Information Sciences 366 (2016) 31–47

Contents lists available at ScienceDirect

Information Sciences

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

# Statistical feature bag based background subtraction for local change detection

Badri Narayan Subudhi [a], Susmita Ghosh [b], Simon C.K. Shiu [c], Ashish Ghosh [d,*]

[a] Department of Electronics and Communication Engineering, National Institute of Technology Goa, Farmagudi, Ponda, Goa 403401, India
[b] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India
[c] Department of Computing, The Hong Kong Polytechnic University, Kowloon 999077, Hong Kong
[d] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

ARTICLE INFO

*Article history:*
Received 21 October 2015
Revised 27 April 2016
Accepted 29 April 2016
Available online 25 May 2016

*Keywords:*
Background subtraction
Object detection
Features bag
Majority voting

ABSTRACT

This article proposes a novel background subtraction (BGS) technique to detect local changes corresponding to the movement of the objects in video scenes. Here we propose an efficient combination of six local features; three existing and three newly proposed. For background modeling and subtraction here a statistical parametric biunique model is proposed. In the proposed BGS scheme, during the background training phase, the multi-valued features corresponding to background pixels are collected. A few simple statistical parameters are used to characterize each feature. For background subtraction, the multi-valued features computed at each pixel location are compared with those of the computed parameters corresponding to that feature. For each pixel location, different labels (either object or background) are obtained due to different features. For assigning a final label to the pixel in the target frame a majority voting based label fusion technique is used. The proposed technique is successfully tested over several video sequences and found to be providing better results compared to various existing state-of-the-art techniques with three performance evaluation measures.



## 1. Introduction

Moving object detection has been widely applied in fields like visual surveillance [3,13], face and gait-based human recognition [48], activity recognition [46], robotics [47], etc. Background subtraction (BGS) is the most commonly used technique for motion detection [14]. BGS techniques include two fundamental steps: background training/modeling and background subtraction [11]. The background training stage is used to construct a model that can efficiently describe the characteristics of the non moving/static objects in the scene. Construction of a robust background model is a prime factor in any BGS scheme as it affects the accuracy. The background subtraction step is used to identify the foreground object from the target frame by comparing it with the constructed background model [28]. Brief surveys on the literature of background subtraction techniques can be found in [2,8,14,28].

It is observed from the literature that, most of the BGS schemes consider non bi-unique model for background subtraction. The concept non bi-unique is to state that the local changes corresponding to the moving objects are obtained by

* Corresponding author. Tel.: +91 3325753110.
*E-mail addresses:* subudhi.badri@gmail.com (B.N. Subudhi), susmitaghoshju@gmail.com (S. Ghosh), csckshiu@inet.polyu.edu.hk (S.C.K. Shiu), ash@isical.ac.in (A. Ghosh).

http://dx.doi.org/10.1016/j.ins.2016.04.049
0020-0255/

making a combination of multiple features rather than combination of decision on individual features. Non bi-unique models are mere to produce results which are biased by the major or minor intensity value of a particular color. To avoid this, in change detection literature [12] mostly researchers suggested the use of the bi-unique model [4] where individual spectral properties are combined or results of the spectral channels are fused to obtain better results.

To the best of the authors' knowledge, in the literature of BGS the use of bi-unique model is very rare. In this article we propose a robust statistical feature based bi-unique model for BGS. To accomplish the task, the following six local features are taken into consideration: three existing features (*brightness (Bs)*, *inverse contrast ratio (ICR)* and *average sharpness (AS)*) and three new (proposed) features (*absolute relative height (ARH)*, *local weighted variance (LWV)* and *integrated modal variability (IMV)*). During the training phase, we have used multiple bags for storing the features of a particular pixel instance. For background subtraction, we compare the feature values computed at each pixel location with that of the constructed (feature) training bags. Hence, for each pixel location, we will get different labels (either object or background) due to different features. For assigning a pixel in the target frame as background or object, majority voting based fusion technique is considered for combining the decision based on different features. To validate the proposed scheme, the results obtained by it are compared with those of twelve existing state-of-the-art techniques. The effectiveness of the proposed scheme is evaluated by three performance evaluation measures namely: precision, recall and F-measure [5,8,14].

The organization of the remaining portion of this article is as follows. Section 2 gives a brief description on the related literature. The motivation behind the work is provided in Section 3. Section 4 represents a brief description of the considered features. A description on the proposed background subtraction technique is provided in Sections 5 and 6, respectively. Section 7 provides experimental results with discussion and future work. Conclusions are drawn in Section 8.

## 2. Related literature

The simplest way of performing motion detection is the use of frame differencing based or manual thresholding based BGS [44]. However, such an approach provides less accuracy in moving object detection. This is due to complex scenarios. The complexity in a video scene may arise due to noise and illumination variations which is very common in daily life captured sequences [2]. Hence most of the research directions were diverted to the developments of BGS technique which can deal with such complex scenarios [37]. Wren et al. in [50] have studied a running Gaussian average based BGS technique to detect moving objects. However, such a technique is unstable against noise and does not produce satisfactory results without availability of sufficient number of reference frames and hence produces ghosts in the scene.

The problem gets critical when moving objects in the scene are present with non-static background. In this regard, Stauffer and Grimson [41] proposed a multi valued background modeling scheme where the pixel at a particular location of the video frame is modeled with mixture of Gaussian (MoG) *pdfs*. A modification of the above model by incorporating dynamic texture (DT) is proposed for detecting moving objects from non-static background [6]. Recently, a new BGS scheme using Dirichlet process based Gaussian mixture models is proposed by Haines and Xiang [18], which estimates per-pixel background distributions followed by probabilistic regularization. A non-parametric Bayesian method is adhered which estimates the number of mixture components required to model the pixels based background color distribution. Cheung and Kamath [8] provided an extensive study where the performances of several BGS algorithms were made for detecting moving vehicles and pedestrians in urban traffic videos.

It is observed that the *pdfs* associated with the background or the foreground parts always do not fit to any known parametric form, and kernel estimation methods are quite popular in this regard. Elgammal et al. [10] proposed a kernelized Gaussian mixture model (GMM) to model the background of the scene, and have efficiently detected moving objects in the target frames. A non parametric BGS technique is also studied by Kim et al. in [21]. Here the authors have used the concept of codebook for construction of the background model and hence does not need estimation or fixation of any parameter. Guo et al. [16] proposed a hierarchical BGS scheme where both block and pixel based codebooks are explored for object detection. Recently, Guo et al. [15] have proposed a multilayer codebook-based BGS scheme where the adaptive features from the blocks of various sizes were combined to model nonstationary background.

A non-parametric dictionary based learning scheme is also explored by Zhao et al. [54]. The authors have modeled the background and performed background subtraction by optimizing L1-measure. It is required to mention here that most of these techniques are pixel based and do not take into account the spatio-contextual information for object detection. Again camera jitter is an important factor to be considered in the video scenes. In this context most of the researchers are found to adhere to the concept of using region/block based background modeling and frame comparison. A comprehensive study of different linear algebra based BGS techniques are studied by Durucan and Ebrahimi [9], where the Wronskian change detection model is found to be efficient. Recently, Subudhi et al. [42] proposed a Gaussian modeled Wronskian function for BGS. It is found to be working better for non-static background, gradual variations of light conditions, camera jitter, etc.

Spagnolo et al. [39] proposed a radiometric similarity based BGS technique, where influence of neighboring pixels information is used to preserve the spatio-contextual changes and also found to be robust against non-static background. A local similarity based multiscale local features, obtained by vector quantization for some video processing applications is proposed by Manzanera [27], where the object semantics in the scene are analyzed by motion estimation followed by a background subtraction technique. Recently, a computationally efficient visual surveillance tool, namely ViBe [1], is designed based on the non-parametric pixel based classification strategy, where a random sampling strategy is used and the neighborhood pixel statistics is considered to build the background model.