Neurocomputing 168 (2015) 382–400

Contents lists available at ScienceDirect

Neurocomputing

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

Moving object detection based on incremental learning low rank representation and spatial constraint

Jianfang Dou [a,*], Jianxun Li [b], Qin Qin [a], Zimei Tu [a]

[a] Department of Automation and Mechanical and Electrical Engineering, School of Intelligent Manufacturing and Control Engineering, Shanghai Second Polytechnic University, Shanghai 201209, PR China

[b] Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, PR China

ARTICLE INFO

Article history:
Received 8 November 2014
Received in revised form
2 March 2015
Accepted 25 May 2015
Communicated by "Qi Li"
Available online 9 June 2015

Keywords:
Background subtraction
Low rank representation
Moving object detection
Dictionary learning
Mean shift
Graph cut

ABSTRACT

Background modeling and subtraction, the task to detect moving objects in a scene, is an important step in video analysis. In this paper, we present a novel moving object detection method based on Online Low Rank Matrix Recovery and graph cut from monocular video sequences. First, use the K-SVD method to initialize the dictionary to construct the background model, perform foreground detection with augmented Lagrange multipliers (ALM) and refine foreground values by spatial smooth constraint to extract the background and foreground information; Then obtain the clusters of foreground and background respectively using mean shift clustering on the background and foreground information; Third, initialize the S/T Network with corresponding image pixels as nodes (except S/T node); Calculate the data and smoothness term of graph; Finally, use max flow/minimum cut to segmentation S/T network to extract the motion objects. Online dictionary learning is adopted to update the background model. Experimental results on indoor and outdoor videos demonstrate the efficiency of our proposed method.

1. Introduction

In many image processing and computer vision scenarios, an important preprocessing step is to segment moving foreground objects from a mostly static background. Major application scenarios are in the field of mobile devices, video games and visual surveillance, e.g. for detection of unattended luggage, person counting, face recognition and gait recognition.

Background subtraction is a challenging task, especially in complex dynamic scenes that might contain moving trees, rippling water, etc. Many approaches have been presented in the literatures to deal with such multimodal scenes [1–4]. The popular idea is to model temporal samples in multimodal distributions, in either parametric or nonparametric way, so that the learned background model is able to tolerate the variations of the background scene. Other major challenges in background subtraction are illumination variation, camera jitter, intermittent object motion, shadows and thermal signatures.

However, most of these methods focus on single frame detection without considering the correlations in temporal sequences, and therefore they are very sensitive to corruptions, such as noise, motion blur and occlusion. Meanwhile, in many applications, images are collected consecutively from videos. These temporally consecutive frames are highly correlated to each other, which inspires us to exploit these correlations to improve the robustness of background modeling.

Very recently, the low-rank decomposition [5] attracts a lot of attentions from the communities of machine learning and signal processing. The objective function of the model is to minimize recovered matrix rank while keeping the sparsity of corruption matrix. Based on the low-rank matrix recovery and completion, robust Principal Component Analysis [6] was proposed to recover the underlying low-rank structure in the data.

In this paper, we present a novel moving object detection method based on Online Low Rank Matrix Recovery and graph cut from monocular video sequences. First, use the *K*-SVD [7] (*K*-means singular value decomposition) method to initialize the dictionary to construction the background model, perform foreground detection with augmented Lagrange multipliers [8] (ALM) and refine foreground values by spatial smoothing constraint to extract the background and foreground information; Then obtain the clusters of foreground and background respectively using mean shift clustering on the background and foreground information; Third, initialize the S/T Network with corresponding image pixels as nodes (except S/T node); Calculate the data and smoothness term of graph; Finally, use max flow/minimum cut to

* Corresponding author.
E-mail address: jfdou@sspu.edu.cn (J. Dou).

http://dx.doi.org/10.1016/j.neucom.2015.05.088
0925-2312/

segmentation S/T network to extract the motion objects. Online dictionary learning is adopted to update the background model. Experimental results on indoor and outdoor videos demonstrate the efficiency of our proposed method.

The paper is organized as follows: Section 2 reviews the previous work related to moving object detection algorithms. Section 3 describes the proposed method KSVD_LRR. The experimental results are presented in Section 4. Finally, the conclusions are presented in Section 5.

2. Related work

The common approach for background subtraction is to give an appropriate background model, which aims to deal with the challenges such as illumination variant, dynamic backgrounds, camouflage, shadows, etc.

Over the years, increasingly complex pixel-level algorithms have been proposed. Wren et al. [9] model the background as a single Gaussian in the system called "Pfinder", which aims to detect people indoors. But this method cannot handle the outdoor scenes well, since the distribution of gray-level value outdoors is multimodal. Stauffer et al. [10] present Gaussian Mixture Model (GMM), which consists of modeling the distribution of the values observed over time at each pixel by a weighted mixture of Gaussians. Its strong assumptions are that the background is more frequently visible than the foreground and that its variance is significantly lower. None of these is valid for each time window. Moreover, the update of GMM's parameters cannot accommodate with the rapidly changing scenes, such as sudden illumination changes, dynamic backgrounds. Oliver et al. [11] first modeled the background by Principal Component Analysis (PCA). This method models the background by projecting the high-dimensional data into a lower dimensional subspace, i.e., the eigen-background model, which can handle the global illumination changes to a certain degree. But they cannot deal with local illumination changes, and fail to distinguish slow moving foreground objects. Elgammal [12] proposed a non-parametric Model for Background Subtraction, which relies on sets of pixel values observed in the past to build pixel model. For each pixel, it builds a histogram of background values by accumulating a set of real values sampled from the pixel's recent history, then estimates the probability density function with this histogram to determine whether or not a pixel values of the current frame belongs to the background. It is more flexible for the rapid variation of the background at the price of heavy computation. Li et al. [13] used a Bayesian framework that incorporates spectral, spatial, and temporal features to characterize background appearance to cope with dynamic backgrounds. Zivkovic [14] proposed a method for adapting the scene to light changes by adding new samples and discarding the old ones a reasonable time period. A constant describes an exponentially decaying envelope that is used to limit the influence of the old data. Nevertheless, this approach uses an even higher number of heuristic thresholds featured by the others, while featuring a slow convergence. Some methods attempted to perform background subtraction in feature space. Notably, local binary patterns (LBP) were used by Heikkil et al. [15], which performed reasonably well, but the features used can only capture changes in texture, not in intensity. To deal with the uncertainty in the detection caused by the cited background maintenance issues, in [16], the authors modeled the background by the Type-2 fuzzy mixture of Gaussian model. Unfortunately, It failed to consider spatial-temporal constraints. The self-organizing approach for background subtraction proposed by Maddalena et al. [17] learns the background in a neural network framework, and the background update at each pixel is influenced by the labeling decision at neighboring pixels. PBAS (pixel-based adaptive segmenter) [18] uses a feedback-based control scheme to continually adjust a background model and its segmentation thresholds. In [19] Barnich and Droogenbroeck also present a really fast method that can cope with background in motion and bootstrapping problems. The method adopts the idea of sampling the spatial neighborhood for refining the per-pixel estimation. The model updating relies on a random process that substitutes old pixel values with new ones. However, it cannot cope with camouflages and shadows.

Some recent works, such as those about sparse representation-based algorithms, have gained improved recognition performance. Wright et al. [20] proposed a method that used a sparse representation for robust face recognition. Such a sparse representation-based classification (SRC) scheme achieved high classification accuracy. A crucial question when applying SRC method is how to choose a proper dictionary for sparse representation. A more generalized method is to learn the dictionary, which has been proven to be an effective scheme for improving the robustness of sparse representation. Aharon et al. [7] presented the K-means singular value decomposition (KSVD) algorithm to iteratively alternate the update of the sparse coefficients of the training samples, based on the current dictionary and the dictionary columns, in order to better fit the data. Liu et al. [5] established a low-rank representation (LRR) method to remove the grossly sparse corruptions of data samples that are approximately drawn from a mixture of multiple low-rank subspaces. Based on the [30], it has been proven that the Augmented Lagrange Multiplier Method has three attractive properties: (1) superior convergence properties; (2) the parameter tuning is much easier; (3) converging to the exact optimal solution. Most importantly, it is suitable for solving the low rank matrix recovery problem. So the convex optimization problem for LRR can be efficiently solved in polynomial time using inexact augmented Lagrange multiplier (ALM) algorithm, which guarantees convergence to the global optimum, and also comes with a (fairly slow) rate guarantee. It works well to recover a low-rank matrix from an observed data matrix that is low rank and contains sparse errors. However, considering a new coming frame, LRR essentially calls for a recalculation over all samples. This can incur a high computational cost, which does not generalize well for online computation. Zhou et al. [31] proposed a Moving object detection algorithm-(DECOLOR) by detecting contiguous outliers in the low-rank representation. While DECOLOR not only has problems in setting a correct regularizing parameter that can handle regions or motions of varying scales, but also works in a batch mode. Thus, it is not suitable for real-time object detection. As a result, a high computational cost seriously limits the practical applicability of LRR, especially for the moving object detection. In order to adapt to the problem of background subtraction, we propose an online version of low-rank representation to detect the moving object, using K-SVD to learning the dictionary.

3. Proposed algorithm

In this section, we present the proposed moving object detection KSVD-LRR algorithms. The flowchart of framework is shown in Fig. 1. The proposed framework mainly contains four phases. In the first phase, for the input video, collect N samples frames for KSVD Dictionary Learning. The Background Model consists of a Dictionary D. In the second phase, for the new coming frame, perform low rank matrix recovery to get the sparse outlier as foreground candidate, meanwhile adopt spatial smooth constraint to smooth the foreground candidate to get the initial foreground mask. In the third phase, refine the foreground mask obtained at the phase of three. Finally, for every M frames, use the incremental low rank representation to update the background model. We will introduce the four phases in detail during the following sub sections.