Pattern Recognition 58 (2016) 110–120

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust individual and holistic features for crowd scene classification

Wenxi Liu [a,b], Rynson W.H. Lau [b], Dinesh Manocha [c]

[a] College of Mathematics and Computer Science, Fuzhou University, China
[b] Department of Computer Science, City University of Hong Kong, Hong Kong
[c] Department of Computer Science, University of North Carolina, Chapel Hill, USA

ARTICLE INFO

Article history:
Received 13 October 2015
Received in revised form
4 March 2016
Accepted 24 March 2016
Available online 16 April 2016

Keywords:
Crowd analysis
Crowd scene classification
Crowd modeling

ABSTRACT

In this paper, we present an approach that utilizes multiple exemplar agent-based motion models (AMMs) to extract motion features (representing crowd behaviors) from the captured crowd trajectories. In the exemplar-based framework, we propose an iterative optimization algorithm to measure the correlation between any exemplar AMM and the trajectory data. It is based on the Extended Kalman Smoother and KL-divergence. In addition, based on the proposed correlation measure, we introduce the novel individual feature, in combination with the holistic feature, to describe crowd motions. Our results show that the proposed features perform well in classifying real-world crowd scenes.

1. Introduction

In recent years, researchers in many fields are interested in understanding and analyzing pedestrians' behaviors in crowds, because of their practical applications in event recognition, traffic flow estimation, and crowd motion prediction. One of the fundamental problems in all these crowd analysis applications is the recognition of crowd scenes. This is a difficult research problem because crowd motion patterns are complicated (e.g., dynamic crowd density, scenarios configuration, and crowd psychology). In computer vision area research, some of the prior works formulate the crowd scene recognition problem as event recognition or anomaly detection [1–5], which aims to extract local-motion patterns from crowd videos. Some other prior works model the interactions among a small number of persons in order to perform action recognition [6,7], while [8] detects abnormal behaviors in dense crowd interaction scenarios. Despite the development of these crowd scene recognition methods, they are not robust enough for application across different scenarios – they require much effort in removing the negative impact of the vision-related factors (e.g., background noise and perspective transformation) and in learning a scene-specific model.

In order to investigate the underlying principles of crowd motions, our work aims to learn features from the crowd trajectories. Recently, pedestrian tracking techniques have progressed to a state that they can reliably capture the trajectories of crowd motions. Therefore, we believe that crowd scene recognition could take advantage of the captured crowd trajectories. Many prior vision-based works rely on optical-flow or key-point tracking, so that researchers have to deal with the unassociated tracklets or background noise when analyzing the crowd motion. The results of the analysis can be affected by factors such as camera positions and angles. Comparatively, given the pedestrian trajectories, the results of the analysis may be more informative. In addition to the crowd's holistic feature, we may also consider the motion feature of the individuals. Previous trajectory-based crowd analysis works are generally based on trajectory clustering [9] or semantic region inference [10,11]. Few works have been based on trajectories for crowd scene recognition, possibly due to the difficulty of regularizing the spatial and temporal trajectory data. In particular, for crowd scenes, the numbers of pedestrians may be different, the duration of the captured crowd motion sequences may be different, and the crowd trajectories are usually compounded with unknown noise. However, the crowd trajectories are informative for studying individuals' interactions within a crowd, e.g., how pedestrians react to oncoming opponents, which provide more insight into crowd scene understanding.

In this paper, we propose a crowd scene recognition algorithm that can handle the difficulty of regularizing the trajectories. The given trajectories allow us to study the individual interactions as the individual motion feature. As prior works in visual pedestrian tracking [12,13] suggest, such interactions, especially collision avoidance and grouping motion, are crucial in crowd motions. However, directly quantifying the interactions from crowd-motion data is a challenging task. To address this problem, we apply the agent-based motion models, denoted as AMMs, which have been demonstrated to be effective at modeling crowd interactions. In order to bridge the gap between AMMs and crowd trajectories, we propose a novel iterative optimization algorithm based on the Extended Kalman smoother and KL-divergence, which estimates how well the AMMs will model a specific crowd motion at both the individual level and the holistic level. The intuition is that,

http://dx.doi.org/10.1016/j.patcog.2016.03.031
0031-3203/

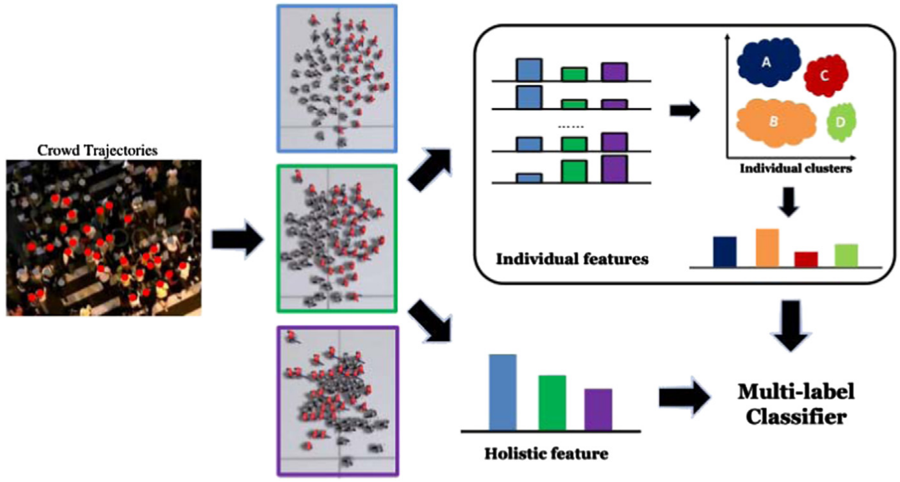

Fig. 1. Computing the proposed features (best viewed in color). The crowd trajectories (red and gray dots indicate the tracked persons in the video) and multiple exemplar agent-based motion models are given. We show the screenshots of the rendered simulated crowd motions using different AMMs (highlighted in three different colors). Comparing the crowd trajectories with the AMMs, the holistic features and the individual features are computed and then jointly applied for multi-label classification. Technical details are explained in Section 3.5. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

since we assume that the AMM-based state transition and the captured noisy sequential crowd states are Gaussian distributions, we can apply KL-divergence to measure the distance between these two distributions.

The prior works on crowd analysis usually aim at learning a general model, which requires a large amount of training data. Hence, using a single AMM to recognize different types of crowd motion is not robust. The fact that most AMMs are non-linear models controlled by several parameters further increases the difficulty of training a robust model. To obviate this difficulty in training a suitable agent-based motion model, we are inspired by prior works on recognition (e.g., [14,34]) that leverage exemplar models to infer unknown models. We apply a similar exemplar-based method that leverages multiple different AMMs to jointly evaluate the input crowd-motion data. Fig. 1 shows the framework. The query crowd trajectories are compared with multiple exemplar AMMs using our proposed iterative optimization algorithm. These exemplar AMMs can model/simulate different types of motion (e.g., the blue AMM models repulsive motions while the purple AMM models penetrative motions). The more similar an AMM to the real data, the lower the score it will get in the holistic feature.

In addition, the proposed optimization algorithm can evaluate a crowd motion not only holistically but individually. The evaluation of the individual motion can serve as a feature, or an *individual motion descriptor*. After collecting the individual motion descriptors for different individuals in various crowd scenes, we can categorize the individuals. In this work, we make an assumption that the distribution of different kinds of individuals in a crowd scene determines the style of the crowd motion. For instance, a crowd scene with most of the individuals belonging to the same motion category should be a coherent motion, while a crowd scene with a lot of heterogeneous individuals should probably be a random motion. Based on such an assumption and individual categorization, we propose an individual feature to describe crowd motion, as shown in Fig. 1 and explained in Section 3.5.

*Contributions***:** In this paper, we present a framework that extracts motion features from the given crowd trajectories based on the use of multiple exemplar AMMs, and propose an algorithm that measures the KL divergence between the crowd trajectories and any of the AMMs. To describe the crowd trajectories, we learn a holistic feature and an individual feature. While the holistic feature is a direct descriptor of the crowd motion sequences with regard to different AMMs, the individual feature stems from the distribution of different kinds of individuals, which is inferred from the individual clusters collected across crowd scenes. To evaluate our feature, we perform the multi-label classification on real-world crowd data.

The rest of this paper is organized as follows. Section 2 summarizes the relevant works. Section 3 introduces our exemplar-based framework and presents how to compute the individual and holistic features. Section 4 presents the multi-classification formulation. Finally, Section 5 evaluates the proposed approach and Section 6 briefly concludes this work.

2. Related works

In this section, we review existing works on visual analysis of crowds and agent-based motion models.

2.1. Crowd analysis

Existing works on crowd analysis can be categorized into holistic methods, particle/tracklet-based methods, and trajectory-based methods.

Holistic methods treat a crowd as an aggregated whole for analysis [15,3,1,2]. Chan and Vasconcelos [15] present a method based on dynamic textures, which represent video sequences as observations from a linear dynamical system. Mahadevan et al. [3] extend the dynamic texture technique to anomaly detection in crowd scenes. There are also works that utilize low-level visual features, e.g., optical flow, to build up models for discovering the motion patterns in crowded scenes [1,2]. These methods can effectively analyze global visual features, but their models require tedious training for different scenes. In addition, as they use low-level visual features, they are not able to analyze individuals' motions in the crowd.

Trajectory-based methods consider a crowd as a collection of individuals and model the interactions among the individuals. They analyze the complete trajectories [10,11,6,7]. Choi and Savarese [6] present a hierarchical activity model to recognize the interaction between individual trajectories and collective activities. Wang et al. [10] propose a model for trajectory clustering and semantic region detection. These methods investigate motion patterns of individuals' interactions, but the learned models are not generalizable to different scenes. Our work is also based on trajectories, but unlike the others, we leverage exemplar-AMMs to learn features without scene-dependent training.