Image and Vision Computing 30 (2012) 829–842

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

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

Robust detection of moving objects in video sequences through rough set theory framework☆

P. Chiranjeevi *, S. Sengupta

Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, Kharagpur-721302, West Bengal (state), India

ARTICLE INFO

Article history:
Received 3 March 2012
Received in revised form 18 May 2012
Accepted 23 June 2012

Keywords:
Background subtraction
Rough set
3D histon
3D fuzzy histon
3D HRI
3D FHRI

ABSTRACT

Detection of moving objects in the presence of challenging background situations like swaying vegetation, rippling water, camera jitter etc., is known to be a difficult task. Background subtraction is considered to be better than the other approaches in terms of robustness. Its success primarily depends on the proper choice of background model(s) associated with every pixel for its foreground/background labeling. In this work, we have adopted rough-set theoretic measures to embed the spatial similarity around a neighborhood as a model for the pixel. Basic histon and its associated measure Basic Histon Roughness Index (BHRI) have been reported in the literature. It was applied to still image segmentation with impressive performance. Its adoption in video sequences for foreground/background labeling is proposed herein. We extended the histon concept to a 3D histon, which considers the intensities across the color planes in a combined manner, instead of considering independent color planes. Further, we also incorporated fuzziness into the 3D HRI measure. The labeling decision is based on Bhattacharyya distance between the model HRI and the corresponding measure in the current frame. Adoption of rough set theoretic concept into moving object segmentation is nontrivial, as the model updating requires careful consideration so that the pixels associated with gradually changing background or dynamic background are labeled as background and at the same time, slow moving objects are never adopted into the background model. A novel background model update strategy proposed herein takes these into consideration and also eliminates the need of having exclusive ideal background frame initially.

1. Introduction

Research on visual surveillance is gaining momentum due to security requirements in sensitive areas such as airport, offices, railway stations, etc. Moving object detection is the fundamental step in many visual surveillance applications like object tracking, action recognition, high level semantic description, gesture recognition, etc. Out of three major classes of moving object detection techniques, namely, image differencing, optical flow and background subtraction, the last is considered to be somewhat robust, as compared to the others.

Background subtraction approaches rely on constructing a background model, and detect moving objects as those that deviate from the background model, under the assumption of static camera. These approaches work for slow moving objects unlike temporal differencing, and are less noisy and simpler than optical flow. But these are sensitive to the background variations such as shadows, swaying vegetation, rippling water, illumination variations, etc. The background model needs to be robust to overcome such challenging situations. Though several approaches on background subtraction exist, and have been reviewed well in the literature [1–3], problems are yet to be solved under challenging situations. The background subtraction approaches are broadly classified as pixel-based and region-based, depending on the methodology used to establish the background model. In pixel-based approaches [4–9], each pixel is considered to be independent while constructing the model for it using its characteristics. The background perturbations, which affect the pixels' characteristics, will also affect their background model. Hence, these approaches are not effective for environments having fast background variations.

Region or block based approaches divide each frame into overlapped [10–23] or non-overlapped blocks [24–28] and calculate block level properties such as covariance, histogram, correlation, etc., to model the block. In either of these two approaches, since only a few pixels in the block are affected by the background disturbances, block based background modeling provides robustness against background variations. In non-overlapping block-based approaches, the modeling as well as the foreground detection is done at the block level. These approaches segment moving objects too coarsely than the overlapping block-based approaches. For overlapped block-based approaches, every pixel is modeled using the properties of the block, surrounding

☆ This paper has been recommended for acceptance by Massimo Piccardi.
* Corresponding author. Tel.: +91 7602449393.
E-mail addresses: chiru.pojala@gmail.com (P. Chiranjeevi), ssg@ece.iitkgp.ernet.in (S. Sengupta).

0262-8856/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2012.06.015

the pixel, and foreground detection is done at the pixel level. Hence, their shape accuracy is better than the non-overlapping ones, therefore, overlapped block-based approaches are preferred.

Most of the overlapping block based approaches use histograms as block models in many forms — RGB histogram [10], edge histogram [10], Local Binary Pattern (LBP) histograms [11–19], local kernel histogram [20], local dependency histogram [21] etc., which store only the feature without the spatial information. Though LBP captures the local texture information, it is very sensitive to the background disturbances, as it is a relation based feature. Gradient and LBP also fail in uniform regions. Hence, the performance of these histogram-based algorithms falls short of expectations in the presence of complex backgrounds having many challenging situations, such as swaying vegetation, camera jitter, rippling water, camouflage, etc. Moreover, these block-based approaches detect the moving objects coarsely. Block size can be reduced for a better shape localization, but it may cause an increase in false positives. Furthermore, most of these approaches use multiple models per pixel to handle multi-modal distributions. But, prior estimation of the number of models required is difficult in different environments. The computational complexity of the algorithm is directly proportional to the number of models.

We therefore felt the need for a robust spatial modeling for the finer extraction of moving objects (also with less false positives) using only a single model per pixel, in the presence of those challenging situations. Being inspired by the still image segmentation performance of rough-set theoretic features, used by Mushrif and Ray [29], we decided to adopt the concept of histon, proposed by Mohabey and Ray [30,31], in rough set theoretic framework for background modeling. Histon is a new concept to visualize the color information for evaluating the similar colored regions. For each intensity value in the base histogram, the number of pixels falling under the similar color sphere is calculated, and this value is added to the histogram value, to get the histon value of that intensity. Histogram and histon distributions, when used in a combined measure, are indicative of the color as well as the spatial information. Identical distributions of both the histogram and the histon in a region suggest that the region lacks spatial homogeneity or spatial similarity. This property of the region is used to model the pixel existing in the center of the region, and is mathematically formulated by Histon Roughness Index (HRI), after correlating the histon concepts with the rough set theory. The advantage of HRI is that it incorporates both the color and the spatial information.

In the original formulation of histon [30,31], each color channel is considered separately, rather independently, to derive the histon for that color channel. But, having each pixel with its three color component values together and working out a 3D spatial distribution out of it, gives us more information than what the spatial distribution of three independent color planes would give.

With this in mind, an integrated 3D histon is proposed, where the histon distribution is calculated by considering the color value on three channels jointly, and its effectiveness as compared to the basic histon is shown. The 3D HRI distribution for a region, centered at a pixel, is calculated using the 3D histon and the 3D histogram. In 3D histon, whether a pixel is similar to its neighbors or not, is determined crisply. By determining the extent of similarity using Gaussian membership function, 3D fuzzy histon is proposed, which is an extension of the basic fuzzy histon proposed in [29]. 3D Fuzzy histon is subsequently used to compute 3D Fuzzy Histon Roughness Index (3D FHRI).

The rest of the paper is organized as follows. As the roughness index formulation of the rough-set theory is already described in the literature, we directly start with the principle of the proposed method in Section 2. Concepts of basic histon and the motivation and the formulation of integrated 3D histon are explained in Section 3. Section 4 presents the proposed algorithm for moving object detection. Detailed results of moving object segmentation under different challenging situations are given in Section 5. Section 6 concludes the paper.

2. Principle of the method

We use the first frame of the video for initial HRI model construction. Foreground detection in a video sequence is performed by evaluating the Bhattacharyya distance between the model HRI distribution and the HRI distribution computed in the current frame in all three types of histons — basic histon, 3D histon, and 3D fuzzy histon. As we use the first frame of the video for model initialization, our proposed basic HRI or its 3D and fuzzy extensions as block model for foreground/background labeling would not succeed without an efficient model updating strategy in bootstrapping sequences, where the first frame without moving objects is not available in general. As a novel background model updating strategy, explained in details in Section 4.3, we not only update the model of the pixels labeled as background, but also update the model of the foreground labeled pixels with a lower learning rate. Our strategy of updating the foreground pixels' model succeeds, because the foreground labeling of a pixel is usually short-lived. This strategy facilitates us to initialize our model with moving objects also. Furthermore, many approaches in the literature require a series of ideal background frames for model initialization, whereas ours is simply initialized using the first frame of the video. Fig. 1 illustrates the block diagram of the proposed approach.

3. Concepts of basic and 3D histons

3.1. Basic histon

Basic histon, by definition [30,31], is a contour plotted on the top of the histograms of three primary color components of a region in a manner that the collection of all points falling under the similar color sphere of predefined radius, called similarity threshold (S_{th}), belongs to one single value. The similar color sphere is the region in RGB color space such that all the colors falling in that region can be classified as one color. For every intensity value in the base histogram, the number of pixels falling under similar color sphere is calculated, and this value is added to the histogram value to get the histon value of that intensity.

Construction of basic histon is given below:

$$h_c(i) = \sum_{m=1}^{M}\sum_{n=1}^{N}\delta(I(m,n,c)-i) \text{ for } 0 \le i \le l-1 \text{ and } c \in \{R,G,B\} \quad (1)$$

$$H_c(i) = \sum_{m=1}^{M}\sum_{n=1}^{N}(1+X(m,n))\delta(I(m,n,c)-i) \text{ for } 0 \le i \le l-1 \text{ and } c \in \{R,G,B\} \quad (2)$$

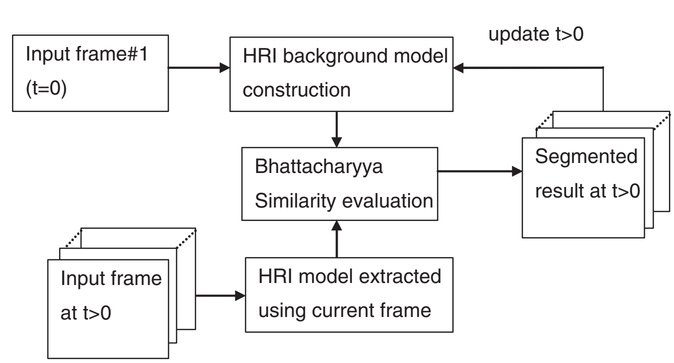

Fig. 1. Block diagram of the proposed approach.