Image and Vision Computing 32 (2014) 14–26

Contents lists available at ScienceDirect

Image and Vision Computing

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

Tracking in dense crowds using prominence and neighborhood motion concurrence ☆

Haroon Idrees [a,*], Nolan Warner [b,1], Mubarak Shah [a]

[a] Center for Research in Computer Vision (CRCV), University of Central Florida, Orlando, United States
[b] Department of Computer Science and Engineering, University of Nevada, Reno, United States

ARTICLE INFO

Article history:
Received 24 January 2013
Received in revised form 25 September 2013
Accepted 30 October 2013

Keywords:
Crowd analysis
Dense crowds
Tracking
Prominence
Neighborhood motion concurrence
Hierarchical tracking

ABSTRACT

Methods designed for tracking in dense crowds typically employ prior knowledge to make this difficult problem tractable. In this paper, we show that it is possible to handle this problem, without any priors, by utilizing the visual and contextual information already available in such scenes.

We propose a novel tracking method tailored to dense crowds which provides an alternative and complementary approach to methods that require modeling of crowd flow and, simultaneously, is less likely to fail in the case of dynamic crowd flows and anomalies by minimally relying on previous frames. Our method begins with the automatic identification of prominent individuals from the crowd that are easy to track. Then, we use Neighborhood Motion Concurrence to model the behavior of individuals in a dense crowd, this predicts the position of an individual based on the motion of its neighbors. When the individual moves with the crowd flow, we use Neighborhood Motion Concurrence to predict motion while leveraging five-frame instantaneous flow in case of dynamically changing flow and anomalies. All these aspects are then embedded in a framework which imposes hierarchy on the order in which positions of individuals are updated. Experiments on a number of sequences show that the proposed solution can track individuals in dense crowds without requiring any pre-processing, making it a suitable online tracking algorithm for dense crowds.

1. Introduction

Crowd analysis is an active area of research in Computer Vision [1]. Over the past few years, methods have been proposed that estimate density and number of people in a crowd [2,3], find group structures within a crowd [4], detect abnormalities [5–8], find flow segments [9,10], and track individuals in a crowd [11–13].

Density is an important feature which can be used to classify different kinds of crowds [1]. From the computer vision perspective, videos of high density crowds can be divided into groups based on the number of pixels on target. High density crowds with extremely small object size permit only holistic approaches for scene understanding, such as finding motion patterns and segmentation of crowd flows [14,15,9,10]. However, if individuals in a crowd are distinguishable, then tracking of individuals may be possible, which is important in the context of safety and surveillance [1].

Tracking in dense crowds [16,12,13] is a challenging problem. Non-crowd methods, which track individuals in isolation do not perform well on dense crowds [16] because the large number of objects in close proximity poses difficulty in establishing correspondences across frames. Furthermore, human motion in crowds is influenced by social constraints [17] which force individuals to follow more complex, non-linear patterns that need to be taken into account for successful tracking of dense crowds.

Methods specifically designed for dense crowds generally require some learning of motion priors, which are later employed for tracking. For instance, Ali and Shah [16] proposed an algorithm which is based on the assumption that all individuals in a crowd behave in a manner consistent with global crowd behavior and learn the direction of motion at each location in the scene. The floor fields they learn severely restrict the permitted motion that individuals in a particular scene can have. This restriction on the motion of individuals due to time-invariant priors would cause the tracker to fail when, (1) the crowd flow is dynamic, (2) the crowd flow shifts or moves to a new region which was not learned before, and (3) when there are anomalies. Furthermore, camera motion and jitter can make learning the crowd flow difficult, if not impossible. Though learning, whether online or offline, certainly helps in tracking dense crowds when these issues are not present, our goal in this paper is to emphasize the use of visual and contextual information available in such

☆ This paper has been recommended for acceptance by Xiaogang Wang.

* Corresponding author.
E-mail address: haroon@eecs.ucf.edu (H. Idrees).

[1] This work was initiated when the author visited University of Central Florida, Orlando, for the National Science Foundation's REU program.

0262-8856/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.imavis.2013.10.006

crowded scenes to track in an online manner, without any pre-processing, learning or crowd flow modeling.

At the core of our approach lies template-based tracking, which is used to obtain the probability of observation. However, the simplicity of a template-based tracker demands more than just appearance to perform well in high density crowds. We supplement the tracker with novel visual and contextual sources of information, which are particularly relevant to crowds and reduce the confusion in establishing correspondences.

The first idea we explore is prominence of individuals which is similar to saliency (generally used for features and points). In any crowded scene with a large number of people, the appearance of some individuals will be markedly different from the rest (Fig. 1). The probability of confusing such individuals with the rest of the crowd will be low. Thus, the prominence of such individuals provides extra information which should be leveraged in tracking.

The second idea is to employ influence from neighbors to make better prediction for an individual's position. This idea is based on the observation that individuals in dense crowd experience social forces that bound their movement. For instance, an individual cannot jump across its neighbors in a single time instance. The restriction on movement that each individual experiences is proportional to the density of the crowd. Social force models, both in computer graphics and vision, are generally geared towards collision avoidance, where the goal is to predict positions such that subjects or individuals don't collide with each other. Our model, on the other hand, exploits the fact that movement of individuals in a dense crowd is similar to their neighbors, and therefore can be used to make better predictions.

Combining prominence and influence from neighbors, our method imposes an order on the way positions of individuals are updated. Individuals with prominent appearance are updated first, which subsequently guide the motion of the rest of the crowd. While updating, if the underlying patch-based tracker gives weak measurement for an individual, then position of the individual is updated based on appearance-based dense instantaneous flow. Thus, the framework we introduce incorporates these ideas as well as their inter-relationships. Our contributions in this paper can be summarized as,

- An alternative approach to dense crowd tracking which highlights the significance of prominence and spatial context for tracking dense crowds without requiring crowd flow modeling,
- Introduction of the notion of prominent individuals, its relevance to tracking in dense crowds, and a method to detect prominent individuals,
- Incorporation of influence from neighbors, prominent or not, to better predict and estimate an individual's position,
- A tracking framework which imposes an order in the way individuals are tracked, where positions of prominent individuals are updated first and individuals with low probability of observation from underlying tracker are updated last.

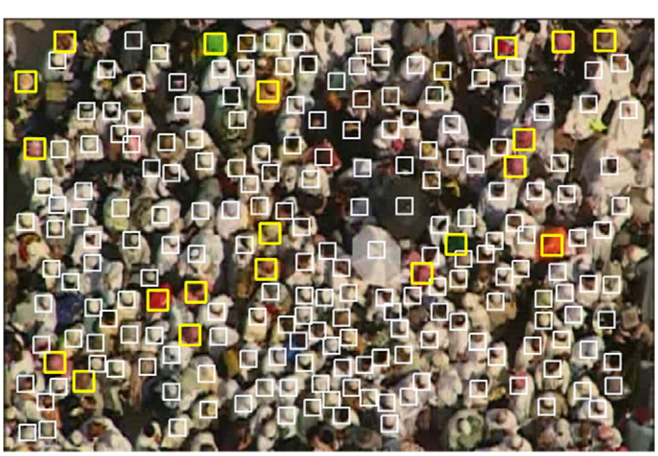

Fig. 1. An example of a dense crowd where individuals that are in yellow squares stand out from the crowd and, therefore, should be easier to track than rest of the individuals, marked with white squares.

Since space is complementary to time, both the visual information (prominence) and spatial context (influence from neighbors) are complementary to temporal constraints (crowd flow, motion patterns) introduced in previous works on tracking people in dense crowds. Our goal in this work is to emphasize the first two, which when coupled together allow tracking in an online fashion, without modeling crowd flow and without looking at observations from the future.

2. Related work

There are a few papers that have used context for tracking, however, there is currently no such work that utilizes it for dense crowds. Yang et al. [18,19] used contextual information to improve the tracking performance of a few objects. Through color segmentation of the image, they find auxiliary objects, which are easier to track and whose motion is correlated with the target. The auxiliary objects are then tracked; they also aid in tracking the target, which occurs simultaneously. The method was streamlined for non-crowd scenarios, with results containing a maximum of three objects per sequence. Furthermore, due to hundreds of people frequently occupying the entire screen in crowd videos, the definition and discovery of auxiliary objects is not applicable to crowd sequences. Khan et al. [20] also capture interaction between targets using particle filters in an MRF framework. However, they do not consider prominence and anomalies while tracking, and the particle filters are not suitable for crowd sequences due to fewer pixels per target.

The methods proposed for multi-target tracking include Park et al. [21] who sped up belief propagation using mean shift by sparsely sampling the belief surface instead of using parametric methods or non-parametric methods that require dense sampling. They do not assume prominence and pass messages in all directions, therefore presuming absence of anomalies.

Next, we review papers relevant to different aspects of our method. For an in-depth analysis of crowd literature, interested readers are referred to the survey by Zhan et al. [1].

2.1. Prominence

Discriminative features were used for tracking by Collins and Liu [22], who rank the foreground features online and track objects using only those features which discriminate foreground from background. A similar idea was explored by Mahadevan and Vasconcelos [23] who, given a pool of features from foreground and background, select the most informative features for classification between the two. In relation to our method, prominence can be seen as a *collection* of salient features which discriminates one foreground object from the rest.

2.2. Social force models

Static motion models (such as linear velocity or constant acceleration) have long been used for tracking in computer vision. Dynamic models, as opposed to static ones, account for the dynamic structure of the scene and objects, and are based on the fact that individuals are driven by goals and respond to changes in their environments by adjusting their paths. Methods that model [24–26] and simulate crowds [27] incorporate this crucial information to produce realistic results.