Pattern Recognition 61 (2017) 266–281

Contents lists available at ScienceDirect

Pattern Recognition

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

Social network model for crowd anomaly detection and localization

Rima Chaker, Zaher Al Aghbari, Imran N. Junejo*

University of Sharjah, 27272, U.A.E.

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form
9 May 2016
Accepted 19 June 2016
Available online 16 July 2016

Keywords:
Crowd modeling
Social network model
Crowd analysis
Anomaly detecting
Anomaly localization
Scene understanding
Video surveillance

ABSTRACT

In this work, we propose an unsupervised approach for crowd scene anomaly detection and localization using a social network model. Using a window-based approach, a video scene is first partitioned at spatial and temporal levels, and a set of spatio-temporal cuboids is constructed. Objects exhibiting scene dynamics are detected and the crowd behavior in each cuboid is modeled using local social networks (LSN). From these local social networks, a global social network (GSN) is built for the current window to represent the global behavior of the scene. As the scene evolves with time, the global social network is updated accordingly using LSNs, to detect and localize abnormal behaviors. We demonstrate the effectiveness of the proposed Social Network Model (SNM) approach on a set of benchmark crowd analysis video sequences. The experimental results reveal that the proposed method outperforms the majority, if not all, of the state-of-the-art methods in terms of accuracy of anomaly detection.

1. Introduction

Crowd is defined as a collection of large number of people in a confined space. Socio-psychological studies [49,50] have shown that people in a crowd tend to walk in groups, thus forming collective entities [31] each of which has a specific goal and similar characteristics like speed and trajectory. Early detection, or prediction, of abnormal behaviors occurring in surveillance scenario scenes is of utmost significance. By alerting human operators, potential dangerous consequences can be reduced, or prevented. However, the analysis of crowded scenes is a very challenging task, due to the fact that the analysis of human actions is still not a fully solved problem. The significance of understanding crowd scenes is due to its potential in applications such as crowd management [41], video surveillance [3], public space design [2], etc. Recently, crowd motion segmentation [5,42], crowd density estimation [7,8], and identifying individuals' behavioral goals within a crowd [6], have all been subject of active research from different disciplines. This problem presents challenges of great complexity due to: (1) occlusion between individual objects, (2) random variations in the density of people over time, (3) low resolution videos with dynamic background, and (4) the inherent difficulty in accurately modeling the crowd behavior. What is needed is an automatic systems for analyzing crowd scenes and alerting human operators once anomalous activities are detected so that dangerous situations can be prevented.

Anomaly detection refers to modeling the normal scene behavior and then to detect the behavior that does not confirm to it. Thus, behavior patterns that appear frequently, are referred to as normal behaviors and those appearing rarely are referred to as abnormal behaviors. In [10], anomaly detection is broadly classified into two types, namely local and global. Local abnormal behavior corresponds to the behavior of a group of objects in a localized region that is different from that of their neighbors in spatio-temporal terms [16]. On the other hand, global abnormal behavior corresponds to the abnormal behavior of a group of objects in the whole scene. The key to accurate detection of abnormal behavior is the selection of an appropriate model that properly models both the local and the global behavior.

Fig. 1(a) denotes a typical scenario. The red circle represents the detected region of instable flow around Kabba in Mecca. Another example is illustrated in Fig. 1(b). The appearance of the bicyclist, circled in red, represents an anomaly with respect to the overall behavior of its surrounding neighbors. In this paper, we aim at detecting local and global abnormal behaviors in crowd scenes using a social network model: a data structure consisting of nodes and links between the nodes. In the crowd scene context, nodes can represent people and links reflect the social relationship among the people. First, the unsupervised approach extracts dense tracklets from the crowd motion data in a scene. Second, the video scene is partitioned at spatial and temporal levels; as a result, a set of spatio-temporal cuboids are constructed. The granularity of

* Corresponding author.
E-mail addresses: rkhalid@sharjah.ac.ae (R. Chaker), zaher@sharjah.ac.ae (Z.A. Aghbari), ijunejo@sharjah.ac.ae (I.N. Junejo).

http://dx.doi.org/10.1016/j.patcog.2016.06.016
0031-3203/

Fig. 1. A typical scenario (anomalies circled in red). (a) The region of instability flow of Pilgrims circling around kabba is detected. (b) Sample frame of anomaly detected (bicycle) in the UCSD dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

scene partitioning is proportional to the crowd density. Third, we cluster the objects in each cuboid based on the unique features of their tracklets, such as velocity, curvature, direction, etc., to build the local social networks, which model the objects' local behavior. Fourth, for each of the subsequent time windows, the global social network is updated incrementally using its local social networks and the previous window's global social network. By analyzing these social networks (local and global), a normal, or dominant, behavior and abnormal behavior can be identified. An earlier version of this work appeared in [51].

2. Related work

Crowd behavior analysis comprises of motion information extraction and behavior modeling. The model is then used to distinguish between normal and abnormal behavior. Basharat et al. [5] use object tracking [43] to detect unusual events in image sequences. Similarly, Ali et al. [12,13] track subjects in high density crowd scenes that are captured from a distance. They learn the direction of motion as a prior information based on a force model (floor fields). However, their method requires a manual selection of individuals to be tracked in the crowd which hinders automatic unexpected behavior recognition. Also, floor fields is chaotic in crowded scenes as they result in highly inconsistent trajectories.

For motion modeling, features, such as optical flow [11], tracklets [26], or Mixture of Dynamic Textures [16], are extracted at the pixel level. Different models are then built to solve the perplexities of occlusion and clutter. These models include Gaussian Mixture Model [21], Social Force Model [10], etc. For example, Mehran et al. [10] explore the socio-psychological concept “social force” in combination with optical flow to compute interaction forces that are later combined with Latent Dirichlet Allocation to model normal behaviors and detect abnormal ones. This method is further extended in [11] using Particle Swarm Optimization, in addition to social force model, to optimize the computed interaction force and thus detect global abnormal activities. Ali and Shah [13] utilize the idea of coherent structures in fluid dynamics for segmenting dominant crowd flows and flow instability detection. Gaidon et al. [38] structure a video as a tree of nested motion components composed of short duration point trajectories, tracklets. Chongjing et al. [26] analyze motion patterns by clustering the extracted tracklets in a dynamic crowd scenes. [46] use spatio-temporal Laplacian eigenmap to extract different crowd activities from videos.

Despite the many different representations of video events, many of the existing works ignore the importance of “contextual” anomaly in the field of crowd analysis. Contextual anomaly arises when an individual behavior exhibits behavior similar to others but it is anomalous in a specific context (e.g. neighborhood) [15]. Jiang et al. [15] focus on detecting contextual anomalies in the context of motion using statistical analysis. Leach et al. [18] detect subtle context-dependent behavioral anomalies based on contextual information. Beside the motion information, other works include important object features such as appearance or size. Mahadevan et al. [16] apply Mixture of Dynamic Textures (MDT) to jointly model the appearance and dynamics of crowded scenes. Their approach investigates both temporal and spatial abnormalities. Due to the reported heavy computational cost of [16], Reddy et al. [17] propose a more robust anomaly detection algorithm with relatively low complexity, while analyzing the size, motion and texture.

An important aspect in crowd behavior analysis is event/behavior recognition. Regular motion patterns such as direction and speed [24,25,40] can be used to estimate the behavior of a crowd in a given environment. A deviant behavior from the normal behavior is considered abnormal behavior. Two types of approaches are commonly used: object-based approach and holistic-based approach [10]. In object-based approaches, the crowd is considered as a collection of individuals. Ozturk et al. [24] propose an approach for clustering a set of flow vectors into local dominant motion flows. The local dominant motion flows are later combined to determine the global dominant motion flows in a crowd scene. In holistic-based approaches, a crowd, or a portion of a crowd, is treated as a single entity to estimate the regular and abnormal motions. For example, Mehran et al. [10] explored the social force model, which is based on socio-psychological studies, to model the behavior of a crowd.

Anomaly detection techniques: To ensure public safety, the main objective of crowd analysis involves modeling the crowd dynamics and the detection of video anomalies in the scene. However, detecting anomalies in crowd scenes is a challenging task due to the followings [1,2]:

- The large number of moving objects in crowd scenes easily weakens the local anomaly detector.
- It is difficult to model the abnormal events, as they are rare and last for a short period of time.
- It is difficult to obtain a training dataset that covers every possible normal behavior.

[48] propose an informative structural context descriptor (SCD), in addition to the 3-D discrete cosine transform (DCT), for describing the crowd individual, Ullah et al. [20], Mehran et al. [10] and Cui et al. [22] detect abnormal events in scenes of *escape panics*. Ullah et al. [20] initialized a fixed grid of particles that extracted the crowd motion features, and Gaussian Mixture Model [27] was adopted to learn the crowd behavior. The closest works to the proposed method are [10,22] in terms of considering people social behaviors. Mehran et al. [10] attempt to detect abnormal