Expert Systems With Applications 66 (2016) 32–41

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Fast illumination-robust foreground detection using hierarchical distribution map for real-time video surveillance system

Jongin Son, Seungryong Kim, Kwanghoon Sohn*

The School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Korea

ARTICLE INFO

Article history:
Received 11 April 2016
Revised 22 August 2016
Accepted 24 August 2016
Available online 1 September 2016

Keywords:
Foreground detection
Illumination invariant color space
Background modeling
Background subtraction

ABSTRACT

Foreground detection is one of the most important and fundamental tasks in many computer vision applications such as real-time video surveillance. Although there have been many efforts to find solutions to this problem, many obstacles such as illumination changes, noises, dynamic backgrounds, and computational complexities have prevented them from being used in real surveillance systems. In this paper, to alleviate these inherent limitations of conventional methods, we propose a fast illumination-robust foreground detection (FIFD) system that provides robustness against illumination variations and noises from various real circumstances with an efficient computational scheme. In contrast to the conventional approaches, our method focuses on efficiently formulating the foreground object detection system by leveraging a foreground candidate region detection and hierarchical distribution map. Specifically, our approach consists of three parts. First, for a query image, foreground candidates are detected by fusing multiple methods. The existence and the block size of the foreground object are determined through the use of the foreground continuity. Second, the foreground block is found from the estimated distribution map and then detected from the extracted valid blocks. Finally, with a labeling scheme, the foreground is detected. To intensively evaluate our approach compared to the conventional methods, we use the publicly available I2R and traffic datasets, and we build a novel electron multiplying charge-coupled device foreground detection benchmark taken in an environment with light lower than 10lux. Experimental results show that our approach provides satisfactory performance compared to the state-of-the-art methods even under very challenging circumstances. Furthermore, our approach is very efficient in that it takes only approximately 31 ms per frame.

1. Introduction

Since the demand for surveillance cameras has been dramatically increasing, the foreground detection problem has become an interesting and important research area in real-time video surveillance (Lee & Hedley, 2002). Most computer vision applications such as object detection (Bouwmans, 2014), tracking (McFarlane & C., 1995), action recognition (Koller, Weber, & Huang, 1994), and context awareness (Stas, Zelnik-Manor, & Tal, 2012) are basic, essential components of a surveillance system. Among these, foreground detection is a particularly important problem. In general, a surveillance system (Bouwmans, 2014; Koller et al., 1994; Lee & Hedley, 2002; McFarlane & C., 1995; Stas et al., 2012) follows the steps of foreground extraction, foreground detection, and high-order analysis. In many surveillance applications, the unnecessary background regions or scenes are not considered. On the other hand, scenes that are actually important in surveillance systems are considered, e.g., automobiles or people entering the scene or something appearing suddenly. That is, in foreground detection, scenes without any changes can be dismissed, where they can be considered as background, and regions or parts that contain specific objects must be considered as foreground.

To achieve reliable foreground detection in surveillance cameras, most approaches have attempted to eliminate unnecessary foreground regions or frames using background or background differences (Piccardi, 2004). Background subtractions require modeling and learning the background model from sequential frames. A number of methods have been proposed to improve performance of background subtraction under various illumination conditions (Javed, Oh, Bouwmans, & Jung, 2015; Tu, Karstoft, Pedersen, & Jorgensen, 2015; Wang & Yagi, 2012; Xie, Ramesh, & Boult, 2004). Specifically, in most prevision approaches, various illumination changes were handled by re-initializing the background frame according to illumination differences. A number of methods

* Corresponding author. Fax: +82-2-2123-2879.
E-mail addresses: go3son@yonsei.ac.kr (J. Son), srkim89@yonsei.ac.kr (S. Kim), khsohn@yonsei.ac.kr (K. Sohn).

http://dx.doi.org/10.1016/j.eswa.2016.08.062
0957-4174/

have been proposed by focusing that foreground detection should be designed to background modeling under various illumination changes. In Javed et al. (2015); Xie et al. (2004), they utilze the assumptions that rank order of the image intensity is preserved under illumination changes. Tu et al. (2015) and Wang and Yagi (2012) employ the approach measuring the illumination changes. However, these methods have some limitations in challenging scenarios, especially, such as blur and low sun angle situations and shadow. Well-known methods for background modeling are the mean, median, histograms, pixel intensity classification Hou and Han (2005); Wren, Azarbayejani, Darrell, T., and Pentland (1997), and pixel change classification Fu, Jiang, and Yu (2010) , and they can be divided into pixel and block levels according to their capacity to deal with the background regions. Pixel-based methods assume that the time series of observations is independent at each pixel. However, their computational complexity is excessively high, and they are very sensitive to irregular and sporadic noises. On the other hand, region-based methods are used on an entire group of spatially close pixels. The background differences provide a basic means of eliminating background by subtracting one frame from another, extracting parts or regions that were actually changed. This is a basic and efficient way to detect foreground contour, but the quality of extracting still images under illumination changes and noises has low computational complexity. A sampling method is usually used when distinguishing and tracking the extracted foreground or objects, but it has limited discriminative power.

To alleviate the limitations of conventional foreground detection, this paper presents a fast and illumination-invariant foreground detection method for real-time video surveillance systems. Our approach solves illumination variation and noise problems while maintaining a low computational complexity. The main contributions of this paper can be summarized as follows:

- **Robustness in outdoor environments:** Our approach are designed to provide robustness against noise and illumination variations, through novel multi-method fusion with an illumination-invariant direction model and hierarchical block distribution maps.
- **Low computational complexity:** Based on adaptive block selection schemes, our system is very efficient when applied to a real-time surveillance system.

We evaluated our proposed foreground detection for intensive experimental datasets of 31,374 frames from I2R Li, Huang, Gu, H., and Tian (2004), BMC 2012 Vacavant (2012), and DIML-EMCCD Kang, Yang, and Sohn (2011) datasets.

The remainder of the paper is organized as follows. We introduce related work for conventional foreground detection in Section 2, and the background of illumination invariant color space in Section 3. In Section 4, we propose a novel foreground detection framework that solves illumination variation problems and provides very low computational complexity. Section 5 shows the experimental results in challenging environments for both indoors and outdoors, followed by conclusions in Section 6.

2. Related works

In this section, we summarize related work on foreground detection. In general, foreground areas are selected through one of two means, pixel-by-pixel, in which an independent decision is made for each pixel, and region-based, in which a decision is made on an entire group of spatially close pixels. We briefly overview several notable papers in both categories.

The majority of the methods described in the literature belong to the pixel-by-pixel category. Stauffer et al. used several Gaussian probability distribution functions to obtain the values of each pixels statistical characteristics (Stauffer & Grimson, 1999) and proposed the use of a distribution of background pixels with a mixture of Gaussians. Some methods based on Gaussian modeling correspond to the background, and some such as MGM (multiple Gaussian model) are associated with active regions (Stauffer, Eric, & Grimson, 2000).

This allows the representation of multimodal distributions that occur in natural scenes and fluttering trees. However, there are many problems in that the number of functions used for each pixel and several parameters such as the average or standard deviation must be determined in advance. Elgammal et al. proposed a means of obtaining the value of a probability distribution function without estimating each parameter (Elgammal, Duraiswami, harwood, & Davis, 2002). This method must save several frames including the background information in advance and obtains the value of a probability distribution function through the use of a Gaussian kernel. However, the frames used for the modeling of a probability distribution function utilize several frames of early models that include only background information, and the speed of adaption is very slow because only the frames without any movement can be used. Further, several frames must always be saved, and the amount of calculation is too great, making it difficult to use practically. Buttler et al. models each pixel as several clusters, with each cluster expressed as a centroid and weight coefficient (Buttler, Sridharan, & Bove, 2003). When the input pixel is provided, the cluster that is most similar to each clusters centroid is selected, and values greater than the clusters weight coefficient are all summed and binary-coded into thresholds. In the situation in which a cluster does not exist, a new cluster with input pixels being the centroid is included. However, since the information related to the moving object is not easy to determine, the function is largely decided by the threshold.

In comparison to the pixel-by-pixel category, relatively little research has been conducted in the region-based category. In those approaches, each frame is typically split into blocks (or patches), and the classification is conducted at the block level (i.e., effectively taking into account contextual information). Kim et al. used vector normalization to express each pixels statistical characteristics using several codebooks (Kim, Chalidabhongse, & Harwood, 2005). This is similar to the method used in Buttler et al. (2003) in that it models each pixel into several clusters but is much more adaptable to patterns that involve moving objects or changing illumination. The methods in Buttler et al. (2003) and Kim et al. (2005) are efficient in that they do not need to save several frames of the background, allowing them to calculate much more rapidly. In addition, they adapt very well, since they can even use frames with moving objects. Migdal et al. conceptualizes energy using a Gibbs distribution (Migdal & Grimson, 2005), uses this to model MRF, and mixes it together with (Stauffer & Grimson, 1999) and a mixture of Gaussians. Boult et al. proposed LOTS (Lehigh Omnidirectional Tracking System) (Boult, Micheals, Gao, & Eckmann, 2001). This is tailored to detect small noncooperative targets such as snipers. To cope with abrupt changes, multiple model techniques, and predictive stochastic models, Maddalena et al. proposed a background model automatically generated by a self-organizing method without prior knowledge regarding the involved patterns (Maddalena & Petrosino, 2008). The models are updated using a stochastic approximation technique. Probabilistic self-organizing maps have also been examined to model the background. To extract foreground exactly, Reddy et al. use a block-based classifier cascade with PDI (probabilistic decision integration) (Reddy, Sanderson, & B.Lovell, 2013). A classifier cascade consists of three methods, probability measurement, cosine distance, and temporal correlation. It is used for generation of an integrated probabilistic foreground mask. These methods extract an exact result of foreground segmentation without post processing. However, where noise and foreground are not well distinguished,