video frames: backward frame, frame of interest and forward frame. First, iterative camera motion compensation and background estimation is carried out on backward and forward frames based on the optical flow. Differences between camera motion compensated backward and forward frames with the frame of interest are then tested against the estimated background models for intensity change detection. Estimate of the background is generated during the compensation process. Next, these change detection results are combined together for acquiring approximate shape of the MO. The moving area information is then merged with region information in terms of region boundary to obtain the final result. This approach effectively solved the problem of uncovered background. However, the iterative optical flow based method used for camera motion compensation is computationally intensive. Also, estimating and generating background model complicates the segmentation process.

2.4. Summarized Findings of Literature Review

Complex computation of watershed video segmentation algorithms leads to non appliance of video segmentation for real-time. An absolute background from video sequence is required by the edge-feature based segmentation algorithms and involve a computation-intensive processing. Spatial-temporal segmentation approach, suffer over-segmentation due to noise. Though this problem can be solved by smoothing the image previously, it will reduce the performance of the algorithm. Hierarchical Segmentation Algorithm does not consider the boundary information of the object.

The limitation of histogram based methods, appears when we compare two different images having a similar histogram. In clustering based Segmentation, partition of the boundary is often not precise enough. In Model-based Segmentation, we have to maintain two versions of the background model, one in the lower and one in the full resolution.

Good results are obtained by the mean shift algorithm but the iterative color clustering for region segmentation may not converge easily, and increases computation intensity. Also, this approach deals only for MOs with still background, and cannot be applied for video sequences with moving background. Graph labeling approach do not consider motion of background. So if global motion is to be considered to make them more generic, the complexity will escalate. Hence turning to methods with reduced complexity of algorithms involved is necessary.

Only the frame difference information of two successive frames (the current and the previous frame) is used in Change detection based methods. One of the problems of change detector is that the object may stop moving temporarily or move very slowly. In these cases, the motion information disappears if we check the frame difference only. However, if we have background difference information, we can see very clearly that these pixels belong to the object region and should be included in the object mask. Unlike other change-detection-based approaches in pixel based method, judge criterion for motion does not come directly from the frame difference of two consecutive frames. Instead an up-to-date background information is maintained from the video sequence and compare each frame with the background. Any pixel that is significantly different from the background is assumed to be in object region. The uncovered background is another region where the Pixel-based algorithm outperforms the traditional change detection algorithms. Since both the uncovered background region and the moving object region have significant luminance change, distinguishing the uncovered background from the object is not very easy if only the frame difference is available.

The problem of slow movements and temporary poses is ignored by most of the algorithms. Gradient-based methods such as optical flow have shown high performance but generally come with increased computational overhead. Block-based algorithms seem to be tolerant to slow and small object motion from frame to frame. There are various methods of video object segmentation, but the faster video object segmentation techniques are based on change detection approach.

Automatic segmentation is simplified by approaches based on change detection as compared to GMOB approaches, they

also have their own problems. Specifically, the complexity of background generation and updating requires more attention in these methods. If there is a static reference background frame, the problem will be easier and accurate results may be obtained. However, when the background itself is in motion, and when no initial background reference (a frame prior to appearance of any moving foreground object in the video sequence) is present, the problem will be more complex, and the results of segmentation may not be accurate. This shows that there is still a lot to be done to obtain better segmentation system.

3. Proposed System

On the backdrop of the afore-mentioned review of literature and subsequent gaps identified from the findings of the literature review, the proposed work aims at contributing to develop a system to segment video objects automatically from the background given a sequence of video frames. In the proposed system three types of videos are considered with fixed and moving backgrounds and whether there is an initial background or not. Block diagram of the entire proposed system is shown in Figure 1, and the description of overall working is given below:

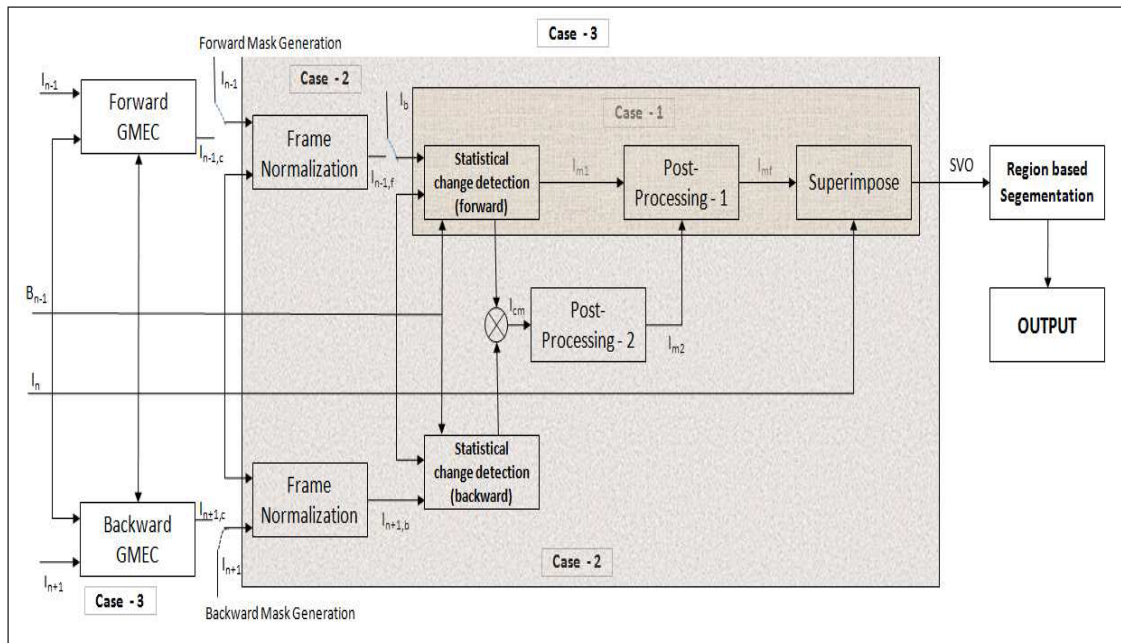


Figure 1. Block Diagram Of Proposed System

3.1 Videos with Initial Background

Videos with an initial background GLOB can be easily obtained. The system uses the background GLOB (I_b), the current frame (I_n), and the previous background frame B_{n-1} to obtain the SVO in the current frame (the dark-gray shaded region Figure 1). Then the Statistical change detection is applied. The goal here is to identify the set of pixels that are significantly different between an image and another reference image in video sequences. A comparison is applied between the image under concern and the reference image to detect those pixels. The changes may result from (1) camera movement, (2) appearance/disappearance of objects, (3) motion of object relative to background, (4) shape changes of objects, or (5) changes in illumination. These changes manifest as temporal intensity variations at each corresponding pixel between the two images. A basic change detection algorithm takes the image sequence as input and generates a binary Map $C(x, y, n)$. This binary map represents the pixels a $t(x, y)$ of an image $I(x, y, n)$, which is the current frame n under test with respect to a another image $I(x, y, r)$ which is the reference frame r , and is defined as:

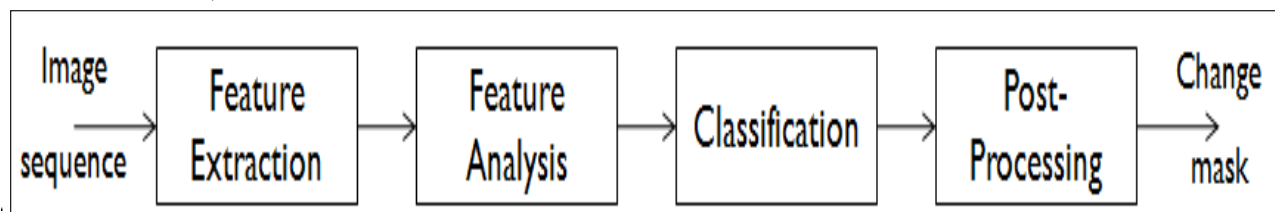


Figure 2. Block diagram of general structure for change detection

In the feature extraction stage, the input frame sequence is transformed into a most appropriate feature space. This feature space may represent luminance, color components, or more complex feature spaces. Luminance is the most commonly used feature space. In color video, luminance can be computed as weighted combination of color components. In feature analysis, the current image and reference image derived after feature extraction are compared to detect areas of change.

To obtain the binary map, based on the results of comparison in the above step, a pixel under test is classified into one of two classes: changed or unchanged based. However, since the results of comparison step contain noise, the decision is done by comparing the results against a threshold.

The result of change detection algorithms is not accurate. Post-processing using binary mask only is the simplest approach to remove irregularities. It preserves contours and reduces spurious regions of the mask. It does not use any information from the original image. It consists of simple morphological opening or closing or a more complex composition of morphological filters.

3.2 Videos without Initial Background

In the case of the Videos without initial background the system aims at removing uncovered background by integrating two change detection masks. These change detection masks are combined by a logical operator. The operator removes all areas except the foreground object detected which is the region that overlaps in the two change detection masks.

Perhaps the most critical issue in comparison is selection of the GMOB. Two choices are possible: the previous frame relative to the current frame in the sequence, or an image representing background of the scene. The background frame can be fixed, or can be updated Periodically. Fixed background frames can be obtained from the sequence prior to entrance of MOs to the scene. Background can be updated periodically by temporal integration of previous and next frames, to produce large background image. However, those approaches are not always practical for the following reasons: For using fixed backgrounds, there should be an initial frame of a sequence with no foreground objects which is not always available. Construction of large background image complicates the system and introduces delay, since it requires integration of a number of previous and next images to obtain optimal background image..

3.3 Videos with Camera in Motion

For video sequences that contain camera motion, which causes GMOB, this global motion is compensated by the Motion estimation and compensation blocks. These are the Forward GMEC and Backward GMEC blocks in Figure 1. There are three phases of the GMEC process proposed in this design. First is Dense motion vector estimation and Parameter estimation then Motion Compensation or Frame warping.

For the dense motion vector estimation, a three-level Hierarchical block based algorithm is used for its compromised performance in computational complexity, speed and accuracy. Mean pyramids are constructed using the simple averaging equation. Then the pyramidal images are searched from top to bottom using MAD for block matching criterion. A full search within two pixels distance around the concerned block is done for better accuracy of matching block. For the parameter estimation the six parameter motion model, the affine model is used for the global motion.

For estimation of the affine parameters, the least squares minimization criterion over the background frame is used to minimize the error between the dense and parametric model motion vectors. The previous background mask B_{n-1} is used to estimate the background part of the current frame where the global motion is valid. The parameters estimated from the above step are used to align the previous/next frame to the current frame.

In the frame warping process, a new frame is calculated from previous frame by transforming the coordinates of the pixels of the previous frame into a new coordinate system.

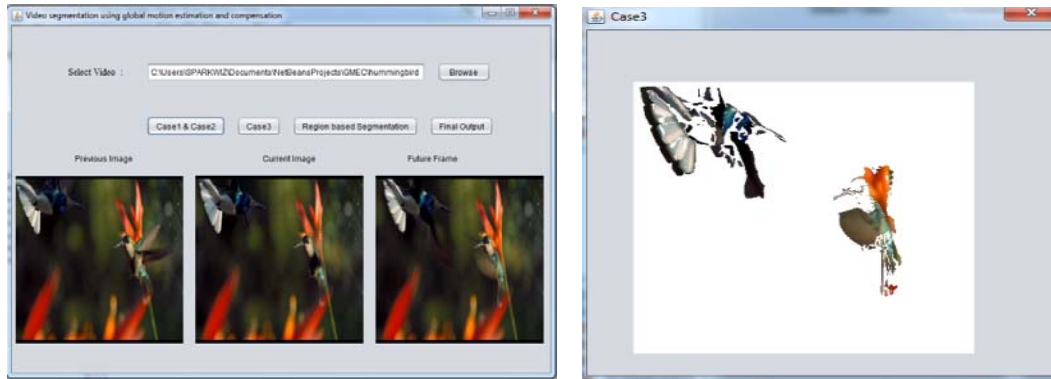
4. Experimental Results

4.1 Dataset

Dataset consist of different standard test video sequences such as segtrack dataset of Hamming bird, birdfall, girl, cheetah which has a ground truth attached to it.

4.2 Forward and Backward Mask Result(Uncovered Background)

Forward and backward mask is applied on humming bird video. Background is removed except the region in which both masks overlap.



4.3 Change Detection Method Results

The change detection algorithm is applied on a sample video. The results includes the detected object, Initial object mask and the resultant frame after applying the post processing.

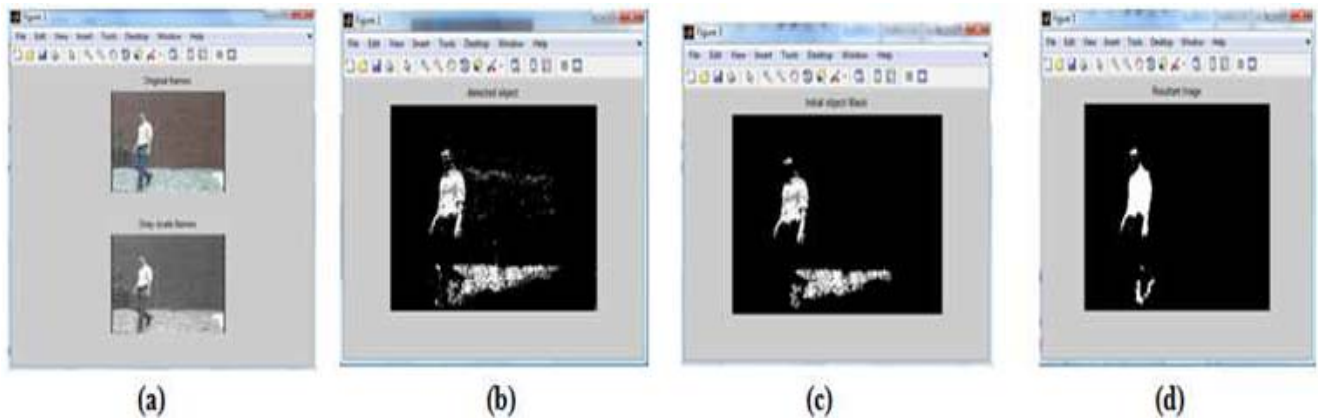


Figure 3. (a) Change detection method on “sample video”.(b) detected Object (c) IOM (d) Resultant Frame after morphological processing

4.4 Performance Analysis

The comparison of the proposed method with the traditional methods is shown in the Fig 4 as follows.

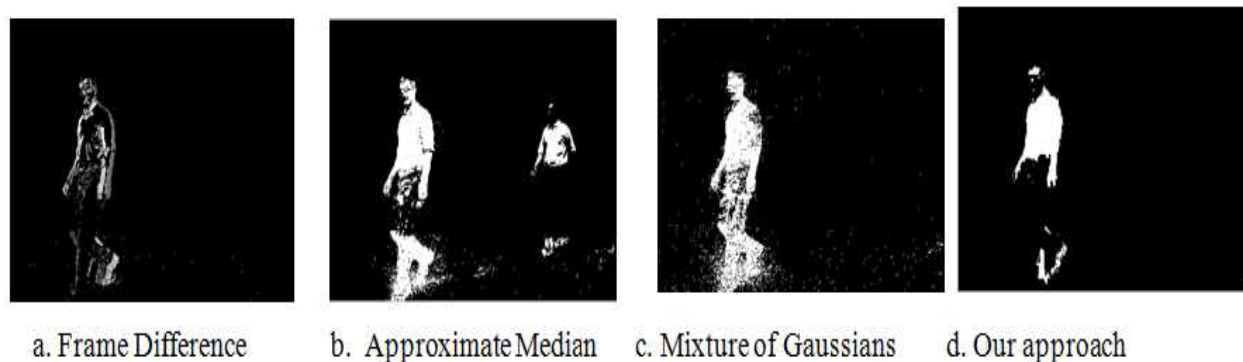


Figure 4. Comparison of proposed Video Segmentation method result with Frame Difference, Approximate median and Mixture of Gaussians (MGM) methods

In Figure 4 (b) object shadows of the object can be seen. But in Fig 4 (c), it is clearly seen occlusion in an image and accuracy is low. All the outputs in Figure 4 (a), (b), and (c) have noise in their frames except the (d), which is the result of the proposed method.

It can be observed that frame difference method is fastest and approximate median method takes almost double frame processing time, moreover other drawback of approximate median is the need for a background model, often unavailable from most videos in uncontrolled settings. Gaussian mixture provides far more accurate results as it is less susceptible to noise. But this method is expensive as its processing time is large and the numerous parameters for configuration make the method very difficult to tune. Contrastingly, frame differencing requires less memory and is more rapid in its simulations. Although frame difference method is also fast and easy to implement, it is very susceptible to noise and highly dependent on continuous movement of the objects, otherwise they will be recognized as background in subsequent frames. Hence the method used proposed work is not just using successive frames instead, we construct and maintain up-to-date background information from the video sequence and compare each frame with the background. Any pixel that is significantly different from the background is assumed to be in object region.

	Processing Time(Sec)	Mean	Variance	Frame Rate (Frame per Second)	Accuracy
Frame Difference (FD)	24.43	4.2	3.5	4.16	H
Approximate Median (AM)	15.56	11.84	10.2	4.06	M
Mixture of Gaussians (MGM)	267.63	11.8	2.0	0.30	L
Proposed method	23.67	0.03	0.04	3.62	H

Table 1. Performance Analysis of Proposed method for Video Segmentation

4. Conclusion

The proposed work solved the issue of Moving camera which adds the unwanted disturbance in the video and also the problem of uncovered background is solved which is visible only when the object moves. The proposed work also solved the issue of Temporary poses when parts of the objects stop moving. The work will improve the accuracy of Video segmentation.

References

- [1] Shao-Yi Chien., Yu-Wen Huang., Bing-Yu Hsieh., Shyh-Yih Ma., Liang-Gee Chen. (2004.). Fast Video Segmentation Algorithm with Shadow Cancellation, Global Motion Compensation, and Adaptive Threshold Techniques, *IEEE Trans. on Circuits and System for Video Technology.*, 6, 732- 748, 5, October.
- [2] Shao-Yi Chien., Shyh-Yih Ma. (2002). Efficient Moving Object Segmentation Algorithm Using Background Registration Technique, *IEEE Trans. on Circuits Syst. Video Technol.*, 12 (7) 577-586.
- [3] Tung-Chien Chen. *Video Segmentation Based on Image Change Detection for Surveillance Systems.*
- [4] Kim, C., Hwang, J. N. (2002). Fast and Automatic Video Object Segmentation and Tracking for Content-Based Application, *IEEE Transaction on Circuits and System for Video Technology*, 12 (2) 122-129.
- [5] Munchurl kim., Jae Gark Choi., Daehee Kim., Hyung Lee., Myoung Ho Lee., Chieteuk Ahn., Yo-Sung Ho. (2002). A VOP Generation Tool: Automatic Segmentation of Moving Object in Image Sequences Based on Spatio-Temporal Information, *IEEE Trans. on Circuits Syst. Video Technol.*, 9 (8) 1216-126, December.
- [6] Neri A. (1998). Automatic moving object and background separation. (*Signal Processing*), 66 (2) 219-232.
- [7] Qingsong Zhu., Guanzheng Liu., Yaoqin Xie. (2011). Dynamic Video Segmentation via a Novel Recursive Kernel Density Estimation, *Sixth International Conference on Image and Graphics.*

- [8] Yang, Lei., Guo, Yiming., Shaobin, Xiaoyu Wu. (2011). An interactive video segmentation approach based on grabcut algorithm 2011. 4th International Congress on Image and Signal Processing IEEE Transactions On Circuits And Systems For Video Technology, 21 (8) August 1163.
- [9] Jiang, H., Helal, A. S., Elmagarmid, A. K., Joshi, A. Scene change detection techniques for video database systems. *multimedia Systems*, 6 (3) 186–195.
- [10] Dailianas, A., Allen, R. B., England, P. Comparison of automatic video segmentation algorithms. *In SPIE Conference on Integration Issues in Large Commercial Media Delivery Systems*, 2615, 2–16, Philadelphia, PA.
- [11] Mandal, M. K., Idris, F., Panchanathan, S. A critical evaluation of image and video indexing techniques in the compressed domain. *Image and Vision Computing*, 17 (7) 513–529.
- [12] Chien, S. Y., Huang, Y. W., Chen, L. G. (2003). Predictive watershed: a fast watershed algorithm for video segmentation,
- [13] Zabih, R., Miler, J., Mai, K. (1999). A feature-based algorithm for detecting and classifying production ejects, *Multimedia Systems* 7, 119-128.
- [14] Zhang, K., Kittler, J. (1998). Using background memory for efficient video coding, In: *Proceedings IEEE Int. Conf. Image Processing*, 944–947.
- [15] Haralick, R. M., Shapiro, L. G. (1992). *Computer and Robot Vision*. Reading, MA: Addison-Wesley, 28–48.
- [16] Chen, T-H., Liao, H-S., 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)
- [17] Thakoor, N., Gao, J., Chen, H. (2004). Automatic Object Detection in Video Sequences with Camera in Motion. *In: Proc. Advanced Concepts for Intelligent Vision Systems*, 2004.
- [18] Tsai, Y., Averbuch, A. (2002). Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach. *IEEE Transactions on Circuits and Systemsfor Video Technology*, 12 (7) 597-612. *Cding Symposium (PCS2004)*, San-Francisco, USA.