Computer Vision and Image Understanding 122 (2014) 84–91

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Object detection based on spatiotemporal background models

Satoshi Yoshinaga *, Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi

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

ARTICLE INFO

Article history:
Received 19 April 2013
Accepted 2 October 2013

Keywords:
Background model
Object detection
Illumination changes
Dynamic changes
Background Models Challenge

ABSTRACT

We present a robust background model for object detection and its performance evaluation using the database of the Background Models Challenge (BMC). Background models should detect foreground objects robustly against background changes, such as "illumination changes" and "dynamic changes". In this paper, we propose two types of spatiotemporal background modeling frameworks that can adapt to illumination and dynamic changes in the background. Spatial information can be used to absorb the effects of illumination changes because they affect not only a target pixel but also its neighboring pixels. Additionally, temporal information is useful in handling the dynamic changes, which are observed repeatedly. To establish the spatiotemporal background model, our frameworks model an illumination invariant feature and a similarity of intensity changes among a set of pixels according to statistical models, respectively. Experimental results obtained for the BMC database show that our models can detect foreground objects robustly against background changes.

1. Introduction

One of the fundamental problems in computer vision is detecting regions or objects of interest from an image sequence. Background subtraction, which removes a background image from the input image, is widely used for detecting foreground objects in practical applications because it enables us to detect foreground objects without any prior knowledge. However, when background subtraction is applied to outdoor surveillance, the "long shot" scenes of cameras often include not only foreground objects but also background changes related to illumination conditions or disturbances in the scenes because the cameras are often installed in high locations to obtain a large field of view. In general, background changes that occur in outdoor scenes can be classified into two types; typical examples are shown in Fig. 1.

- *Illumination changes* – changes relating to lighting conditions such as the sun rising, setting, or being blocked by clouds (see Fig. 1(a)),
- *Dynamic changes* – changes relating to the swaying motion of tree branches, leaves and grass, fleeting cloud, waves on water and so on (see Fig. 1(b)).

To robustly detect foreground objects, we need to be able to handle these background changes. Many researchers have proposed background modeling approaches for dealing with these effects [1,2].

Because illumination changes affect not only a target pixel but also its neighboring pixels, local feature-based approaches that use this characteristic have been proposed to cope with illumination changes [3–7]. The Local Binary Pattern (LBP) [5,6] is a well-known local feature for background modeling. The LBP is defined by the signed differences between a target pixel and neighboring pixels. The LBP is unaffected by local intensity changes caused by illumination changes because it is a binary pattern describing lower or higher intensity relations between neighboring pixels. The distance to a neighboring pixel depends on the scene context and should be decided on a case-by-case basis. The Radial Reach Filter (RRF) [7] extends the LBP to adaptively determine the distance. Both approaches assume that local features are unaffected by background changes. However, surveillance scenes also often include dynamic changes that significantly affect the local features in the background. It is therefore difficult for local feature-based background models to handle dynamic changes in the background.

To cope with dynamic changes, statistical methods [8–13] have been used. In these approaches, the background is modeled by a probability distribution of the previously observed intensity values of each pixel. Background pixel values are usually observed with higher probabilities if we assume foreground objects are moving. When we use multiple distributions for the pixels, we can treat multi-modal backgrounds caused by dynamic changes. A Gaussian mixture model is used to represent the multiple distributions in the literature [8,9]. Non-parametric statistical methods [10–13] that use kernel density estimation have also been proposed. However, it is difficult for statistical background models to handle

* Corresponding author. Fax: +81 92 802 3579.
E-mail addresses: yoshinaga@limu.ait.kyushu-u.ac.jp (S. Yoshinaga), atsushi@limu.ait.kyushu-u.ac.jp (A. Shimada), nagahara@limu.ait.kyushu-u.ac.jp (H. Nagahara), rin@limu.ait.kyushu-u.ac.jp (R. Taniguchi).

1077-3142/$ - see front matter
http://dx.doi.org/10.1016/j.cviu.2013.10.015

/ Procedia Computer Science 00 (2013) 1–12

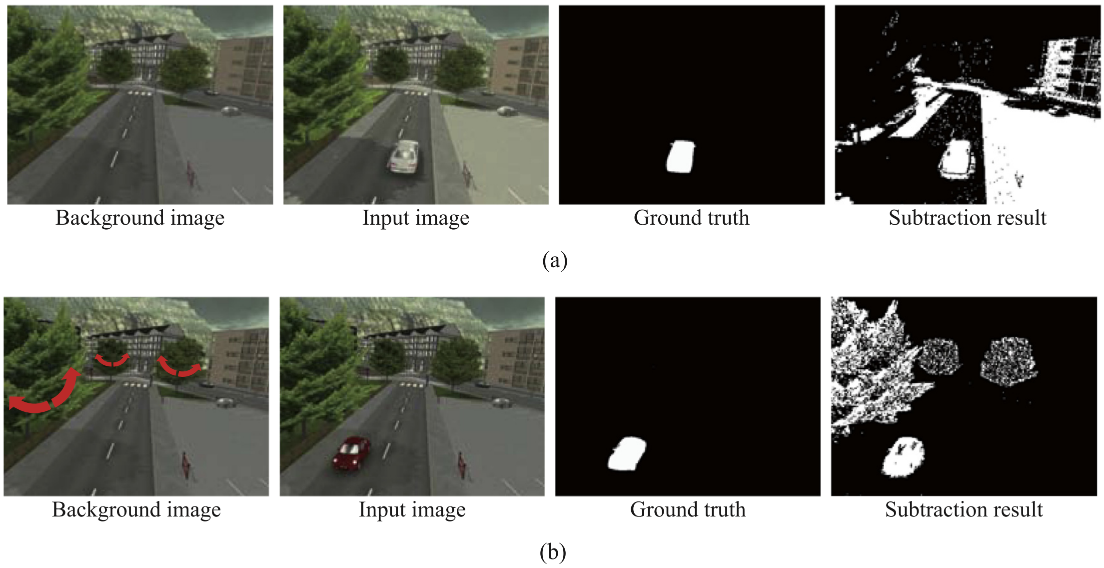

Fig. 1. Typical examples of "illumination changes" and "dynamic changes" with the results of the background subtraction employing a fixed background image. (a) Illumination changes. (b) Dynamic changes.

illumination changes, which vary intensity values rapidly and significantly.

Hybrid methods [14–16], which use multiple different background models, have also been proposed. To avoid falsely classifying the object regions as background, Yoshimura et al. [14] used a local feature-based background model in addition to a model focused on each pixel, and combined the results using a logical OR operation. In contrast, to cope with both illumination and dynamic changes in the background, Shimada et al. [15] and Tanaka et al. [16] used both local feature-based and statistical background models, and combined the results using a logical AND operation. However, the methods cannot adapt to particular regions that are affected by both illumination and dynamic changes at the same time because they assume at least one of the background models employed by the hybrid methods can adapt to background changes correctly. These methods are a kind of tandem system, and a logical combination of the detection results does not improve the accuracy of the foreground detection. Zhao et al. [17] used a local feature defined by multiple point pairs that exhibit a stable statistical intensity relationship as a background model. However, their method is not suitable for online surveillance because it needs to scan the entire input sequence to analyze the stability between point pairs.

To solve these problems, we integrate the methodologies of statistical and local feature-based approaches into a single framework. In this paper, we propose two types of spatiotemporal background modeling frameworks suitable for outdoor surveillance.[1] The first type [18,19] is based on a spatiotemporal local feature, where a statistical framework is introduced for an illumination-invariant local feature. The second type is based on a spatiotemporal feature, where similarity of intensity changes among a set of pixels is modeled using a statistical framework. By considering the similarity of intensity changes among the pixels, the second type can use the spatiality to handle not only illumination but also dynamic changes, while the first model uses spatial information only in defining an illumination-invariant feature to adapt to illumination changes. Our background modeling frameworks have properties of both statistical and local feature-based approaches because they use spatiotemporal information. Therefore, our spatiotemporal models can adapt to various background changes, even if some regions are affected by different types of background changes at the same time. To verify the effectiveness of our approaches, we report evaluation results obtained using the database of the Background Models Challenge (BMC[2]).

2. Background model based on a statistical local feature

We present a spatiotemporal background model by applying a statistical framework to a local feature-based approach [18,19]. In practice, we apply a Gaussian mixture model (GMM) to a local feature called the *local difference* (LD) to get a statistical local feature called the *statistical local difference* (SLD). Finally, we define the *statistical local difference pattern* (SLDP) [18,19] for the background model using several SLDs.

In most cases where illumination changes, there are small changes in the difference between a target pixel and its neighboring pixel because the values of pixels in a localized region increase or decrease proportionally. Owing to the invariance of the difference value with respect to illumination changes, the SLDP has the ability to tolerate the changes as shown in Fig. 2(a) because it uses the difference value as a local feature. Furthermore, our proposed method can also cope with dynamic changes because the SLDP can learn the variety of the changes as shown in Fig. 2(b). This is because a GMM, which can handle a multi-modal background, is applied to the LD, which is an important component of the SLDP. Thus, our background model can combine the concepts of statistical and local feature-based approaches into a single framework.

2.1. Construction of local difference

A target pixel and its neighboring pixel in an observed image are described by the vectors $\boldsymbol{p}_c = (x_c, y_c)^T$ and $\boldsymbol{p}_j = (x_j, y_j)^T$, respectively.

[1] Our target scenes are mainly "long shot" scenes in the outdoors, and our proposed method is not intended for "close-up shot" scenes in which a foreground object is very large.

[2] 1st ACCV Workshop on Background Models Challenge: http://bmc.univ-bpclermont.fr/.