Available online at www.sciencedirect.com

SCIENCE DIRECT

Image and Vision Computing 24 (2006) 1192–1201

image and vision COMPUTING

www.elsevier.com/locate/imavis

Motion detection and tracking using belief indicators for an automatic visual-surveillance system

Cina Motamed

Laboratoire LASL, Université du Littoral Côte d'Opale, 50 R. F. Buisson Bat. B, 62228 Calais, France

Received 7 April 2003; received in revised form 6 June 2005; accepted 7 June 2005

Abstract

A motion detection and tracking algorithm for human and car activity surveillance is presented and evaluated by using the Pets'2000 test sequence. The proposed approach uses a temporal fusion strategy by using the history of events in order to improve instantaneous decisions. Normalized indicators updated at each frame summarize the history of specific events. For the motion detection stage, a fast updating algorithm of the background reference is proposed. The control of the updating at each pixel is based on a stability indicator estimated from inter-frame variations. The tracking algorithm uses a region-based approach. A belief indicator representing the tracking consistency for each object allows solving defined ambiguities at the tracking level. A second specific tracking indicator representing the identity quality of each tracked object is updated by integrating object interaction. Tracking indicators permit to propagate uncertainties on higher levels of the interpretation and are directly useful in the tracking performance evaluation.

Keywords: Video-surveillance; Motion detection; Tracking; Belief indicator; Belief updating; Uncertainty management

1. Introduction

The development of vision systems carrying out the monitoring of sites is an interesting field of investigation. Indeed, motivations are multiple and concern various domains as the monitoring and surveillance of significant protected sites, the control and estimation of flows (car parks, airports, ports, and motorways). Because of the fast evolution in the fields of data processing, communications and instrumentation, such applications become possible.

An automatic visual-surveillance system generally contains several main hierarchical modules. The motion detection, the tracking procedure, the object classification and finally the high-level motion interpretation. Main surveillance applications are object motion indexing, activity recognition and incidents detection [30].

The majority of real applications use a fixed camera. Vision modules for visual surveillance are optimised in order to contribute to the real time response of the global system. The performance of such a system is generally context-dependent and may be compared to a specific application with respect to:

- the quality of decisions,
- trade-off between quality and computing time,
- robustness with respect to small perturbations.

In fact, visual-surveillance is a vision application with a high degree of dependence to the context. In many applications a top-down approach has to be envisaged by modelling what we want to observe and the contextual knowledge has to be well highlighted in order to be integrated at each level of the interpretation system [4,5,20].

Tracking is an important task of computer vision with many applications in surveillance, scene monitoring, navigation, sport scene analysis, and video database management. In vision community, two main tracking approaches are proposed: the model-based and feature-based approaches.

The objective of the model-based approach is to exploit a model of the tracked object. This model can be a 2D template of the object [21], such as active contours [13] or a statistical appearance representation [7,23] or a CAD model [15,16]. Active contour models, or snakes, permit an efficient representation of the updated bounding contour of

E-mail address: motamed@lasl.univ-littoral.fr

0262-8856/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2005.06.005

the object. The advantage of having an active contour-based representation is its robustness with respect to low-level segmentation and its efficient data reduction. However, it requires a good initialisation. In the scope of visual surveillance applications, many models of humans and vehicles have been proposed. The human can be modelled by means of a stick figure, a 2D contour or a specific volumetric model [1]. For vehicles, we can use a deformable grey-scale template [3], or a 3D deformable model of vehicle [15,11]. The disadvantages of a 2D model, with respect to a 3D model, are firstly its restriction to the camera's angle and secondly its sensitivity to occlusion and lighting conditions. The main draw back of a 3D model is that it requires more parameters and leads to more expensive computation during the matching process.

The feature based tracking method uses low level sub-features such as distinguishable points or lines on the object [24]. The advantage of this approach is in the presence of partial occlusion when some of the sub-features of the tracked objects remain visible [22]. The principal difficulty of low level sub-features is related to the problem of their grouping for the object level elaboration. The region-based approach is widely used in the visual surveillance field, because it brings many advantages [18,19,31]. Moving regions are easily identified by using motion information when a fixed background is present. Detected regions contain important information as colour, texture, position orientation, etc. They are efficient for object tracking because they are associated to the global position and the area of objects.

This work is concerned with the detection and tracking of objects in an outdoor car park environment by using a fast region-based approach. Many uncertain events are likely to occur when the system has to operate outdoors and in unconstrained scenes containing multiple objects. One important aspect is firstly the need to understand uncertainty in terms of sources of uncertainty. Because generally an appropriate way of uncertainty management is to focus on its source. The main difficulties concern: illumination changes and shadow artefacts at the detection level, and occlusions at the tracking level. These problems are very harmful for surveillance tasks because they induce target missing, false detection and object identity confusion.

It is important for visual-surveillance algorithms to have the capacity to manage efficiently the uncertainty of their decisions. A common technique for representing uncertainty is to qualify all decisions by attaching a belief indicator representing a confidence factor. Two main principles of uncertainty management exist. The principle of minimum uncertainty is commonly used in estimation theory. The strategy is to discard the less likely conclusions in order to highlight the best ones. On the other hand, the principle of maximum uncertainty is essential for any problem involving secure reasoning. It guarantees that the ignorance has been fully recognized and propagated. Our approach is mainly based on the second principle. In fact, for a surveillance system, it is crucial to incorporate the notion of 'positive security' by an autonomous recognition of its possible inability to decide. In the context of visual tracking, if the tracker can not maintain efficiently the identity of a set of objects over their tracks, it is more cautious for the high level interpretation to propagate all possible alternatives in order not to lose real critical events. A major drawback of such secure approaches is that it may increase the false alarm rate.

Physical events observed by visual surveillance systems are typically continuous and by accumulating evidence from successive observations, decisions may be improved by focusing on consistent events. This evidence accumulation approach is also called temporal fusion. We have tried to integrate it into the motion detection and tracking modules.

In Section 2, we present the motion detection process. It is based on a difference between the current image and a reference image. In Section 3, we expose the tracking algorithm. All presented algorithms are evaluated over the 'test sequence' proposed in the framework of the PETS'2000 (IEEE Workshop on the Performance Evaluation of Tracking and Surveillance). This test sequence contains some of the typical ambiguities often appearing in video surveillance applications. Algorithms are tested with an image resolution of 384×288 pixels. Global tracking results are summarized and discussed in Section 4.

2. Motion detection process

2.1. Introduction

For fixed cameras, the standard motion detection approach is to model the stationary background [12,10, 23]. In this case, the moving objects representing the foreground are easily extracted by a simple difference between the current image and the background.

A motion detection algorithm is typically composed of three tasks:

• Reference background generation

The background reference model R is updated at each new acquisition. Generally, it uses a recursive filter, which limits image storage. Classical background updating is computed for each pixel p as:

$$R_{k+1}(p) = a \times R_k(p) + (1-a) \times I_k(p) \quad (1)$$

R^k and I^k represent respectively the background and the current observation at frame k. The parameter a controls the updating process. This updating is made generally over the whole image and permits to absorb progressive illumination changes. This technique is widely used in highway traffic monitoring where objects continuously move.

• Combination operator

It computes a distance between the observation and the background reference.