Expert Systems with Applications 38 (2011) 12791–12806

Contents lists available at ScienceDirect

Expert Systems with Applications

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

OCULUS surveillance system: Fuzzy on-line speed analysis from 2D images

J. Albusac [a,*], D. Vallejo [b], J.J. Castro-Schez [b], L. Jimenez-Linares [b]

[a] School of Technical Engineering, Plaza Manuel Meca 1, 13400 Almaden, Ciudad Real, Oreto Research Group, University of Castilla-la Mancha, Spain
[b] School of Computer Science, Department of Information Technologies and Systems, Paseo de la Universidad 4, 13071 Ciudad Real, Oreto Research Group, University of Castilla-la Mancha, Spain

ARTICLE INFO

Keywords:
Behaviour analysis
Speed classification
Anomaly detection
Surveillance system
Machine learning

ABSTRACT

This paper presents an independent component integrated into a global surveillance system named as OCULUS. The aim of this component is to classify the speed of moving objects as normal or abnormal in order to detect anomalous events, taking into account the object class and spatio-temporal information such as locations and movements. The proposed component analyses the speed of the detected objects in real-time without needing several cameras, a 3D representation of the environment, or the estimation of precise values. Unlike other works, the proposed method does require knowing the camera parameters previously (e.g. height, angle, zoom level, etc.). The knowledge used by this component is automatically acquired by means of a learning algorithm that generates a set of highly interpretable fuzzy rules. The experimental results demonstrate that the proposed method is accurate, robust and provides a real-time analysis.

1. Introduction

In the last two decades, the new technologies have greatly influenced the design of surveillance systems deployed in security control centers. As time passes, these systems are more and more robust when monitoring tasks, reducing the human workload and avoiding dangerous situations. Computer Vision (Forsyth & Ponce, 2002) and Artificial Intelligence (Russell & Norvig, 2003) techniques are playing a key role in the evolution of this kind of systems (Blauensteiner & Kampel, 2004; Bloisi, Iocchi, Remagnino, & Monekosso, in press; Haritaoglu, Harwood, & Davis, 2000; Valera & Velastin, 2005; Velastin, Khoudour, Lo, Sun, & Vicencio-Silva, 2004). A good example can be found in the video analysis field, where one of the main challenges is to develop security expert systems with the autonomy and ability required to automatically understand events and behaviours in order to improve the productivity and effectiveness of surveillance tasks.

In complex environments where multiple situations take place simultaneously, human agent operators have to deal with all of them, being affected by negative factors such as fatigue or tiredness after a prolonged period of observation (Smith, 2004). Nevertheless, artificial expert systems do not have these limitations due to their processing capabilities. Furthermore, this kind of systems can be more effective than people when recognising certain classes of events, such as the detection of suspicious or unattended objects (Dee & Velastin, 2008).

Anomaly detection in real environments may imply the analysis of multiple factors such as trajectory, speed, location, and spatial relationships among objects. In this work, we face the problem of speed analysis from 2D images captured by surveillance cameras. The speed of an object is an important aspect to take into account because a high speed usually represents an abnormal situation. For instance, a vehicle quickly leaving a particular area may be an indication that something is going wrong, or a person running through an area where people normally walk may also be a problem to be detected.

The speed of an object at a particular environment is considered as normal whether its value belongs to the interval in which the allowed limits for that environment have been defined. These limits depend on the characteristics of the environment to be analysed, and they can be upper and lower limits. An object normally behaves according to the speed aspect, whether it does not exceed the upper limit and moves at a speed greater than the lower one. Both kind of limits are not always present in all environments. For example, a lower limit is not common in urban areas where people and vehicles can be stopped without involving an abnormal situation. However, this kind of limit exists in highways where vehicles must overtake a minimum speed.

Many authors have addressed the analysis of speed from 2D images. An easy alternative consists on using aerial views to study the objects displacements in a concrete time interval (Liu, Yamazaki, & Maruyama, 2007). However, it is often difficult to get this kind of images. A second alternative is focused on building a 3D representation of the environment from several cameras in order to know the exact position and movement of each moving object (Cheung, Kanade, Bouguet, & Holler, 2000; Hu, Wang, & Uchimura,

* Corresponding author. Tel.: +34 617342824; fax: +34 926 29 53 54.
E-mail address: JavierAlonso.Albusac@uclm.es (J. Albusac).
URL: http://www.esi.uclm.es/www/jalbusac/ (J. Albusac).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.04.071

2008; Lee, Romano, & Stein, 2000). Many of these methods are computationally expensive and the use of several cameras is an expensive solution. Other authors have proposed methods to overcome the same problem by using a single camera (Beymer, McLauchlan, Coifman, & Malik, 1997; Cathey & Dailey, 2005; Maduro, Batista, Peixoto, & Batista, 2008; Palaio, Maduro, Batista, & Batista, 2009). In this kind of methods, a calibration process is normally carried out in which parameters such as height, zoom level or angle camera must be previously known. Afterwards, the correspondence between the location of each object in the 2D image and its location in the real world is established.

In this work, we are not interested in measuring the exact speed of each detected object, but to provide a fast response when distinguishing between normal and anomalous speeds in the same way that humans do (Zadeh, 2001). A person does not need to know precise values or perform complex calculations when monitoring a concrete scene in order to determine whether an object is moving fast or slow. In this case, the displacements in the 2D image are analysed by taking into account the perspective of the camera and the static elements of the environment as reference. Medium displacements made by objects in areas which are far away from the camera may represent a high speed. On the contrary, the same displacements in areas close to the camera may involve a normal speed. In the next sections, we propose fuzzy methods (Zadeh, 1965) to learn and analyse the normal displacements of moving objects in 2D images, considering the perspective of a single camera and without previously knowing the camera parameters. This approach allows the surveillance system to distinguish between normal and anomalous speeds.

The rest of this paper is organised as follows: Section 2 discusses several approaches and previous work related to speed analysis from visual information. Section 3 summarises the architecture of the surveillance system in which the speed analysis is carried out and a formal model to define new analysis components. In Section 4, we describe a novel method to classify the speed of moving objects as normal or abnormal, and how to integrate it into the global system. The experimental results are presented in Section 5. Finally, this paper is concluded in Section 6.

2. Previous work

As mentioned before, there are three relevant approaches to estimate the speed of moving objects from video information: (i) analysis of speed from aerial pictures, (ii) generation of a 3D representation of the monitored environment and, finally, (iii) use of a single camera in calibration processes and geometric methods to establish correspondences between 2D images and a portion of the real world.

Possibly, the first one is the simplest of these three approaches since it is not necessary to deal with the perspective problem and the number of occlusions is reduced. In this case, the movements of moving objects detected in a 2D image can be easily matched with the actual movements in the real world. Within this context, Liu et al. (2007) described a method to estimate the speed of vehicles on highways. To do that, the authors manually built a scene model in which it is possible to distinguish the highways from the rest of elements. Secondly, they make use of two consecutive aerial frames to establish correspondences between objects and calculate their speeds on the basis of their movements. Although this approach is an effective way to determine the speed, using aerial cameras it is not always possible, so other methods are required.

On the other hand, when the cameras do not provide aerial views, it is necessary to address the perspective problem, i.e., there is no direct equivalence between movements made in the 2D image and the actual ones. The approaches (ii) and (iii) deal with this problem. An expensive solution, but very accurate, is to build a 3D environment representation by using several 2D images obtained from different cameras. Cheung et al. (2000) proposed a system composed of five cameras which is able to perform 3D reconstruction of moving humans in real time. The system individually extracts the silhouettes of the moving people in each image and uses this information for generating a 3D reconstruction. Lee et al. (2000) proposed a method for monitoring activities from multiple video streams. Once the 3D model is built, the authors compare the speed and distance between objects in different parts of the scene. Hu et al. (2008) developed a system for traffic monitoring by data fusion from multiple stationary cameras. A probability fusion map is proposed to estimate the speed of vehicles.

Finally, many single-camera methods have been proposed as an alternative to the previously mentioned techniques. Cathey and Dailey (2005) presented a method for estimating the speed of vehicles on highways in Seattle. The cameras can be remotely controlled and every time their parameters are modified, a fast calibration process is performed. This process automatically detects the boundary lines of the road, which are used to determine the perspective of the scene. The proposed method is able to reduce the problem of perspective since it generates aerial views of the monitored scene by means of geometric methods. Once these images have been generated, the displacements of the vehicles are analysed to estimate their speeds in a third phase. Other similar methods in which the boundary lines of a highway are automatically detected, are discussed in Beymer et al. (1997), Palaio et al. (2009). Maduro et al. described in Maduro et al. (2008) how the previously cited methods lose their effectiveness when there are occlusions between objects, long distances covered by vehicles in few frames or low image resolution, among other reasons. They made a new proposal based on two previous methods (Cho & Rice, 2006; Magee, 2004). This new method can estimate the speed of a vehicle although it was not correctly detected by the tracking process.

The above methods have in common the calculation of a precise speed. From this calculated value, it is possible to know if an object moves between the allowed limits. A person watching a video is able to quickly determine whether a particular object is going too slow or fast depending on the area where it is located. However, this person would not be able to determine a precise value of the speed. In other words, a person usually refers to the speed and location of an object in terms of linguistic labels such as very left, very right, center, up, very down, etc. in the case of location, and very slow, slow, medium speed, fast or very fast, in the case of speed. This is the way in which we want our expert system to behave.

In the next section, we will briefly summarise the architecture of OCULUS surveillance system and the model on which it is based. OCULUS is a scalable and flexible surveillance system consisting of several independent normality components, that increases its analysis capability when new components are designed and included (Albusac, Vallejo, Jimenez-Linares, Castro-Schez, & Rodriguez-Benitez, 2009). A normality component specifies how each kind of object must behave according to a surveillance aspect such as trajectory or speed. The system includes a normality component for each surveillance aspect to be monitored and they can be enabled or disabled depending on the surveillance requirements. Besides, it combines the output of each component by means of aggregation operators in order to obtain a global interpretation of the current state of the environment. After describing the OCULUS architecture and the model on which it is based, the normality component designed to analyse the speed of objects will be presented in Section 4.