A Survey of Moving Object Segmentation Methods

Merin Antony A, J. Anitha

Abstract—Segmentation of moving objects in video sequences is important for multimedia aspects. Moving object segmentation is the extraction of foreground (moving object) from background. Moving object segmentation includes different steps as object detection and motion detection. The object detection and motion detection are done using different methods. This survey deals with three approaches of segmentation as region-based, boundary-based and combination of region and boundary-based methods. A comparative study is done for these methods using strengths, weakness and computational complexity.

Index Terms—background subtraction, boundary based segmentation, motion detection, object detection, region based segmentation.

I. INTRODUCTION

Segmentation is the key concept in the image processing field especially in the image analysis process. Segmentation is used to simplify the representation of an image or video into a more relevant or informative meaningful partitions or segments or to decompose a scene into its components. The input of segmentation is a raw data including an image or a video sequence. The output is a much simpler one in which the homogenous parts are partitioned into much more simpler parts or segments. Segmentation extracts relevant information about the structure of objects from a given image or video sequence and discerns various attributes of interest from the data. These measurements or features are used for the qualitative analysis process. Segmentation is a goal dependent and subjective set-up.

Segmentation of the moving objects in a video sequence is the most important and basic technology used in many real time applications. Segmentation can be useful in areas of video surveillance, security (path detection, target tracking, dynamic scene analysis), object recognition, medical field, navigation system and communication. Different visual features as color, texture, and motion, are used for achieving segmentation.

Segmentation techniques are subdivided into different approaches [7] as:

- Amplitude thresholding or window slicing
- Component labeling segmentation
- Boundary based segmentation
- Region based segmentation
- Template matching
- Texture segmentation

When an object is characterized by the amplitude features, amplitude thresholding is useful. The thresholding technique can be used in various methods as background subtraction, frame difference, etc. Component labeling is an efficient method in case of segmentation of binary images. This includes pixel labeling and connectivity analysis. Boundary based segmentation and region based segmentation are most common techniques used for the segmentation process. Boundary based approach deals with the discontinuities in the images. Region based approach partition images into connected regions by grouping neighboring pixels of similar intensity. Region growing and region merging are used in this segmentation technique. In template matching, the segmentation is done by matching an image against templates from a given list. This technique is mainly used to segment busy images as, journal pages where the text detection is done using the template matching. When the objects to be segmented have textured background, texture segmentation can be used.

The segmentation of objects in a video sequence is a difficult task to accomplish. Here the motion information is important. The slow moving object segmentation process is much more difficult than the segmentation of fast moving video objects. This is due to the semantic gap between the low-level visual cues as color, edge, texture, etc and high-level human interpretation of video semantics.

Moving object segmentation includes the detection, tracking and extraction of the objects in motion. Detection or correspondence is keeping an account of the object that is in motion, about its course, properties, etc. Extraction is the meaningful segmentation of the moving objects from the scene.

Video segmentation is used to identify the regions in a frame of video that are homogeneous with respect to any given parameter. Different features and homogeneity criteria leads to different segmentation of the same data. For example, segmentation with respect to color, texture, motion etc.

Steps in moving object detection

The steps for the detection of moving objects include Object Detection and Motion Detection or Motion Correspondence.

Object Detection
Object Detection is the main step in applications where the human detection is applied as in the video surveillance [2]. The goal of Object Detection is to establish similarity of objects and object parts between consecutive frames of video. The moving object detection process classifies the pixels into two categories: foreground pixels and background pixels. The foreground pixels are considered as of the moving object and the background pixels are considered as of the static background. The detection of moving objects can suffer a lot of problems as noise, sudden illumination changes, shadows, etc. There are many approaches proposed for the successful detection of the moving objects in a video sequence.

The segmentation of moving objects can be classified into several techniques. This survey paper deals with region-based approach, boundary based and combination of region-based and boundary-based approach [1]. In the region based approach, the image is partitioned into connected regions by grouping neighbouring pixels of similar intensity levels. Adjacent regions are then merged under some criterion involving perhaps homogeneity or sharpness of region boundaries. The region based approach includes the traditional background subtraction and optical flow methods. The papers [2]-[4] are using the region based techniques for the detection of the moving objects. The various methods used in the region based approach are background subtraction [2], Markov Random Field [14], Change Detection Mask (CDM) [27], clustering [8] and modified statistical methods [4].

The boundary based approach is an important feature for the image analysis and processing applications. This approach deals with the discontinuities in the images. The boundary approach center around the contour methods. The boundary based approach uses active contour method, edge based and optical flow methods. Many algorithms are proposed for the detection of moving objects which uses various methods. [1] deals a boundary based detection method which uses the optical flow method.

In this survey, a third classification is also introduced. This is the combination of region-based and boundary-based segmentation classifications. [5], [6] and [14] uses both the concepts of region based approach and boundary based approach in equal priorities. This combined approach uses the advantages of both methods and tries to eliminate the disadvantages up to an extent.

Object detection can be classified into fast moving object detection and slow moving object detection. The slow moving object segmentation is difficult compared to the fast moving object segmentation [3].

- **Object correspondence or Motion detection**

Motion detection is the first step in the segmentation process. Object correspondence is performed as an aid for higher-level applications. It is a significant and difficult problem in the area of computer vision. The main goal of the motion detection is to achieve high sensitivity in the extraction of the moving objects with lowest possible false detection rate. It is a significant task in most of the surveillance applications since it provides cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation and to enable higher level data extraction such as activity analysis and behavior recognition. Object extraction has been a difficult task to apply in congested situations due to inaccurate segmentation of objects.

Common problems of erroneous segmentation are long shadows, partial and full occlusion of objects with each other and with stationary items in the scene. Thus, dealing with shadows at motion detection level and coping with occlusions at segmentation level is important for robust extraction. Object extraction in video can be categorized according to the needs of the applications. It is used according to the methods used for its solution.

There are mainly two ways for the tracking step. They are semi-automatic tracking and automatic tracking. Semi-automatic approach is the process of segmenting a video or an image with the user interventions. The user has the provision to specify the attributes or measure dimensions for an image or video. This approach has different methods some of which uses spatio-temporal gradients [9], [12], [13], watershed method [10] and region based approach [11].

The automatic approach do not need the supervision of an individual, it can perform on its own using different predefined methods. The methods used in automatic tracking and segmentation includes Markov Random Field method [14], [15], active contour method [16], [25], region adjacency graph [17], background subtraction [20], [21], [23], modified frame difference [26], thresholding [18] and fast marching algorithm [22].

II. MOVING OBJECT DETECTION AND SEGMENTATION

The segmentation of the moving objects is classified into three approaches including the region-based techniques, boundary-based technique and the combination of region-based and boundary-based technique.

A. Region-Based Methods

The traditional region-based methods are background subtraction and frame difference techniques. The background and frame difference approaches is used for discriminating moving object from the background scene. The idea is to subtract the consecutive images.

- **A geodesic framework for fast interactive image and video segmentation and matting** [9]

The interactive framework is used for the tracking and extraction of foreground and background in an image. Here a method is proposed which uses the geodesic distance of the weight function of a pixel that is included in the user provided scribble region which is to be segmented. For getting the foreground and background details, the gradient of likelihood of a pixel is calculated. In this method the user adds new scribbles for getting the accurate result. The transparency factor is also computed using the feature vector. This geodesic framework is extended for the segmentation of the video images. The scribble is made and the weighted distance is calculated in spatio-temporal space. The spatial and temporal gradients of likelihood function give geodesic weights. Finally, an interactive refinement method is used to extract the foreground from the background.
This approach can be used in case of both images and videos and has an increased computational efficiency as this method speeds up the extraction of foreground from the background.

This method should not work in case of occlusion. This approach mainly exploits the geodesic computation weights rather than the foreground-background image details. This should need user interference as this is not automatic. As the weight of pixels is calculated, the work is tedious.

- **Interactive tracker - A semi-automatic video object tracking and segmentation system [11]**

  This semi-automatic strategy describes a combination of automatic segmentation along with manual segmentation. This method combines the efficiency of the automatic segmentation and accuracy of the manual segmentation. A single frame (mostly the initial one) of the input video sequence is taken as the input. The first step, the interactive segmentation is done by specifying the contour of the object with a computer-aided tool (the improved intelligent scissors). This tool enables the users to extract the objects accurately and also reduces the lag which occurs when the user is adding a key point for the segmentation. The second step is the automatic segmentation which extracts a homogenous region using region extraction and then estimates its motion parameters and finally classifying the result to verify for overlapping. This step helps to smooth the object contour. A user interface is also added to make the segmentation more accurate according to the user concepts.

  More accurate and efficient result is obtained as this method deals with both the automatic and the manual ideologies. The computational time is reduced due to the improved intelligent scissors technique. The presence of a user interface also makes the process more efficient.

  The method does not work accurately for the long period processes.

- **Semiautomatic segmentation using spatio-temporal gradual region merging for MPEG-4 [12]**

  This technique takes the advantage of Video Object Planes (VOP). This method combines low level automatic segmentation with the interactive method for defining and tracking high level semantic video objects. This method uses gradual region merging and bi-directional temporal boundary refinement for the extraction of the object boundaries. The two overall steps include the intra-frame segmentation and inter-frame segmentation. The intra-frame segmentation includes the region partitioning of a frame using the gradual region merging with user interaction and the result is refined by the boundary refinement process. The user interaction is used to identify and portray the initial object boundary easily and accurately. The gradual region merging deals with region labeling, color and frame difference similarities and edge strength. The output of this step is the accurate object boundary of the initial frame. The second step (inter-frame segmentation step) uses the same methods used in the first step for different frames.

  This system accurately extracts object boundaries. The noise and the over-segmentation are reduced. This approach enables the localization of edges and segments.

  The object boundary of the initial frame alone is used for the entire segmentation process. This approach is more compatible in case of still background. This method is not so good in case of fast moving objects.

- **A noise robust method for segmentation of moving objects in video sequences. [28]**

  This method uses the CDM for the automatic segmentation of the moving objects. The segmentation algorithm is subdivided into two steps: change detection mask calculation and object mask calculation. In the change detection mask calculation step, every pixel is calculated where the image luminance is changed. In the object mask calculation, the uncovered backgrounds are detected and eliminated. After the computation of the initial CDM, the relaxation step is being performed which extends the memory of the previous CDM and OM (object mask). The simplification and elimination of the small regions are done. The elimination of the background region is also done. By iterative process, the moving object extraction is done.

  This algorithm reduces the amount of noise from the video in a considerable amount. The boundaries are detected accurately. The false detection is reduced. The temporal coherency is improved.

  The background is not preserved. This algorithm is not perfectly working in case of moving cameras.

- **A robust moving object segmentation algorithm [26]**

  This technique introduces a modified frame difference method for the segmentation of the moving objects. This is a combination of ordinary Reference Frame Subtraction Algorithm (RFSA) and Double Difference Algorithm (DDA). Double Difference Algorithm is used to outline the upper boundary of moving areas and the actual moving object segmentation is performed by reference frame subtraction algorithm. For each new image, an intermediate moving object segmentation image is calculated using DDA. This intermediate moving object segmentation image is dilated by a relatively large dilation core to make sure all fragments belong to the same moving object silhouette are well connected and any holes in the middle of this fragments are filled. The reference frame is updated either replaced by the new value at the same position in or remain the same according to whether intermediate image shows that this pixel location belongs to static or moving region.

  This algorithm is simple. This combined method is robust to background variations and sensitive to slow moving objects. The noise detection is done.

  The segmentation of fast moving objects is not accurate.

- **Image segmentation in video sequences using modified background subtraction [20]**

  Background subtraction technique is a common technique for the detection and segmentation of images in the area of image processing. But there are many defects in the traditional background subtraction algorithm as shadow detection, slow moving object recognition dealing with the poor quality images and videos, etc. So a mixture of GMM and EM algorithms are introduced to modify the background subtraction method. The input videos are converted to gray
scale mode and these in turn generate a number of frames. By using these frames the segmentation is done by Exception Maximum technique. Then the traditional background subtraction method is used for detecting the moving object boundaries. This will give the results as segmented moving objects.

The shadows are detected and eliminated more effectively. Slow moving objects and poor quality videos are handled perfectly. The algorithm is robust against changes in illumination.

The size of the video should be limited. The defined formats can only be used in this case.

- **Unsupervised segmentation of color-texture regions in images and video [24]**

Here an unsupervised segmentation method for the color images and videos is proposed. A new approach is proposed as Joint Segmentation Scheme (JSEG) is used for the tracking and segmentation of the moving objects. This method proposes two steps: color quantization and spatial segmentation. In color quantization step, the colors in the images are quantized to different classes for the differentiation of regions in the images. The color-class labels replace the regions forming a color map of the image. In this method a spatial segmentation is done using the color-map. The result for the spatial segmentation is a J-image which has high and low values for boundaries and interiors of the color-texture regions. Then the multi-scale J-images are segmented using a region growing method where a region tracking method is embedded to the region growing method for achieving consistent segmentation and tracking results.

The main advantage is the robustness of the proposed method for segmenting accurately.

The demerit of this method is the low percentage colors are ignored. This method is not good for a uniform image.

- **A hybrid algorithm for automatic segmentation of slowly moving objects [3]**

This paper uses the idea of frame difference for the detection phase and the spatial segmentation for the tracking and the extraction of the slow moving objects. This paper has three main steps: 1) Motion analysis 2) Spatial Segmentation and 3) Fusion Operation. Motion analysis deals with the analysis of the initial moving regions along with the removal of noise from the images. Spatial segmentation is used for getting the details about the homogenous regions from the images. Fusion operation fuses the results from the motion analysis step and the spatial segmentation step for extracting the perceptually consistent stable moving object as the output.

The major advantages of this paper includes: i) reduces the noise from the images by removing the zero-mean Gaussian distribution. ii) An improved Expectation Maximum (EM) algorithm is been introduced to segment the moving objects to provide more compact image content representation and to get a perceptually consistent output. iii) To reduce the challenging gap between the visual features and the human interpretation, the spatial structure information is incorporated to extract a complete segmentation result. The main disadvantage is it is time consuming.

### B. Boundary-Based Methods

The boundary-based methods are used for detecting the edges or boundaries of the moving objects. The several techniques used for the boundary-based methods are contour modeling techniques and edge detecting techniques.

- **A contour-based moving object detection and tracking [1]**

This is a fast approach for detecting the moving objects. This algorithm uses a boundary based approach. This algorithm uses gradient-based optical flow and an edge detection method. The moving edges are detected. For detecting the moving edges, a two step approach is used. This approach includes a gradient-based optical flow technique and an edge detection technique which uses the canny edge detector for detecting the moving edges.

The main advantage of this approach is it gives more accurate information about the edges than the techniques which uses the rectangle or ellipse contour-patterns for edge detection of moving objects. The computation velocity is made accurate. This technique is more reliable. This method is robust as it uses edge-based features which are insensitive to illumination changes.

The main demerit of this method is the objects extracted by these methods are regions which follow the same motion model. Articulated (jointly) moving objects are split into more than one object. So for extracting the meaningful objects additional processing is required. Here the computational complexity is more which makes this algorithm not a desirable one for real-time uses.

- **Semi-automatic video object segmentation basing on hierarchy optical flow [13]**

This semi-automatic segmentation technique deals with the spatio-temporal techniques. The spatial segmentation uses a user interface (point-based graphic user interface) which helps the user for inputting the measurement points. After this, the active contour modeling and bug tracking algorithm is used for defining the user points. The result of the spatial segmentation is used for the temporal step. This involves non-rigid object contour tracking and rigid object tracking by optical flow algorithm. The tracking performance is improved by the tracking point selection algorithm.

The user can input the parameters with the help of Graphical User Interface (GUI). The user satisfaction is attained. The slow movements are captured accurately.

The fast moving object segmentation is not efficient.

- **Fast moving object segmentation based on active contours [16]**

The active contour method is the useful method for the detection of the edges. This algorithm is an approach for the segmentation of the fast moving objects. There are mainly two steps for this algorithm. The first step comprises of the motion detection based on the frame difference. The frame difference is calculated by using the binarization method on the difference between the frames and merging these results. After motion detection, the initial contours are detected. The second step is spatial segmentation which uses the improved active contour method. Track the object in the next frame by
the current calculated contour as the initial value. This gives
the result as the extraction of the fast moving object.

The computational load is reduced. The algorithm does not
need human supervision. This will improve the fastness and
efficiency of the algorithm.

If there are multiple objects the method is not appropriate.
This algorithm gives no successful results for the complex
background.

- **A moving object segmentation technique using dynamic
  programming [25]**

  The method deals with the edge detection and contour
  linkage of a moving object. This includes three phases:
motion region identification, motion edge detection and
contour linkage. The motion region identification includes
the computing the frame difference and the removal of the
noise. The motion edge detection step uses a laplacian filter
to extract motion edge of current frame. And the Gaussian
low pass filter is used for the presmoothing. In the third step,
the contour linkage in motion regions is calculated. The
thinning algorithm is used to thin the blur images. And using
the start and end points, the outline is calculated. Now the
contour linkage outside the motion regions is calculated by
computing color and brightness difference.

  This efficiently segments video streams with good visual
effect. The spatial and temporal coherency is present. This
tracks the objects motion edges in motion regions. This
detects the existing edges of the still objects in the scene. The
computational efficiency is high and the reliability is high.

  When the motion edge accounts for only a small part of the
whole contour, the technique will reduce to an image
segmentation method. This method provides no occlusion
detection.

**C. Combination of region-based and boundary-based
methods**

The region-based and boundary-based techniques are
combined and these combined methods are used for the
segmentation of moving objects. These methods use the
region-based techniques as background subtraction and the
boundary-based techniques as edge detection methods.

- **Automatic object detection in video sequences with
  camera in motion [6]**

  Most video detection algorithms focus on difference
between two consecutive frames. Here another approach is
taken into consideration. This algorithm uses the frame
difference of three consecutive video frames a backward
frame, a frame of interest and a forward frame. Here the
optical flow method is used for the camera motion
compensation. Afterwards, detection of intensity change in
Displaced Frame Differences (DFD) is calculated. This helps
to generate an estimate for the background. This moving area
information is merged with the region information using
region boundary to obtain the final result.

  This algorithm reduces the limitations of the stationary
cameras. It is more efficient in detecting the moving
foreground region by using the optical flow method. This
combines the frame difference method with the optical flow
method for estimating the moving object.

  The boundaries of the objects are not well defined as of the
moving object interiors. This method does not detect
shadows separately and does not eliminate them to get the
accurate moving object without the presence of any
deformities. This will yield good accuracy if the texture
information is given. This method is computationally
complex and cannot be used in real-time video streams
without a specialized hardware. It is not useful in case of
dynamic background.

- **An approach for efficient real time moving object
detection [5]**

  Real time object detection is used in more sophisticated
areas such as classification and visual tracking. This real time
moving object detection method deals with reducing the short
comings of traditional background subtraction method using
the combination of sobel edge detection method. The image
is captured and the frames are being extracted for the
background modeling process. The resulting image is then
processed by the background subtraction method. Then the
final result is obtained by the sobel filtering method with
more accuracy by detecting the edges.

  The computational time is low for this method. This
method is simple compared to the other existing methods.
The edges and their orientations are detected more efficiently.

  The main disadvantage of this method is sensitive to noise
and thus making this method an inaccurate one.

- **Accurate moving object segmentation by a hierarchical
  region labeling approach [14]**

  This technique proposes a method for segmentation of
color sequences accurately by using the hierarchical Markov
Random Field model. This is a fast moving object
segmentation method. This method partition the input frame
into homogenous regions using the watershed algorithm.
Foreground detection is done as the outliers of the estimated
background and the motion vector is estimated. The
segmentation of the moving object is done by using the
hierarchical MRF models which uses large-scale spatial
partition and small-scale spatial partition. The hierarchical
MRF uses to initialize the object mask on large-scale spatial
partition. The object boundary is then refined by the
small-scale spatial partition.

  The main advantages are the object boundary is refined
accurately and gives a consistent result.

  This method is not appropriate for the slow moving object
segmentation. In this method the background is not
preserved.

  Table 1 gives a comparison of different methods used in
region based, boundary based and in the combination of
using strength, weakness and computational complexity.
<table>
<thead>
<tr>
<th>Different Techniques</th>
<th>Methods</th>
<th>Strength</th>
<th>Weakness</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region Based</strong></td>
<td>CDM by local thresholding technique [28]</td>
<td>reduces noise, false detection is reduced, boundaries are detected accurately</td>
<td>Time consuming</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Spatial segmentation [24],[12],[9]. [3]</td>
<td>increased computational efficiency, noise is reduced</td>
<td>useful for still background, work is tedious</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Intelligent scissors [11]</td>
<td>Computational time is reduced, more efficient</td>
<td>not accurate for long period processes</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>RFSSDA [26]</td>
<td>robust to background variations, noise is detected</td>
<td>Inaccurate segmentation results, complex</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Modified background subtraction method [20]</td>
<td>shadows are detected and eliminated</td>
<td>size of the video should be limited, defined formats can only be used</td>
<td>High</td>
</tr>
<tr>
<td><strong>Boundary based</strong></td>
<td>Optical flow method [13],[1]</td>
<td>High Computational velocity, robust, accurate</td>
<td>False detection</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Contour linkage [25],[16]</td>
<td>computational load is reduced</td>
<td>Inappropriate for complex background</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Region and boundary based</strong></td>
<td>Frame difference method [6]</td>
<td>Efficient foreground detection, efficient edge detection</td>
<td>does not detect and eliminate shadows, not useful in case of dynamic background</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Hierarchical MRF method [14]</td>
<td>edges and their orientations are detected more efficiently</td>
<td>sensitive to noise</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Combination of background subtraction and sobel filtering [5]</td>
<td>object boundary is refined accurately</td>
<td>not appropriate for the slow moving object segmentation, background is not preserved</td>
<td>High</td>
</tr>
</tbody>
</table>
III. CONCLUSION

Video processing is the basic method in many image processing applications. By using video processing there are many revolutionary ideologies developed in the areas of image processing. These ideas are applied in various real-time applications. For the meaningful video manipulation, videos are to be segmented into meaningful objects (semantic video objects). Many applications are requiring automatic segmentation of semantic video objects. But it is considered as difficult task.

As the computational complexity of the region based is more than the boundary based methods, the emphasis should be given to the real-time implementation of the region based methods. Compared to the region based approaches, the boundary based approaches are better in computational complexity but they are not so efficient in presence of occluded objects. A combination of region based and boundary based methods are also considered which can reduce the short comings up to an extent.

The future works should be concentrated in the combined methods which reduce the computational complexity.

REFERENCES


Ms. Merin Antony A completed her B.Sc in Physics with First class in 2008 from Calicut University, India and her MCA with First Class in 2011 from CUSAT, India. Currently she is pursuing M.Tech Research in the department of Computer Science and Engineering, Karunya University. Her research interests include the area of image processing.

Mrs. J. Anitha completed her B.E. in Information Technology with First class distinction in 2004 from Manonmaniam Sundarnar University, India. She completed M.E. in Computer Science with first class distinction in 2006 from Manonmaniam Sundarnar University. She started her teaching career in Noorul Islam college of Engineering from March 2006. Currently she is pursuing Ph.D in the area of Image processing. Her area of interest is in the field of medical image segmentation, compression and watermarking. She is the life member of Computer Society of India.