Empirical Evaluation of Detecting Moving Objects Using Graph Cut Segmentation

A .Ramya, P.Raviraj

Dept of Computer Science and Engineering, Kalaignar Karunanidhi Institute of Technology, Coimbatore, India.

Kalaignar Karunanidhi Institute of Technology, Coimbatore, India.

Abstract—Real-time moving object detection, classification and tracking capabilities is presented with its system operates on both color and gray scale video imagery from a stationary camera. It can handle object detection in indoor and outdoor environments and under changing illumination conditions. Object detection in a video is usually performed by object detectors or background subtraction techniques. Our proposed new background model updating method and adaptive thresholding are used to produce a foreground object mask for object tracking initialization. The proposed method to determine the threshold automatically and dynamically depending on the intensities of the pixels in the current frame. In this method update the background model with learning rate depending on the differences of the pixels in the background model of the previous frame. The graph cut segmentation based region merging algorithm approach achieves both segmentation and optical flow computation accurately and they can work in the presence of large camera motion. The algorithm makes use of the shape of the detected objects and temporal tracking results to successfully categorize objects into pre-defined classes like human, human group and vehicle.

Keywords— Object Detection, Background Subtraction, Object Tracking, Graph-Cut, Region Merging

I.INTRODUCTION

Automated video analysis is important for many vision applications, such as surveillance, traffic monitoring, augmented reality, vehicle navigation, etc. [1], [2], [3]. As pointed out in [1], [2] there are three key steps for automated video analysis: object detection, object tracking, and behavior recognition. As the first step, object detection aims to locate and segment interesting objects in a video. Then, such objects can be tracked from frame to frame, and the tracks can be analyzed to

critical role in practical applications. Object detection is usually achieved by object detectors or background subtraction [2]. An object detector is a classifier that scans the image by a sliding window and labels each sub image defined by the window as either object or background. Generally, the classifier is built by offline learning on separate datasets [4], [5] or by online learning initialized with a manually labelled frame at the start of a video [6], [7]. Alternatively, background subtraction [8] compares images with a background model and detects the changes as objects and also assumes that no object appears in images when building the background model [9], [3]. Such requirements of training examples for object or background modelling actually limit the applicability of above-mentioned methods in automated video analysis. The detecting of moving target in video sequence is of importance in many applications, such as intelligent transportation, safety monitoring, etc. At present, the main existed approaches of moving target detection are background difference, time difference (frame difference) and optical flow. In the past, there have been researchers investigating kinds of methods for segmenting moving objects in real time to achieve these vision-based applications. Another category of object detection methods that can avoid training phases are motion-based methods [2], [3], which only use motion information to separate objects from the background. The problem can be rephrased as follows: Given a sequence of images in which foreground objects are present and moving differently from the background, can we separate the objects from the background automatically? Fig. 1 shows such an example, where a walking lady is always present and recorded by a handheld camera. The goal is to take the image sequence as input and directly output a mask sequence of the walking lady.

Fig.1. Examples to illustrate the problem. (a) A sequence of 40 frames, where a walking lady is recorded by a handheld camera. From left to right are the first, 20th, and 40th frames.

The most natural way for motion-based object detection is to classify pixels according to motion patterns, which is usually named motion segmentation [10], [11]. These approaches achieve both segmentation and optical flow computation accurately and they can work in the presence of large camera motion. However, they assume rigid motion [10] or smooth motion [11] in respective regions, which is not generally true in practice. In practice, the foreground motion can be very complicated with non-rigid shape changes. Also, the background may be complex, including illumination changes and varying textures such as waving trees and sea waves. Fig. 2 shows such a challenging example.

Fig.2. A sequence of 48 frames clipped from a surveillance video at the airport. From left to right are the first, 24th, and 48th frames. Notice that the escalator is moving. The objective is to segment the moving people automatically without extra inputs.

The video includes an operating escalator, but it should be regarded as background for human tracking purpose. An alternative motion-based approach is background estimation [12], [13]. Different from background subtraction, it estimates a background model directly from the testing sequence.

II. RELATED WORK

A. Motion Segmentation

In motion segmentation, the moving objects are continuously present in the scene, and the background may also move due to camera motion. The target is to separate different motions. A common approach for motion segmentation is to partition the dense optical-flow field [16]. This is usually achieved by decomposing the image into different motion layers. The assumption is that the optical-flow field should be smooth in each motion layer, and sharp motion changes only occur at layer boundaries. Dense optical flow and motion boundaries are computed in an alternating manner named motion competition [11], which is usually implemented in a level set framework. A similar scheme is later applied to dynamic texture segmentation. While high accuracy can be achieved in these methods, accurate motion analysis itself is a challenging task due to the difficulties raised by aperture problem, occlusion, video noises, etc. Moreover, most of the motion segmentation methods require object contours to be initialized and the number of foreground objects to be specified [11].

B. Background Subtraction

In background subtraction, the general assumption is that a background model can be obtained from a training sequence that does not contain foreground objects. Moreover, it usually assumes that the video is captured by a static camera [8]. Thus, foreground objects can be detected by checking the difference between the testing frame and the background model built previously. A considerable number of works have been done on background modelling, i.e., building a proper representation of the background scene. Typical methods include single Gaussian distribution, Mixture of Gaussian (MoG), kernel density estimation, block correlation, codebook model, Hidden Markov model, and linear autoregressive models. Learning with sparsity has drawn a lot of attention in recent machine learning and computer vision research, and several methods based on the sparse representation for background modelling have been developed. One pioneering work is the eigen backgrounds model, [14] where the principal component analysis (PCA) is performed on a training sequence. When a new frame arrives, it is projected onto the subspace spanned by the principal components, and the residues indicate the presence of new objects. An alternative approach that can operate sequentially is sparse signal recovery. Background subtraction is formulated as a regression problem with the assumption that a new-coming frame should be sparsely represented by a linear combination of preceding frames except for foreground parts. These models capture the correlation between video frames. Thus, they can naturally handle global variations in the background such as illumination change and dynamic textures.

III. PROPOSED SYSTEM

The proposed method aims to extract moving objects from an input image by utilizing their background. The proposed method consists of four steps: (a) Video Frame Extraction (b) Static Background Subtraction (c) Dynamic Background Subtraction (d) Graph cut segmentation.

A. Video Frame Extraction

Video summarization is a compact representation of a video sequence. It is useful for various video applications such as video browsing and retrieval systems. A video summarization can be a preview sequence which can be a collection of key frames which is a set of chosen frames of a video. Key-frame-based video summarization may lose the spatio-temporal properties and audio content in the original video sequence; it is the simplest and the most common method. When temporal order is maintained in selecting the key frames, users can locate specific video segments of interest by choosing a particular key frame using a browsing tool. Key frames are also effective in representing visual content of a video sequence for retrieval purposes. Video indexes may be constructed based on visual features of key frames, and queries may be directed at key frames using image retrieval techniques. Video frames reduce the amount of data required in video indexing and provides framework for dealing with the video content. In Video Frame extraction method involves converting .avi video file in to number of frames. In frame extraction there are two type's methods to be implemented; .JPG file format, Cdata (Character Data) conversion. Processing each frame by taking the histogram difference between them and the frames above the some average will be collected. The collected frames are called as key frames. The Color image data array is convert into indexed array data (i.e., the corresponding

data is convert into gray scale value (0 and 1)). The Converted Frames are in 4-Dimension array, so first reshape the matrix into 3D array. The Pre-Alignment process is used to converting 4-D array into 3-D array image data with help of image total pixels and number of frames.

B. Static Background Extraction

The basic scheme of background subtraction is to subtract the image from a reference image that models the background scene. Typically, the basic steps of the algorithm are as follows:

- Background modeling constructs a reference image representing the background.
- Threshold selection determines appropriate threshold values used in the subtraction operation to obtain a desired detection rate.
- Subtraction operation or pixel classification classifies the type of a given pixel, i.e., the pixel is the part of background (including ordinary background and shaded background), or it is a moving object.

Background subtraction is particularly a commonly used technique for motion segmentation in static scenes. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period. The pixels where the difference is above a threshold are classified as foreground. After creating a foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt to dynamic scene changes. In background subtraction, the general assumption is that a background model can be obtained from a training sequence that does not contain foreground objects. Background model is a static image (assumed to have no objects present). Pixels are labelled as object (1) or not object (0) based on thresholding the absolute intensity difference between current frame and background. Background subtraction does a reasonable job of extracting the shape of an object, provided the object intensity/color is sufficiently different from the background. Background model is replaced with the previous image. The background intensity should be unchanged over the sequence except for variations arising from illumination change or periodical motion of dynamic textures [18]. Thus, background images are linearly correlated with each other, forming a low-rank matrix B. Besides the low-rank property, we don't make any additional assumption on the background scene. Thus, we only impose the following constraint on B: Rank (B) $\leq$ K where K is a constant to be predefined. Intrinsically, K constrains the complexity of the background model.

C. Dynamic Background Extraction

The video training frames must be repeated for each scene where the algorithms are deployed, but training information may not always available, and the background parameters may need to be continuously updated if the scene is dynamic. For instance, it is often assumed that the foreground moves in a consistent direction (temporal persistence), with faster appearance changes than the background. Such assumptions are not always valid, and are particularly questionable when there is ego motion (e.g. a camera that tracks a moving object). To address these limitations, we propose a novel paradigm for background subtraction. This paradigm is inspired by biological vision, where background subtraction is inherent to the task of deploying visual attention. This can be done in multiple ways but frequently relies on motion saliency mechanisms, which identify regions of the visual field where objects move differently from the background. The background subtraction is formulated as the complement of saliency detection algorithm.

D. Graph cut Region merging Motion Segmentation

The graph cut region merging is the basic terminology in the context of our segmentation method. An undirected graph G = ⟨V, E⟩ is defined as a set of nodes (vertices V) and a set f undirected edges (E) that connect these nodes [15]. Each edge in the graph is assigned a nonnegative weight (cost) we, there are two special nodes called terminals. A cut is a subset of the edge such that the terminals become a cut of graph as G = ⟨V, E⟩, to find a minimum cut is well suited for segmentation of images. A graph cut with s/t on a graph is to set two disjoint subsets S and T such that the source s is in S and the sink t is in T. the minimum cut problem on a graph is to find a cut that has the minimum cost among all cuts. One of the fundamental results in combinatorial optimization is that the minimum s/t cut can be solved by find the maximum flow from the source s to the sink t. generally speaking, maximum flow is the maximum amount of energy that can be sent from the source s to the sink t with the graph edge is the pipe and edge weight is the capacity of the pipe. In fact, it is appropriate for a combinatorial problem, the nodes of the graph can represent the pixels (or voxels) and the edges of the graph stand for the pixels neighbourhood relationship between pixels. The minimum cut of the graph will generate an optimal segmentation in the image. A robust and novel approach to automatically extract a set of projective transformations induced by these frame regions, detect the occlusion pixels over multiple consecutive frames, and segment the scene into several motion layers. First, after determining a number of seed regions using correspondences in two frames, to expand the seed regions and reject the outliers employing the graph cuts region merging method integrated with salient motion representation. Next, these initial regions are merged into several initial layers according to the motion similarity. Third, an occlusion order constraint on multiple frames is explored, which enforces that the occlusion area increases with the temporal order in a short period and effectively maintains segmentation consistency over multiple consecutive frames.

IV. EXPERIMENTAL RESULTS

The performance of the proposed video segmentation algorithm is tested with many video sequences. Both the Precision and Recall and F-Measure quality evaluations are applied on our algorithm. For quantitative evaluation, we measure the accuracy of outlier detection by comparing a foreground occlusion with support S_0 with S_1

energy. We regard it as a classification problem and evaluate the results using precision and recall, which are defined as
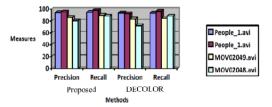
$$precision = \frac{TP}{TP + FP} \qquad (1)$$

$$recall = \frac{TP}{TP + FN} \qquad (2)$$

where TP, FP, TN, and FN mean the numbers of true positives, false positives, true negatives, and false negatives, respectively. Precision and recall are widely used when the class distribution is skewed [17].

Table1: Quantitative Evaluation Using the Sequences from the Berkeley Motion Segmentation Dataset [17] and Real sequence.

Sequence	Proposed		DECOLOR	
	Precision	Recall	Precision	Recall
People_1.avi	94.3%	94.8%	93.6%	93.3%
People_1.avi	95.2%	98.3%	92.5%	96.5%
MOV02049.avi	86.2%	89.9%	83.7%	85.1%
MOV02048.avi	80.75	88.3%	72.6%	88.4%

Precision and Recall Measure

For simplicity, instead of plotting precision/recall curves, we use a single measurement named F-measure that combines precision and recall:
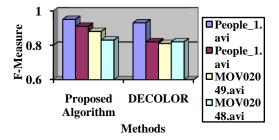
$$Fmeasure = 2\frac{precision \cdot recall}{precision + recall} \qquad (3)$$

The higher the F-measure is, the better the detection accuracy is.

Table 2: Quantitative Evaluation (F-Measure) on the Sequences

Sequence	Proposed Algorithm	DECOLOR
People_1.avi	0.95	0.93
People_1.avi	0.91	0.82
MOV02049.avi	0.88	0.81
MOV02048.avi	0.83	0.82

Quantitative Evaluation (F-Measure)

V. CONCLUSION

Motion segmentation is an important in video indexing, traffic monitoring security. There are many challenging problems in studying real traffic scenes within a complex background. In this work, efficient graph cut based region merging algorithm techniques are applied to extract high quality motion detection from an input video. This research is divided into segment the moving object from the static background, motion segmentation is done by updating background subtraction. This algorithm can generate segmentation results with low computation complexity and high efficiency compare to other change detection based video segmentation algorithm. Experimental results indicated that the proposed algorithm is simple and effective in segmenting moving objects. Since the background update was performed only in the changed areas where the moving objects occurred too frequently, the computational load is reduced significantly. Moreover, the proposed methods are based on general scenes, so it is suitable for other surveillance and real sequence.

REFERENCES

[1] Xiaowei Zhou, Can Yang, and Weichuan Yu," Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation" IEEE Trans. On Pattern Analysis And Machine Intelligence, Vol. 35, No. 3, March 2013.

[2] A.Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," ACM Computing Surveys, vol. 38, no. 4, pp. 1-45, 2006.

[3] T. Moeslund, A. Hilton, and V. Kruger, "A Survey of Advances in Vision-Based Human Motion Capture and Analysis," Computer Vision and Image Understanding, vol. 104, nos. 2/3, pp. 90-126, 2006.

[4] C. Papageorgiou, M. Oren, and T. Poggio, "A General Framework for Object Detection," Proc. IEEE Int'l Conf. Computer Vision, p. 555, 1998.

[5] P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," Int'l J. Computer Vision, vol. 63, no. 2, pp. 153-161, 2005.

[6] H. Grabner and H. Bischof, "On-Line Boosting and Vision," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp. 260-267, 2006.

[7] B. Babenko, M.-H. Yang, and S. Belongie, "Robust Object Tracking with Online Multiple Instance Learning," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1619- 1632, Aug. 2011.

[8] M. Piccardi, "Background Subtraction Techniques: A Review," Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics, 2004.

[9] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and Practice of Background Maintenance," Proc. IEEE Int'l Conf. Computer Vision, 1999.

[10] R. Vidal and Y. Ma, "A Unified Algebraic Approach to 2-D and 3-D Motion Segmentation," Proc. European Conf. Computer Vision, 2004.

[11] D. Cremers and S. Soatto, "Motion Competition: A Variational Approach to Piecewise Parametric Motion Segmentation," Int'l J. Computer Vision, vol. 62, no. 3, pp. 249-265, 2005.

[12] D. Gutchess, M. Trajkovics, E. Cohen-Solal, D. Lyons, and A. Jain, "A Background Model Initialization Algorithm for Video Surveillance," Proc. IEEE Int'l Conf. Computer Vision, 2001.

[13] V. Nair and J. Clark, "An Unsupervised, Online Learning Framework for Moving Object Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 317-324, 2004.

[14] N.Oliver, B.Rosario and A.Pentland,"A Bayesian Computer Vision System for Modeling Human Interactions",IEEE Trans.Pattern Analysis and Machine Intelligence vol 22,no.8,pp.831-843,Aug, 2000.

[15] Zhayida Simayijiang, Stefaine Grimm,"Segmentation with Graph Cuts".

[16] M. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," Computer Vision and Image Understanding, vol. 63, no. 1, pp. 75-104, 1996.

[17] T. Brox and J. Malik, "Object Segmentation by Long Term Analysis of Point Trajectories," Proc. European Conf. Computer Vision.2010.

[18] G.Doretto, D.Cremers, P.Favaro and S.Soatto," Dynamic Texture Segmentation", Proc. IEEE Int'l Conf. Computer Vision 2003.