Foreground-Background Separation from Video Clips using RMAMR Method Along with Ground Truth Extraction

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

Prasad J. C.
Associate Professor
Department of Computer Science and Information Systems Engineering
Federal Institute of Science and Technology, Mookkannoor
P O, Angamaly, Ernakulam, Kerala 683577, India

Abstract

Foreground-Background Separation from video clips using RMAMR method along with ground truth extraction plays a great role in video surveillance systems. The method proves to be a fully and important technique, making recognition, classification, and scene analysis more efficient. Initially ground truth images are extracted for the given video. A motion-assisted matrix restoration (MAMR) model for foreground background separation in video clips is proposed. 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. A robust MAMR model (RMAMR) is proposed against noise for practical applications. Experiment shows that the proposed methodology affirms a good performance in separation of foreground and background from video clips.

Keywords- Background subtraction, Foreground Detection, ADM, Motion Detection, Ground truth extraction

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 behaviours. 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 centres, schools, and hospitals), and efficient management of transport networks and public facilities (e.g., traffic lights, railroad crossings). In these applications, robust tracking of objects in the scene calls for a reliable and effective moving object detection that should be characterized by some important features: high precision, with the two meanings of accuracy in shape detection and reactivity to changes in time; flexibility in different scenarios (indoor, outdoor) or different light conditions; and efficiency, in order for detection to be provided in real-time. In particular, while the fast execution and flexibility in different scenarios should be considered basic requirements to be met, precision are another important goal. In fact, a precise moving object detection makes tracking more reliable (the same object can be identified more reliably from frame to frame if its shape and position are accurately detected) and faster (multiple hypotheses on the objects identity during time can be pruned more rapidly).

The main tasks in automatic 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. There are two conventional approaches to moving object segmentation with respect to a static camera: temporal differencing and background subtraction. Temporal differencing is very adaptive to dynamic environments, as only the current frames are used for analysis, but generally does a poor job of extracting all the relevant object pixels corresponding to object motion. Temporal differencing would particularly fail if the object motion between successive frames is small. This happens particularly for moving non-rigid objects, where certain parts of the object may experience almost zero motion between successive frames. Background subtraction provides the most complete object data but is extremely sensitive to dynamic scene changes due to lighting and extraneous events. More recent adaptive back grounding methods can cope much better with environment dynamism. However, they cannot handle multi-modal backgrounds and have problems in scenes with many moving objects. A more advanced adaptive background modelling method was proposed then. Here, each pixel is modelled using a mixture of Gaussians and is updated by an on-line approximation. The adaptive background model based segmentation method would alone suffice for applications where a rough estimate of the moving

foreground, in the form of irregular space blobs, is sufficient. Here the exact shape of the moving object need not be determined and only some post processing of the segmentation output using appropriate filters would give the desired blobs of interest. Recently, the level set method has become popular for object shape extraction and tracking purposes. The central idea is to evolve a higher dimensional function whose zero-level set always corresponds to the position of the propagating contour. There are several advantages of this level set formulation. Topological changes such as splitting and merging of contours are naturally handled.

The final extracted contour is independent of the curve initialization, unlike other active contour models like the snakes, where the final object contour is very much determined by the contour initialization. Lastly, quite a stable numerical approximation scheme is available. The only major drawback of this level set method is that by embedding the evolving contour as the zero level set of a higher dimensional function, a one-dimensional curve evolution problem has been transformed into a two-dimensional problem. This adds to the computational complexity and renders the standard level set method incapable of real time implementations.

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. The general idea of background segmentation is to automatically generate a binary mask which divides the set of pixels into the set of foreground and the set of background pixels. In the simplest case, a static background frame can be compared to the current frame. Pixels with high deviation are determined as foreground. This process is easy and it acts in accordance to a simple protocol. It accurately aids in obtaining features of target data, nevertheless, it had sensitivity towards small changes within the external environment. Hence, its only usage is within situations in which the background can be predicted or is known. Typically the BS process includes the following steps: a) Background model initialization, b) background model maintenance and c) foreground detection.

The BS initialization consists in creating a background model. In a simple way, this can be done by setting manually a static image that represents the background. The main reason is that it is often assumed that initialization can be achieved by exploiting some clean frames at the beginning of the sequence. Naturally, this assumption is rarely encountered in real-life scenarios, because of continuous clutter presence. The main challenge is to obtain a first background model when more than half of the video frames contain foreground objects. Some authors suggest the initialization of the background model by the arithmetic mean (or weighted mean) of the pixels between successive images. Practically, some algorithms are: batch ones using N training frames (consecutive or not), incremental with known N or progressive ones with unknown N as the process generates partial backgrounds and continues until a complete background image is obtained. Furthermore, initialization algorithms depend on the number of modes and the complexity of their background models.

There is a crucial distinction between the background detection stages and background modelling, which constitute the complete subtraction process. The two stages of background subtraction are overlapping and co-relating. The modelling 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 modelling process so as to avoid coincidence of foreground objects and background model. The simplest way to model the background is to acquire a background image which doesn't include any moving object. In some environments, the background isn't available and can always be changed under critical situations like illumination changes, objects being introduced or removed from the scene. To take into account these problems of robustness and adaptation, many background modelling methods have been developed. These background modelling methods can be classified in the following categories: Basic Background Modelling, Statistical Background Modelling, Fuzzy Background Modelling and Background Estimation. Other classifications can be found in term of prediction, recursion, adaptation or modality. All these modelling approaches are used in background subtraction context.

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.

In practice, the foreground motion can be very complicated with non-rigid shape changes. Also, the background may be complex, including illumination changes and varying textures such as waving trees and sea waves. An alternative motion-based approach is background estimation. Different from background subtraction, it estimates a background model directly from the testing sequence. Generally, it tries to seek temporal intervals inside which the pixel intensity is unchanged and uses image data from such intervals for background estimation. However, this approach also relies on the assumption of static background. Hence, it is difficult to handle the scenarios with complex background or moving cameras.

II. RELATED WORKS

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 a representation 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 colour feature, and also the moving background object is diagrammatic by the colour 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 background 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 Autoregressive 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 sup-position 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 neighbouring 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 mea-sure 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 outlier's 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 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 behaviour of the neurons. The learning process of the active neurons is seen to be spatially restricted which is founded on their local neighbourhood. 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]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 modelled by a low-rank matrix, while the foreground objects are modelled 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. PROPOSED WORK

An overview of the proposed method is explained in Fig.1. Initially the input video is taken. Its ground truth video is extracted. Then the input video and ground truth video are combined. The main idea is to incorporate motion information into the matrix recovery framework to facilitate the separation of the foreground and the background .A dense motion field is first estimated for each frame against an anchor frame, and mapped into a weighting matrix which indicates the likelihood that each pixel belongs to the background. The separation problem is then formulated into an MAMR model with the weighting matrix. The model is solved by the alternating direction method under the augmented Lagrangian multiplier (ADM-ALM) framework. Then, we

estimate the foreground using our background subtraction technique. In addition, we extend our model to a robust MAMR (RMAMR) model for noise related practical applications.

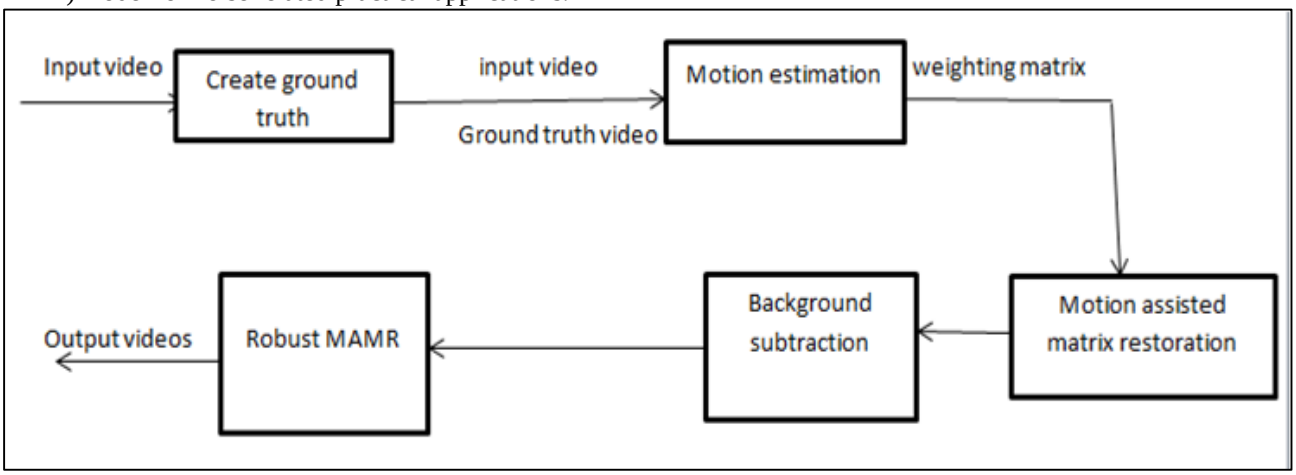

Fig. 1: Block diagram of the proposed method

A. Ground Truth Extraction

The ground truth is the known location and identity of objects that the user intends on detecting (or ignoring). Ground truth data plays a central role not only in learning tasks but also in quantitative performance evaluation. Ground truths on videos typically consist of a list of the objects with information on their bounding box, the contour, the recognition class and associations to other appearances in past or following frames. By reviewing a fragment of the recording frame by frame, the operator is able to generate ground truth. The current frame of the input video sequence is scanned in parallel by multiple detectors while all the objects already present in the previous frame are tracked. Each of the detectors produces as output a set of detected system targets, while the tracker produces a set of tracked objects. The use of multiple detectors allows to compare their output and to implement a confidence assessment mechanism which makes the final annotation output more reliable. When no detections are associated with a track for several consecutive frames, the example assumes that the object has left the field of view and deletes the track.

B. MAMR Model

The key idea of MAMR method is to assign to each pixel likelihood that it belongs to the background based on the estimated motion at that pixel. The background is to be extracted from K frames of a surveillance video clip denoted by i_k where k=0 to K-1 and of size M*N. The frame sequences is represented with matrix D = [i_0; i_1; :::; i_{k-1}] of size MN*K. The recovered background component and foreground component in D are denoted by B and F, respectively. The aim is to separate B and F from D.W is the weighting matrix whose elements represent the confidence levels that corresponding pixels in D belong to the background. The foreground - background separation problem can be solved by using the following optimization formula:

$$\min_{B,F} \| B \|_* + \lambda \| F \|_1 \quad (1)$$

$$W \circ D = W \circ (B + F) \quad (2)$$

By incorporating motion information, areas dominated by slowly moving objects are suppressed, while the background that appears in only a few frames has more chances to be recovered in the final results.

C. Weighting Matrix Construction from Motion Information

Optical flow gives a description of motion and can be a valuable contribution to image interpretation even if no quantitative parameters are obtained from motion analysis. Optical flow reflects the image changes due to motion during a time interval which must be short enough to guarantee small inter-frame motion changes. The optical flow field is the velocity field that represents the three-dimensional motion of object points across a two-dimensional image. The optical flow between two adjacent frames is not sufficient to determine it belongs to foreground or background, resulting in misclassification. To remedy this problem, for each frame, find a proper reference frame (called anchor frame) that differs from the current frame even in regions containing slowly moving foreground objects or even camouflages. Then, motion information for each frame referring to its nearest anchor frame is estimated. Finally, map the motion field into a weighting matrix.

1) Dense Motion Estimation with Anchor Frame Selection

For a single video, we set the first frame i_0 as the initial anchor frame. The difference between the current frame i_k and the previous nearest anchor frame i_{anchor} is calculated for each frame. The difference e_k is defined as mean absolute difference between two frames:

$$e_k = \frac{\sum_{meM,neN} | i^{m,n}k - i^{m,n}anchor |}{M * N} \quad (3)$$

where m and n are the 2-D pixel indexes in a frame. If the difference is larger than a threshold T, this frame is selected as a new anchor frame. For each frame, use the optical ow method to extract a dense motion field (o_k^x, o_k^y) between current video frame i_k and its previous nearest anchor frame, where o_k^x and o_k^y are the horizontal component and vertical component of the motion field, respectively.

2) Motion-to-Weight Mapping

The sigmoid function is used to map the motion field (o_k^x, o_k^y) into the weighting matrix. A sigmoid function is a bounded differentiable real function that is defined for all real input values and has a positive derivative at each point. The weighting matrix W is constructed as follows:

$$w_{j,k} = 1 - \frac{1}{1 + \exp(\alpha(-\sqrt{(o_{jk}^x)^2 + (o_{jk}^y)^2} + \beta} \quad (4)$$

Where o^x of size MN K as the matrix form of horizontal motion fields for all frames in D by stacking o_k^x k = 0, 1, . . . , K 1 as columns. Similarly, o^y is defined for vertical motion fields. α and β are the parameters of the sigmoid function which control the fitting slope and phase, respectively. β is chosen according to the average intensity of the motion field. As α increases, the slope of sigmoid function becomes steeper; when α takes very large values, the sigmoid function will become approximately a step function, while W also turns into a binary matrix. The weighting matrix W is degraded to the following binary mask:

$$w_{j,k} = \begin{cases} 0, \sqrt{(o_{jk}^x)^2 + (o_{jk}^y)^2} \geq \beta \\ 1, otherwise \end{cases} \quad (5)$$

With such weighting, (1) becomes the following matrix completion model:

$$\min_{B,F} \| B \|_* + \lambda \| F \|_1 \quad (6)$$

Such that

$$P_\Omega(D) = P_\Omega(B + F) \quad (7)$$

D. ADM-ALM Algorithm to Solve the MAMR Model

Augmented Lagrangian methods (ALM) are a certain class of algorithms for solving constrained optimization problems. Given a constrained problem, this problem can be solved as a series of unconstrained minimization problems. The alternating direction method of multipliers (ADM) is a variant of the augmented Lagrangian scheme that uses partial updates for the dual variables. The idea of ALM framework is to convert the original constrained optimization problem (1) to the minimization of the augmented Lagrangian function:

$$L(B, F, Y, \mu) = \| B \|_* + \lambda \| F \|_1 + \langle Y, W \circ (D - B - F) \rangle + \frac{\mu}{2} \| W \circ (D - B - F)_F^2 \quad (8)$$

Where μ a positive constant, Y is the Lagrangian multiplier. The ADM solves B, F, and Y alternatingly as:

$$F_{j+1} = \arg\min_F \lambda \| F \|_1 - \langle Y_j, W \circ F \rangle + \frac{\mu_j}{2} \| W \circ (D - B_j - F) \|_F^2 \quad (9)$$

$$B_{j+1} = \arg\min_B \| B \|_* - \langle Y_j, W \circ B \rangle + \frac{\mu_j}{2} \| W \circ (D - B - F_{j+1}) \|_F^2 \quad (10)$$

$$Y_{j+1} = Y_j + \mu_j W \circ (D - B_{j+1} - F_{j+1}) \quad (11)$$

$$\mu_{j+1} = \rho \mu_j \quad (12)$$

In the ADM-ALM framework, the sub problems are not necessarily solved exactly as long as the approximated solutions reduce the cost of Lagrangian function, which is therefore called inexact ALM. Allowing inexact approximation of the sub problems actually reduces overall computational complexity as the inner-loop iterations require considerable amount of computation to reach convergence. Here, the inner loop for solving B_{j+1} has only one iteration for acceleration. The solution of (1), denoted by (B^*, F^*), is obtained after the convergence of the iterative procedure: B^* contains a background component for

each frame, while F^* provides a foreground component for each frame. F^* contains not only the desired foreground components but also noise leaked from background areas.

E. Foreground Separation with Background Subtraction

Denote by $\overline{f_k(x)}$ the foreground image for frame i_k. The intensity value of $\overline{f_k(x)}$ at pixel x, denoted by $\overline{f_k(x)}$, is determined as

$$\overline{f_k}(x) = i_k(x), \frac{\sum_{xeN_x} | i_k(x) - \bar{b}(x)}{| N_x |} \succ \tau + \sigma \quad (13)$$

And 0 otherwise. Nx is the neighborhood of size $\omega * \omega$ around x. |Nx| is the number of pixels in Nx.

F. Robust MAMR

Noise is quite inevitable in real time applications. The data matrix is seriously damaged in some elements, while all of the elements would receive some lightweight noise pollution. Though the l1 norm can separate the intensive sparse errors from the intrinsic low-rank data matrix, it cannot deal with dense noise distributed over the whole frames. Therefore, RMAMR model is being proposed. The Frobenius norm is used to model dense noise.

Denote by G the error matrix of dense noise, the model can be formulated as follows:

$$\left(\overset{\min}{B,F,G}\right) \| B \|_* + \lambda \| F \|_1 + \gamma \| G \|_F^2 \quad (14)$$

Such that

$$W \circ D = W \circ (B + F + G)$$

The augmented Lagrangian function of problem is given by:

$$L(B,F,G,Y,\mu) = \| B \|_* + \lambda \| F \|_1 + \gamma \| G \|_F^2 + \langle Y, W \circ (D - B - F - G) \rangle + \frac{\mu}{2} \| W \circ (D - B - F - G) \|_F^2 \quad (15)$$

The solution of G-sub problem is:

$$G_{j+1} = \arg\min_G \gamma \| G \|_F^2 - \langle Y_j, W \circ G \rangle + \frac{\mu_j}{2} \| W \circ (D - B_j - F_{j+1} - G) \|_F^2 \quad (16)$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

For comprehensive evaluation, we test our method on 10 video clips from change detection dataset (CDnet). CDnet contains six video categories with four to six video clips in each category. We pick continuous 60 frames from each dataset in the experiment. Each of these datasets may include various kinds of motions, lighting variations, camera jitter, camouflages, shadows, dynamic backgrounds, or the combination of them. We use the peak signal-to-noise ratio (PSNR) to measure the quality of extracted backgrounds against their ground truth. The complete system was implemented in MATLAB 2014 prototype.

A. Experimental Results

An input video is being taken. Any video from the database or any alternative videos can also be selected. The input videos are then converted to their corresponding ground truths.

Fig. 2: A snapshot from the input video

Create objects for reading a video from a file, drawing the tracked objects in each frame, and playing the video. The method displays an object after it was tracked for some number of frames. When no detections are associated with a track for several consecutive frames, the method assumes that the object has left the field of view and deletes the track. Motion segmentation using the foreground detector is being done. It then performs morphological operations on the resulting binary mask to remove noisy pixels.

Fig. 3: Extracted ground truth video clip

Then both the input video set and the ground truth video set is used together for the further processing. A dense motion field is first estimated for each frame against an anchor frame, and mapped into a weighting matrix which indicates the likelihood that each pixel belongs to the background. The separation problem is then formulated into an MAMR model with the weighting matrix. The model is solved by the alternating direction method under the augmented Lagrangian multiplier (ADM-ALM) framework. The figure below shows the extracted background video. The background video contains the static background only which is compared against the ground truth.

Fig. 4: (a) Detected background (b) detected foreground

Noise is quite ubiquitous in video clips. So it is necessary to remove noise from the video clips and obtain a result that is very close to the input video. Robust motion-assisted matrix restoration is being used to deal with noisy dataset.

When it comes to slowly moving objects, for example, the running boats in Boats and Canoe, results produced by most of the methods present severe smearing artifacts. This is because the slowly moving objects occlude the scene across many frames, which may be considered as a part of background, resulting in the failure of background extraction. On the contrary, our method achieves promising results for all the evaluation datasets. With the help of motion information, we can prevent the slow moving objects (e.g., motionless man and running boat) from leaking into backgrounds, and recover the accurate backgrounds without smearing and ghosting artifacts. We tested the method with slow movement videos.

The most difficult category on detecting foreground is the Dynamic Background. Due to the motions in the background, such as the running water in Boats and Canoe. For example, in Boats, all the methods fail to detect the body of the boat, while MAMR model is able to faithfully separate the boat.

B. Experimental Analysis

The PSNR block computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is often used as a quality measurement between the original and a compressed image. The higher the PSNR, the better the quality of the compressed or reconstructed image. The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics used to compare image compression quality. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error. To compute the PSNR, the block first calculates the mean-squared error using the following equation:

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N}$$

In the above equation, M and N are the number of rows and columns in the input images, respectively. Then the block computes the PSNR using the following equation:

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right)$$

Here we perform the analysis for both the cases such as extracted ground truth, and PSNR value for the extracted foreground. The results show that Extracted ground truth videos are having high psnr values which indicates that they are compatible and can be used. The PSNR value for a number of videos that we have tested are calculated and shown in table given below.

Table 1: PSNR Value Computation of extracted ground truth

Name of Video	PSNR Value
Canoy	19.8011
Car	17.4321
Highway	15.8738
Office	14.1998
Pedestrians	15.4332

Table 2: PSNR Value Computation of Foreground

Name of Video	PSNR Value
Canoy	17.9617
Car	16.6451
Highway	19.7403
Office	19.6358
Pedestrians	16.2323

V. CONCLUSION

The proposed method presents an RMAMR model for foreground background separation from video clips. Ground truth dataset can be extracted for the given input video. In the MAMR model, the backgrounds across frames are modelled 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 to as- sign the likelihood of pixels belonging to the background. Anchor frames are selected in the dense motion estimation to overcome the difficulty of detecting slowly moving objects and camouflages. An RMAMR model is also presented here to deal with noisy datasets. The proposed method proved to be efficient in separation of foreground-background from given video clips.

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 Videos Containing 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 , Jun2012.

[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] 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.