Information Fusion 24 (2015) 3–15

Contents lists available at ScienceDirect

Information Fusion

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

Towards crowd density-aware video surveillance applications

Hajer Fradi *, Jean-Luc Dugelay

EURECOM, Campus Sophia Tech, 450 route des Chappes, 06410 Biot-Sophia Antipolis, France

ARTICLE INFO

Article history:
Received 6 February 2014
Received in revised form 4 July 2014
Accepted 9 September 2014
Available online 22 September 2014

Keywords:
Crowd density
Detection
Tracking
Behavior analysis
Privacy protection

ABSTRACT

Crowd density analysis is a crucial component in visual surveillance mainly for security monitoring. This paper proposes a novel approach for crowd density measure, in which local information at pixel level substitutes a global crowd level or a number of people per-frame. The proposed approach consists of generating automatic crowd density maps using local features as an observation of a probabilistic density function. It also involves a feature tracking step which excludes feature points belonging to the background. This process is favorable for the later density estimation as the influence of features irrelevant to the underlying crowd density is removed. Since the proposed crowd density conveys rich information about the local distributions of persons in the scene, we employ it as a side information to complement other tasks related to video surveillance in crowded scenes. First, since conventional detection and tracking methods are hard to be scalable to crowds, we use the proposed crowd density to enhance detection and tracking in videos of high density crowds. Second, we employ the local density together with regular motion patterns as crowd attributes for high level applications such as crowd change detection and event recognition. Third, we investigate the concept of crowd context-aware privacy protection by adjusting the obfuscation level according to the crowd density. In the experimental results, our proposed approach for crowd density estimation is evaluated on videos from different datasets, and the results demonstrate the effectiveness of feature tracks for crowd measurements. Moreover, the employment of crowd density in other applications demonstrate good performances for detection, tracking, behavior analysis, and privacy preservation.

1. Introduction

Studying crowd phenomenon is becoming of great interest mainly with the increasing number of popular events that gather many people such as in markets, subways, religious festivals, public demonstrations, sport events, and high density moving objects like car traffic. In this context, crowd analysis has emerged as a major topic for crowd monitoring and management in visual surveillance field. In particular, the estimation of crowd density is receiving much attention for safety control. It could be used for developing crowd management strategies by measuring the comfort level in public spaces. Also, its automatic monitoring is extremely important to prevent disasters by detecting potential risk and preventing overcrowd. Many stadium tragedies could illustrate this problem, as well as what happened in 2010, in the Love Parade stampede in Germany and the Water Festival stampede in Colombia. To prevent such deadly accidents, early detection of unusual situations in large scale crowd is required and appropriate decisions for safety control have to be taken to insure assistance and emergency contingency plan.

Many recent works in the field of automatic video surveillance have been proposed to address the problem of crowd density analysis. Typically, given a video sequence the objective is to estimate the number of people, or to alternatively estimate the crowd level. For people counting problem, significant progress has been recently made to handle that by using features regression methods [1–3]. This paradigm is proposed as an alternative solution to detection-based methods because of the partial occlusions that occur in the crowd, and that make delineating people a difficult task. In addition to person counts, level of the crowd is another indicator in crowd density analysis. According to the classification introduced in [4], the crowd density can be categorized into 5 levels: free, restricted, dense, very dense, and jammed flow. Early attempts to handle this problem generally made use of local texture features. Especially the use of some variants of Local Binary Pattern (LBP) [5], has been an active topic of research for this problem [6–9].

Although these categories of people counting and crowd level classification are commonly used in the field of crowd analysis, they have the limitation of providing a global information of the whole image, and discarding local information about the crowd. We therefore resort to crowd measure at local level by computing

* Corresponding author.

http://dx.doi.org/10.1016/j.inffus.2014.09.005
1566-2535/

crowd density maps. This alternative solution is indeed more appropriate since it enables both the detection and the localization of potentially crowded areas. The proposed crowd density map is based on using local features as an observation of a probabilistic density function. Also, a feature tracking step is involved in the estimation of crowd density. In fact, considering all extracted local features brings an inconvenience to the density function estimation as a substantial amount of components are irrelevant to the underlying crowd density. Therefore, we propose to use motion information to alleviate this effect.

In addition to the estimation of local crowd density, we intend to explore in this paper the usefulness of such crowd measure as additional information to other video surveillance tasks, mainly because common capabilities of automated surveillance systems are of limited success in high-density scenes. This is due to the challenging characteristics of crowded scenes such as the small size of objects in crowds, the occlusions caused by inter-object interactions, and the constant interactions among individuals in the crowd which make them indiscernible from each other. Given these difficulties, visual analysis of high density scenes remains a challenge compared to scenes with fewer people. As the density of people increases in the scene, a substantial deterioration in performances of automatic video surveillance tasks such as person detection, tracking, and behavior analysis is observed [10]. In this paper, we mainly focus on three major representative set of problems which are: (1) detection and tracking of people in crowded scenes, (2) modeling crowd behaviors and detecting anomaly (or change), and (3) studying privacy aspects in crowds.

The problems of detection and tracking in crowds have been addressed in the literature by learning motion patterns in order to constraint the tracks. In [11], global motion patterns are learned and participants of the crowd are assumed to follow a similar pattern. Rodriguez et al. [12] extended this approach in unstructured environments to cope with different crowd behaviors by studying overlapping motion patterns. Although these solutions have shown promising results, they operate in off-line mode and the learned patterns are tied to a particular scene. Also, they impose constraints to the crowd motion, thus, trajectories not following the common patterns are penalized. Moreover, some of these methods include additional constraints; in [12], Rodriguez et al. employed a limited descriptive representation of target motion by quantizing the optical flow vectors into 10 possible directions. Also, the *floor fields* proposed in [11] impose how a pedestrian should move based on scene constraints, which results in only one single direction at each spatial position in the video.

Crowd behavior analysis is another problem that has attracted research attention in the field. This problem covers different subproblems such as crowd change or anomaly detection [13–15], and crowd event recognition [16–18]. Usually the activity process in video sequence can be categorized into three main steps [16]: (1) detection, (2) tracking, and (3) behavior analysis. Given the difficulties encountered by analyzing crowded scenes, related works to crowd behavior analysis bypass the detection and the tracking of individuals and instead operate on local features [15], or particles [14,17]. In general, these methods aim at detecting and categorizing crowd events using motion information. This latter could correspond to normal (frequent) behavior or abnormal (unusual) behaviors.

The last problem we intend to address in this paper, is about preserving privacy in crowded scenes. Actually, with the widespread growth in the adoption of digital video surveillance systems, several concerns have been raised related to the possibility of infringing the privacy rights of the subjects being monitored [19]. At the same time, the adoption of automated methods for the analysis of video surveillance data has raised additional concerns, since algorithms such as face recognition or people re-identification could potentially expose the identity of any individual under video surveillance at any time [20]. One big challenge related to privacy protection policies in crowded scenes is the identification of the correct trade-off between intelligibility of the video, which should be adequate for crowd monitoring tasks, and privacy protection itself. An attempt to deal with this problem is presented in [21], where a context-aware surveillance system is proposed by combining a number of contextual information (based on the analysis of visual features such as global motion, and person counts) to determine an appropriate level of privacy protection.

To overcome all these problems, in this paper, we propose to incorporate the local crowd density measure in the three aforementioned applications: First, we propose a method for enhancing human detection and tracking in crowded scenes; it is based on applying a scene-adaptive dynamic parametrization using the crowd density measure. Compared to prior works, our approach does not depend on any learning step, and does not impose any direction to the crowd flow. It models the crowd in a temporally evolving system, which enables a large number of likely movements at each space–time location of the video. Second, we propose a novel approach to detect crowd change and to recognize crowd events. It is based on analyzing temporal and spatial distributions of persons using long-term trajectories within a sparse feature tracking framework. Our proposed approach employs the local density together with the commonly used motion patterns (speed and direction). The idea is motivated by the necessity of using local density to determine the ongoing crowd behavior since that helps to characterize the event, and to localize crowded regions. Finally, we investigate the usefulness of applying crowd density in privacy context. The concept of context-aware privacy protection has recently emerged, as the required amount of privacy protection is deeply linked to the context. In particular, we propose adaptive protection filters that select suitable level of privacy preservation according to the crowd density measure.

The remainder of the paper is organized as follows: In the next Section 2, we present our proposed approach for crowd density map estimation. In Section 3, we demonstrate how a prior estimation of crowd density could provide valuable information and could complement other applications in video surveillance. In particular, three applications are investigated: enhancing human detection in crowded scenes is presented in Section 3.1, studying crowd behaviors is presented in Section 3.2, and formulating contextualized privacy preservation filters is presented in Section 3.3. Detailed experimental results of the density map and the different applications follow in Section 4. Finally, we briefly conclude and give an outlook for possible future works in Section 5.

2. Crowd density map estimation

In this paper, we explore a new promising research direction which consists of using crowd density measures to complement some other applications in crowded scenes. For this, generating local crowd density measure is more helpful than computing only an overall density or a number of people in a whole frame. In the following, we present our proposed approach for crowd density estimation [22]. First, local features are extracted to infer the contents of each frame under analysis. Then, we perform local features tracking using the Robust Local Optical Flow algorithm from [23] and a point rejection step using forward–backward projection. To accurately represent the motion within the video, the estimation of the optical flow between consecutive frames is extended to trajectories. The generated feature tracks are thereby used to remove static features. Finally, crowd density maps are estimated using Gaussian symmetric kernel function. An illustration of the density map modules is shown in Fig. 1. The remainder of this section describes each of these system components.