International Journal For Technological Research In Engineering
Volume 2, Issue 8, April-2015 *ISSN (Online): 2347 - 4718*

A COMPREHENSIVE SURVEY ON MOVING OBJECT SEGMENTATION METHODS

Kinan B. Panchsheria[1], Prof. Mehul C. Parikh[2]
CS&E Deparatment, Government Engineering College, Modasa, Arvalli, Gujarat, India

Abstract: Motion detection, the process which segments moving objects in video streams, is a critical task for many computer vision applications. Complete and accurate motion detection in dynamic scenes, such as those containing object such as rippling water, moving curtains, spouting fountains and so on, is still a very difficult task. Moving object segmentation provides a classification of the pixels in the video sequence into either foreground or background. In this paper, we survey many existing schemes in literature for moving object segmentation. We also survey how to measure the performance of any moving object segmentation algorithms.

Index Terms: Background subtraction, dynamic background, foreground detection, neural network, performance evaluation, video surveillance.

I. INTRODUCTION

One of the important and critical tasks in many computer-vision applications is the segmentation of moving object. The accuracy of segmentation can significantly affect the overall performance of the application employing it. Segmentation is useful in applications, such as human activity understanding, traffic monitoring and analysis anomaly detection. Visual surveillance is a key technology for fight against terrorism and crime, public safety and efficient management of transport networks and public facilities [16]. Segmentation is the process of partitioning an image into semantically interpretable regions or divides an image into several meaningful parts based on several characteristics of image: color, shape, edge, and gray level, etc [10]. We can say also that, Image segmentation is the partition of an image into a set of no overlapping regions whose union is the entire image. The purpose of segmentation is to decompose the image into parts that are meaningful with respect to a particular application. Moving object segmentation provides a classification of the pixels in the video sequence into either foreground or background. A common approach used to achieve such classification is background subtraction [1-3], optical flow [2], [5], temporal difference [2], [5], visual background extractor [4]. Moving object segmentation in complex environments is still far from being completely solved. There are typically numbers of challenges such as illumination changes whether gradual changes (time of day) or sudden changes (light switch) [3], a foreground object might have similar characteristics as the background, it become difficult to distinguish between them (camouflage). A foreground object becomes motionless can't be distinguished from a background object that moves and then becomes motionless (sleeping person). Non-stationary background objects e.g. waving trees, and image changes due to camera motion which is common in outdoor applications. Some of these problems can be handled by very computationally expensive methods [3]. Now a day, the rapid development of internet and multimedia, the digital cameras and huge data store hardware cause the need to process the multimedia data such as video. The video object segmentation is a complicated problem and has an important key for essential multimedia applications such as video retrieval, modern video compression, and intelligent surveillance [7]. Video surveillance systems have been required to facilitate a wide range of science and technology applications in computer vision, including human activity understanding, traffic monitoring and analysis.

II. RELATED WORK

In the following section, we describe different stat-of-the-art methods for motion segmentation. Background subtraction, optical flow, Temporal difference, Codebook method, Mixture of Gaussian (MoG), Visual background extractor, Clustering-based method, RBF-Motion Detection (RBFMD), Cerebellar-model-articulation-controller-based motion detection. In general, foreground areas are selected in one of the two ways, pixel-by-pixel, where an independent decision is made for each pixel, and second region-based, where a decision is made on an entire group of spatially close pixels.

A. Background subtraction

Background subtraction [1-3] is a popular method for motion detection. Background subtraction is the process in which an image frame is compared to the background model in order to determine whether individual pixels are part of the background or the foreground (moving object). So it is also referred to as foreground detection. This mode of detection has attracted the most attention due to the ability of this approach to extract moving objects while exhibiting only moderate time complexity. 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. However, motion detection by background subtraction usually provides incomplete detection results of moving objects due to faulty background model generation.

B. Visual Background extractor

The ViBe (Visual Background extractor) [2], [4] detects moving objects by calculating the difference between the background model M and the incoming pixel$p(x, y)$.

Initially, the background model is initialized from the first frame. The tth background sample $M\mathrm{t}(x,y)$ is randomly chosen by N neighboring pixels in the 8-connected neighborhood of location(x,y). Second, a good match occurs when the Euclidean distance between $M\mathrm{t}(x,y)$ and $p(x,y)$ is lower than a predefined threshold R. If the number of occurrence of good matches is larger than or equal to the given threshold $\#min$, then the current pixel $p(x,y)$ is classified as background. Otherwise, $p(x,y)$ is regarded as a foreground pixel. Finally, if $p(x,y)$ is determined to be a background pixel, then two randomly chosen background samples – one at location (x,y) and the other at a location in the 8-connected neighborhood –are replaced by $p(x,y)$.

C. Mixture of Gaussians (MoG)

The background of the scene contains many non-static objects such as tree branches whose movement depends on the wind in the scene. This kind of background motion causes the pixel intensity values to vary significantly with motion. The generalized mixture of Gaussians (MoG) has been used to model complex, non-backgrounds [2]. Stauffer and Grimson [8] allow the background model to be a mixture of several Gaussians. Every pixel value is compared against the existing set of models at that location to find a match. The parameters for the matched model are updated based on a learning factor. If there is no match, the least-likely model is discarded and replaced by a new Gaussian with statistics initialized by the current pixel value. The models that account for some predefined fraction of the recent data are deemed "background" and the rest "foreground". However, the MoG has its own drawbacks. First, it is computationally intensive and its parameters require careful tuning. Second, it is very sensitive to sudden changes in global illumination [2].

D. Clustering –based

Butler et al. [13] propose a moving object segmentation algorithm with a similar premise to that of Stauffer and Grimson [8] but having the capability of processing 320 × 240 video in real-time on modest hardware. The premise of their algorithm is the more often a pixel takes a particular colour, the more likely that it belongs to the background. Therefore, this requires a technique for maintaining information regarding the history of pixel values. They model each pixel by a group of K clusters where each cluster consists of a weight and an average pixel value or centroid ck. Incoming pixels are compared against the corresponding cluster group. The matching cluster with the highest weight is sought and so the clusters are compared in order of decreasing weight. A matching cluster is defined to have a Manhattan distance (i.e. sum of absolute differences) between its centroid and the incoming pixel below a user prescribed threshold T. If no matching cluster is found, the cluster with the minimum weight is replaced by a new cluster having the incoming pixel as its centroid and a low initial weight. If a matching cluster is found, then the weights of all clusters in the group are adjusted. The centroid of the matching cluster must also be adjusted according to the incoming pixel. Previous approaches adjust the centroid based on a fraction of the difference between the centroid and the incoming pixel. Butler chooses instead to accumulate the error between the incoming pixel and the centroid. After adaptation, the weights of all clusters in the group are normalised so that they sum up to one. The normalised clusters are next sorted in order of decreasing weight to aid both the initial cluster comparisons and the final classification step [3]. The algorithm assumes that the background region is stationary is the limitation of method.

D. Temporal difference

Temporal differencing method uses the pixel-wise difference between two or three consecutive frames in video imagery to extract moving regions. It is a highly adaptive approach to dynamic scene changes. Let $I_n(x)$ represent the gray-level intensity value at pixel position x and at time instance n of video image sequence I, which is in the range [0, 255]. T is the threshold initially set to a pre-determined value. Lipton et al.[3] developed two-frame temporal differencing scheme suggests that a pixel is moving if it satisfies the following [3]:

$$| I_n(x) - I_{n-1}(x) | > T \qquad (1)$$

This method is computationally less complex and adaptive to dynamic changes in the video frames. In temporal difference technique, extraction of moving pixel is simple and fast. However, extracted shape of the moving objects is generally incomplete, especially when the moving objects in a scene are stationary or it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly [3].

E. Optical flow

Optical flow [2], [9] can achieve robust detection by projecting motion on the image plane with proper approximation. Optical flow is appearance of the object motion in image represented by the velocity vector distribution. Different methods such as spatio-temporal differentiation and compensation method are use to obtain optical flow. It is based on relative motion rather than absolute motion, as in the case of motion vector search method. Using this method the direction and speed of moving object from one image to another is obtained in terms of velocity vectors. Optical vector flow is used to estimate the motion of pixels in an image sequence within a visual representation. The motion represented by these optical flow vectors originate or terminate at pixels in the sequence of image frames derived from an MPEG movie. This depicts a dense vector field across all moving pixels in each frame. This method is based on the assumption that the displacement between the two frames is relatively small.

However, most of optical flow methods are complex, very sensitive to noise and not computationally affordable for real- time application [10].

F. Codebook Method

Kim et al. [14] used RGB color space in their motion segmentation for codebook formation. A codebook method is

formed to represent significant states in the background using quantization and clustering. The new value observed for a pixel is compared with the prior observed value. If the value is close to a prior value, then it is modeled as a perturbation on that color else a new group of colors is associated with that pixel. The result obtained can be visualized as a group of blobs floating in RGB color space.
The method is adaptive to only the small and gradual changes in the background and result distorts in case of sudden changes.

G. Cerebellar-model-articulation-controller based motion detection (CMACMD)
Video communication over real-world network with limited bandwidth often suffers from either unstable bandwidth or network congestion especially when communication occurs through a wireless network. Motion detection in variable bit-rate video streams produced by the rate control scheme in response to networks with limited bandwidth is a very difficult task for many previous stat-of-the-art methods. CMACMD approach is capable of attaining the complete and accurate motion detection in both low and high bit-rate video-streams [5]. CMAC-based motion detection method having two modules: Probabilistic background generation module (PGB) and Moving object detection (MOD) module. PGB module involves construction of the probabilistic background model that is capable of adapting to the properties of variable bit-rate video streams. The moving object detection module detects moving objects completely and accurately from both high and low bit-rate video streams over real-world limited bandwidth network through the cerebellar-model-articulation-controller network. To accomplish this, each incoming pixel is mapped to the weight memory elements in the weight memory space of CMAC network [5].

H. Radial Basis Function based Motion Detection
With popularity of Artificial neural network (ANN) due to its ability to learn, solution to complex non-linear problem and generalization, attracted much attention. An ANN works as a solid massive parallel processor, which is constituted by several simple units and has a natural propensity to store experimental knowledge and use it to create non-linear relationships between inputs and outputs. Radial basis functions neural network (RBFNN) [15] is one of the most popular neural network because (i) its architecture is very simple, only one hidden layer consist between input and output layer; (ii) in hidden layer localized radial basis function are used to nonlinear transform of feature vector from input space to hidden space; (iii) this network is faster and free from local minima problem etc [17]. RBFMD method involves two important modules: a multi-background generation module and a moving object segmentation module. The multi-background generation module generates a flexible multi-background model automatically by calculating the Euclidean distance from each incoming pixel to the corresponding reference background candidates; it then relays this information to the network as network as hidden layer neuron centers[2], [6]. The flexible multi-background model can express the dynamic Range of each pixel within the background and is used to construct a hidden layer in the RBF network structure. After processing the multi-background generation module, the moving object detection module is use to detect the moving object accurately [6].

III. PERFORMANCE EVALUATION

Performance evaluation allows the appropriate selection of segmentation algorithms as well as adjusts their parameters for optimal performance [3]. Many algorithms have been proposed for moving object detection in many applications, as a primary step towards video segmentation. Now a day, evaluation involves a representative group of human viewers which is subjective, time consuming and expensive process. Depending on the availability, or not, of reference segmentation, two alternatives can be considered for the evaluation of video segmentation quality; *standalone evaluation* when the reference segmentation is not available and *relative evaluation* when the reference segmentation is available for comparison.

A. Evaluation Methodology
Correia and Pereira [11] propose the methodology for performing individual object segmentation evaluation, which consists of three major steps:(1) Segmentation; the segmentation algorithm is applied to the test sequences selected as representative of the application domain in question. (2) Object selection; the object whose segmentation quality should be evaluated is selected. (3) Segmentation evaluation; the objective segmentation evaluation metric, as surveyed later, is computed. This metric differs for standalone and relative evaluation. The methodology for objective overall segmentation evaluation follows a five-step approach as proposed by Correia and Pereiro [12], both for the standalone and the relative evaluation cases. These steps are: (1) Segmentation; the segmentation algorithm is applied to the test sequences selected as representative of the application domain in question. (2) Individual object segmentation evaluation; for each object, the corresponding individual object segmentation quality, either standalone or relative, is evaluated. (3) Object relevance evaluation; the relevance of an object must be evaluated taking into account the context where it is found. (4) Similarity of objects evaluation; the correctness of the match between the objects identified by the segmentation algorithm and those relevant for the targeted application is evaluated. This step is different depending on whether standalone or relative evaluation is being performed. (5) Overall segmentation evaluation; by weighting the individual segmentation evaluation for the various objects in the scene with their relevance values [3].

B. Relative Performance Evaluation
Relative evaluation is expected to provide more reliable evaluation results as it has access to ground truth information. Three approaches have been recently considered [3]: pixel-based, template-based and object-based

methods. Pixel based methods assume that we wish to detect all the active pixels in a given image. Moving object detection is therefore formulated as a set of independent pixel detection problems. This is a classic binary detection problem provided that we know the ground truth. The algorithms can therefore be evaluated by standard measures used in Communication theory such as false alarm rate and receiver operating characteristic (ROC) [3]. The ROC curve approach has several weaknesses. One significant problem is that once an ROC plot is generated, it is impossible to infer the amount or nature of the data considered when creating the plot-information that is clearly important for ascertaining the relevance of the curve.

C. Quantitative Evaluation

For quantitative evaluation, Recall, Precision, F1, and Similarity metrics to video sequences are used [2]. The Recall metric provides the percentage of detected true positives as compared with the total amount of true positives in the ground truth. The Precision metric provides the percentage of detected true positives as compared with the total amount of items detected in the detected binary object mask. They can be described as follows:

$$Recall = tp/(tp + fn) \quad (2)$$

$$Precision = tp/(tp + fp) \quad (3)$$

Where tp, fn, and fp indicate total amount of true positive pixels, false negative pixels, and false positive pixels, respectively. Because Recall and Precision selectively measure the incorrect association of lost true positive pixels and external true positive pixels, satisfactory comparison results cannot obtained through the use of Recall and Precision alone. Therefore, two other accuracy metrics—F1 and Similarity— are used which are derived from Recall and Precision and are considered a significant measurement of accuracy. They can be expressed as follow [2]:

$$F_1 = 2(Recall)(Precision)/(Recall + Precision) \quad (4)$$

$$Similarity = tp/(tp + fp + fn) \quad (5)$$

D. Qualitative Evaluation

In qualitative evaluation the evaluation of the object extraction results is done for different test sequence by each method via visual inspection [2]. Through this subjective examination, the effects generated by the detected binary masks of each method are assessed.

IV. CONCLUSION

A perfect system should solve many problems, such as moved objects, shadows, gradually and suddenly change of illumination, waving tree and so on. But some of these cannot be solved simultaneously because differentiating of them needs semantic understanding of motion of foreground and of background, and it is impossible if we have no information from the ultimate purpose. The RBF neural network possesses the strong nonlinear mapping ability and the local synaptic plasticity of neurons with a minimal network structure. This allows it to be suitable for motion detection application in either dynamic or static scenes. A good system should use the knowledge derived from its purpose as possible as enough to solve the problems encountered.

REFERENCES

[1] Reddy, V.; Sanderson, C.; Lovell, B.C., "Improved Foreground Detection via Block-Based Classifier Cascade With Probabilistic Decision Integration," Circuits and Systems for Video Technology, IEEE Transactions on , vol.23, no.1, pp.83,93, Jan. 2013

[2] Shih-Chia Huang; Ben-Hsiang Do, "Radial Basis Function Based Neural Network for Motion Detection in Dynamic Scenes," Cybernetics, IEEE Transactions on , vol.44, no.1, pp.114,125, Jan. 2014

[3] Shireen Y. Elhabian, Khaled M. El-Sayed and Sumaya H. Ahmed, "Moving Object Detection in Spatial Domain using Background Removal Techniques – State-of-Art," Recent Patents on Computer Science, vol. 1, No. 1, 2008

[4] Barnich, O.; Van Droogenbroeck, M., "ViBe: A Universal Background Subtraction Algorithm for Video Sequences," Image Processing, IEEE Transactions on , vol.20, no.6, pp.1709,1724, June 2011

[5] Shih-Chia Huang; Bo-Hao Chen, "Highly Accurate Moving Object Detection in Variable Bit Rate Video-Based Traffic Monitoring Systems," Neural Networks and Learning Systems, IEEE Transactions on , vol.24, no.12, pp.1920,1931, Dec. 2013

[6] Ben-Hsiang Do; Shih-Chia Huang, "Dynamic background modeling based on radial basis function neural networks for moving object detection," Multimedia and Expo (ICME), 2011 IEEE International Conference on , vol., no., pp.1,4, 11-15 July 2011

[7] Thach-Thao Duong; Anh-Duc Duong, "Moving Objects Segmentation in Video Sequence Based on Bayesian Network," Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2010 IEEE RIVF International Conference on , vol., no., pp.1,6, 1-4 Nov. 2010

[8] Stauffer, Chris; Grimson, W.E.L., "Learning patterns of activity using real-time tracking," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.22, no.8, pp.747,757, Aug 2000

[9] Wixson, L., "Detecting salient motion by accumulating directionally-consistent flow," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.22, no.8, pp.774,780, Aug 2000

[10] Zheng-Wei Huang; Yeung, D.S.; Ng, W.W.Y.; Jiang Ding; Jin-Cheng Li, "Image segmentation with color and texture using RBFNN minimizing the L-GEM," Machine Learning and

Cybernetics, 2009 International Conference on , vol.6, no., pp.3221,3226, 12-15 July 2009

[11] Correia PL, Pereira F. Objective evaluation of video segmentation Quality. IEEE Trans Image Proc 2003; 12(2): 186-200.

[12] ITU-T. Recommendation - Subjective video quality assessment methods for multimedia applications August 1996; 910.

[13] Butler D., Sridharan S., Bove VMJr. Real-time Adaptive Background Segmentation. Acoustics, Speech, and Signal Processing. 2003. Proceedings. (ICASSP '03). 2003 IEEE Int. Conf. on April 2003: 3: 349-52.

[14] K. Kim, T. H. Chalidabhongse, D. Harwood, L. Davis, "Real Time Foreground Background Segmentation using Codebook Model", Real Time Imaging, Vol. 11, No. 3, June 2005, pp. 172-185

[15] Haykin, S.: 'Neural networks' (Prentice-Hall, 1999, 2nd edn.).

[16] Maddalena, L.; Petrosino, A., "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications," Image Processing, IEEE Transactions on , vol.17, no.7, pp.1168,1177, July 2008

[17] Ghosh, D.K.; Ari, S., "A static hand gesture recognition algorithm using k-mean based radial basis function neural network," Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on , vol., no., pp.1,5, 13-16 Dec. 2011

