A Survey on Moving Object Detection

Roopa Gokul, Dr. J. C. Prasad

Abstract— **Object detection and tracking are the key steps for automated video analysis. Object detection in a video is usually performed by object detectors, background subtraction techniques or motion based methods. Object detectors are binary classifiers it classifies the sub images into object or background. It requires manually labeled examples to train a binary classifier. Background subtraction technique needs a training sequence that contains no objects to build a background model. Then it compares images with a background model and detects the changes as objects. Motion-based methods avoid training phases and use motion information to separate objects from the background. In this work various object detection techniques introduced so far and their merits and demerits are studied.**

Index Terms—**Foreground support, low rank modeling, object detection, background subtraction, background model, motion based method.**

I. INTRODUCTION

Automated video analysis is important for many applications, such as video surveillance, vehicle navigation, traffic monitoring etc. There are three steps for automated video analysis: object detection, tracking, and behavior recognition. Object detection aims to locate and segment objects in a video. Such objects can be tracked from frame to frame, and such tracks can be analyzed which would subsequently help in human activity analysis. Object detection is usually performed by object detectors, background subtraction techniques or motion based methods. Object detectors are binary classifiers which classifies the sub images into object or background. Classifier can be built by offline learning on separate datasets, or by online learning initialized with a manually labeled frame at the start of a video. Background subtraction technique on the other hand compares images with a background model and detects the changes as objects. This method usually assumes that no object appears in images when building the background model, this requirements actually limit the applicability of above mentioned methods in automated video analysis. Motion-based techniques avoid training phases and use motion information to separate objects from the background. The moving objects are present in the scene, and due to camera motion the background may also move. Motion segmentation which falls into the category of motion based method classifies pixels according to motion patterns. These approaches achieve both optical flow computation and segmentation accurately and it works in the presence of large camera motion. But they assume rigid motion or smooth motion. But in practice, the foreground motion can be very complicated. Also, the background may have illumination changes and changing textures such as sea waves and waving trees. Another motion-based approach is background estimation. It estimates a background model directly from the testing sequence. It tries to find temporal intervals within which the pixel intensity is unchanged and then uses image data from such intervals for the purpose of background estimation. However, this approach relies on the assumption of static background. So it is difficult to perform object detection in the videos captured using moving cameras.

In this paper we present a study on various object detection techniques. The survey focuses on three main areas such as object detectors, background subtraction and motion segmentation.

Manuscript received Nov, 2014.

***Roopa Gokul**, MTech Computer Science and Information System, FISAT, Mookkannoor, Ankamaly-683577, Kerala, India*

***Dr. J. C. Prasad**, Head of the Department, Computer Science and Engineering, FISAT, Mookkannoor, Ankamaly-683577, Kerala, India*

II. LITERATURE SURVEY

This study focuses on various object detection techniques including object detectors, motion segmentation and background subtraction.

A. Object Detectors

An object detector is often a classifier that scans the image by a sliding window and labels each subimage defined by the window as either object or background. The classifier can be built by offline learning or online learning. Offline learning can be done on separate datasets [1], [2] and online learning can be initialized with a manually labeled frame at the start of a video [3], [4].

Constantine P. Papageorgiou et al proposed "A General Framework for Object Detection" [3] in 1998. In this work they describe a framework for object detection in static images of cluttered scenes. The detection technique is based on a wavelet representation of an object class derived from a statistical analysis of the class instances. By learning an object class in terms of a subset of an overcomplete dictionary of wavelet basis functions, a compact representation of an object class is derived which is used as an input to a support vector machine classifier. This approach overcomes both the problem of in-class variability and also provides a low false detection rate in unconstrained environments. The strength of this system comes from the expressive power of the overcomplete set of basis functions. This representation effectively encodes the intensity relationships of certain pattern regions that define a complex object class. The encouraging results of the system in two different domains, faces and people, suggest that this approach well generalize to several other object detection tasks.

ISSN: 2278 – 1323

Michael J. Jones et al proposed "Detecting Pedestrians Using Patterns of Motion and Appearance" [4]. In this work they describe a pedestrian detection system that integrates image intensity information with motion information. This approach use a detection style algorithm that scans a detector over two consecutive frames of a video sequence. The detector is trained (using AdaBoost) to take advantage of both motion and appearance information to detect a walking person. Previous approaches have built detectors based on motion information or detectors based on appearance information, but this approach is the first to combine both sources of information in a single detector. This work provides two contributions: (i) development of a representation of image motion which is extremely efficient, and (ii) implementation of a state of the art pedestrian detection system which operates on low resolution images under difficult conditions such as rain and snow.

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. It usually assumes that the video is captured by a static camera. Thus, foreground objects can be detected by checking the difference between the testing frame and the background model built previously. Some background subtraction techniques are discussed below.

Christopher Wren et al proposed "Pfinder: Real-Time Tracking of the Human Body" in 1996 [5]. Pfinder is a real-time system which heips in tracking and interpretation of people. 1st it Builds the scene model by observing the scene without people in it. Then when a human enters the scene it build up a model of that person. Initially the person model is built by detecting a large change in the scene, and then building up a multi-blob model of the user over time. This process is driven by the distribution of color on the person's body, and blobs provide account for each differently-colored region. Separate blobs are required for the person's hands, head, Feet. Pfinder uses a 2D contour shape analysis that attempts to identify the head, hands, and feet locations. Finally the features produced by the blob models and the contour analyzer are integrated. This is more accurate and more general method. The deletion and addition of blobs makes Pfinder very robust to occlusions and strong shadows. This method employs several domain-specific assumptions which degrade the performance. Also it cannot compensate for large, sudden changes in the scene. Another problem is that system expects only one user to be in the space. Multiple users cause significant difficulties with the primitive gesture recognition system included in Pfinder.

Chris Stauffer et al proposed "Adaptive background mixture models for real-time tracking" in 1996[6]. This system models each pixel as a mixture of Gaussians and then uses an on-line approximation to update the model. Pixel values that do not fit the background distributions are considered foreground. Each pixel is classified based on whether the Gaussian distribution which represents it most effectively is considered part of the background model. It is flexible enough to handle repetitive motions of scene elements, introducing or removing objects from the scene, tracking through cluttered regions, lighting changes and slow-moving objects. But performance degrades when the scene become dynamic like waves , moving clouds etc.

J. Rittscher et al developed "A Probabilistic Background Model for Tracking" in 2000 [7]. This system is based on the concept of Hidden Markov Model. The hidden states enable the discrimination between foreground, background and shadow. Probabilistic trackers based on a particle filters are used which can be extended to tracking multiple objects. A novel observation density for the particle filter which models the statistical dependence of neighboring pixels based on a MRF is used. It is no longer necessary to select training data. It is Capable of modelling shadow, foreground & background. But the illumination changes throughout the day affect the updation of background model.

N. Oliver et al proposed "A Bayesian Computer Vision System for Modeling Human Interactions" in 2000 [8]. This approach helps in modeling and recognizing human behaviors. Use supervised statistical learning techniques to recognize normal single person behaviors and common person to person interactions. It provides Bayesian integration of prior knowledge with evidence from data. The Combined use of priori models Hidden Markov Model and Coupled Hidden Markov Model helps to increase accuracies of recognition. The real time computer vision input module detects and tracks moving objects in the scene and for each moving object outputs a feature vector describing its motion and heading and its spatial relationship to all nearby moving objects. These feature vectors constitute the input to stochastic state based behavior models. Both HMMs and CHMMs with varying structures depending on the complexity of the behavior are then used for classifying the perceived behaviors. It can identify novel behaviors. The Combined use of priori models HMM & CHMM increases accuracies of recognition. But output will be affected by the presence of dynamic textures and illumination changes.

Hanzi Wang et al proposed "A Novel Robust Statistical Method for Background Initialization and Visual Surveillance" in 2006 [9]. This method can be used in the places where foreground objects can not be avoided in the training stage. First it will locate all non-overlapping stable subsequences of pixel values. Then choose the most reliable subsequence among them. Later the mean value of either the grey-level intensities or the color intensities over that subsequence are used to model background value. All previous methods require that the training sequence is free of any foreground objects. It can tolerate over 50% of noise in the data. But it does not work with dynamic background.

Junzhou Huang et al proposed "Learning with Dynamic Group Sparsity" in 2009 [10]. It provides an extension of the standard sparsity concept in compressive sensing. A new greedy sparse recovery algorithm which prunes data residues in the iterative process according to both sparsity and group clustering priors rather than only sparsity as in previous methods is used here. The group clustering concept specifies that if a point lives in the union of subspaces, its neighboring points would also live in this union of subspaces with higher probability, and vice versa. It provides a generalized framework for priors-driven sparse data recovery algorithms. It can perform sparse recovery, multi-task sparse recovery,

ISSN: 2278 – 1323

group/block sparse recovery, DGS recovery, and adaptive DGS recovery, respectively. 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 obtain the correlation between video frames. It can handle illumination changes and dynamic textures. This approach has higher accuracy and lower computational complexity. It decreases the minimal number of necessary measurements. It improves robustness to noise and prevent the recovered data from having artifacts. But it requires that the training sequence is free of any foreground objects.

C. Motion Segmentation

In motion segmentation, the moving objects are continuously present in the scene and due to camera motion the background may also move. Our aim is to separate different motions.

Daniel Cremers et al proposed "Motion Competition: A Variational Approach to Piecewise Parametric Motion Segmentation" in 2005 [11]**.** It provides a framework for segmenting the image plane into a set of regions of parametric motion. It uses an explicit spline-based implementation which can be applied to the motion-based tracking of a single moving object, and an implicit multiphase level set implementation which allows for the segmentation of an arbitrary number of multiply connected moving objects. In this method all normalizations in the equation are derived in a consistent manner. The level set formulation permits the segmentation of several multiply connected objects. This approach is based on the assumption of small motion. And it is also assumed that objects do not change their brightness throughout time. But the problem is that it is not well-suited to deal with new objects entering the scene. Computational complexity is high for this method. Another problem is that segmentation of images in terms of piecewise parametric motion are not applicable in several cases.

Antoni B. Chan et al proposed "Layered Dynamic Textures" in 2009[12]. This approach represents the video as a collection of stochastic layers of different dynamics and appearance. Each layer is modeled as a temporal texture sampled from a different linear dynamical system. It includes a collection of hidden layer assignment variables which control the assignment of pixels to layers and a Markov random field prior on these variables which encourages smooth segmentations. An EM algorithm is derived for the estimation of the model parameters from a training video. It avoids boundary uncertainty. It can effectively segmenting real video sequences depicting different classes of scenes like various types of crowds, highway traffic, Scenes containing a combination of globally homogeneous motion and highly stochastic motion (e.g., rotating windmills plus waving tree branches, or whirlpools). But Aperture problem, occlusion, video noises, etc. affects the result.

Sandor Fazekas et al proposed "Dynamic texture detection based on motion analysis" in 2009 [13]. Motion estimation is usually based on the brightness constancy assumption. But this would work for rigid objects not for fluids and gas. So this approach examines three possible alternatives namely color constancy, gradient constancy and brightness conservation. Accurate segmentation into regions of static and dynamic texture is achieved using a level set scheme. It separates each image into regions that obey brightness constancy and regions that obey the alternative assumption. Use of a threshold value in real time application limits the robustness of this method.

Yaser Sheikh et al proposed "Background Subtraction for Freely Moving Cameras" in 2009 [14]. This approach applies to videos captured from a freely moving camera. It segments the objects by analyzing point trajectories. All trajectories corresponding to static areas in the scene lie in a 3Dsubspace. RANSAC (Random sample Consensus) algorithm is used to estimate the background trajectory basis using this rank constraint, and to classify trajectories as background or foreground. Then these trajectories are used to build background and foreground appearance models. It is based on 2 assumptions 1) An orthographic camera model is used, 2) The background is the "rigid" entity in the image. Problem is that use of an affine camera model over a more accurate perspective camera model.

Thomas Brox et al proposed "Object Segmentation by Long Term Analysis of Point Trajectories" in 2010 [15]. Long term point trajectories based on dense optical flow are used for object detection. Given the pairwise distances between trajectories, we can build an affinity matrix for the whole shot and run spectral clustering on this affinity matrix which results in temporally consistent segmentations of moving objects in a video shot. This method can deal with the large motion of limbs or the background on the other. Occlusion and disocclusion is naturally handled by this approach. This is more general than most previous techniques. But the problem is that it require point trajectories as input and only output a segmentation of sparse points. The performance depends on the quality of point tracking. Post processing is needed to obtain the dense segmentation. They are limited when dealing with noisy data and nonrigid motion. Spectral clustering of dense point trajectories is too slow.

Peter Ochs et al proposed "Object Segmentation in Video: A Hierarchical Variational Approach for Turning Point Trajectories into Dense Regions" in 2011 [16]. It is the first continuous hierarchical model. Method to obtain dense segmentations from such sparse trajectory clusters. In homogeneous areas of the image there are no such structures, This results in point trajectories to be sparse. Main idea is to do segmentation based on color in homogeneous areas. It is a Hierarchical variational model where we have continuous labeling functions on multiple levels. Each level corresponds to a super pixel partitioning at a certain granularity level. It has additional auxiliary functions at coarser levels which are optimized in a coupled diffusion process. Structure-aware label propagation can be obtained by this approach. Final solution and is free of metrication errors or block artifacts. Also motion in homogeneous areas can be estimated accurately. But the problem is that the performance relies on the quality of point tracking. Also they are limited when dealing with noisy data and nonrigid motion.

Xiaowei Zhou et al proposed "Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation" [17] in 2013. In this approach the correlated background images are represented as a low rank matrix and moving objects are detected as outliers in the low rank representation. This allows us to get rid of many assumptions on the behavior of foreground. The low-rank representation of background makes helps to accommodate the global

ISSN: 2278 – 1323

variations in the background. Also, DECOLOR performs object detection and background estimation simultaneously without training sequences. DECOLOR has some disadvantages it may misclassify unmoved objects or large texture less regions as background since they are prone to entering the low-rank model. By incorporating additional models such as object appearance or shape, this problem can be solved to some extend. DECOLOR works in a batch mode so it is not suitable for real-time object detection.

III. CONCLUSION

Different object techniques were discussed. Mainly three areas were studied: object detectors, background subtraction and motion segmentation. It was found that motion segmentation technique provides best results compared to other two methods. The recent motion based algorithm DECOLOR [17] introduced by Xiaowei Zhou et al found to provide best result compared to preexisting methods. It avoids complex motion computation by implementing the problem as outlier detection and it makes use of the low-rank modeling to handle complex background. Thus it helps to get rid of many assumptions on the behavior of foreground. The low-rank representation of background helps to accommodate the global variations in the background. Also DECOLOR performs object detection and background estimation simultaneously without training sequences.

REFERENCES

[1] Papageorgiou, M. Oren, and T. Poggio, "A General Framework for Object Detection," Proc. IEEE Int'l Conf. Computer Vision, p. 555, 1998.
[2] 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.
[3] H. Grabner and H. Bischof, "On-Line Boosting and Vision," Proc.IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp. 260-267, 2006.
[4] 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.
[5] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland,"Pfinder: Real-Time Tracking of the Human Body," IEEE Trans.Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780-785,July 1997.
[6] Stauffer and W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1999.
[7] J. Rittscher, J. Kato, S. Joga, and A. Blake, "A Probabilistic Background Model for Tracking," Proc. European Conf. Computer Vision, 2000.
[8] 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.
[9] H. Wang and D. Suter, "A Novel Robust Statistical Method for Background Initialization and Visual Surveillance," Proc. AsianConf.Computer Vision, 2006.
[10] J.Huang, X. Huang, and D. Metaxas, "Learning with Dynamic Group Sparsity," Proc.IEEE Int'l Conf. Computer Vision, 2009.
[11] A.Chan and N. Vasconcelos, "Layered Dynamic Textures," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 10, pp. 1862-1879, Oct. 2009.
[12] S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati, "Dynamic Texture Detection Based on Motion Analysis," Int'l J. Computer Vision, vol. 82, no. 1, pp. 48-63, 2009.
[13] Y. Sheikh, O. Javed, and T. Kanade, "Background Subtraction for Freely Moving Cameras," Proc. IEEE Int'l Conf. Computer Vision, 2009.
[14] T. Brox and J. Malik, "Object Segmentation by Long Term Analysis of Point Trajectories," Proc. European Conf. Computer Vision, 2010.
[15] R. Vidal, "Subspace Clustering," IEEE Signal Processing Magazine, vol. 28, no. 2, pp. 52-68, Mar. 2011.
[16] P. Ochs and T. Brox, "Object Segmentation in Video: A Hierarchical Variational Approach for Turning Point Trajectories Into Dense Regions," Proc. IEEE Int'l Conf. Computer Vision, 2011.
[17] Xiaowei Zhou; Can Yang; Weichuan Yu, "Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.35, no.3, pp.597,610, March 2013

Roopa Gokul, BTech in Computer Science and Engineering from Viswajyothi college of Engineering and Technology, Vazhakulam. Currently doing MTech at Federal Institute of Science and Technology, Mookkannoor. Her research interests are Image Processing and Database Management System. Have published 5 papers in international journals.

Dr.J.C.Prasad is presently Associate Professor and Head of the CSE Department, FISAT. He was graduated in Mathematics from University of Calicut in 1998, had his post graduation in Computer Applications from Bharathiar University in the year 2001, and gained second post graduation in M.Tech and Ph.D in Computer Science and Engineering from Dr.M.G.R.University in the year 2006 and 2012 respectively. He has published 8 research papers in the International Journals and 16 papers in National and International Conferences.

ISSN: 2278 – 1323