Future Generation Computer Systems 90 (2019) 198–210

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Simultaneous denoising and moving object detection using low rank approximation

Shijila B., Anju Jose Tom *, Sudhish N. George

Department of Electronics & Communication Engineering, National Institute of Technology Calicut, Kerala, India

HIGHLIGHTS

- Denoising and Moving object detection based on Low-rank approximation and l_1-TV regularizations.
- Denoising is done by using nuclear norm and l_2-norm regularization.
- Spatial continuity constraint was effectively utilized by the TV regularization.
- Detection performance of the method outperformed the compared existing techniques in terms of F_{joint} measures.

ARTICLE INFO

Article history:
Received 19 March 2018
Received in revised form 16 July 2018
Accepted 28 July 2018
Available online xxxx

Keywords:
Moving object detection
Low rank recovery
Background subtraction
Robust principle component analysis

ABSTRACT

Moving Object Detection (MOD) and Background Subtraction (BS) are the fundamental tasks in video surveillance systems. But, one of the major challenges which badly affects the accuracy of detection is the presence of noise in the captured video sequence. In this work, we propose a new moving object detection method from noisy video data named as De-Noising and moving object detection by Low Rank approximation and l_1-*TV* regularization ($DNLRl_1TV$). In general, background of videos are assumed to lie in a low-rank subspace and moving objects are assumed to be piecewise smooth in the spatial and temporal direction. The proposed method consolidates the nuclear norm, l_2-norm, l_1-norm and Total Variation (TV) regularization in a unified framework to obtain simultaneous denoising and MOD performance. The nuclear norm exploits the low-rank property of the background, the sparsity is enhanced by the l_1-norm, *TV* regularization is used to explore the foreground spatial smoothness and noise is detected and removed by the l_2-norm regularization. Extensive experiments demonstrate that the proposed method outperforms the state-of-the-art approaches in terms of denoising capability and detection accuracy.

1. Introduction

The detection of moving objects is an important step in computer vision to develop numerous kinds of systems, such as intelligent video surveillance and motion capture. These systems are used in a wide range of applications, including retail, home automation, security, traffic monitoring, control applications, safety etc. In visual surveillance systems, the detection of moving objects is an important task to identify the useful insights from video data, such as intrusion detection, sidelined objects, traffic data collection etc. Object detection in real time environment has several important applications. This provides better sense of security using visual information and helps to automatically recognize people and objects. In retail space, it helps to analyze shopping behavior of customers. In traffic management system, it helps to find the flow of vehicles and detect/warn accidents. In video editing scenarios, it aids to eliminate cumbersome human operator interaction, for designing futuristic video effects. These are some of the applications of object detection in real time environment.

The top level applications of computer vision such as visual surveillance systems, object tracking etc. demand Background Subtraction (BS) followed by detection of moving objects. The simplest method for background modeling is to obtain a background image without including any moving objects. The main conditions which can ensure good results using background subtraction are static camera, constant illumination and static background. However, the background can be affected under some situations, such as illumination changes, noise and dynamic background. Therefore, it is a mandatory requirement that the background representation model must be robust and adaptive with respect to these challenges.

In addition to above mentioned problems, there are many challenges associated with general Moving Object Detection (MOD) to decide its accuracy and precision in performance. Noise, illumination changes, occlusion, camouflage, clutter, camera jitter etc. are some of them. A video signal is generally contaminated by noise

* Corresponding author.
E-mail address: anju_p160075ec@nitc.ac.in (A.J. Tom).

https://doi.org/10.1016/j.future.2018.07.065
0167-739X/

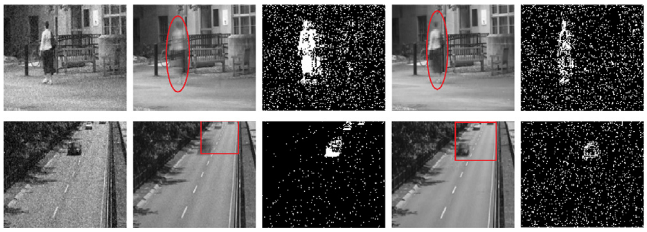

Fig. 1. Illustration of IBT [1] and GRASTA [2] against Gaussian noise: 1–6th column illustrates the input noisy video data, background and foreground masks estimated using GRASTA [2] and IBT [1] method respectively. Top row: Backdoor sequence, Bottom row: Highway sequence. The circled portions correspond to the misclassified background components.

during the recording process. That is mainly caused during signal transmission, acquisition, coding and between the processing steps. The different types of noises like sensor noise, speckle noise, compression artifacts etc. degrade the quality of videos. Usually, these noises can produce undesirable effects or artifacts in the background scenes.

Many algorithms on BS and MOD have been proposed in response to the rising demands for surveillance cameras. Most of the existing methods are unable to achieve good performance due to the above mentioned issues. Usually, these methods classify moving objects as well as background pixels which are near to moving objects, as foreground pixels which reduces the precision badly. Even if there exists many methods for MOD, no method is capable of detecting a moving object from the noisy video by removing the noise from each frame in an effective manner. The noise affected frames stand against the efficient working of the classical methods, as those methods result in holes in the foreground as well as misclassifications in the background. Fig. 1 shows the visual performance of the IBT [1] and GRASTA [2] against noise. Both methods fail in the accurate background and foreground extraction. The considerable amount of foreground components is captured in the background, which reduces the precision and recall value of the background component. From Fig. 1, it can be seen that these methods detect noise as foreground pixels. Hence there is an immense requirement of developing a low complexity scheme for simultaneous denoising and moving object detection so that the accuracy of the detection can be maximized. This work is concentrated on extracting moving objects from a noisy background.

1.1. Literature review

Nowadays, computer vision systems are commonly employed in many applications. Surveillance systems, traffic monitoring and robotics are some of the common areas where computer vision systems are employed for automation. Detection of moving objects is fundamental and crucial task in these systems. MOD methods are mainly classified as pixel based methods and frame-wise methods. Adaptive Median Filtering (AMF) [3], Running Gaussian Average (RGA) method, Mixture of Gaussian (MoG) [4] etc. are some of the examples of pixel based methods. The mixture of Gaussian is a method that can handle multi-modal distribution. In this method, all objects are filtered out and each pixel location is represented by a mixture of Gaussian functions that come together to form a probability distribution function. Pixel based methods result in comparatively poor visual quality for latest high definition (HD) videos and are computationally expensive.

The methods developed later [5–8] are concentrated on frame-wise statistics. In 1999, Oliver et al. proposed a background model based on Principal Component Analysis (PCA) [9]. This technique changes each frame to a PCA-transformed space and uses eigenvector to reconstruct the background image. Foreground detection is then achieved by thresholding the difference between the generated background image and the current image. PCA provides a robust model of the probability distribution function of the background, but not for the moving objects as they do not have a significant contribution to the model. Recently, Robust Principal Component Analysis (RPCA) has been extensively used in many algorithms in computer vision problems. However, it can handle only simple indoor and outdoor scenarios. It works well only with static background having few moving object with uniform movements. But in real life scenarios, the background is not always static in nature, and the objects (foreground) movements are not uniform. Recently, to improve the performance of conventional RPCA on moving object detection, many methods have been derived.

Detecting Contiguous Outliers in the Low-Rank Representation (DECOLOR) [10] is based on RPCA template, where regularized non-convex l_0 penalty and Markov Random Field (MRF) are used for extracting moving objects and background. This method gives better result than RPCA, but in some other contexts DECOLOR fails. This method detects some regions near to moving objects (greedy property) and they are misclassified as foreground pixels. Hence the exact shape of the foreground is lost.

In recent years, low-rank subspace learning models [11,12] and sparse models [13,14] represent a new trend and achieve better performance than the state-of-the-art techniques. Cao et al., proposed a framework named Total Variation Regularized RPCA (TVRPCA) [15] which deals with dynamic backgrounds and long-standing or slowly moving objects. This method is based on two assumptions, namely, the moving foreground objects have spatial and temporal continuity and the dynamic background is sparser than foreground objects. However, it works only if the dynamic natured background is sparser than the moving foreground having smooth boundary and trajectory and becomes ineffective in addressing the special cases similar to camouflage and far away objects. Zhao et al. proposed Low Rank and Sparse Decomposition (LRSD) [16] to detect outliers, which prefers the regions that are relatively solid and contiguous. LRSD is able to tolerate dynamic background variations, without losing the sensitivity to detect real foreground objects. This method is weak in case of noise robustness performance.

Tensor-based robust PCA [17] approach for background subtraction from compressive measurements, in which Tucker decomposition is utilized to model the spatial and temporal correlation of the background in video streams. 3D-TV [18] is employed to characterize the smoothness of video foreground. The convergence requirement for this method is about 250 iterations even for a video sequence with 20 frames. Hence this method is not suitable for real-time applications.

Iterative Block Tensor Singular Value Thresholding (IBTSVT) [1] method developed by Chen et al., considered the video data as a 3D tensor. They proposed a new method called Tensor Principal Component Analysis (TPCA) to extract the principal components of the data based on tensor singular value decomposition. This is a block based approach, and hence the main limitation of this method is the selection of the block size. The block size depends on the number of frames and hence the computational complexity increases with the increase in block size. Sajid et al. proposed Online Tensor Decomposition (OSTD) method [19] to address this problem and is efficient for large size data. But the main problem with this method is that it considers single frame per each time instant and it is not a suitable choice for real time applications.

Hu et al. proposed [20] another method for MOD based on Saliently Fused Sparse (SFS) decomposition for the low rank tensor [21]. This method combines 3D Locally Adaptive Total Variation (LATV) with l_1-norm to construct the fused-sparse regularization