Pattern Recognition 48 (2015) 1374–1390

Contents lists available at ScienceDirect

Pattern Recognition

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

Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes

Dong Liang [a,*], Shun'ichi Kaneko [a], Manabu Hashimoto [b], Kenji Iwata [c], Xinyue Zhao [d]

[a] Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
[b] School of Information Science and Technology, Chukyo University, Aichi, Japan
[c] National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan
[d] Department of Mechanical Engineering, Zhejiang University, Zhejiang, China

ARTICLE INFO

Article history:
Received 5 April 2013
Received in revised form
4 July 2014
Accepted 9 October 2014
Available online 24 October 2014

Keywords:
Object detection
Sudden illumination fluctuation
Burst motion
Background modeling
Co-occurrence probability

ABSTRACT

An illumination-invariant background model for detecting objects in dynamic scenes is proposed. It is robust in the cases of sudden illumination fluctuation as well as burst motion. Unlike the previous works, it uses the co-occurrence differential increments of multiple pixel pairs to distinguish objects from a non-stationary background. We use a two-stage training framework to model the background. First, joint histograms of co-occurrence probability are employed to screen supporting pixels with high normalized correlation coefficient values; then, K-means clustering-based spatial sampling optimizes the spatial distribution of the supporting pixels; finally the background model maintains a sensitive criterion with few parameters to detect foreground elements. Experiments using several challenging datasets (PETS-2001, AIST-INDOOR, Wallflower and a real surveillance application) prove the robust and competitive performance of object detection in various indoor and outdoor environments.

1. Introduction

Detecting moving objects plays a very important role in an intelligent surveillance system. It is often integrated with various tasks, such as tracking objects [1,2], recognizing their behaviors [3,4] and alerting when abnormal events occur [5]. However, object detection suffers from non-stationary scenes in surveillance videos, especially in two potentially serious cases: (1) sudden illumination variation, such as outdoor sunlight changes and indoor lights turning on/off; (2) burst physical motion, such as the motion of indoor artificial objects, which include fans, escalators and auto-doors. If the actual background includes a combination of any of these factors, it becomes even more difficult to perform detection. State-of-the-art algorithms [6–10] can handle gradual illumination changes by updating the statistical background models progressively as time goes by. In practice, however, this kind of model update is usually relatively slow to avoid mistakenly integrating foreground elements into the background model, making it difficult to adapt to sudden illumination changes and burst motion.

In this study, we propose a novel framework to build a background model for object detection, which is brightness-invariant and able to tolerate burst motion. We name it Co-occurrence Probability-based Pixel Pairs (CP3). It is inspired by the previous work in [11,12]. In the work of Haralick et al. [11], gray-level co-occurrence matrices (GLCM) were employed to measure the spatial co-occurrence of pixels to produce an image texture feature (Haralick feature). In the work of Hashimoto and Saito [12], pixels with low spatial co-occurrence probability and with high temporal co-occurrence probability were preferentially extracted as spatially distinctive and temporally stable features to reduce computational complexity for template matching. In this study, in order to model the dynamic background, spatial pixel pairs with high temporal co-occurrence probability are employed to represent each other by using the stable intensity differential increment between a pixel pair which is much more reliable than the intensity of a single pixel, especially when the intensity of a single pixel changes dramatically over time. A pixel pair consists of each pixel itself (called *target pixel* hereafter) and a selected pixel (called *supporting pixel* hereafter). As a pixel-wise background model, the target pixel P refers to all pixels in a scenario. The supporting pixels are neither arbitrary pixels in the scene, nor predefined fixed local structures around each target pixel; instead, the supporting pixels are selected based on their statistical stability with the target pixels.

* Corresponding author.
E-mail addresses: liang@ssc.ssi.ist.hokudai.ac.jp, donglcn@hotmail.com (D. Liang).

http://dx.doi.org/10.1016/j.patcog.2014.10.020
0031-3203/

The remainder of this paper is organized as follows. In the next section, some related works are discussed. Section 3 details the background model. Section 4 presents the object detection procedure. Section 5 presents the experimental results, and Section 6 concludes the main contributions of this work.

2. Related work

Since observations of the background in image sequences can be considered stochastic events, many statistical approaches have been employed to model effective backgrounds. The former background modeling approaches can be classified into two categories: (1) independent pixel-wise modeling, which employs the statistical processing of time-domain observations to each pixel. (2) Spatial-dependence modeling, which employs principles to exploit spatial-dependence among pixels to build a local or global model.

Most of the earlier background modeling approaches tend to fall into the first category. Wren [6] modeled the observations (YUV) of each pixel as a single Gaussian probability density function. To cope with periodic moving background patterns, the Gaussian mixture model (GMM) [7,13] was proposed. Elgammal [8] employed kernel density estimation (KDE) as a data-driven modeling method. Since KDE is a non-parametric model, it is closer to the real probability distribution than GMM. Hidden Markov models (HMMs) [14,15] have also been applied to model the background; topology free HMMs were described and several state splitting criteria were compared in the context of background modeling in [14], and a non-adaptive three-state HMM was used to model the background in [15]. The recent notable pixel-wise method by Kim [9] presented a real-time algorithm, which sampled background pixel values and quantized them into compressed codebooks (CBs). To improve the processing efficiency of the codebooks, Guo [16] presented a hierarchical scheme. All the above methods use a learning rate function for updating the background model online. However, because none of these methods is free from erroneous updating, they have a well-known trade-off problem: with a low learning rate, they can not adapt to sudden changes of illumination, e.g., turning on/off a light, while with a high learning rate, slowly moving objects or temporarily stopped objects will be detected as background.

The second category uses spatial information to exploit the spatial dependencies of pixels in the background. Matsuyama [17] proposed a regional block matching method against varying illumination, and Seki [18] proposed a co-occurrence-based block correlation method. The above two methods can only yield coarse region-level detection. Toyama et al. [19] proposed a three layers algorithm in which Weiner filters were employed. It used region and frame-level information to verify the pixel-wise background model. Oliver [20] employed eigenspace decomposition in which the background was modelled by the eigenvectors corresponding to the largest eigenvalues. Sheikh [10] used the joint representation of image pixels in a local spatial distribution (proximal pixels) and colour information to build both background and foreground KDE models competitively in a decision framework. Monnet [21] and Zhong [22] built an auto-regressive moving average (ARMA) model in dynamic scenes, which is used to incrementally learn (using PCA) and then predict motion patterns in the scene. Heikkilä and Pietikäinen [23] used a local binary pattern (LBP) to subtract the background and detect moving objects in real time. This method models each pixel as a group of adaptive LBP histograms that were calculated over a predefined circular region around the pixel. Similarly, the statistical reach feature (SRF) [24] builds a local texture model for each target pixel to be brightness-invariant. A recent spatial-dependence approach [25] utilized a tensor subspace learning algorithm to represent spatial correlations between pixel values, and modeled appearance changes by incrementally learning a tensor subspace representation by adaptively updating the sample mean and an eigenbasis for each unfolding matrix of the tensor.

In our previous research, we proposed a background model called grayscale arranging pairs (GAP) [26,27] which falls into the second category. GAP employed an alignment of supporting pixels for the target pixel which held a stable intensity subtraction in training frames without any restriction of locations. The intensity subtraction of the pixel pairs allowed the background model to tolerate noise and be illumination-invariant. However, this fixed intensity subtraction influenced the sensitivity of the background model, especially when the dynamic range was compressed due to low illumination; it was also not an optimal way to search for supporting pixels by using a fixed intensity subtraction in that most co-occurrence pixels were not considered. In addition, the GAP method mainly focused on illumination-invariance, so that the dynamic background caused by burst motion was not discussed sufficiently. In this study, the proposed method addresses these open problems. Compared with GAP, the proposed method employs a co-occurrence histogram to describe the relationship of a pixel pair, which is free from any intensity differences, and calculates normalized correlation coefficients for measuring the degree of co-occurrence which can deal with a dynamic background. It also introduces a spatial clustering operation to select optimal supporting pixels and then provides a more accurate parameterized detection criterion instead of a fixed double-sided threshold.

3. Background modeling

The algorithm is described for gray-scale imagery; however, it can also be used for colour or multi-modality imagery with minor modification. Fig. 1 shows the fundamental definitions of the image data. Suppose we are given a training image sequence $B = \{I_1, I_2, \dots, I_T\}$ with a total of T images, and each image has $M = U \times V$ pixel positions. In the three-dimensional space $\Gamma = \{(u, v, t) | 1 \le u \le U, 1 \le v \le V, 1 \le t \le T\}$, we have $U \times V \times T$ intensity values within a gray-scale level range $[0, L-1]$. In the following, the intensities over time at each pixel position are regarded as samples from a stochastic process. We define P as a *target pixel* at location (u, v). The location of P varies to cover all pixels of a frame, and its intensity sequence over time is denoted as $\{p_t(u, v)\}_{t=1,2,\dots,T}$. In the same way, we define $Q(u', v')$ as an *arbitrary pixel* with intensity sequence $\{q_t(u', v')\}_{t=1,2,\dots,T}$ at location (u', v'). For simplicity, we have omitted most of the (u, v) and (u', v') in the following discussion.

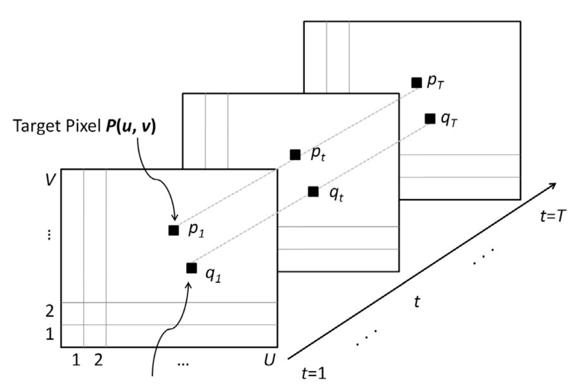

Fig. 1. Fundamental definitions of the image data. Target pixel P and an arbitrary pixel Q with their intensity sequence over time p_t and q_t.