Available online at www.sciencedirect.com

ScienceDirect

CAAI Transactions on Intelligence Technology 1 (2016) 43−60

http://www.journals.elsevier.com/caai-transactions-on-intelligence-technology/

Original article

Background modeling methods in video analysis: A review and comparative evaluation

Yong Xu [a,b,*], Jixiang Dong [a], Bob Zhang [c], Daoyun Xu [d]

[a] *Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, 518055 Shenzhen, China*

[b] *Shenzhen Key Laboratory of Network Oriented Intelligent Computation, Shenzhen, China*

[c] *Department of Computer and Information Science, University of Macau, Avenida da Universidade, Taipa, Macau, China*

[d] *The College of Computer Science and Technology, Guizhou University, Guiyang 550025, Guizhou, China*

Available online 4 June 2016

Abstract

Foreground detection methods can be applied to efficiently distinguish foreground objects including moving or static objects from background which is very important in the application of video analysis, especially video surveillance. An excellent background model can obtain a good foreground detection results. A lot of background modeling methods had been proposed, but few comprehensive evaluations of them are available. These methods suffer from various challenges such as illumination changes and dynamic background. This paper first analyzed advantages and disadvantages of various background modeling methods in video analysis applications and then compared their performance in terms of quality and the computational cost. The Change detection.Net (CDnet2014) dataset and another video dataset with different environmental conditions (indoor, outdoor, snow) were used to test each method. The experimental results sufficiently demonstrated the strengths and drawbacks of traditional and recently proposed state-of-the-art background modeling methods. This work is helpful for both researchers and engineering practitioners. Codes of background modeling methods evaluated in this paper are available at www.yongxu.org/lunwen.html.

Keywords: Background modeling; Video analysis; Comprehensive evaluation

1. Introduction

Foreground detection based on video streams is the first step in computer vision applications, including real-time tracking [1,2] and event analysis [3−6]. Many researchers in the field of image and video semantics analysis pay attention to intelligent video surveillance in residential areas, junctions, shopping malls, subways, and airports which are closely associated with foreground detection [7−9]. Background modeling is an efficient way to obtain foreground objects. Though background modeling methods for foreground detection have been studied for several decades, each method has its own strength and weakness in detecting objects of interest from video streams [10,11]. Therefore a comprehensive evaluation is needed to help researchers and practitioners choose suitable methods under different scenarios.

Over the past few decades, a large number of background modeling methods have been proposed to identify foreground objects in a video. They generally share the same following scheme [2,12]: they utilize the first frame or previous frames to build a background model, and then compare the current frame with the background model to detect foreground objects, and finally they update the background model. Various background modeling methods can be categorized into pixel-based, region-based, and hybrid methods. Background modeling methods can also be categorized into parametric and nonparametric methods. One of the most famous pixel-based parametric methods is the Gaussian model. Wren et al. [13] first proposed modeling the background at each pixel

* Corresponding author. Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, 518055 Shenzhen, China.

E-mail address: yongxu@ymail.com (Y. Xu).

Peer review under responsibility of Chongqing University of Technology.

http://dx.doi.org/10.1016/j.trit.2016.03.005

location with a Gaussian distribution [14,15]. However a single Gaussian function is not able to quickly deal with an actual dynamic background owing to a low updating rate of the background model [14]. In order to eliminate the influence of the background texture caused by waves on the water or trees shaken by the wind [15], Stauffer and Grimson [16,17] proposed the Gaussian mixture model (GMM) which models every pixel with a mixture of K Gaussians functions. After that, an improvement to GMM was proposed by using the online EM-based algorithm to initialize the parameter in the background model, which is time consuming. Zivkovic also [18,43] proposed an adaptive GMM (AGMM) to efficiently update parameters in GMM, and Lee [19] used a new adaptive learning rate to improve the convergence rate without changing the stability of GMM [9]. To improve the accuracy and reduce the computational time, Shimada et al. [20] used a dynamic Gaussian component to control the Gaussian mixture model. In addition, Oliver et al. [22] proposed a Bayesian method to model the background based on the prior knowledge and evidence from the data. Chien et al. [63] proposed a threshold decision method to detect foreground objects. They assumed the camera noise to be the zero-mean Gaussian distribution which is the only factor affecting the threshold. However, this assumption is hard to satisfy in practice.

Unlike parametric background modeling methods, a nonparametric algorithm based on self-organization through artificial neural networks (SOBS) was proposed by Maddalena et al. [30]. Kim et al. [28,29] proposed a codebook method to model the background which initializes codewords of codebooks to store background states. Wang et al. [55] proposed a method computing sample consensus (SACON) of the background samples to estimate a statistical model of the background, per pixel. SACON exploits both color and motion information to detect foreground objects. Barnich et al. [23,24] proposed a pixel-based nonparametric algorithm named Vibe to detect the foreground using a novel random selection strategy. The performance of Vibe is superior to many other state-of-the-art methods and it can represent exact background changes in recent frames [25]. The Vibe method was further studied by Van Droogenbroeck and Paquot [26] and they considered additional constraints to enhance the performance of Vibe. Another pixel-based nonparametric adaptive segmenter (PBAS) method was proposed by Hofmann et al. [27]. PBAS makes the foreground decision by applying a history of recently observed pixel values as the background model. Although, pixel-based background modeling methods can effectively obtain detailed shapes of foreground objects, they are easily affected by noise, illumination changes, and dynamic backgrounds.

Differing from pixel-based methods, region-based methods take advantage of inter-pixel relations to segment the images into regions and identify foreground objects from image regions. Elgammal et al. [21,41] presented a novel method by building a nonparametric background model based on kernel density estimation (KDE). Seki et al. [64] applied co-occurrence of image variations to model background changes in image regions. A heuristic block matching algorithm was proposed by Russell et al. [65] to distinguish foreground object from the background. They compared each image region of incoming frames with typical examples of a fixed-size database of backgrounds. In order to solve the dynamic background modeling in outdoor swimming pool environments, Eng et al. [66] used random homogeneous region movements and pre-filtering of image regions in the ClELab color space to detect foregrounds. In addition to methods featured by color, texture or descriptor-based methods also received much attention among region-based methods. Heikkila et al. [67] employed a discriminative texture feature called local binary pattern (LBP) [77] for modeling the background. They built LBP histograms based on partially overlapping regions for the background, and compared them with LBP histograms of each region of incoming frames via histogram intersection. Liu et al. [68] proposed a binary descriptor-based background modeling method to extract foreground objects under illumination changes. In addition, Huang et al. [69] modeled the background as samples of binary descriptors which can replace parametric distributions. In contrast to pixel-based methods, region-based methods can reduce the effects of noise, however, they can only obtain rough shapes of foreground objects.

Hybrid methods, which integrate both pixel-based and region-based methods, can achieve better background representation and deal with illumination changes and dynamic backgrounds [70]. The Wallflower system proposed by Toyama et al. [1] obtains the background model using pixel-level, region-level, and frame-level information. It applies the Wiener filter to predict background values at the pixel level, fills homogeneous regions of foreground objects at the region level, and handles sudden or global changes of a video sequence at the frame level. Huang et al. [71] integrated pixel-based RGB colors with optical-flow motions to model the background. Though hybrid methods can efficiently retrieve foreground objects from the background, their computational complexity is relatively high. Thus, Tsai et al. [72] proposed to embed hybrid algorithms in hardware to implement foreground detection. Some representative background modeling methods are classified in Table 1.

Table 1
Classification of representative background modeling methods.

Background modeling methods				
Category	Pixel-based methods		Region-based methods	Hybrid methods
Parametric	GMM [16]	AGMM [18,43]	Russell [65]	Huang [71]
	Oliver [22]		Heikkila [67]	Tsai [72]
Nonparametric	Vibe [23,24]	Schick [73]	KDE [21,41]	Cristani [75]
	SACON [55]	CodeBook [28,29]	Seki [64]	Chen [74]
	SOBS [30]	PBAS [27]	Liu [68]	Toyama [1]