Survey on Background Modeling and Foreground Detection Methods

Nifi C Joy
PG Student
Department of Computer Science & Information Systems Engineering
Federal Institute of Science and Technology, Mookkannoor PO, Angamaly, Ernakulam, Kerala 683577, India

Prasad J. C
Associate Professor
Department of Computer Science & Engineering
Federal Institute of Science and Technology, Mookkannoor PO, Angamaly, Ernakulam, Kerala 683577, India

Abstract

In Image Processing, a very trivial task is to detect the changes in the multiple images of the same scene of a real time instant. The task is not only trivial but also very indispensable as it brings into play a great number of diversified subject area applications such as, remote sensing, surveillance, medical diagnosis and treatment, security surveillance, and underwater sensing. This paper reviews the studies involved in background subtraction techniques. Object detection is basically the most important step in video analysis. There are various studies which are aimed at detecting the objects in video sequences. But due to fast illumination change in a visual surveillance system, many of them are not tolerant to dynamic background. There are various background subtraction algorithms for detection of moving objects, but many of them fail with slow-moving objects or in poor image qualities of videos and does not distinguish shadows from moving objects. The real time video surveillance models encompass the predictive and the foreground background modelling techniques.
Keywords: Background Subtraction, Visual Surveillance System, Foreground Detection, Background Modeling

I. Introduction

Visual surveillance is a very active research area in computer vision. The scientific challenge is to devise and implement automatic systems able to detect and track moving objects, and interpret their activities and behaviors. The need is strongly felt world-wide, not only by private companies, but also by governments and public institutions, with the aim of increasing people safety and services efficiency. Visual surveillance is indeed a key technology for fight against terrorism and crime, public safety (e.g., in transport networks, town centers, schools, and hospitals), and efficient management of transport networks and public facilities (e.g., traffic lights, railroad crossings). The main tasks in visual surveillance systems include motion detection, object classification, tracking, activity understanding, and semantic description. The detection of moving objects is the basic low-level operation in video analysis. This detection is usually done by using foreground detection. This basic operation consists of separating the moving objects called foreground from the static information called background.

The BS is commonly used in video surveillance applications to detect persons, vehicles, animals, etc., before operating more complex processes for intrusion detection, tracking, people counting, etc. Typically the BS process includes the following steps: a) Background model initialization, b) background model maintenance and c) foreground detection.

There is a crucial distinction between the background detection stages and background modeling, which constitute the complete subtraction process. The two stages of background subtraction are overlapping and co-relating. The modeling stage maintains and creates a model of the scene in the background. The detection process segments the current image into dynamic (foreground) and static (background) regions based on its background model. The resulting detection masks are then put back into the modeling process so as to avoid coincidence of foreground objects and background model.

Designing an algorithm that is robust under a wide variety of scenes encountered in complex real-life applications remains an open problem. For cameras that are mounted and are more or less stationary, background subtraction method is a major class of technique used to detect moving objects. Essentially in such methods, video frames are compared with a background model; changes are then identified as the foreground. The background models can be estimated via a parametric approach or a non-parametric approach or just in the form of thresholding. The reason why these methods fail to work well in realistic complex situation is that these methods often make overly restrictive assumptions about the background. In reality, the background itself can have complex changes. It might contain motion such as those caused by ripples on a lake, or swaying vegetation, which can cause false alarms. The motion of these backgrounds can be larger than that of the foreground. There could be sudden illumination change caused by cloud cover, causing complex intensity and shadow variation, or more gradual illumination change caused by the movement of the sun. During dawn and dusk hours, the image quality can be poor due to the low light condition.

Recent research on problem formulations based on decomposition into low-rank plus sparse matrices show a suitable framework to separate moving objects from the background. Practically, the background sequence is then modeled by a low-rank subspace that can gradually change over time, while the moving foreground objects constitute the correlated sparse outliers.

Although, many efforts have been made to develop methods for the decomposition into low-rank plus additive matrices that perform visually well in foreground detection with reducing their computational cost, no algorithm today seems to emerge and to be able to simultaneously address all the key challenges that accompany real-world videos. This is due, in part, to the absence of a rigorous quantitative evaluation with synthetic and realistic large-scale dataset with accurate ground truth providing a balanced coverage of the range of challenges present in the real world.

Some of the well-known issues in background maintenance include: a) light changes: the background model should adapt to gradual illumination changes. b) moving background: the background model should include changing background that is not of interest for visual surveillance, such as waving trees; c) cast shadows: the background model should include the shadow cast by moving objects that apparently behaves itself moving, in order to have a more accurate detection of the moving objects shape; d) bootstrapping: the background model should be properly set up even in the absence of a complete and static (free of moving objects) training set at the beginning of the sequence; e) camouflage: moving objects should be detected even if their chromatic features are similar to those of the background model.

II. LITERATURE SURVEY

In [1] Christopher Wren, Ali A, Trevor Darell, Alex Pentland et al.proposes a real time tracking of the people and interpreting their human behavior. Initially representations of the person and surrounding scene is made. It first builds the scene model by observing the scene without people in it, and then when a human enters the scene it begins to build up a model of that person. The person model is built by first detecting a large change in the scene, and then building up a multi blob model of the user over time. Given a person model and a scene model, the algorithm can acquire a new image, update the scene and person models. The first step is to predict the appearance of the user in the new image using the current state of the model. Next for each image pixel and for each blob model, calculate the likelihood that the pixel is a member of the blob. Resolve these pixel by pixel into a support map, indicating for each pixel whether it is part of one of the blobs or of the scene. Spatial priors and connectivity constraints are used to accomplish this resolution. Individual pixels are then assigned to particular classes: either to scene texture class or a foreground blob. A classification decision is made for each pixel by comparing the computed class membership likelihoods and choosing the best ones. Update the statistical models of all blob models. Though it finds application in exploring several different human interface applications, it suffers from many drawbacks. Pfinder assumes several domain specific assumptions to make the vision task tractable. It cannot compensate for large sudden changes in the scene.

In [2] L.Li, W.Huang, I.Gu, and Q.Tian et.al states that for detection and segmentation of foreground objects from a video that contains each stationary and moving background objects and undergoes each gradual and unforeseen once-off change. A Bayes regulation for association of background and foreground from selected feature vectors is developed. Underneath this rule, different types of background objects are classified from foreground objects by selecting a correct feature vector. The stationary background object is delineating by the color feature, and also the moving background object is diagrammatic by the color co-occurrence characteristic. Foreground objects are extracted by fusing the classification results from each stationary and moving pixel. Learning methods for the gradual and unforeseen once-off background changes are projected to adapt to numerous changes in background through the video. The convergence of the training method is tried and a formula to pick out a correct learning rate is additionally derived. Experiments have shown promising ends up in extracting foreground objects from several complicated backgrounds as well as wavering tree branches, unsteady screens and water surfaces, moving escalators, gap and shutting doors, change lights and shadows of moving objects.

In [3] Jing Zhong and Stan Sclarofi proposed a method to segment the foreground objects in video given time varying, textured backgrounds. Foreground background segmentation algorithm that accounts for the nonstationary nature and clutter-like appearance of many dynamic textures. The dynamic quality is modeled by an Auto regressive Moving Average Model (ARMA). A Kalman filter algorithm describes the appearance of the dynamic texture and also the regions of the foreground objects. The foreground object regions are then obtained by thresholding the weighting function used in the robust Kalman filter. Algorithm can successfully segment the foreground objects, even if they share a similar grayscale distribution with the background. However the ARMA model only takes grayscale images as input.

In [4] S. Derin Babacan, Martin Luessi and RafaelMolina et.al proposed a method based on Sparse Bayesian Methods for Low-Rank Matrix Estimation. A novel Bayesian formulation for low-rank matrix recovery is done based on the sparse Bayesian learning principles. Based on the low-rank factorization of the unknown matrix, this method employ independent sparsity priors on the individual factors with a common sparsity profile which favors low-rank solutions. Low-rank constraint is imposed on the estimate by using a sparse representation; starting from the factorized form of the unknown matrix, a common sparsity profile on its underlying components using a probabilistic formulation. The sparse error component in the robust PCA problem is also modeled and effectively inferred by sparse Bayesian learning principles.

In [5] Zoran Zivkovic et.al proposes an Improved Adaptive Gaussian Mixture Model for Background Subtraction. The Gaussian mixture model (GMM) algorithm is based on a supposition that background is more regularly visible than the foreground, and background variance is little. Every pixel in the background is modelled as a mixture of Gaussian. Each and every pixel value is matched with the current set of models to discover the match. If no match is found, the least model that is acquired is rejected and it is substituted by new Gaussian with initialization by the existing pixel value means the pixel values that don't suit into the background are taken to be background. The new algorithm can automatically select the needed number of components per pixel and in this way fully adapt to the observed scene. The processing time is reduced but also the segmentation is slightly improved. However it cannot handle multimodal background and involves rigorous computations.

In [6]Martin Hofmann, Philipp Tiefenbacher, Gerhard Rigoll et.al proposes a novel method for foreground segmentation.Pixel Based Adaptive Segmenter (PBAS) is a model which holds the recently observed pixel values and deigns the background. PBAS model contains a set of divisions. The decision block which is the prime component makes a decision either for or against the foreground biased on the per-pixel threshold of the current image and as well the background. Adding on to the designing process of the background model, the model gets updated over time with a defined procedure to carry out the changes in the background. The per-pixel learning parameter is the one which governs this update. The centroid of innovative fact in the PBAS approach is paved by the two per-pixel threshold which changes the background dynamics. Seemingly, the choice of the foreground decision is made from the foreground threshold value. The foreground decision depends on a decision threshold. Due to these enthralling differences the PBAS outshines almost all the state-of-the art approaches.

In [7] Olivier Barnich and Marc Van Droogenbroeck et.al proposes a technique for motion detection that incorporates several innovative mechanisms. ViBe works on random selection which leads to a smooth exponential decaying lifespan given a sample set which comprises the pixel models. The other novelty of the approach is pitched upon the post-processing which gives spatial consistency with the aid of a faster spatial information propagation technique. The pixel values are distributed in a random order among the neighboring pixels. The other descendent of the novelty in the approach credits from the background initialization done instantaneously. Hence the algorithm can proceed from the progressive second frame. ViBe sums up to a satisfactory outcome in most of the scenarios but when it comes to scenarios with darker or shadowy backdrop it gets intriguing. The performance of the ViBe method happens to seemingly increase when convoyed with other distance measure rather than the ancestral Euclidean (L2) distance measure. The performance increase has been measured based on the reduction in the computation time for processing the image.

In [8] Candles, Xiaodong Li, Yi Ma, and John Wright et.al proposes proposed the robust PCA problem as one of separating a low-rank matrix L (true data matrix) and a sparse matrix S (outliers matrix) from their sum A (observed data matrix).Under minimal assumptions, this approach called Principal Component Pursuit (PCP) perfectly recovers the low-rank and the sparse matrices. The background sequence is then modeled by a low-rank subspace that can gradually change over time, while the moving foreground objects constitute the correlated sparse outliers. PCP presents several limitations for foreground detection. The first limitation is that it required algorithms to be solved that are computational expensive. The second limitation is that PCP is a batch method that stacked a number of training frames in the input observation matrix. In real-time application such as foreground detection, it would be more useful to estimate the low-rank matrix and the sparse matrix in an incremental way quickly when a new frame comes rather than in a batch way. The third limitation is that the spatial and temporal features are lost as each frame is considered as a column vector. The fourth limitation is that PCP imposed the low-rank component being exactly low-rank and the sparse component being exactly sparse but the observations such as in video surveillance are often corrupted by noise affecting every entry of the data matrix. The fifth limitation is that PCP assumed that all entries of the matrix to be recovered are exactly known via the observation and that the distribution of corruption should be sparse and random enough without noise. These assumptions are rarely verified in the case of real applications because of the following main reasons: (1) only a fraction of entries of the matrix can be observed in some environments, (2) the observation can be corrupted by both impulsive and Gaussian noise, and (3) the outliers i.e. moving objects are spatially localized. Many efforts have been recently concentrated to develop low-computational algorithms for solving PCP.

In [9] Lucia Maddalena and Alfredo Petrosino et.al proposed Self-Organizing Background Subtraction (SOBS) algorithm accurately handles scenes containing moving backgrounds, gradual illumination variations, and shadows cast by moving objects, and is robust against false detections for different types of pictures taken with stationary cameras. Even without prior knowledge self-organizing method can detect the moving object based on background model. The neural network based image sequence model, models itself by learning in a self-organizing manner. The variations in the image sequence are viewed as trajectories of pixels along the time domain. The neural network exhibits a competitive win at all times function, this winner-take function is in turn coupled with the local synaptic plasticity behavior of the neurons. The learning process of the active neurons is seen to be spatially restricted which is founded on their local neighborhood. The neural background model can be portrayed as an adaptive one, since it adapts well to changes in the scene and succeeds in capturing most of the prominent change of features in the image sequence. The proposed approach can handle scenes containing moving backgrounds, gradual illumination variations and camouflage, has no bootstrapping limitations, can include into the background model shadows cast by moving objects, and achieves robust detection for different types of videos taken with stationary cameras.

In [10] Tianyi Zhou and Dacheng Tao et.al proposes a method based on Randomized Low-rank and Sparse Matrix Decomposition in Noisy Case. Go Decomposition (GoDec) is a method which efficiently and robustly estimate the low-rank part L and the sparse part S of a matrix X = L + S + G with noise G. GoDec alternatively assigns the low-rank approximation of X S to L and the sparse approximation of X L to S. The algorithm can be significantly accelerated by bilateral random projections (BRP). They also proposed GoDec for matrix completion as an important variant. It was proved that the objective valuejjXLSjj2F converges to a local minimum, while L and S linearly converge to local optimums. Theoretically, the influence of L, S and G to the asymptotic/convergence speeds was analyzed in order to discover the robustness of GoDec.

In [11] Xiaowei Zhou, Can Yang and W Yu et.al proposes a Moving Object Detection method by Detecting Contiguous Outliers in the Low-Rank Representation. It is assumed that the underlying background images are linearly correlated. Thus, the matrix composed of vectorized video frames can be approximated by a low-rank matrix, and the moving objects can be detected as outliers in this low-rank representation. Formulating the problem as outlier detection allows to get rid of many assumptions on the behavior of foreground. The low-rank representation of background makes it flexible to accommodate the global variations in the background. DECOLOR performs object detection and background estimation simultaneously without training sequences. However DECOLOR may misclassify unmoved objects or large texture less regions as background since they are prone to entering the low-rank model.

In [12]Xinchen Ye, Jingyu Yang,Xin Sun, Kun Li et.al proposes a motion-assisted matrix restoration (MAMR) model for foreground background separation in video clips. In the proposed MAMR model, the backgrounds across frames are modeled by a low-rank matrix, while the foreground objects are modeled by a sparse matrix. To facilitate efficient foreground background separation, a dense motion field is estimated for each frame, and mapped into a weighting matrix which indicates the likelihood that each pixel belongs to the background. Anchor frames are selected in the dense motion estimation to overcome the difficulty of detecting slowly moving objects and camouflages. In addition, a robust MAMR model against noise for practical applications is also been introduced. The method was found to be quite versatile for surveillance videos with different types of motions and lighting conditions.

III. Conclusion

This paper summarizes the most popular and accepted methodologies applicable to the various foreground background subtraction methods. Much work has been carried out towards obtaining the best background model which works in real time. In general, many methods have been developed using the framework of sparse representation and rank minimization. However, previous methods are motion-unaware and would introduce smearing artifacts when handling slow motion and motionless foreground (camouflages). When a video or picture is shooting from the camera then various noises, illumination variation, shadow etc poses threat to background subtraction. It is necessary to exploit the incorporation of more complex motion models or other clues into the low-rank and sparse recovery framework for foreground detection.

References

[1] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: Real-time tracking of the human body", in IEEE Trans. Pattern Anal. Mach. Intell., Jul. 1997
[2] L.Li, W.Huang, I.Gu, and Q.Tian, “Foreground Object Detection from VideosContaining Complex Background ", ACM 2003.
[3] Jing Zhong, Stan Sclarofi , “Segmenting Foreground Objects from a Dynamic Textured Background via a Robust Kalman Filter", in Proc. IEEE, 2003
[4] S. Derin Babacan, Martin Luessi and RafaelMolina, “Sparse Bayesian Methods for Low-Rank Matrix Estimation", in Proc. IEEE, Jun. 2012.
[5] Zoran Zivkovic, “Improved Adaptive Gaussian Mixture Model for Background Subtraction" ICPR, Aug. 2004.
[6] Martin Hofmann, Philipp Tiefenbacher, Gerhard Rigoll, “Background Segmentation with Feedback: The Pixel-Based Adaptive Segmenter ", in Proc. IEEE Int., Jan. 2012, pp. 21532160.
[7] Olivier Barnich and Marc Van Droogenbroeck, “ViBe: A Universal Background Subtraction Algorithm for Video Sequences ", in Proc. IEEE, 2011.
[8] Candles, Xiaodong Li, Yi Ma, and John Wright, “Robust Principal Component Analysis", ACM,2011.
[9] Lucia Maddalena and Alfredo Petrosino, “A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications", IEEE Trans., Jul. 2008.
[10] Xiaowei Zhou, Can Yang and W Yu , “GoDec: Randomized Low-rank and Sparse Matrix Decomposition in Noisy Case", International Conference on Machine Learning, 2011.
[11] K. E. A. Van de Sande, J. R. R. Uijlings, T. Gevers, and A. Smeulders, “Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation", in Proc. IEEE, Nov. 2013.
[12] Xinchen Ye, Jingyu Yang,Xin Sun, Kun Li, “Foreground Background Separation From Video Clips via Motion-Assisted Matrix Restoration", in Proc. IEEE , 2015, pp. 25772580.