ARTICLE IN PRESS
JID: PATREC [m5G;March 31, 2017;13:56]
Pattern Recognition Letters 000 (2017) 1–10

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

# Adaptive background model registration for moving cameras

Tsubasa Minematsu*, Hideaki Uchiyama, Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi

*Kyushu University, W2-828, 744, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan*

ARTICLE INFO

*Article history:*
Available online xxx

*Keywords:*
Moving object detection
Background subtraction
Moving camera
Re-projection error

ABSTRACT

We propose a framework for adaptively registering background models with an image for background subtraction with moving cameras. Existing methods search for a background model using a fixed window size, to suppress the number of false positives when detecting the foreground. However, these approaches result in many false negatives because they may use inappropriate window sizes. The appropriate size depends on various factors of the target scenes. To suppress false detections, we propose adaptively controlling the method parameters, which are typically determined heuristically. More specifically, the search window size for background registration and the foreground detection threshold are automatically determined using the re-projection error computed by the homography based camera motion estimate. Our method is based on the fact that the error at a pixel is low if it belongs to background and high if it does not. We quantitatively confirmed that the proposed framework improved the background subtraction accuracy when applied to images from moving cameras in various public datasets.



## 1. Introduction

Background subtraction is an essential technique in image processing and computer vision and is used in various types of applications such as moving object detection and background inpainting [3,10]. Previous research has considered a stationary camera with a fixed view [1,6,12,30]. This means that the background is the region of pixels that do not change over time, and the foreground is the moving objects. Background subtraction for moving cameras has also been recently proposed [7]. In an image sequence captured with moving cameras, both background and foreground move according to the camera motion. Therefore, we need a framework that makes the background stationary.

To register background models with an image for moving cameras, pixel correspondences between two consecutive images are computed using geometric transformations and optical flow [15,18,28,36,38]. In these methods, a background model of a pixel in an image is simply derived from a corresponding pixel in the previous image. Background subtraction using only these pixel correspondences results in many false detections because inaccurate pixel correspondences directly lead to inaccurate registrations of background models [36]. Even if the pixel correspondence errors are small, the accumulated error over time becomes large enough to cause failures. To improve the registration of background models, a standard solution is to search background models within a window instead of simply using a corresponding pixel. This search based registration is more robust to errors because there is an increased likelihood of finding an appropriate background model. However, the search typically uses a fixed window size and many false detections can occur when the size is inappropriate. The appropriate window size depends on various parameters of the target scene, and should be automatically adjusted to improve the accuracy of the background subtraction process.

In this paper, we propose a framework for automatically controlling the parameters in a background subtraction method for moving cameras. These parameters were heuristically determined in existing research. We adaptively search for background models within neighboring pixels using motion differences between the background and foreground. These differences can be represented by the re-projection error computed using homography based motion estimation, where the target environment can be considered planar or the camera motion is only rotational. Our method is based on the fact that the error at a pixel is low if the pixel belongs to the background. Additionally, we automatically control the threshold for foreground detection based on the movement of foreground and background. In our evaluation of the proposed method using various public datasets, we quantitatively confirmed that our adaptive framework effectively suppressed false detections when compared with other existing methods. Note that our method does not work for scenes containing translation and very complex background because we use homography for motion es-

* Corresponding author.
*E-mail address:* minematsu@limu.ait.kyushu-u.ac.jp (T. Minematsu).

http://dx.doi.org/10.1016/j.patrec.2017.03.010
0167-8655/© 2017 Published by Elsevier B.V.

timation. In many cases such as PTZ cameras and aerial images, however, camera motion is efficiently represented by homography. In this point, our proposal is meaningful.

A preliminary version of our method was initially presented at the International Conference on Image Processing (ICIP) 2015 [21]. In this paper, we have improved the procedure for updating the background models. We select an updated background model based on similarities between an observed pixel value and background models in the neighborhood. In addition, we compared our method with several state-of-the-art methods using more datasets than in our ICIP paper.

## 2. Related work

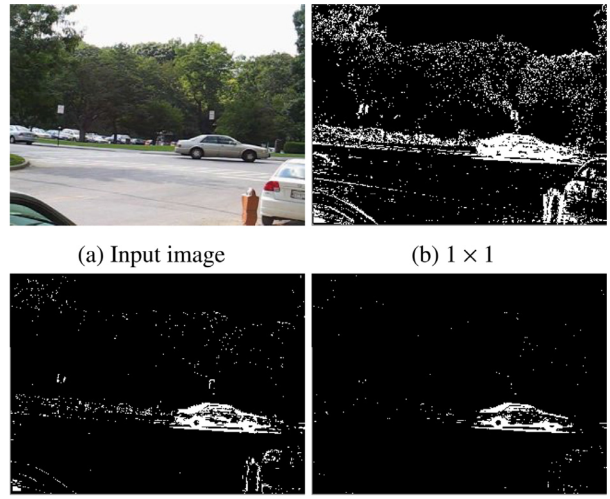
Background modeling for stationary cameras has been discussed for over 20 years (see surveys [3,26]). Most existing background models can be easily applied to background subtraction with moving cameras, if the background motion can be estimated and cancelled. The frame differences between consecutive frames is applied to moving cameras for detection of moving objects [20,33]. These methods use one displacement vector and an affine transformation as camera motion between consecutive frames. The frame differences is performed after a current frame is warped to the previous frame for compensating background motion.

Generally, in surveillance systems for PTZ cameras and aerial images, background motion generated by camera motion is efficiently cancelled by homography that represents the perspective transformation between two planes. These methods extend traditional background models such as Gauss model [14,36] and neural network-based models [8] for moving cameras. A pixel $(x, y)$ at a current frame is classified using a background model on $(x+\delta_x, y+\delta_y)$, where $\delta$ is generated by camera motion and computed by homography. These methods accurately approximate background motion only if the environment is planar or the camera motion is only rotational.

For wide area background modeling, background models have been constructed for panoramic images from pan-tilt cameras [11,35]. A Pixel $(x, y)$ at a current frame can be classified using the GMM [30] at location $(x', y')$ in the panoramic coordinates. To compute a panoramic coordinate, the methods use camera parameters [11] and match features between a current frame and a stitched panoramic image [35].

In other situations, multiple homographies for different planes [13,38], epipolar constraints [37] or optical flow [18] have been used. The methods [13,38] relax a limitation of homography by using one homography for each plane. In addition, the authors [38] choose appropriately one homography or multiple homographies for each frame based on the scenes and the camera motion. The method [37] combines homography and epipolar constraints. If motion at a pixel does not satisfy the constants that camera motion is rotational and translational, the pixel is a part of moving objects. Lim et. al. [18] and Kwak et. al. [15] uses optical flow and is not affected by a limitation of geometrical transformations such as homography and an affine transformation. The methods construct motion models from optical flows between consecutive frames and appearance models from pixel values at frames using belief propagation.

Other works [7,24,28] first classify pixel trajectories with salient features into either background and foreground, and then use them to calculate the background and foreground models in terms of the appearances. Sheikh et. al. uses factorization for sparse point trajectories based on an assumption of an affine camera model. The trajectories are divided to the ones satisfying camera motion and caused by moving objects. The limitation can not be applied to a perspective camera model. Ochs et. al. and Ali et. al. reduce dimensions of long term trajectories using Laplacian-Eigenmap [2]. In the low dimension space, background trajectories and foreground ones can be clearly separated. The methods classify the long term trajectories on the low dimension space based on a heuristic energy function.



(a) Input image (b) $1 \times 1$

(c) $3 \times 3$ (d) $5 \times 5$

**Fig. 1.** Comparison of background subtraction using different window sizes for the same scene. A white (black) pixel indicates foreground (background). An increase in the window size corresponds to a decrease in false foreground detections. However, false background detections increase when the window size is increases.

These methods we explained are online methods which sequentially obtain a new frame. Offline methods which have already obtain all frames can simultaneously assign a foreground or background label for each pixel at every frame. Lee et. al. obtains candidates of foreground regions based on object proposals and constructs foreground and background models from the candidates [16]. Papazoglou et. al. estimates roughly initial foreground regions based on motion boundaries [25] These methods construct foreground and background models from the candidates or the initial foreground regions and optimize an energy function defined on a space-time Markov Random Field in order to assign a foreground or background label for each pixel. The methods has no limitations for camera motion and scenes. However, in surveillance systems, online methods are more desirable. In our paper, we focus on online methods.

These online methods use the same framework for registering background models of an image in a pixel-by-pixel manner. However, it is difficult to realize accurate pixel correspondences because of feature tracking errors or image noise. The estimation errors gradually accumulate and then cause inappropriate registration of background models. For this reason, background subtraction for moving cameras often does not work well [36]. To tolerate the error and improve the robustness of the registration process, background models can be searched for within a window [23]. In another approach, background models are constructed in a region such as a rectangle or a super pixel [18,36]. However, it is difficult to appropriately set the size for the search window, as illustrated in Fig. 1. If the window contains 5 × 5 pixels and the environment has foreground regions, there are more false detections of the background.

## 3. Adaptive registration of background models for moving cameras

As mentioned in Section 2, background motion generated by camera motion in many cases, efficiently represented by homog-

Please cite this article as: T. Minematsu et al., Adaptive background model registration for moving cameras, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.03.010