Computer Vision and Image Understanding 144 (2016) 166–176

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Online real-time crowd behavior detection in video sequences

Andrea Pennisi, Domenico D. Bloisi*, Luca Iocchi

Department of Computer, Control, and Management Engineering Sapienza University of Rome via Ariosto 25, Rome,00185 Italy

ARTICLE INFO

Article history:
Received 20 December 2014
Accepted 21 September 2015

Keywords:
Event detection
Crowd analysis
Image segmentation
Intelligent surveillance

ABSTRACT

Automatically detecting events in crowded scenes is a challenging task in Computer Vision. A number of offline approaches have been proposed for solving the problem of crowd behavior detection, however the offline assumption limits their application in real-world video surveillance systems. In this paper, we propose an online and real-time method for detecting events in crowded video sequences. The proposed approach is based on the combination of visual feature extraction and image segmentation and it works without the need of a training phase. A quantitative experimental evaluation has been carried out on multiple publicly available video sequences, containing data from various crowd scenarios and different types of events, to demonstrate the effectiveness of the approach.

1. Introduction

Event detection in the field of automatic video surveillance has gained a growing interest [1]. The huge amount of data generated by existing surveillance systems in public areas requires the development of intelligent solutions that can avoid information overload for the users [2]. In particular, in the context of a crowd image analysis problem, it is desirable to develop online algorithms that reliably detect abnormal events in real-time. As an example, the automatic detection of anomalies in crowded scenes can be used to avoid crowd related disasters and ensure public safety [3].

An anomaly can be defined as "*something that deviates from what is standard, normal, or expected*"[1]. This means that *abnormal events* can be identified as irregular situations with respect to usual normal ones. Thus, the abnormal detection becomes the identification of abnormal events given some sample normal events. Zhan et al. [4] point out that conventional Computer Vision can be ineffective when dealing with the analysis of very crowded video sequences. Indeed, in a high-density situation the presence of severe occlusions consistently limits the performance of traditional methods for visual tracking [3]. Additional factors that can limit the effectiveness of existing approaches aiming at detecting abnormal events are: (1) Offline computation and (2) Need of a training phase. The offline assumption can limit the application of anomaly detection methods in practice [5]. For instance, it is desirable to detect panic situations as soon as possible in order to avoid damage to people. The methods that rely on the training of a classifier are limited by the possible lack of well-suited training data. Indeed, since it is not easy to find data about real emergency situations in crowded scenes, the resulting classifier could be suitable only for dealing with particular video sequences.

In this paper, we propose an online and real-time method for automatic anomaly detection in crowded scenes, which does not need any training stage. The method is inspired by the concept of Shannon entropy [6]. Entropy characterizes the uncertainty about the source of information and it increases for more sources of greater randomness. The idea is that the less likely an event is, the more information it provides when it occurs. Fig. 1 shows an example where the Shannon entropy value increases in the case of an abnormal event.

The two main contributions of the proposed approach are:

1. The use of two different metrics, namely *instant entropy* and *temporal occupancy variation*, to detect online abnormal situations in crowded scenes;
2. An unsupervised segmentation algorithm for images containing crowds.

Furthermore, we provide:

- A novel video sequence annotated with ground truth data, containing images of hundreds of runners at the start of a marathon, as an example of crowd video with locally steady optical flow.
- Ground truth annotations for two well-known video sequences containing abnormal events in crowded scenes, namely PETS 2009 [7] and AGORASET [8].
- The C++ source code and all the data used for the experimental evaluation, thus allowing for reproducing the results described in this paper and to compare other similar approaches.

The remainder of the paper is organized as follows. Related work is analyzed in the next Section 2, while our method is presented in

* Corresponding author.
E-mail address: bloisi@dis.uniroma1.it (D.D. Bloisi).
URL: http://www.dis.uniroma1.it/~bloisi/ (D.D. Bloisi)

[1] Definition from the Oxford Dictionary.

http://dx.doi.org/10.1016/j.cviu.2015.09.010
1077-3142/

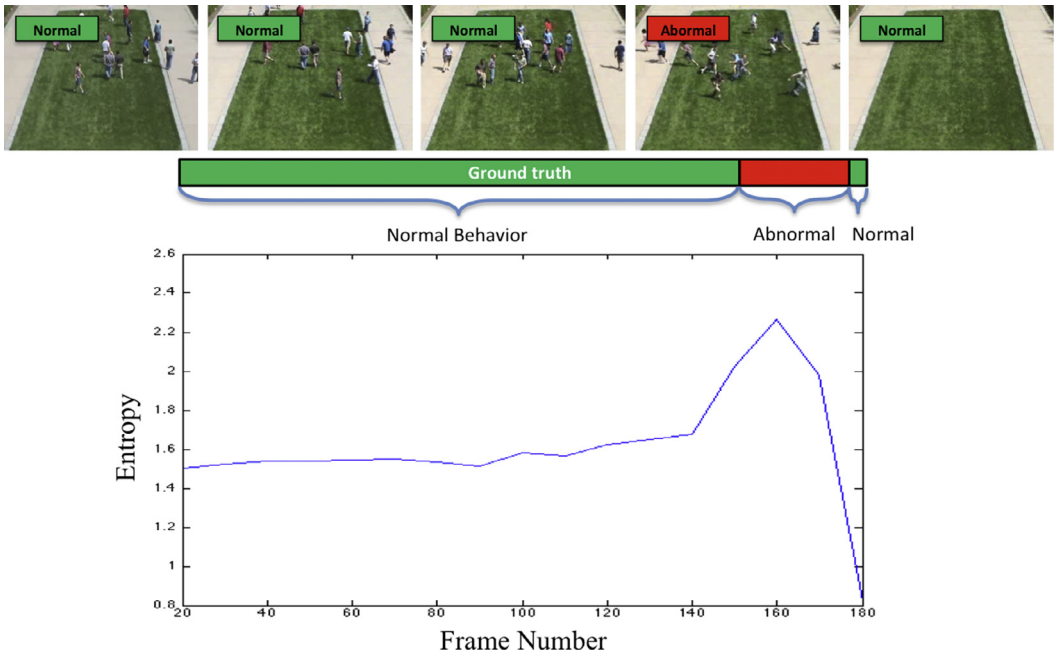

Fig. 1. Shannon entropy variation in the sequence Lawn from the UMN data set.

Section 3. Section 4 describes the qualitative and quantitative experimental results, providing also a comparison with other online approaches in the literature. Finally, conclusions are drawn in Section 5.

2. Related work

The techniques for crowd behavior analysis are usually grouped into two main categories [3,9]: (1) object-based and (2) holistic approaches. In the object-based methods the analysis is carried out at an individual level. For example, it can be of interest to detect if a single person is trying to enter a restricted area or if an individual is moving against the dominant flow. On the other hand, holistic techniques treats the crowd as a single entity, trying to extract global information, such as the main flow of the crowd, instead of analyzing single trajectories.

We propose a different classification, based on the nature of the methods used for detecting abnormal situations. According to our classification, existing approaches can be grouped into:

- Statistical analysis;
- Background subtraction;
- Image segmentation;
- Classification.

Statistical analysis. Methods in this category are based on the collection of particular features representing the flow of the crowd. For example, Mehran et al. [9] propose a method for localizing abnormal behaviors by using a social force model. A grid of particles is placed over the image for analyzing the space-time average of optical flow. The moving particles are treated as individuals and the flows in the scene are estimated by employing the social force model. The interaction forces are then mapped into the image plane to obtain a Force Flow for every pixel in the current frame. Spatio-temporal volumes of Force Flow are randomly selected for modeling the normal behavior of the crowd. Then, the normal and abnormal behaviors are classified by exploiting an approach based on a bag of words. The regions of anomalies in the abnormal frames are localized using interaction forces. Social force model algorithms only consider the temporal characteristic, i.e. local velocity, and ignore the spatial information such as local density for detecting the crowd behavior.

Zhu and Saligrama [1] propose a probabilistic framework that takes into consideration local spatio-temporal anomalies to characterize the observed scene by optimal decision rules. If anomalies are local optimal decision, they are local as well, even if the behavior exhibits global spatial and temporal statistical dependencies. This helps to collapse the large ambient data dimension space to detect local anomalies. Consistent data-driven local empirical rules with provable performance can be derived with limited training data. The empirical rules are based on scores functions derived from local nearest neighbor distances. These rules aggregate statistics across spatio-temporal locations and scales and produce a single composite score for video segments. However, this method is scene-specific.

Chang et al. [10] describe a statistical framework able to recognize group-level activity in many scenarios, by combining a soft grouping metric and track-based motion analysis. The approach recognizes group interactions without making hard decisions about the underlying group structure. In particular, a path-based grouping scheme is used to understand if an individual belongs to a group. The method is bottom-up and thus suffers when the tracking output is not reliable.

Zhang et al. [11] describe a social attribute-aware force model for abnormal crowd pattern detection in video sequences. An unsupervised method is used to estimate the scene scale and two attributes, namely social disorder and congestion, are introduced to describe the realistic social behaviors by means of statistical context features. Through the semantic attribute-aware enhancement, it is possible to improve the model on the basis of social forces. Even if the method has good results, it is an offline method.

Kratz and Nishino [12] describe a statistical framework for modeling the motion pattern behavior of extremely crowded scenes in order to detect unusual events. The authors model the dense activity of the crowd using a 3D Gaussian distribution of spatio-temporal gradients, by capturing the local spatio-temporal motion patterns through a distribution-based Hidden Markov Model. The results demonstrate that the proposed approach provides a suitable representation for analyzing crowded scenes, detecting unusual motion patterns in pedestrian behavior including movement against the normal flow of traffic. The method is appropriate for crowded scenes of a very high density, but it cannot handle videos containing middle to low density crowds, which is often the case in video surveillance systems.