Information Fusion 39 (2018) 154–167

Contents lists available at ScienceDirect

Information Fusion

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

Real-time foreground detection approach based on adaptive ensemble learning with arbitrary algorithms for changing environments

Yi-Tung Chan [a,*], Shuenn-Jyi Wang [b], Chung-Hsien Tsai [b]

[a] *School of Defense Science, Chung Cheng Institute of Technology, National Defense University, No. 75, Shiyuan Rd., Daxi Dist. Taoyuan City 33551, Taiwan, ROC*

[b] *Department of Computer Science and Information Engineering, Chung Cheng Institute of Technology, National Defense University, No. 75, Shiyuan Rd., Daxi Dist., Taoyuan City 33551, Taiwan, ROC*

A R T I C L E I N F O

Article history:
Received 29 May 2016
Revised 31 January 2017
Accepted 4 May 2017
Available online 5 May 2017

Keywords:
Arbitrary algorithm
Change detection
Ensemble learning
Foreground segmentation
Real-time applications
Stopped object detection

A B S T R A C T

Foreground detection technologies have emerged as an important research area with increasing popularity of computer vision and camera devices. Even though several foreground detection approaches have been proposed, they cannot address various challenges in actual complex scenes owing to their applicability and restrictions. This study proposes a method that can integrate arbitrary detection technologies to detect foregrounds in real time, thereby improving overall detection performance of video-based systems. Moreover, the proposed approach can be fully initialized with initial foreground results, requires no training, and performs dynamic adjustments online, for every new frame. In this approach, critical weighted values are automatically calculated over time based on observed scenes for optimal flexibility and parameterization. Thus, the proposed method has the flexibility to accommodate any new technology to overcome the challenging problems of foreground detection in changing environments. Experimental results demonstrate that the performance of the proposed method is comparable to that of state-of-the-art methods and satisfies the requirements of real-time practical applications.

1. Introduction

Computer vision has become an integral part of our lives owing to its current and potential applications in industries, military, space exploration, medical diagnosis, and multimedia systems. Foreground detection, which is an important research area in computer vision, is a critical step in change detection, video tracking, pattern recognition, and behavior analysis using image sequences. Furthermore, it plays a significant role in intelligent surveillance systems, elderly healthcare systems, childcare systems, infrastructure protection systems, and smart transportation systems, and has a significant and direct effect on promotion of video systems. Several foreground detection methods have been proposed to overcome problems in various applications. The most representative schemes include background subtraction, optical flow, and frame difference methods [1].

Background subtraction is the most commonly used method in video surveillance systems with fixed cameras. According to Borges et al. [2], background subtraction is a key approach that enables video surveillance systems to address problems such as occlusion, shadows, illumination changes, and dynamic backgrounds. Nevertheless, conventional background-model-based approaches cannot eliminate these problems completely. Therefore, Stauffer et al. [3] proposed a Gaussian mixture model (GMM)-based background subtraction scheme to overcome the drawbacks of shadows, illumination changes, and dynamic backgrounds through an adaptive learning mechanism. Hence, GMM-based background subtraction has been regarded as a standard step in video surveillance systems [3,4]. However, this method cannot completely overcome drastic illumination changes, dynamic backgrounds, and other challenges in actual environments because of the problems of initialization, update strategies, and parameter optimization in different scenes [4,5].

The strong temporal and spatial characteristics of the optical flow method allow for segmentation of moving objects through calculation of motion information in scenes [4,6]. The following two methods are commonly used to solve the optical flow equation: the Horn–Schunck method [7] and the Lucas–Kanade (LK) method [8]. Even though these methods can detect subtle pixel movements, they are considerably sensitive to changes in environments and produce a large amount of noise owing to the constant brightness assumption [6,9]. Thus, the optical flow method cannot segment a moving foreground completely [4].

* Corresponding author at: Department of Computer Science and Information Engineering, No. 75, Shiyuan Rd., Daxi Dist., Taoyuan City 33551, Taiwan, ROC.

E-mail addresses: eadown92@gmail.com (Y.-T. Chan), sjwang@ndu.edu.tw (S.-J. Wang), keepbusytsai@gmail.com (C.-H. Tsai).

http://dx.doi.org/10.1016/j.inffus.2017.05.001
1566-2535/

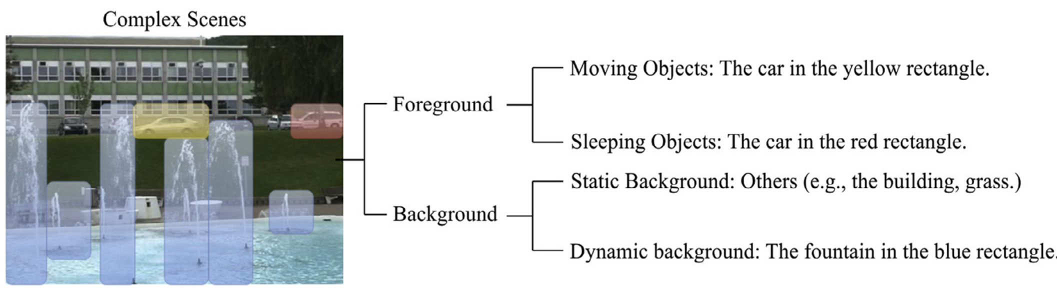

Fig. 1. Analysis of complex scenes in frame 1130 in the fountain01 sequence.

As discussed above, each foreground detection method has its applicability and restrictions because it has been developed under various original conditions and problems [10]. Applying these methods to real-world scenarios remains a challenging task because of variations in monitored scenes, such as illumination changes, dynamic backgrounds, and sleeping foreground objects. Fig. 1 shows an example of complex scenes involving four primary problems, i.e., moving objects, sleeping objects, static backgrounds, and dynamic backgrounds. It is not possible to overcome all of these problems by applying one algorithm to complex scenes because monitored scenes have different spatial and temporal properties [10]. For example, as shown in Fig. 1, model-based approaches can eliminate the dynamic background of the fountain by updating the learning mechanism with a high learning rate; however, this will cause the objects to fade quickly into the background. This is referred to as the stopped object problem [11], which is an inevitable trade-off involved in selecting parameters using one foreground detection technology.

Even though several foreground detection algorithms have been proposed to overcome the above-mentioned problems individually, a flexible integration of individual approaches to simultaneously address all key challenges in complex real-world environments has not been investigated extensively. Therefore, in this study, we propose an approach that integrates arbitrary foreground detection technologies, thereby improving overall accuracy and reducing overall computational load of video based systems. The theoretical foundation of this study is derived from the concept of ensemble systems in machine learning. Similar to different experts, each classification model in ensemble systems has different reasoning abilities for unknown samples. Thus, overall error can be significantly reduced, and the limitations of a classifier can be overcome by combining the outputs of these models, owing to their individual reasoning abilities. Well-known ensemble algorithms, such as bagging, boosting and AdaBoost, can be used to develop a strong classification model by training weak classifiers [12–15]. Polikar et al. [16] suggested that combining the outputs of several classifiers via averaging may reduce the risk of selecting a poorly performing classifier. Therefore, the proposed method can be adopted to combine different foreground detection technologies, such as background subtraction and optical flow, and the final foreground can be effectively detected using a fusion strategy. Misclassified pixels resulting from a particular technology can be prevented and overall detection performance can be improved. In addition, replacement of detection technologies depends on the application or scene, because their weights can be automatically calculated over time for optimal flexibility and parameterization in the proposed method. Experimental results confirm that the proposed method exhibits good performance for a change detection benchmark dataset [17].

The primary contributions of this study are as follows:

1) An innovative and interdisciplinary approach is developed, which emphasizes real-time foreground detection, which is the first step of the most practical importance in video-based systems, aimed at translating advances in ensemble learning and image processing into solutions that will benefit vision-based systems.
2) The proposed method can integrate arbitrary detection technologies to solve problems in a changing environment, as foreground detection in video-based systems should be considered in terms of independent user-specific practical applications. This is in contrast to conventional detection approaches that have insufficient flexibility in complex scenarios.
3) An adaptive weighted fusion mechanism, which can automatically calculate over time based on observed scenes for optimal flexibility and parameterization, is proposed to address problems in fusing heterogeneous algorithms.
4) With respect to practical considerations for applicability and operation in video-based systems, the proposed method can be initialized using initial foreground results, requires no training, and performs dynamic adjustments on-line, for every new frame.
5) The advantages of modular architecture include the ease with which different technologies may be added, updated, and removed from the proposed method.

The remainder of this paper is organized as follows: Section 2 reviews related works. Section 3 describes the proposed approach in detail. Section 4 presents the experimental results. Finally, Section 5 concludes the paper.

2. Related works

Ensemble learning is a significantly important research area for classification in machine learning. Ensemble approaches achieve enhanced performance by combining individual classifiers and utilizing their strengths [14]. This can be explained by the famous “no free lunch’ ’ theorem proposed by Wolpert [18], which states that there is no single classifier modeling approach that is optimal for all pattern recognition tasks, as each approach has its domain of competence [14]. The combination of a diversified pool of classifiers has been shown to improve overall system accuracy [19]. In recent years, ensemble learning algorithms have been successfully applied to video-based systems for video surveillance [20,21], face recognition [15,19,22–24], person re-identification [20,25,26], anomaly detection [27], image segmentation [28,29], video tracking [30], and intelligent transportation systems [31]. However, Visentini et al. [30] showed that the concept of online classification has received significant attention in recent years in the computer vision community. Even though several researchers have investigated the development of ensemble-based methods in video-based systems, there still exist the following four critical issues: selecting diverse classifiers, efficiently fusing heterogeneous information from different modalities, reducing influence of complex environments, and training it with incoming samples in an unsupervised manner and without any prior knowledge of data distribution [27,30].