Information Sciences 430–431 (2018) 414–431

Contents lists available at ScienceDirect

Information Sciences

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

Fusion-based foreground enhancement for background subtraction using multivariate multi-model Gaussian distribution

Thangarajah Akilan, Q.M. Jonathan Wu*, Yimin Yang

University of Windsor, Canada

ARTICLE INFO

Article history:
Received 15 December 2016
Revised 19 November 2017
Accepted 30 November 2017
Available online 2 December 2017

Keywords:
Background-foreground detection
Multivariate multi-model Gaussian distribution
Color distortion
Color similarity
Illumination changes
Feature fusion

ABSTRACT

These days, detection of visual attention regions (VAR) such as moving objects have become an essential pre-processing stage in many computer vision applications. In this paper, we focus on the vital issue of separating moving objects a.k.a. Foreground (FG) in a scene, which has a near-static background (BG). We address the difficulty in setting an adaptive threshold in the multi-model Gaussian-based BG-FG separation through a novel FG enhancement strategy by assimilating color and illumination measures. We formulate the problem mathematically by using a histogram of a fused feature of color and illumination measures. The proposed method improves the BG-FG separation by introducing the following items: (i) A new distance measure to check if a pixel matches a Gaussian distribution. (ii) A new strategy to enhance the results of traditional background subtraction (BGS) with a fusion of color and illumination measures. (iii) A methodology to find appropriate threshold adaptively that separates BG and FG. (iv) A foreground validation process through probability estimation of multivariate Gaussian model distribution (MVGMD). We test the proposed algorithm on five different benchmark video sequences. The experimental results reveal that the proposed approach works well in challenging conditions, at the same time, it performs competitively against state-of-the-art Gaussian-based algorithms and few other non-Gaussian-based methods as well.

1. Introduction

Recent state-of-the-art technology in transistor fabrication has driven CPU clock rate to reach super high frequency; thus, the computational speed of processors has enormously increased. Hence, it has paved the way for larger available physical memory and storage disk space in modern computers. Consequently, it has enabled application of computer vision technology in several fields, for instance, Multimedia video-based surveillance, Industrial automation [1], Automotive, and Transportation. In such applications, BG-FG separation module plays an integral role in guiding a vision-based system to ignore unwanted region of a scene being monitored by utilizing visual properties and bringing attention to moving objects, such as running cars and walking people. It is crucial when it comes to priority specific data compression on TB size multimedia (mainly video) surveillance footage. For example, consider Fig. 1, where the sub-figures show: (a) an actual scene[1] and (b)

* Corresponding author.
E-mail addresses: thangara@uwindsor.ca (T. Akilan), jwu@uwindsor.ca (Q.M.J. Wu), eyyang@uwindsor.ca (Y. Yang).

[1] A scene in Railway data sequence downloaded from http://www.cs.cmu.edu/~yaser/new_backgroundsubtraction.htm .

https://doi.org/10.1016/j.ins.2017.11.062

(a) (b)

Fig. 1. A scene and its useful region.

useful FG region, which is about 25–30 % of the actual scene. This piece of information is valuable for a video compression algorithm so that it can use particular techniques to compress the FG and BG regions differently, whereby details of the useful FG region remain intact while the BG region experiences higher compression rate. Besides, once the FG region is defined, the fruitful low-level visual cues of the current scene can be extracted, and it can be employed for high-level analysis. For instance, object indexing and retrieval, traffic monitoring (detecting, counting or tracking of objects) [2], human activity recognition (run, walk, jump, squat, etc.) [3,4], human-machine interaction (in general, human-machine interface- HMI) [5], moving object tracking (many live sports telecasting channels have adopted this), scene classification, digital forensics [6,7], and so forth.

The core issue in BG-FG separation is, what exactly models the BG. In recent literature, for this task, many algorithms have been proposed either with pixel-based, region-based, or a hybrid strategy using cues of both pixel and region. There are several surveys can be found on these methods; among them the articles by Bouwmans [8] and Xu et al. [9] are very comprehensive and good to refer when more detail is required. The following section, however, briefly overviews the three methodologies.

2. Related work

2.1. Pixel-based background modeling

Pixel-based methods have received great acclamation since Stauffer and Grimson [10] introduced multi-model Gaussian or the GMM for real-time tracking for video surveillance. Due to its applicability, the GMM is utilized in many applications such as object detection, recognition, and tracking. The GMM is a statistical model, where each BG pixel is represented as a mixture of k number of Gaussian models; then based on persistence and variance of each distribution, m distributions are chosen to represent the BG. Following the work of Stauffer and Grimson, several researchers in this domain have proposed various techniques in order to improve the statistical model. Among them, some researchers focus on techniques that to update the model parameters, for instance, Zivkovic [11] implements an adaptive algorithm to choose a required number of Gaussian components per pixel and update the model parameters. Similarly, Dawei et al. [12] and Zhou et al. [13] take advantage of expectation-maximization (EM), k-means clustering, kernel density estimation (KDE) and Markov random field (MRF) for parameter updates and to refine the estimated FG respectively. Since the MRF-based FG refinement is an iterative process, it generally consumes higher processing time. On the other hand, Yang et al. [14] utilizes a spatial relationship between neighboring pixels through a conditional random field (CRF), which influences the learning process of the GMM-based classification for in-door video segmentation. This model faces a significant toll in computational speed and implementation complexity. To address this issue, Mukherjee et al. [15] propose a wavelet-based decomposition and a variable number of clustering technique (referred as WavGMM hereafter). This approach decomposes the input scene into multi-resolution sub-bands so that useful features in different scales can be incorporated in the learning step as temporal data with the consideration that the sub-bands are relatively independent. At the same time, this method achieves better computational speed because the sub-bands are smaller compared to the original input image. Lee [16] proposes a technique to balance GMM convergence speed and stability through computing appropriate learning rate when parameters of a Gaussian are updated and integrates a Bayesian framework to isolate the most-likely BG Gaussian distributions and generate an intuitive representation of the believed-to-be-BG. This approach, however, falls into the cost of additional computation and updates.

Besides the GMM based approaches, Bayesian-based BG modeling frameworks also have frequently been adapted for BG-FG separation. In [17], Zhu et al. take a multilayered Gaussian distribution to initialize each pixel for dissimilar BG contents and a recursive Bayesian learning based on texture correlation to refine the detected FG. Likewise, Li et al. [18] use Bayesian decision rule that incorporates spectral, spatial, and temporal features to define BG. Vosters et al. [19] compute an Eigenstate model from a training set of BG images that are recorded under various lighting conditions from a fixed viewpoint and reconstruct a generalized BG.

Hence, there are more pixel-based algorithms proposed; notably with MRF [20], non-parametric models, such as ViBe (visual background extraction) [21] that rely on the sample consensus, non-parametric Bayesian models [22], algorithms