Pattern Recognition 47 (2014) 1626–1641

Contents lists available at ScienceDirect

Pattern Recognition

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

Modeling local behavior for predicting social interactions towards human tracking

X. Yan, I.A. Kakadiaris, S.K. Shah *

Quantitative Imaging Laboratory, University of Houston, Department of Computer Science, Houston, TX 77204-3010, USA

ARTICLE INFO

Article history:
Received 7 March 2013
Received in revised form
27 September 2013
Accepted 19 October 2013
Available online 31 October 2013

Keywords:
Human interaction
Social cues
Interactive Markov Chain Monte Carlo
Multi-object tracking

ABSTRACT

Human interaction dynamics are known to play an important role in the development of robust pedestrian trackers that are needed for a variety of applications in video surveillance. Traditional approaches to pedestrian tracking assume that each pedestrian walks independently and the tracker predicts the location based on an underlying motion model, such as a constant velocity or autoregressive model. Recent approaches have begun to leverage interaction, especially by modeling the repulsion forces among pedestrians to improve motion predictions. However, human interaction is more complex and is influenced by multiple social effects. This motivates the use of a more complex human interaction model for pedestrian tracking. In this paper, we propose a novel human tracking method by modeling complex social interactions. We present an algorithm that decomposes social interactions into multiple potential interaction modes. We integrate these multiple social interaction modes into an interactive Markov Chain Monte Carlo tracker and demonstrate how the developed method translates into a more informed motion prediction, resulting in robust tracking performance. We test our method on videos from unconstrained outdoor environments and evaluate it against common multi-object trackers.

1. Introduction

Multiple pedestrian tracking in unconstrained environments is an important task that has received considerable attention from the computer vision community in the past two decades. A number of approaches that address this problem have been proposed [1,2]. Accurate multiple pedestrian tracking can greatly improve the performance of activity recognition and analysis of high level events through a surveillance system. However, the complexity of human motion poses several challenges to the accuracy and precision of any tracking system. In the context of video surveillance, human motion can be thought of as blob motion in which arms and legs are difficult or unnecessary to localize. At this scale, the study of human motion predominantly involves cues related to space and environment, and we can expect to recover how people move from place to place. Accordingly, the recovery of motion pattern of people facilitates a measure of social phenomena among interacting individuals [3]. Interpersonal distance cues have their basis in the seminal findings that people tend to organize the space around them in four concentric zones associated with different degrees of intimacy [4]. The spatial organization of people within these concentric zones is dominated by relationships between interacting individuals [5]. Hence, it is the encoding of social relationships along with tracking methods that has been most commonly exploited in recent years to model human motion.

The integration of social relationships to address the dynamics of human motion has its origin in the social force model [6] that applies a fluid flow analogy to the dynamics of pedestrians. It is primarily a physical model that captures a continuous phenomenon where humans are considered to react to energy potentials caused by other pedestrians and static obstacles, while trying to keep a desired speed and motion direction. Recently proposed local motion models such as linear trajectory avoidance (LTA) model [7] or human motion prediction model [8] demonstrate that leveraging social relationships can improve tracking performance. Typical social relationships can be envisioned through simple interaction effects that can take forms such as: (1) attraction effects, (2) repulsion effects, and/or (3) no social effect. The attraction and repulsion effect can be characterized as the tendency to move toward or away from objects. Repulsion effect has been leveraged in most existing tracking methods, but modeling of multiple effects of social relationships simultaneously remains challenging. Modeling motion based on repulsion effects alone excludes the possibility of people's intent to meet and only captures the intent of avoiding collisions. Nevertheless, unconstrained environments would typically involve people with

* Corresponding author. Tel.: +1 713 743 3360.
E-mail addresses: xyan5@uh.edu (X. Yan), ioannisk@uh.edu (I.A. Kakadiaris), shah@cs.uh.edu, sshah@central.uh.edu (S.K. Shah).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.10.019

motion dynamics explained under the combination of several basic social effects. In this paper, we present a model that embeds social relationships in terms of linear combination of predefined basic social effects.

Generally, the intent of pedestrians produces different social relationships in which the intent of avoidance is explained by the repulsion effect and the intent of approach is explained by the attraction effect. The intent varies over time, thus motion prediction of corresponding trackers should be adjusted dynamically depending on the current interaction environment. A specific limitation of many trackers is that the motion model used to predict the dynamics of a target is based on a fixed motion model, typically a first-order approximation. Thus, it fails to model the complex motion that is affected by elaborate pedestrians' intent and corresponding interactions. Our approach focuses on how to incorporate the temporally varying pedestrian interaction or intent into a dynamic motion model without explicit knowledge of local social relationships. Although the desired mode of interaction is unknown, the intent of pedestrians can be assumed to belong to a finite set which combines the intent of avoidance, approach, or non-interaction [9]. The finite set of intent generates a finite set of interactions. We propose to decompose complex pedestrian interaction into a finite set of interactions, where the decomposition is motivated by the work of Kwon and Lee [10].

Consider a simple scenario with two pedestrians as illustrated in Fig. 1, wherein pedestrians can either decide to meet and interact with others or choose their motion direction to avoid colliding with others. By modeling their intents in this case (interaction modes), local interactions can be hypothesized to guide tracking. Conversely, the tracking output validates the mode of social interactions. If we model the local interaction between them under the intent of either avoidance or approach, the approach predicts two possible motions for each pedestrian. Then it searches the best tracking result by sampling pedestrians' state space. On the other hand, the best tracking result validates the intent under which local interaction effects contribute more accurately to prediction using a linear search strategy.

The key contributions of our work are as follows:

1. *Local interaction model that explicitly includes repulsion, attraction, and non-interaction*. We model repulsion, attraction, and non-interaction effects in pedestrian dynamics. Such interactions are more common in unconstrained environments and can be leveraged to capture various interaction behaviors such as people meeting, people following, and/or group interactions.
2. *A decomposed social interaction model*. We propose a decomposed motion model that approximates complex social interactions by tracking all the possible combination of basic interaction effects among multiple pedestrians. It enables motion prediction without the knowledge of instantaneous interaction modes.
3. *A dynamically adjusted state space*. The algorithm adjusts the number of basic trackers dynamically based on the exact interaction among pedestrians, which expands or shrinks the joint state space to facilitate the search of tracking results.

This paper is an extension of our work in [11] that details and generalizes our proposed approach along with additional experiments to evaluate the benefits of the developed tracker. Specifically, (1) synthetic experiments are presented and analysis performed to evaluate the accuracy of social interaction mode prediction and its impact on tracking performance; (2) various parameters of the proposed framework are evaluated and results presented to better understand their impact on tracking performance; and (3) a more detailed comparison is presented to validate the advantage of modeling multiple basic social effects including approach, avoidance, and non-interaction as compared to existing trackers that incorporate social effects to model motion dynamics. The rest of this paper is organized as follows. Section 2 describes related work. Section 3 presents the proposed social interaction model and describes its decomposition into multiple models. The incorporation of the proposed model within a Bayesian tracking framework and the design of the compound tracker is presented in Section 4. Section 5 presents the experiments performed and a qualitative and quantitative assessment of the tracker performance. Comparative analysis against multiple existing trackers is also presented. Finally, conclusions are presented in Section 6.

2. Related work

Previous tracking algorithms mainly exploit two aspects including coping with targets' appearance variance and modeling complex targets' motion. To account for appearance variation of the target caused by change of illumination, deformation and pose, a large amount of work has been proposed [12–17] and these methods perform well and get good results. However, the dynamics of target and interaction between targets is much less explored. The state space of targets is affected by motion of target and interaction of targets. While a dynamic model is used mainly to reduce the search space of state space, it affects the tracking results especially when multiple targets undergo complex interacting motion. The early tracking algorithms adopt constant velocity model or acceleration model, such as [18] which leverages second-order auto-regressive dynamics. Most recent tracking algorithms incorporate random walk model [15,16] which models the targets' dynamics as Brownian motion. Recently, various approaches that account for interaction among people to improve visual tracking have been developed. Motion behaviors are distinctly encoded dependent on the tracking environment, primarily differentiated in terms of densely crowded scenes and more

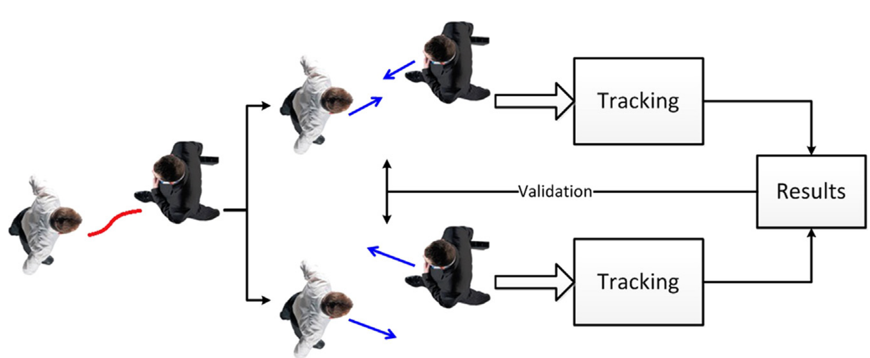

Fig. 1. Depiction of the process flow.