iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** V **Month of publication:** May 2014

DOI:

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

A Survey on Techniques used for Detection and Tracking of Video Objects

Sukanyathara J[#1]

[#1]*Computer Science, Viswajyothi College of Engineering & Technology, Kerala, India*

Abstract— Segmentation and tracking are two important aspects in visual surveillance systems. Many barriers such as cluttered background, camera movements, and occlusion make the robust detection and tracking a difficult problem, especially in case of multiple moving objects. Object detection in the presence of camera noise and with variable or unfavourable luminance conditions is still an area of active research. In this paper, a survey of various techniques or methods that are used to segment, detect and track objects in the surveillance videos with stationary and complex backgrounds, crowded area, multi-modality background, occluded object, and deformable based objects is provided. For handling complex backgrounds, multi-background registration based segmentation is available. Various techniques used for segmentation based on frame differencing and background modelling are included.

Keywords— Surveillance, segmentation, multi-background registration, threshold decision, energy minimization, tracking.

I. INTRODUCTION

Object detection is the task of finding and identifying objects in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes / scale or even when they are translated or rotated. It is commonly used in applications such as security, surveillance, and automated vehicle parking systems.

Tracking, generally finds its use in videos. It is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging and video editing. It also estimates the trajectory of an object in an image plane as it moves around a scene.

Automated systems which help security personals to detect and track people are gaining importance. Video Surveillance allows us to remotely monitor a live or recorded video feed which often includes people. There has been a tremendous proliferation of video surveillance cameras in public locations such as stores, ATMs, highways, traffic signals, schools, buses, subway stations, and airports in order to detect and track the moving objects. Detection and tracking of moving objects is very important to monitor public transportation, and critical assets.

The previous methods for detection and tracking based approaches for video object tracking are unreliable in complex surveillance videos due to problems like occlusions, lighting changes, and other factors. Segmentation based on single background model is suitable for only stationary backgrounds. There are several methods which includes frame differencing and background modelling for the segmentation and detection of moving objects.

This paper provides the different techniques that are being used for segmentation and tracking and analyse their advantages and disadvantages. An object segmentation and tracking method using multi-background registration and particle filter is given in [1]. A background registration based segmentation method is used in [2] which is suitable only for stationary backgrounds. In [3], a method which usesjoint registration and active contour segmentation is discussed. A method for segmenting video objects with background registration is discussed in [4]. These techniques have its own merits and limitations.

Researchers are still working hard for creating the most robust detection and trackingtechniques for visual

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

surveillance. Reference [5] shows a multi-modal Gaussian MixtureModel based method for segmentationin video sequences. Multi-target tracking methodsare provided in [6]. It makes use of particle filters to track multiple objects in a video or an image sequence. Tracking in a compressed domain is dealt in [7].

II. VIDEO OBJECT SEGMENTATION AND TRACKING FRAMEWORK WITH IMPROVED THRESHOLD DECISION AND DIFFUSION DISTANCE

In this work [1], a foreground mask is constructed by segmentation using multi-background registration. Then, a single object tracking is implemented using particle filter. The use of Diffusion distance makes it more accurate for tracking non-rigid object motion under drastic changes of illumination.

There are four parts in the segmentation algorithm: Multi-Background Registration, Background Update and Release, Object Detection and Post-Processing. In the dynamic background region, such as waving trees, the observed pixel value of each pixel might be up to N values, which show up randomly. For this reason, it assumes each pixel can have up to N candidate background values at the corresponding positions of background images.

In the environments under surveillance monitoring, backgrounds are changed as environment changes and old backgrounds may no longer be active. For this reason, we need to update the values in background images and release the background image locations occupied by non-active background values. The method of updating the background image values and releasing the non-active background values are done in Background update and release blocks. The method of detecting foreground object mask and de-noising the initial object mask are done in Object detection and Post-processing respectively.

Before measuring the parameters of camera noise, the background of the input frame should be automatically indicated because it is very difficult to correctly measure the camera noise from the foreground. The Gaussianity test which indicates whether a group of values is Gaussian distributed or not can be used to find the background.The frame is divided into a number of non-overlapped $M_b * N_b$ blocks. The Gaussianity test is then applied to these blocks to determine if the minimal background differences in the block are Gaussian distributed or not.

The most important contribution of this framework is that it makes the use of Diffusion Distance (DD) as a cross-bin colour histogram distance function to improve the tracking performance, especially when the object being tracked has appearance variations due to changes in illumination. The outputs of particle filter are the updated object states. Those object masks that have large area but are not included by any trackers belong to new incoming objects. The trackers for these new objects are initialized with the object masks and added to the object state buffer.

In video object tracking, different kinds of features, such as colour, texture, and gradient, can be utilized to measure the importance of particles, where the distance functions, which are defined for those features, between the tracked object of the previous frame and each assumption (particle) in the current frame are calculated. For tracking non-rigid objects, a colour histogram is usually utilized since the stability of a colour histogram is better than that of other features under the non-rigid motion. However, a colour histogram suffers from appearance variations due to changes in illumination. Therefore, a histogram distance calculation scheme that is robust to changes of illumination is critical for object tracking.

III. JOINT REGISTRATION AND ACTIVE CONTOUR SEGMENTATION FOR OBJECT TRACKING

In this work [2], a novel object tracking framework is presented which uses joint registration and active contour segmentation (JRACS), which can robustly deal with the non-rigid shape changes of the target. The target region, which includes both foreground and background pixels, is implicitly represented by a level set.

A Bhattacharyya similarity based metric is proposed to locate the region whose foreground and background distributions best match those of the tracked target. Based on this metric, a tracking framework that consists of a registration stage and a segmentation stage is then established. The registration step roughly locates the target object by modelling its motion as an affine transformation, and the segmentation step refines the registration result and computes the true contour of the target.

To track robustly the target in the video sequence, it is important to represent the target as robustly as possible. The regionwhere the target locates can be represented as an ellipse, arectangle, or an arbitrary contour. Once the region is fixed,there are various features that can be used to describe

ISSN: 2321-9653

thetarget defined in it, such as colour, edge, and texture features,or the combination of them. Among these, colourhistogram is selected to model the target because of its merits, such asindependence of scaling and rotation, robustness to partialocclusions, low computational cost, etc. Of course, whenobject and background differ much from each other on texturefeatures, the texture histogram can also be used to representthe target.

In joint registration and segmentation method, the Bhattacharyya similarity isadopted for target matching. Based on this metric, a registration formula and a segmentation formula are derived. Theregistration formula is used to estimate the affine deformationof the target, and the segmentation formula is used to refinethe registration results so that the contour of the target can beobtained.

The method of joint registration seamlessly handles the topological changes of the target. It also overcomes some limitations of previous works, including EM-shift tracker, SOAMST tracker.

IV. Efficient Moving Object Segmentation Algorithm Using Background Registration Technique

In this method [3], an efficient moving object segmentation algorithm suitable for real-time content-based multimedia communication systems is introduced. First, a background registration technique is used to construct a reliable background image from the accumulated frame difference information. The moving object region is then separated from the background region by comparing the current frame with the constructed background image.

Finally, a post-processing step is applied on the obtained object mask to remove noise regions and to smooth the object boundary. In situations where object shadows appear in the background region, a pre-processing gradient filter is applied on the input image to reduce the shadow effect. In order to meet the real-time requirement, no computationally intensive operation is included in this method.

The algorithm is divided into five major steps. The first step is to calculate the frame difference mask by thresholding the difference between two consecutive input frames.

Then, according to the frame difference mask of past several frames, pixels which are not moving for a long time are considered as reliable background in the background registration step. This step maintains an up-to-date background buffer as well as a back ground registration mask indicating whether the background information of a pixel is available or not.

By the third step, the background difference mask is generated by comparing the current input image and the background image stored in the background buffer. This background difference mask is our primary information for object shape generation.

In the fourth step, an initial object mask is constructed from the background difference mask and the frame difference mask. If the background registration mask indicates that the background information of a pixel is available, the background difference mask is used as the initial object mask. Otherwise, the value in the frame difference mask is copied to the object mask.

The initial object mask generated in the fourth step has some noise regions because of irregular object motion and camera noise. Also, the boundary region may not be very smooth. In the last step, these noise regions are removed and the initial object mask is filtered to obtain the final object mask. This is suitable for real-time applications, but it does not support dynamic backgrounds.

V. Real-Time Memory-Efficient Video Object Segmentation in Dynamic Background with Multi-Background Registration Technique

Background subtraction video segmentation is the important first step for video surveillance applications with fixed camera. There are many existing, methods in the literature. However, most of them are either too simple to handle complex environment, suchas dynamic background, or too complex to be executed in real-time. Based on the proposed multi-background registration technique, this work presents a real-time video objectsegmentation algorithm. The proposed algorithm can better handle non static backgroundcases compared with the original single background registration segmentation. Comparedwith other works that can handle dynamic background cases, it is efficient in memoryusage.

It provides a method for segmenting and tracking vehicles on highways using a camera that is relatively low to the ground [4]. Among the many technologies, vision-basedsystems are emerging as an attractive alternative due to

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

their ease of installation, inexpensive maintenance,and ability to capture a rich description of the scene. Here,an automatic method for segmenting and tracking vehicles in video taken by a camerabeside the road at a height of approximately 9 m, where occlusion is a serious issue isproposed.The approach is based on sparse features that are incrementally tracked throughout avideo sequence using local gradientbased search techniques. Background subtraction is apopular technique used by many vehicle- tracking systems to detect and track vehicleswhen they are well separated in the image.

Feature points are automatically detected and tracked through the video sequence,and features lying on the background or on shadows are removed by background subtraction, leaving only features on the moving vehicles. These features are then separated into two categories. Stable and unstable. The 3-D coordinates of the stable features arecomputed, these stable features are grouped together to provide a segmentation of thevehicles, and the unstable features are then assigned to these groups. The final step involves eliminating groups that do not appear to be vehicles, establishing correspondencebetween groups that are detected in different image frames to achieve long term tracking.

It uses background subtraction for separating the foreground and the background. Thebackground of the scene is learned by storing the average grey level of each pixel over a fixed period of time.The technique not only segments objects of interest but also ignores dynamic backgrounds. It is more efficient than single background registration methods, and also efficient in memory usage.

VI. MULTI-MODAL BACKGROUND SUBTRACTION USING GAUSSIAN MIXTURE MODELS

Background subtraction is a common first step in the field of video processing and it is used to reduce the effective image size in subsequent processing steps by segmenting themostly static background from the moving or changing foreground. This method [5] extendsthe previous approaches towards background modelling to handle videos accompanied byinformation gained from a novel 2D/3D camera. This camera contains a colour and aPMD chip which operates on the Time-of-Flight operating principle.

The background isestimated using the widely spread Gaussian mixture model in colour as well as in depthand amplitude modulation. A new matching function is presented that allows for bettertreatment of shadows and noise and reduces block artefacts. Problems and limitationsto overcome the problem of fusing high resolution colour information with low resolutiondepth data are addressed and the approach is tested with different parameters on severalscenes and the results are compared to common and widely accepted methods.

The ordinary colour based GMM background subtraction cannot distinguish betweenforeground and background when the colour difference is small due to its pixel based nature.The depth values gained from a ToF camera provide the ability for a correct classificationof all image blocks with depth values different from those of the background as long asthere are valid depth measurements for the background.

Classificationbased only on low resolution depth values will result in an unnaturalforeground contour due to block artefacts. Practically, this drawback cannot be resolvedin typical situations, because in such areas the background usually continues with thesame colour so that there is no edge that would allow gradient based methods to smooththe contour of the foreground mask correctly.

To resolve the general case contour estimation methods that incorporate knowledgeof the object given a priori or learned through time are necessary, but it does not seem tobe possible to achieve good results in a not strictly defined setting. Complexity is not high. It provides favourable segmentation with stationery backgrounds.

VII. MULTI-TARGET TRACKING OF PEDESTRIANS IN VIDEO SEQUENCES BASED ON PARTICLE FILTERS

Video target tracking is a critical problem in the field of computer vision. Particle filters have been proven to be very useful in target tracking for nonlinear and non-Gaussian estimation problems. Although most existing algorithms are able to track targets well in controlled environments, it is often difficult to achieve automated and robust tracking of pedestrians in video sequences if there are various changes in target appearance or surrounding illumination.

To surmount these difficulties, this work [6] presents multi-target tracking of pedestrians in video sequences based on particle filters. In order to improve the efficiency and accuracy of the detection, the algorithm firstly obtains target regions in training frames by combining the methods of background subtraction and Histogram of Oriented Gradient (HOG) and then establishes discriminative appearance model by

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

generating patches and constructing codebooks using super pixel and Local Binary Pattern (LBP) features in those target regions.

During the process of tracking, the algorithm uses the similarity between candidates and codebooks as observation likelihood function and processes severe occlusion condition to prevent drift and loss phenomenon caused by target occlusion. This method provides good segmentation results, but the computational cost is high.

VIII. Video Object Tracking in the Compressed Domain Using Spatio-Temporal Markov Random Fields

This presents an approach to track a moving object in a H.264/AVC-compressedvideo. The only data from the compressed stream used in the proposed method are themotion vectors and block coding modes. As a result, the proposed method has a fairlylow processing time, yet still provides high accuracy.

After the pre-processing stage, which consists of intra-coded block motion approximation and global motion compensation, it employs a Spatio-Temporal Markov RandomField model to detect and track a moving target. Using this model, an estimate of thelabelling of the current frame is formed based on the previous frame labelling and currentmotion information.

The aim of this method [7] is to provide a framework for tracking moving objects incompressed domain based on MVs and associated block coding modes alone. The objectof interest is selected by the user in the first frame, and then tracked through subsequentframes.

In each frame, the proposed method first approximates MVs of intra-coded blocks,estimates global motion (GM) parameters, and removes GM from the MV field. Theestimated GM parameters are used to initialize a rough position of the object by projecting its previous position into the current frame. Eventually, the procedure of IteratedConditional Modes (ICM) updates and refines the predicted position according to spatialand temporal coherence under the MAP criterion, defined by the ST-MRF model. It takes fairly low processing time, still provides accuracy. It is unable to distinguish between foreground and background when the colourdifference is small due to its pixel based nature.

IX. Conclusions

Detection and tracking in video based surveillance systems, is still an area of active research. There are several techniques related with tracking objects in a video. A survey of the existing methods used for detection and tracking in surveillance videos is provided in this work. The previous methods proposed for detection and tracking has several shortcomings. For the meaningful video manipulation, videos are to be segmented into meaningful objects (semantic video objects). Many applications require automatic segmentation of semantic video objects. But it is considered as difficult task. The real-time usability of the segmentation and tracking algorithms must be considered as an important parameter in selecting the algorithm. The advantages and disadvantages of several methods have been discussed.

Acknowledgment

The author would like to thank the Department Head, Project Guide and Group Tutor, for their constructive comments and informative suggestions that have helped to improve this paper.

References

[1] Shao-Yi Chien, Wei-Kai Chan, Yu-Hsiang Tseng, and Hong-Yuh Chen, "Video Object Segmentation and Tracking Framework with Improved Threshold Decision and Diffusion Distance,"*IEEE Transactions On Circuits and Systems for VideoTechnology*, vol. 23, no. 6, June 2013.

[2] JifengNing, Lei Zhang, David Zhang, and Wei Yu, "Joint Registration and Active Contour Segmentation for Object Tracking," *IEEE Transactions On Circuits and Systems For Video Technology*, vol. 23, no. 9, Sep. 2013.

[3] Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen, "Efficient Moving Object Segmentation Algorithm Using Background Registration Technique," *IEEE TransactionsOn Circuits And Systems For Video Technology*, vol. 12, No. 7, July 2002.

[4] W.-K. Chan and S.-Y. Chien, "Real-time memory-efficient video object segmentation in dynamic background with multi-background registration technique", *IEEE Multimedia Signal Process., Workshop*, Oct. 2007.

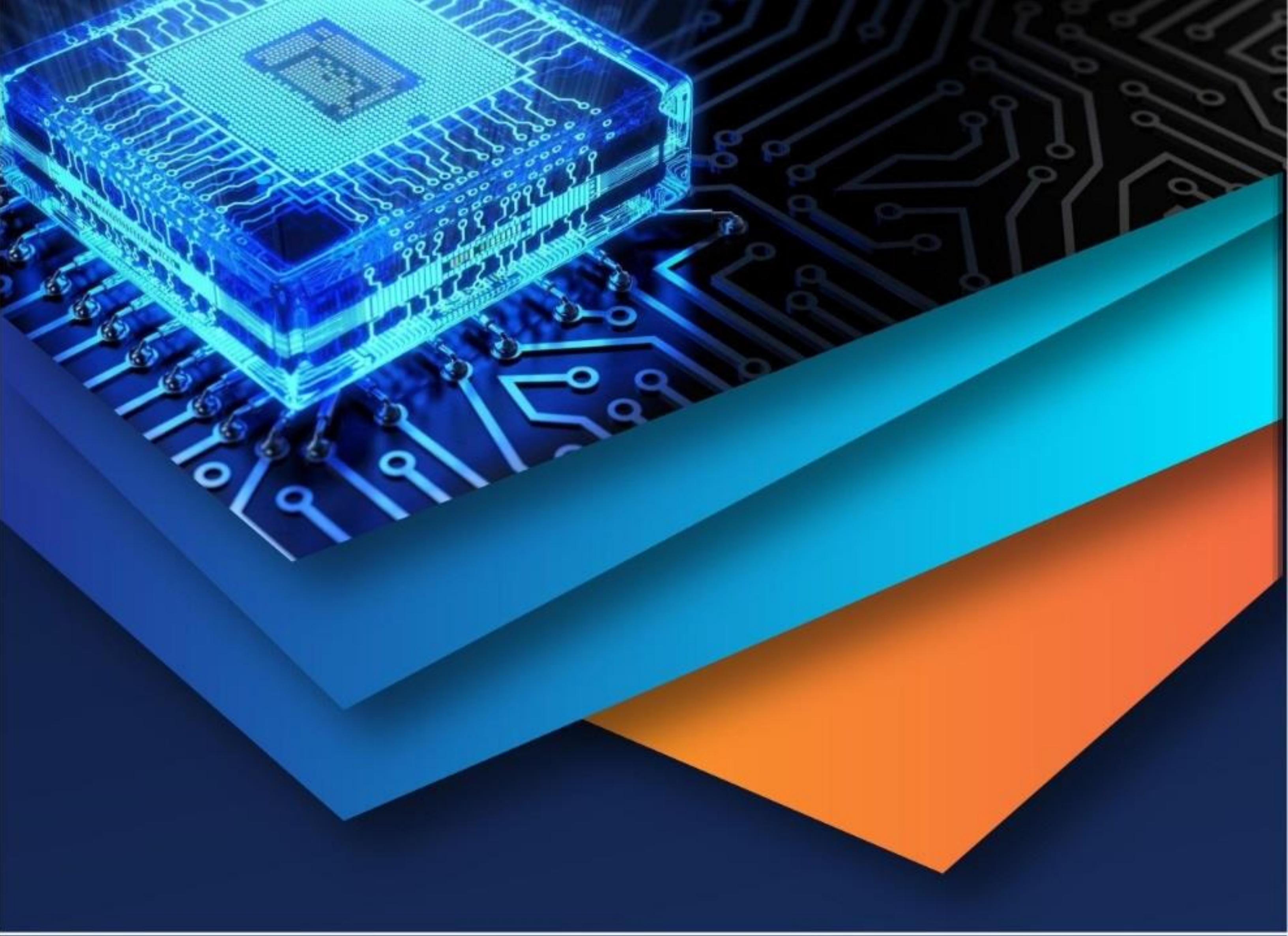
ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

[5] Benjamin Langmann, Seyed E. Ghobadi, Klaus Hartmann, OtmarLoffeld, "Multi-Modal Background Subtraction Using Gaussian Mixture Models," *IAPRS, Vol.XXXVIII, Part 3A,* Saint-Mandé, France, September 1-3, 2010.

[6] Hui Li, ShengwuXiong, PengfeiDuan, and Xiangzhen Kong, "Multitarget Tracking of Pedestrians in Video Sequences Based on Particle Filters," *Hindawi Publishing Corporation Advances in Multimedia* vol. 2012.

[7] SayedHosseinKhatoonabadi, Ivan V. Baji, "Video Object Tracking in the Compressed Domain Using Spatio-Temporal Markov Random Fields ," IEEE Transactions On Image Processing, Vol. 22, No. 1, January 2013.



10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY