J. Vis. Commun. Image R. 39 (2016) 181–195

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Robust techniques for abandoned and removed object detection based on Markov random field ☆

Chih-Yang Lin [a,b], Kahlil Muchtar [c], Chia-Hung Yeh [c,*]

[a] Dept. of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan
[b] Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan
[c] Dept. of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 11 January 2016
Accepted 30 May 2016
Available online 31 May 2016

Keywords:
Abandoned object detection
Background modelling
GMM
Markov random field

ABSTRACT

This paper presents a novel framework for detecting abandoned objects by introducing a fully-automatic GrabCut object segmentation. GrabCut seed initialization is treated as a background (BG) modelling problem that focuses only on unhanded objects and objects that become immobile. The BG distribution is constructed with dual Gaussian mixtures that are comprised of high and low learning rate models. We propose a primitive BG model-based removed object validation and Haar feature-based cascade classifier for still-people detection once a candidate for a released object has been detected. Our system can obtain more robust and accurate results for real environments based on evaluations of realistic scenes from CAVIAR, PETS2006, CDnet 2014, and our own datasets.

1. Introduction

Recently, there have been a number of terrorist attacks in public areas such as railway stations, bus stations, airports, and outdoor competition venues. Therefore, smart surveillance systems have become a hot topic within computer vision research. Many methods have been proposed to detect suspicious objects in crowded environments. In general, these objects are classified into two main groups: suspicious moving objects (e.g., a person trying to harm others, a person behaving abnormally, etc.), and immobile dangerous objects (e.g., abandoned luggage, backpack, etc.). In this paper, we focus on the latter group, which is typically referred to as the abandoned object (AO) detection system.

Methods that are used to solve the AO problem can be grouped into two classes: tracking-based approach and detection-based approach [13]. The tracking-based approach often fails in crowded scenes due to problems like merging, splitting, entering, corresponding, leaving, and occlusion. In recent years, the detection-based approach has become more widely used than the tracking-based approach. For instance, Porikli et al. [16], proposed a pixel-based method that uses dual foregrounds. Their rationale is that adjusting the background learning rate can help determine how long a static item should be regarded as background. Unlike [20] approaches that mainly utilize Gaussian mixtures, Porikli et al. used an online Bayesian approach that calculates the probability distributions of mean and variance. Compared to tracking-based approaches, which suffer from failures in crowded scenes, their method can segment the *AO* more accurately. However, the method has no specific handling to distinguish whether the detected static region corresponds to an abandoned or removed item. Another detection framework, this time designed by Tian et al. [21], employs abandoned and removed object detection based on BG subtraction and foreground analysis. Object tracking is then utilized as an additional task to reduce false alarms. Nevertheless, fragmented static foreground masks are often generated due to the imperfections in BG modelling. Ferryman et al. [7] integrated several existing methods consisting of person detection, luggage detection [16], and object tracking to infer whether something is regarded as a threat or not, while also reducing the number of false alarms. Their proposed technique, however, combines BG subtraction and object tracking, which incur a relatively high computational cost.

The detection-based-only approach is prone to noise and misleading results. A common problem in BG modelling methods is that the shadow and quick-light changes are incorporated during the detection phase. Inaccurate results are likely when directly performing connected component processing on an *AO* mask. Moreover, in some circumstances, only enabling common shadow removal can weaken the texture and degrade the details of an

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: Dept. of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan.

E-mail address: yeh@mail.ee.nsysu.edu.tw (C.-H. Yeh).

http://dx.doi.org/10.1016/j.jvcir.2016.05.024
1047-3203/

object. As illustrated in Fig. 1(a) and (b), when overlapping *AO*s exist, the connected component fails to get a precise interpretation of each object. In addition, Fig. 1(c) and (d) shows a failure case where two adjacent objects are detected together (only the red box is an *AO*), because shadow is also included in the region of interest (ROI). Apart from noise, still-people close to the observed object and an *AO* that later is removed by the owner can further contribute to inaccurate results. To solve the aforementioned issues that greatly influence the robustness and accuracy of an *AO* system, we propose a fully-automatic GrabCut framework that is specifically designed for *AO* problems. GrabCut, an MRF-based segmentation model, has shown a high degree of robustness in computer vision labelling problems. The Markov random field (MRF) model [8] is a graphical model that represents a joint probability distribution. Since a pixel and its adjacent pixel are likely to have similar probabilities, the joint probability function is built over the pixels and neighbourhood system. First, we define the set of GrabCut seed initialization by means of a BG modelling-based *AO* detection system. To distinguish an *AO* from other objects, the removed object validator and still-people checker are then utilized. Finally, we propose an extension of temporal coherence to preserve the historical GMM and cost during iterative energy minimization. By replacing the connected component process with fully-automatic GrabCut segmentation, the ROI can be precisely extracted.

Recently, the CDnet 2014 benchmark dataset [9] has been widely used to evaluate video-based change detection algorithms. Instead of only providing test videos, it carefully provides a hand-annotated foreground, background and shadow area for each frame. As shown in [10,15,24,25], the authors present the state-of-the-art and precise comparisons comprised of qualitative and quantitative measurements.

This paper is organized as follows: Section 2 presents a brief review of GMM, Section 3 explains the detail of the proposed algorithm and framework, Section 4 demonstrates our experimental results and Section 5 summarizes our work.

2. Related work

2.1. Overview of GMM

The Gaussian Mixture Model (GMM) was proposed by Grimson and Stauffer [20]. The authors presented a pixel-based method to model each pixel (regarded as background) into a mixture of Gaussians. The number of Gaussians K is typically set from 3 to 5. In addition, each Gaussian has its own weight to represent the portion of the data accounted for the corresponding distribution. The probability that a pixel regards a value z at a certain time Z_t is given as follows [20]:

$$P(Z_t) = \sum_{j=1}^{K} \omega_{j,t} * \eta\left(Z_t, m_{j,t}, \sum_{j,t}\right), \quad (1)$$

where K is the number of Gaussian distributions, $\omega_{j,t}$ is the weight estimation of the j_{th} Gaussian in the mixture at time t, $m_{j,t}$ and $\sum_{j,t}$

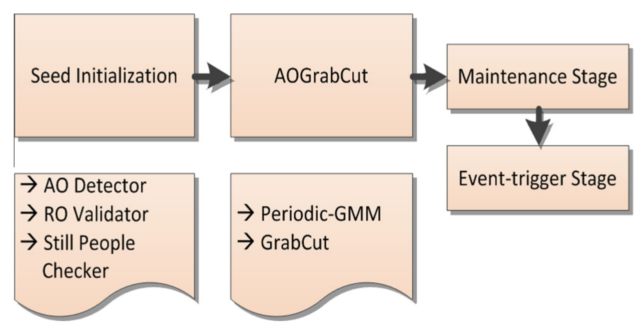

Fig. 2. System diagram of the proposed method.

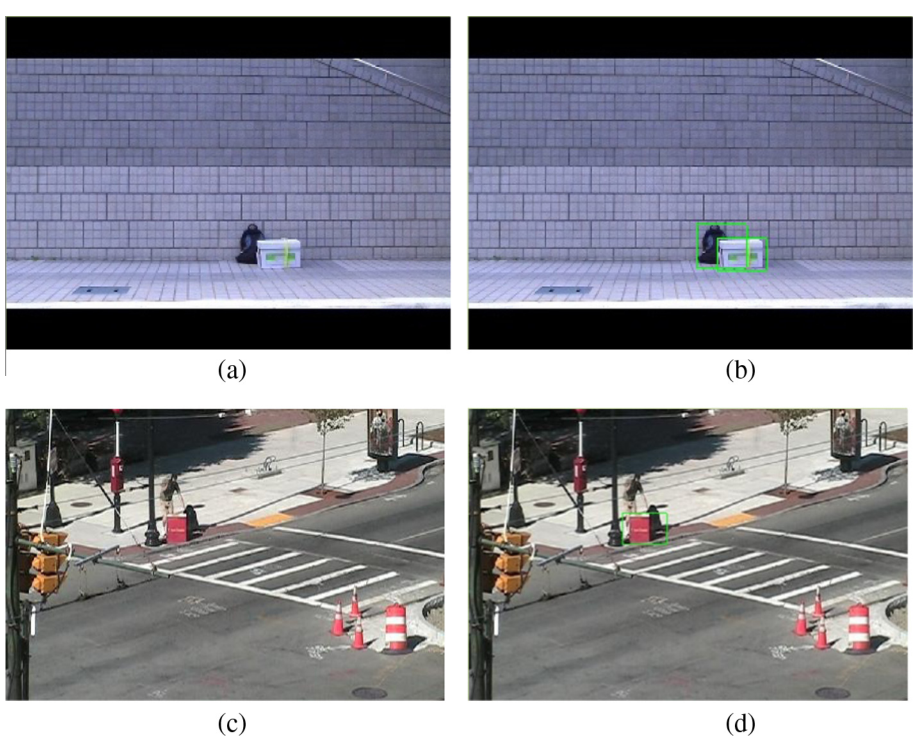

Fig. 1. The illustration of inaccurate detection cases, where dual pure GMM and connected component are used to obtain the ROI (region of interest). The (a)–(b) shows the case of two overlapped objects detection whereas (c)–(d) is the detection of abandoned box.