Pattern Recognition Letters 33 (2012) 1558–1568

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Enhanced foreground segmentation and tracking combining Bayesian background, shadow and foreground modeling

Jaime Gallego [a,*], Montse Pardàs [a], Gloria Haro [b]

[a] Universitat Politècnica de Catalunya (UPC), Department of Signal Theory and Communications, Barcelona, Spain
[b] Universitat Pompeu Fabra, Department of Information and Communications Technologies, Barcelona, Spain

ARTICLE INFO

Article history:
Received 12 July 2010
Available online 23 May 2012

Communicated by G. Borgefors

Keywords:
Foreground segmentation
Space-color models
Shadow model
Objects tracking
GMM

ABSTRACT

In this paper we present a foreground segmentation and tracking system for monocular static camera sequences and indoor scenarios that achieves correct foreground detection also in those complicated scenes where similarity between foreground and background colours appears. The work flow of the system is based on three main steps: An initial foreground detection performs a simple segmentation via Gaussian pixel color modeling and shadows removal. Next, a tracking step uses the foreground segmentation for identifying the objects, and tracks them using a modified mean shift algorithm. At the end, an enhanced foreground segmentation step is formulated into a Bayesian framework. For this aim, foreground and shadow candidates are used to construct probabilistic foreground and shadow models. The Bayesian framework combines a pixel-wise color background model with spatial-color models for the foreground and shadows. The final classification is performed using the graph-cut algorithm. The tracking step allows a correct updating of the probabilistic models, achieving a foreground segmentation that reduces the false negative and false positive detections, and obtaining a robust segmentation and tracking of each object of the scene.

1. Introduction

Accurate and robust segmentation and tracking of moving objects in dynamic and cluttered visual scenes is a big challenge in computer vision. It is needed in video applications where the output depends completely or partially on the visualization of the segmentation. For instance, in video surveillance applications in order to allow a correct identification and tracking of the objects of interest. In 3D multi-camera environments, robust foreground segmentation allows a realistic 3-dimensional reconstruction without background artifacts, which can be used, for example, for free-viewpoint video. In addition, such systems are the building blocks of higher-level intelligent vision-based or assisted information analysis and management systems with a view to understanding the complex actions, interactions, and abnormal behaviors of objects in the scene. Therefore, the better the foreground segmentation, the higher the quality and performance of these kind of applications.

In this paper we focus on applications with a fixed camera and indoor smart-room scenarios used for a posterior 3-dimensional reconstruction. Our objective is to obtain an accurate segmentation and tracking of the foreground objects robust to shadow effects and camouflage problems. For this purpose, we propose a system that combines an accurate foreground objects detection with a tracking algorithm, which allows the segmentation of several objects of interest within the scene in analysis.

1.1. Techniques based on background modeling

Over the recent years there have been extensive research activities in proposing new ideas, solutions and systems for robust object segmentation and tracking to address the above situations. Most of them adopt the background subtraction as a common approach for detecting foreground moving pixels, whereby the background scene structures are modeled pixel-wise by various statistically-based learning techniques on features such as intensities, colours, edges, textures, etc. A pixel is classified as background when its value is not correctly modeled by the background model, in the so called exception to background analysis. The models employed include mono-modal Gaussians (Jabri et al., 2000), Gaussian mixture model (GMM) (Stauffer and Grimson, 2000), nonparametric Kernel density estimation (Elgammal et al., 2000), or simply temporal median filtering (Zhou and Aggarwal, 2001).

Optionally, shadow removal techniques can be incorporated in the background subtraction/modeling step to improve the segmentation removing the false positives detections that illumination problems produce. Prati et al. (2003) have presented an in-depth

* Corresponding author.
E-mail addresses: jgallego@upc.edu, jgallego@gps.tsc.upc.edu (J. Gallego), montse @upc.edu (M. Pardàs), gloria.haro@upf.edu (G. Haro).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.05.004

survey of these algorithms while Xu et al. (2005) propose the hybrid shadow removal method that we have used as the initial step for shadow modeling.

After foreground detection, a connected component analysis (CCA) is usually performed in order to cluster and label the foreground pixels into meaningful object blobs, from which some inherent appearance and motion features can be extracted. Finally, there is a blob-based tracking process aiming to find persistent blob correspondences between consecutive frames. Several authors employ this kind of solution (Porikli and Tuzel, 2003; Gabriel et al., 2003; Chen et al., 2005; Xu et al., 2004;Gallego et al., 2008).

The main problem of these algorithms is that false negatives appear when foreground and background present color similarities. False positives can also be observed when an external agent modifies the configuration of the scene (illumination changes, shadow effects or spatial alterations of the background objects configuration). The trade-off between false positive and false negative detections, makes it difficult to solve this problem using only the techniques explained above. Furthermore, none of these proposals uses feedback between the foreground detection and the tracking process, to improve the updating of the models in order to avoid the propagation of wrong detections along the sequence.

1.2. Techniques based on foreground modeling

Background subtraction techniques only require the construction of a background model. However, if a foreground model is available, a Bayesian approach for foreground segmentation and tracking can be performed with the objective to improve the segmentation of the foreground object. In order to create the models, an initial segmentation is usually performed using an exception to background method, and once there is sufficient evidence that the foreground entities are in the scene, foreground models are created.

Several foreground models have been proposed in the past for different purposes including the foreground segmentation task (Khan and Shah, 2000; Mittal and Davis, 2003; Li et al., 2004), or object and person trackers where the foreground has been previously segmented (McKenna et al., 1999; Elgammal et al., 2000). As with background models, foreground models are Gaussian-based in most of the cases. Different alternatives are: single-Gaussians (Wren et al., 2002), GMM (McKenna et al., 1999; Khan and Shah, 2000), and nonparametric models with Gaussian kernels (Mittal and Davis, 2003; Sheikh and Shah, 2005). In (Khan and Shah, 2000) people are first segmented with the exception to background approach and tracked by segmenting them into classes of similar color (initialized by expectation maximization (EM) (Dempster et al., 1977)). Each pixel is assigned in the following frames to the class that maximizes the probability of the pixel to belong to that class (including a class for the background). Means and variances of the classes are updated after classification. However, the partition of the object in regions modeled by independent Gaussians is too rigid and prone to errors. The work in (McKenna et al., 1999) uses a GMM to model the color distribution of the objects to track and EM to update its distribution. Since the objective is to track a single object, a background model is not used and thus a complete segmentation is not achieved. In (Yu et al., 2007) a GMM for modeling both the foreground and background, in spatial and color domains, is used. The models are first initialized using a reference frame and the background and foreground models are adjusted using the EM algorithm. This kind of algorithms, with iterative processes, present a high computational cost that doesnot allow a real time sequence analysis. Moreover, in case of a complex background, and even using a GMM with a very high number of Gaussians, the foreground can occupy background regions of similar color which become close to its position as the object moves along the scene.

In the literature, none of these systems propose to combine this approach with tracking methods, because it is assumed that foreground modeling allows a good segmentation and tracking for itself. However, as it has been said above, there is certain difficulty to correctly maintain a good foreground model in some scenarios where foreground and background present color similarities.

Moreover, a specific model for the shadow of each object can be constructed using the tracking information and an initial shadow detection. This allows to make the foreground/background segmentation within a Bayesian framework, using a background model and specific foreground and shadow models for each object and its shadow.

The segmentation system that we propose, combining foreground detection with an object tracking algorithm, follows the work flow of Fig. 1. It consists of three main blocks: pixel-wise foreground segmentation, objects tracking, and foreground segmentation based on spatial-color models.

1.3. Proposed method

All the system runs as a complex implementation of the simple concept of surveillance: be aware for external changes, detect and track objects and refine the object detection improving the knowledge about it and focusing the attention in its region.

- **Pixel-wise foreground segmentation:** This initial step is used as a first glance at the foreground objects that appear in the scene. It also segments shadow pixels to create a shadow model for each detected object.
- **Objects tracking:** It is used to detect and track those objects that appear in the scene, matching the blobs detected in the first segmentation with the objects that are being tracked. It assigns the detected blobs to objects with a label that characterizes them along the time and brings us the valuable spatial information about the position and size of the object in the scene. A region of interest (ROI) is obtained for each object to track, and it is used for appropriate background and foreground models updating and for associating, in the next step, each foreground model with its corresponding object. The method proposed uses a classical mean-shift tracking method with the following improvements: several connected components association to each object (it avoids false positive detections when an object is segmented in several connected components), detection and solving of objects occlusion (analyzing the connected components detected in each frame), focus the position estimation in those regions that belong to the foreground and incorporation to the background of all the foreground detections, not belonging to the objects in analysis, which appear outside the defined ROI.
- **Foreground segmentation based on spatial-color model:** Here a final enhanced foreground segmentation of each object is obtained, combining in a Bayesian framework spatial-color models of the foreground and shadows regions with the pixel-wise color model of the background. The foreground and shadow models are obtained using preliminary shadows and foreground masks, the position of each object, and the background model, all obtained in the previous two steps. The novelty of this approach resides in the combination of a pixel-wise background model with foreground and shadow spatial models within a MAP-MRF framework. We associate a spatial-color GMM foreground and shadow models to each object that is being tracked in the scene, assuming that the shadow effect that each object produces is an attribute of the object that produces the background color change. Novel updating techniques