Pattern Recognition Letters 68 (2015) 161–168

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Foreground detection for moving cameras with stochastic approximation ☆

Francisco Javier López-Rubio, Ezequiel López-Rubio*

Department of Computer Languages and Computer Science, University of Málaga, Bulevar Louis Pasteur 35, Málaga 29071, Spain

ARTICLE INFO

Article history:
Received 5 May 2015
Available online 11 November 2015

Keywords:
Background modeling
Moving camera
Covariance matrix interpolation
Stochastic approximation

ABSTRACT

Most foreground detection algorithms do not perform well with pan-tilt-zoom (PTZ) cameras for video surveillance and static cameras that experience vibration, since they rely on the assumption that the background does not move. Here a novel approach based on stochastic approximation learning of probabilistic mixtures is proposed. It assumes that the camera can zoom and move in both horizontal and vertical planes, and it is also adequate for egomotion sequences without abrupt changes. In other words it is a non panoramic model for moving cameras, where the camera movement allows to reuse enough background information from the previous frame. Two pixel models are used, one to follow the camera movement and the other to detect foreground objects. A procedure is developed to transform and interpolate the covariance matrices of the Gaussian mixture components as the camera moves and zooms. Moreover, a background extrapolation method is presented in order to generate new mixture models for previously unseen regions. The proposal is compared with some state-of-the-art alternatives, with competitive results both quantitatively and qualitatively.

1. Introduction

The vast majority of foreground detection algorithms build background models and identify foreground pixels or regions because their features are different from those of the background, but they are based on the assumption that the background does not change significantly with time or at least does not move. However in real situations this condition is rarely met even for static cameras due to camera shake, which can be produced by wind or other external factors. This kind of problem is called jitter and is very common, e.g. in street or highway surveillance cameras. Moreover, with the spread of video surveillance technology, it is increasingly common to find moving cameras of the PTZ kind (pan, tilt and zoom) covering a larger area than static cameras. Finally, foreground detection algorithms are also necessary for hand-held cameras that allow free movement. Again the usual background segmentation methods for static cameras are not capable of dealing with these situations.

There are several proposals to address background segmentation of a scene recorded by a moving camera. A joint representation of pixel color and spatial structures is used in [21] to build background and foreground models. This idea is extended in [9] by finding trajectories of key points in the scene, which requires the analysis of several consecutive frames.

Some proposals are based on building a panoramic model of the scene [2,5]. That is, they put each frame in the proper position of the model of the scene, and then foreground objects are detected using a traditional segmentation approach for static cameras. Panoramic methods can only be applied to cameras whose movement is restricted to a scene of fixed size, such as those used to broadcast sports events.

Non-panoramic methods like Kim et al. [13] are suitable for free moving cameras, since they only store a background model of the current frame. Their main challenge is how to recycle the information from the model of the current frame to build the model of the next frame. This is commonly done by estimating the camera motion, which usually involves the search for corresponding keypoints in the current and next frames [23].

Here a non panoramic method for foreground detection with a single camera is proposed. It overcomes the limitations of previous algorithms such as in [13] by employing a more realistic background model for the pixels, based on our previous works [15,17]. This requires new ways to pass the background information from the current frame to the next frame, as we will see. Moreover, our method does not require any knowledge of the internal camera parameters. Hence it is applicable to a wide range of situations. Nevertheless, it is assumed that there is a substantial fraction of the background model

☆ This paper has been recommended for acceptance by R. Davies.

* Corresponding author. Tel.: +34 95 213 71 55.
E-mail addresses: xavierprof@hotmail.com (F.J. López-Rubio), ezeqlr@lcc.uma.es, elr@uma.es (E. López-Rubio).
URL: http://www.lcc.uma.es/%7Eezeqlr/index-en.html (E. López-Rubio)

http://dx.doi.org/10.1016/j.patrec.2015.09.007
0167-8655/

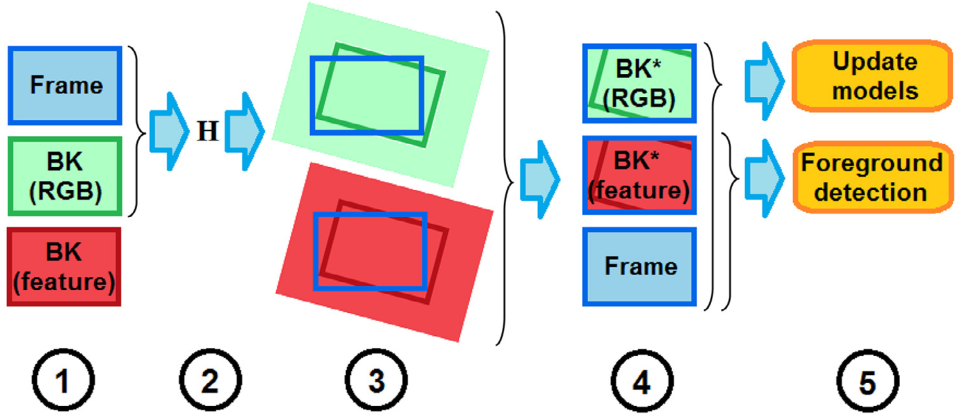

Fig. 1. Flowchart of our proposal. 'BK(RGB)' stands for the first background model and 'BK(features)' stands for the second background model. In the first step we have the input frame to be processed and the current state of the algorithm represented by 'BK(RGB)' and 'BK(features)'. Step two estimates the transformation matrix **H** using the input frame and 'BK (RGB)'. In the third step the background models are projected using **H**, which leaves a previously unseen region $\mathcal{U}$. The fourth step determines the area corresponding to the current frame in the projected models. In the fifth and last step 'BK*(feature)' is used to perform the foreground detection and both models are updated.

previous frame which can be reused for the current frame, so very fast camera movements are not managed. Consequently it is best suited for PTZ cameras, jitter compensation, and egomotion without abrupt background changes.

The main novelties of this work with respect to [17] and previous literature are twofold:

1. On one hand, a procedure to interpolate the full covariance matrices of the pixel models is proposed (Section 3.3). To the best of our knowledge, full covariance matrix interpolation has not been researched in background modeling literature.
2. On the other hand, following the camera movement and detecting foreground objects are distinct problems with their own complexities. Here we propose to have two pixel models, one for each task so that each of them can be tuned to their specific goals (Section 2).

This paper is organized as follows. Section 2 outlines the proposed foreground detection method, which is presented in detail in Section 3. Section 4 presents some experimental results to demonstrate the ability of our approach to manage complex scenes. The main features and properties of our proposal are discussed in Section 5. Finally, Section 6 is devoted to conclusions.

2. Overview

Our system considers two probabilistic background models for each pixel at each frame:

- The first model is aimed to match the key points of the scene between two successive frames. Its background features are the plain RGB values. The probability density function of the RGB values at the pixel with frame coordinates $\mathbf{x}$ is noted $p_{\mathbf{x},\,RGB}$.
- The second model is used to classify the pixel into background or foreground. A set of background features $\mathcal{F}$ must be chosen, as seen in Section 3.1. The probability density function of the selected background features at the pixel with frame coordinates $\mathbf{x}$ is noted $p_{\mathbf{x},\mathcal{F}}$.

Both background models have the same mathematical structure, so the model definition and the learning procedure described in Section 3 applies to both of them equally.

Each time that a new frame arrives, it is necessary to estimate the transformation matrix $\mathbf{H}$ which transforms from the old pixel coordinate system to the new pixel coordinate system. Then new background models for its pixels are built using the information from the models of the previous frame (Section 3.3). That is, no panorama is maintained. The interpolation procedure to obtain the parameters for the new background models includes the use of correlation matrices in order to produce accurate approximations of the covariance matrices of the Gaussian mixture components.

Some of the background models of the previously unseen regions of the new frame can be initialized with the help of the models of the pixels near the borders of the old frame. A Bayesian decision is carried out to determine whether a previously unseen pixel is suitable to be initialized in this way (Section 3.4). Otherwise, a neutral initialization is used.

3. Methodology

In this section a model is proposed to address the foreground detection problem for mobile cameras. It is based on our earlier models [15,17], which were oriented to fixed cameras. The probabilistic mixture model is defined in Section 3.1, and the features are specified in Section 3.2. The camera motion compensation mechanism is described in Section 3.3, and finally the background extrapolation procedure is given in Section 3.4. Fig. 1 shows the flowchart of our proposal for a frame.

3.1. Mixture model

The model is based on a stochastic approximation algorithm which is used to train probabilistic mixtures which model distributions of pixel feature values $\mathbf{t}(\mathbf{x}) \in \mathbb{R}^D$ at frame coordinates $\mathbf{x} = (j, k)$, where D is the dimension of the pixel feature vector. The set of features of interest will be noted $\mathcal{F}$. There is one Gaussian component $p(\mathbf{t}|Back)$ for the background and one uniform component $p(\mathbf{t}|Fore)$ for the foreground. It has been found that a single Gaussian with a full covariance matrix is flexible enough for most background distributions, while it speeds up the operation. On the other hand, the uniform distribution for the foreground models any incoming object equally well, no matter how unexpected their characteristics are. In contexts where some degree of redundancy in the foreground objects is expected, other foreground models could be used [6]. These considerations lead to the following probabilistic model $p(\mathbf{t})$ for the distribution of feature values at any given position $\mathbf{x}$ for the feature set $\mathcal{F}$, where we drop $\mathbf{x}$ and $\mathcal{F}$ for the sake of simplicity:

$$\begin{aligned} p(\mathbf{t}) &= \pi_{Back}\, p(\mathbf{t}|Back) + \pi_{Fore}\, p(\mathbf{t}|Fore) \\ &= \pi_{Back} K(\mathbf{t}|\boldsymbol{\mu}_{Back}, \mathbf{C}_{Back} + \boldsymbol{\Psi}) + \pi_{Fore} U(\mathbf{t}) \end{aligned} \quad (1)$$

$$K(\mathbf{t}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-D/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left((\mathbf{t}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{t}-\boldsymbol{\mu})\right) \quad (2)$$