Inspection of Various Object Tracking Techniques

Ann Maria Jacob, Anitha J

Abstract—Object tracking finds its application in several computer vision applications, such as video compression, surveillance, robotics etc. This paper presents a survey on the various existing object tracking techniques. Besides using either region or boundary features, tracking techniques that combine both region and boundary features are also considered. In region based object tracking, color and texture features are considered. These features can quickly and roughly locate objects. In boundary based object tracking, edge and frame difference features are included. Boundary features provide more accurate shape information. This survey shows that object tracking techniques using both region and boundary features significantly improves the performance under different situations.

Index Terms—Boundary Based Object Tracking, Object Representations, Object Tracking, Region and Boundary Based Object Tracking.

I. INTRODUCTION

Object tracking is a famous technique involved in many computer applications. The rapid increase of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has introduced a great level of interest in object tracking. Object tracking is defined as the process of segmenting an object of interest from a video scene and keeping track of its motion, orientation, and occlusion etc., to extract useful information [10]. The first relevant step of information extraction is the detection of the moving objects in video streams. Next steps are the tracking of such detected objects from frame to frame and analysis of object tracks to recognize their behavior. The information extraction is mainly in computer vision applications like traffic monitoring, automated remote video surveillance and people tracking. The three basic approaches in object tracking are feature-based methods, differential methods and correlation [10]. Feature based methods extracts characteristics such as points, line segments from image sequences. Then the tracking stage is ensured by a matching procedure at every time instant. Differential methods use optical flow computation. The correlation is used to measure interimage displacements. Selection of each method largely depends on the domain of the problem. The use of object tracking is applicable in the tasks of motion-based recognition, automated surveillance, video indexing, human-computer interaction, traffic monitoring, vehicle navigation etc. In a process of tracking, an object can be defined as anything that a person wants to track. It can be like, fish inside an aquarium, vehicles on a road, planes in the air, people walking on a road, or bubbles in the water are a set of objects etc. Objects can be represented by their shapes and appearances. The object shape representations commonly employed for tracking are the following [11]:

- Points: The object is represented by a point, that is, the centroid (Fig. 1(a)), generally the point representation is suitable for tracking objects that occupy small regions in an image.
- Primitive geometric shapes: Object shape is represented by a rectangle, ellipse (Fig. 1(c), (d)). Object motion for such representations is usually modelled by translation, affine, or projective transformation.
- Object silhouette and contour: Contour representation defines the boundary of an object (Fig 1(g), (h)). The region inside the contour is called the silhouette of the object (see Fig 1(i)). Complex non-rigid shapes can be tracked by using Silhouette and contour representations
- Articulated shape models: Articulated objects are composed of body parts that are held together with joints as shown in Fig.1 (e).

Fig 1 Object representations from [11] (a) Centroid, (b) Multiple Points, (c) Rectangular Patch, (d) Elliptical Patch, (e) Part-Based Multiple Patches, (f) Object Skeleton, (g) Complete Object Contour, (h) Control Points on Object Contour, (i) Object Silhouette.

- Skeletal models: Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation for recognizing objects. Skeleton representation can be used to model both articulated and rigid objects (see Fig.1. 1(f)).

There are different ways to represent the appearance features of objects. The shape representations and appearance representations can be combined together for tracking. Some common appearance representations in the context of object tracking are probability densities of
II. TRACKING METHODS

Earlier approaches of object tracking consider either region features or boundary features. There are also methods that consider both the region and boundary features. In this chapter, a survey is performed considering object tracking using either region or boundary features and considering both features together.

A. Region Based Object Tracking

Reference [1] proposed a method to track the complete object regions thus making it adaptable to changing visual features. It is also able to handle occlusions. By evolving the contour from frame to frame, tracking is achieved. Contour can be evolved by minimizing some energy functional. This energy functional is evaluated in the contour vicinity defined by a band. Color and texture comprises the visual features. It is modelled by a semi-parametric models and they are fused using independent opinion polling. Shape priors consist of shape level sets which are used to recover the missing object regions during occlusion. In this approach, the performance and effectiveness are demonstrated on real sequences with and without object occlusions. Here a Bayesian framework is used for contour tracking. It is formulated as a variation calculus problem. There are two energy terms in this contour energy functional, the image energy and the shape energy. Image energy is based on colour and texture observations and is evaluated in a band around the contour. The shape energy is preserves the shape of the object during partial and full occlusions. Here, an online shape model is considered during the course of tracking. Tracking can be achieved by contour evolution. Level sets can be used to represent the contour, by minimizing energy in the gradient descent direction. Object tracking can be treated as two-class discriminant analysis of pixels, where the classes correspond to the object and the background regions. In order to handle the occlusions, the first step is to detect the occlusion, and the second step is to recover the shape of the occluded objects. This method works well for mobile cameras and does not require camera motion estimation. It is well adapted to changing color and texture features. It is capable of tracking the complete region of the non-rigid objects. The occluded object parts can be easily recovered by this method. Several ambiguities may be caused by the complete occlusion of similar looking objects. Using area and distance heuristics object tracking can be performed. It can be ambiguous when the object size changes due to swift zooming, while it is close to the other objects. This method shows the robust tracking performance with occlusion in video acquired from moving cameras. Reference [2] proposed a method that introduces a real-time tracker based on color, texture and motion information. This approach uses multiple image features for frame-to-frame correspondence matching. Color cues include RGB color histogram and correlogram (autocorrelogram). Local Binary Patterns (LBP) is used to represent the texture features. LBP is chosen for the texture, and geometric location and the smoothness of trajectory provide the motion support. The merging and splitting of objects are handled using the same set of features. Object location and trajectory can be considered in object tracking. The extracted features are used to build a unifying distance measure. This measure is used in tracking and in the classification event, in which an object is leaving a group. Using texture-based background subtraction algorithm, the initial object detection are performed. This method works well with low illumination conditions and low frame rates which are very common in large scale surveillance systems. Color, texture, shape and temporal properties are the main sources of descriptors. The tracker consists of two main elements: background subtraction (detection) and tracking. The subtraction on the video data, that is first processed with a Gaussian filter to remove noise, is done by an adaptive algorithm which is based on LBP texture distributions. The algorithm was chosen because of its good performance in most environments and the fact that it exploits the same texture properties as the object matching part of the tracker, so re-use of features is possible. The tracking is done by matching features extracted from the subtracted foreground shapes. The color, texture and motion are based on histogram distributions: RGB color histogram and correlogram and LBP. The method is able to track multiple objects in diverse conditions. This approach is also less sensitive to color due to the use of a versatile collection of cues. The system is also robust in performance. Reference [3] proposed a fast texture-segmentation approach. This is suitable for real-time or interactive applications. This approach is based on detecting texture boundaries in the direction normal to the contour boundaries. A Hidden Markov model (HMM) used to link these boundary points in the other direction. It is used to bind and exploit local texture information in form of texture crossing...
probabilities on a bundle of scan lines that is referred to as scan stripes. These yields the most likely connected contour(s) of the object. The probabilities in this computation are strictly related to texture entropy and Kullback- Leibler divergence. It is the relative entropy between the target texture and the proposed target texture model is measured in the course of tracking and used to dynamically update the texture model. The method yields more robust tracking behavior, for example under lighting changes. This method is also applicable for other applications that require near real-time performance, such as the interactive drawing of image boundaries that are becoming increasingly popular in image-processing systems such as Photoshop. This approach is for contour-based object tracking and texture segmentation that relies on detecting texture boundaries in a given search region and direction. The time taken to calculate the texture boundary is typically less than a second depending on the length of the initial curve. Here the implementation for tracking is so fast and it can be optimized to process several frames per second.

B. Boundary Based Object Tracking

Reference [4] proposed a non-parameterized object contour-based method in a single video stream is used. The border between a tracked object and background areas forms the contour. In the process of motion tracking, it is required to compute the object motion and edge motion. The object motion can be computed by taking the inside area of the previous contour and computing the offset to match the current image best. The edge motion corresponds to the magnitude and direction of an optical flow for every edge pixel of an image edge map. To create a new contour, the water snake model is used in this approach. Two edge indicator functions are defined to decide a new contour in the stage of new contour detection. The process of boundary edge selection consists of two steps. First, background edge removal is achieved using edge motions. Second, boundary edge pixel selection is achieved using the gradient in the normal direction of the contour. The background edges are removed by comparing motions of object and background. The tracked subject motion and background edge motions are computed. The background edges, whose motion directions differ from that of the tracked subject, are removed. Edge motion is computed using optical flow from edge maps generated by the canny edge generator. After removing the background edge, the boundary edge pixels are selected. The boundary edge selection can handle the changes in object shape in complex scene. The method can select the appropriate edges and contour without interference from the texture of object and background. This method cannot preserve the contour which has high curvature. Reference [5] proposed a fast and robust approach to the detection and tracking of moving objects. The method is based on using lines computed by a gradient-based optical flow and an edge detector. The gradient based optical flow and edges are well matched for accurate computation of velocity, here much attention is not paid to creating systems for detecting and tracking objects using this feature. In this method, extracted edges by using optical flow and the edge detector are restored as lines, and background lines of the previous frame are subtracted. Object contour can be obtained by using snakes to clustered lines. Detected objects are tracked, and each tracked object is able to handle occlusion and interference. This method is very fast and has robust performance in outdoor-scenes. This method is for detecting and tracking moving objects which includes non-rigid objects. This method is robust because edge-based features which are insensitive to illumination changes are used. The method is also fast because as the area of edge-based features are less than region based features. Here features having strong magnitudes of the gradients are used. Here the features are extracted using Canny Edge Detector. The method for extracting contours of moving objects consists of 4 steps: line restoration, line-based background subtraction, clustering, and active contour. This method is tested on various sequences including cars, pedestrians, approaching cars, occluded cars and interfering objects. All sequences are recorded on outdoor scenes that include the sky, trees, buildings, grounds, and snow. They include several kinds of noise which are caused by illumination changes, small movement in the background, and reflection. The results showed remarkable robustness against these environments. The method also succeeded in detecting and tracking moving objects very accurately in all video sequences even if these sequences had many causes of noise. This method is a contour-based detection, which allows users to obtain more accurate information using rectangles or ellipses.

Reference [6] proposed a method using the concepts of dynamic template matching and frame differencing. This method is usually implemented for a robust automated single object tracking system. Here a monochrome industrial camera is used. Normally it is used to grab the video frames and track an object. The frame differencing calculated on each frame-by-frame of the moving object, and is detected with high accuracy and efficiency. By employing an efficient template matching algorithm, detected object is tracked. The templates for the matching purposes can be generated dynamically. So that any change in the pose of the object will not cause any hindrance to the tracking procedure. In order to automate the tracking process the camera is mounted on a pan-tilt arrangement. This is synchronized with a tracking algorithm. When the object being tracked moves out of the viewing range of the camera, then the pan-tilt setup will automatically get adjusted to move the camera to keep the object in view. This method is capable of handling entry and exit of an object. So this tracking system is very cost effective. It is also useful as an automated video conferencing system. This also has application as a surveillance tool. The system first analyses the images which is being grabbed by the
camera, in order to detect any moving object. The Frame Differencing algorithm is used for this purpose. Its output will be the position of the moving object in the image. This information can then be used to extract information from the video or image. The template is generated using the appearance changes of the object. The template which is newly generated is then passed on to tracking module, which starts tracking the object taking this template as the reference input. The module uses template-matching to search for the input template in the scene which is grabbed by the camera. If the object is lost while tracking, a new template is generated and used. As the image templates are generated dynamically, the process is called Dynamic Template Matching. For automation of the Pan-Tilt mechanism, the movement of the object is analyzed. The pan-tilt mechanism is operated to keep the object in the camera's view. This is proven to be quite accurate and effective in detecting a single moving object even under bad lighting conditions or occlusions. Using a pan-tilt setup, the system has been automated. Such an automated object tracking system can be used in applications where accurate tracking is required but good lighting conditions cannot be provided. The system is also very much applicable to areas like surveillance and video conferencing.

C. Region and Boundary Based Object Tracking

Reference [7] proposed an approach based on both region and boundary information approaches. This uses feed-forward, feedback and motion based methods. This approach uses feed-forward to use region-based information to propagate boundary estimates, feedback to use boundaries to improve motion estimation, and finally uses motion-based warping to compare image appearance between frames in order to provide additional information for the boundary estimation process. When either the region or boundary information is considered, there are so many limitations. Motion constraints can be weak in areas of limited texture, while boundary constraints can be weak in areas of limited contrast. Here both the region and boundary information are considered to detect, track and recover the shape of a moving object. So feed-forward and feed-back approaches are used to combine region-based information and boundary-based information. Motion constraints and intensity-consistency constraints are considered as region-based information. Gradient-based motion constraints are used to compute a parametric, layered flow model to estimate the local image motion. The object boundary is considered as boundary-based information determined by an active contour. The distance-transform active contour is used here. In addition to these a motion model is used to wrap the contour between the frames, thus it provides a more accurate starting location. This introduces a framework that allows for a moving background or observer, and multiple moving objects. The feed-forward and feed-back mechanisms lead to a synergistic segmentation and tracking algorithm. In the case of articulating objects that are composed of rigid parts, this approach will allow for decomposition of objects based on these parts, also facilitate the tracking of objects under occlusion. Using this approach it is difficult to incorporate non-parametric flow estimation. Reference [8] proposed a method to address the problem of tracking several non-rigid objects over a sequence of frames acquired from a static observer using boundary and region based information under a coupled geodesic active contour framework. On a given current frame, a statistical analysis on the observed frame difference frame is performed. This provides the observed probability density function of the difference frame that distinguishes between the static and mobile regions. In this two information's to a coupled geodesic active contour model is incorporated. Using a gradient descent method we are minimizing the objective function. A Level-Set approach is used to implement the obtained probability density function. Using a very fast front propagation algorithm, the level-set propagation is performed. In this approach, several moving objects can be tracked simultaneously. Another contribution consists of creating a tracking model that integrates different types and sources of information. In this approach, a system in which both the boundary and the region module operates simultaneously is developed. The main disadvantage of this model is that in cases of occlusion, the output is a single curve for both objects. Reference [9] proposed an active-contour method which is capable of tracking multiple moving objects on non-static and cluttered background. This also efficiently handles inter-object occlusions. The multiple cues from the image is used to track the objects. The colour, texture, the image edge map and the shape of the objects are the cues that are considering here. The tracking is performed by minimizing energy functional. This combines region, boundary and shape information about the objects to find their boundaries in all the video frames. By minimizing the distance between local and global statistics of the object and background is used to form the region information. Multi-band edge detector is used to formulate the boundary information. Similarly shape information is formulated using the properties of level set contours. In this approach, only segmentation of the objects in the first frame of the sequence is required. Energy functional is obtained by combining the region, boundary and shape modules in one module. The minimization of the above energy is performed using Euler–Lagrange equations and implemented using level set functions. This method was successful for sequences with noise, cluttered background and inter-object occlusions. The drawback is that the accuracy begins to drop when the change of brightness is severe. When the occlusion is not short lived, the object may lose. When performing the tracking in a current frame, a part of the tracked object must initially be contained within the contour; otherwise, the contour may disappear and the object be lost forever. This condition is not always
satisfied, e.g. when the movement of the object is very fast. The experimental results demonstrated good tracking performance despite the presence of occlusions and the possibility of contour distractions. In this approach the segmentation could be performed automatically, by combining detection and tracking.

III. COMPARISONS OF VARIOUS OBJECT TRACKING TECHNIQUES

This section compares the various object tracking techniques using either region or boundary features and using both region and boundary features and thus identifying which approach provides the best object tracking result for the given input video. The comparison is given in Table 1 which is shown in Appendix.

IV. CONCLUSION

In this paper, a comparison of some existing tracking techniques using either region or boundary feature and those combining both region and boundary features is considered. The region features, including color and texture, determined the correct location of the object. Boundary features included frame difference and edge, which determined the moving object boundaries. For a low computational cost, motion prediction was used to initialize the tracking contour in each frame. The experimental results showed the improved performance of the object tracking method using both region and boundary Features when compared with either region feature based or boundary feature based model in many complex circumstances.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who helped us in completing this work.

REFERENCES


AUTHOR’S PROFILE

Miss Ann Maria Jacob completed her B.Tech in Computer Science with First class from Mahatma Gandhi University, 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 digital image processing.

Mrs Anitha J completed her B.E in Information Technology with First class distinction in 2004 from Manonmaniam Sundaranar University, India. She also completed M.E in Computer Science with First class distinction in 2006 from Manonmaniam Sundaranar University, India. 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 includes image segmentation, compression and watermarking. She is the life member of Computer Society of India.
<table>
<thead>
<tr>
<th>Tracking Technique</th>
<th>Method Description</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| Region Based                        | [1] Tracking is based on the features like color and texture and uses the Bayesian framework for object tracking. | • Handles occlusions.  
• It works for mobile cameras and do not require camera motion estimation. | • For similar objects with different shapes, detected contour will be wrong.                      |
|                                     | [2] Tracking is done by matching features extracted from the subtracted foreground shapes | • Works well for low illumination conditions and low frame rates.  
• The system is able to track multiple objects in diverse conditions.  
• The method is less sensitive to colour. | • Cars and other interesting objects of arbitrary shapes cannot be detected. |
|                                     | [3] A fast texture-segmentation approach that relies on detecting texture boundaries in a given search region and direction is used for tracking the object. | • This technique does not rely on a training database.  
• The method of tracking is fast and can be optimized to process several frames per second. | • The application is restricted when the contrast is sufficient.  
• It is difficult to find the intensity values and structures available in images. |
| Boundary Based                      | [4] Selected only tracked object boundary edges in a video stream with a changing background and moving camera. | • There is no great loss of information.  
• The selection of contour edges are not affected by noisy edges or small cross striped textures | • The tracking method is restricted to the space-variant images.  
• The method cannot preserve the contour which has high curvature. |
|                                     | [5] Edge based features are used for detecting and tracking moving objects. | • The method is robust and fast.  
• Handles occlusion and interference. | • Difficult to track the object when the noise is very high. |
|                                     | [6] The concepts of frame differencing and dynamic template matching is used to implement the object tracking system. | • The system is capable of handling entry and exit of object .  
• The method is cost effective and can be used as an automated video conferencing system  
• It also acts as a surveillance tool. | • It is difficult to track multiple objects at the same time.  
• Lack of tracker accuracy during camera motion. |
<p>| Region and Boundary based           | [7] The method uses a feed-forward and feed-back approach to combine the region based and boundary based | • This allows for a moving background or observer, and multiple moving objects. | • The shape of the contour may change as an object is tracked over time. |</p>
<table>
<thead>
<tr>
<th>Information</th>
<th>This method allows the decomposition of objects based on rigid parts.</th>
<th>The presence of small and very narrow concavities is problematic.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[8] This addresses the problem of tracking using region and boundary based information under a coupled geodesic active contour framework.</td>
<td>Several moving objects can be tracked simultaneously.</td>
<td>The objects cannot be tracked correctly in the case of occlusion.</td>
</tr>
<tr>
<td>[9] In this paper, a robust model for tracking in video sequences with non static backgrounds.</td>
<td>Tracking multiple moving objects on non-static and cluttered backgrounds.</td>
<td>Accuracy begins to drop when the brightness is severe.</td>
</tr>
<tr>
<td></td>
<td>Handles inter-object occlusions</td>
<td>When the occlusion is short-lived, it is difficult to track the object.</td>
</tr>
</tbody>
</table>

Table I Comparison of Various Tracking Techniques