Pattern Recognition Letters 44 (2014) 49–57

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Activity analysis in crowded environments using social cues for group discovery and human interaction modeling

K.N. Tran, A. Gala, 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:
Available online 3 October 2013

Communicated by Rita Cucchiara

Keywords:
Group activity recognition
Social signaling
Graph clustering
Local group activity descriptor
Crowd analysis

ABSTRACT

This paper presents a novel and efficient framework for group activity analysis. People in a scene can be intuitively represented by an undirected graph where vertices are people and the edges between two people are weighted by how much they are interacting. Social signaling cues are used to describe the degree of interaction between people. We propose a graph-based clustering algorithm to discover interacting groups in crowded scenes. Two social signaling cues are presented and compared for group discovery. The grouping of people in the scene serves to isolate the groups engaged in the dominant activity, effectively eliminating dataset contamination. Using discovered interacting groups, we create a descriptor capturing the motion and interaction of people within it. A bag-of-words approach is used to represent group activity and a SVM classifier is used for activity recognition. The proposed framework is evaluated in its ability to discover interacting groups and perform group activity recognition using two public datasets. The overall recognition system is compared to a baseline top-down model to understand the impact of social cues for activity recognition. The results of both the steps show that our method outperforms state-of-the-art methods for group discovery and achieves recognition rates comparable to state-of-the-art methods for group activity recognition.

1. Introduction

Human activity analysis is one of the most challenging problems that has received considerable attention from the computer vision community in recent years. Its applications are diverse, spanning from its use in activity understanding for intelligent surveillance systems to improving human–computer interactions. Recent approaches have demonstrated great success in recognizing actions performed by one individual (Ryoo and Aggarwal, 2011; Tran et al., 2012). However, a vast number of activities involve multiple people and their interactions. This poses a far more challenging problem due to variations in the number of people involved, and more specifically the different human actions and social interactions exhibited within people and groups.

Group activities are characterized by actions of individuals within their group and their interactions with each other as well as the environment. The environment in which these groups exist provide important contextual information that can be invaluable in recognizing the group activities. These activities can be described by location and movement of individuals. However, understanding groups and their activities is not limited to only analyzing movements of individuals in group. Most of the current work that has gone into group activity recognition is based on a combination of actions of individuals and contextual information within the group (Lan et al., 2010, 2011; Choi et al., 2011). The contextual information is most often encoded by the inter-person interaction within the group. In addition, there might exist more than one group in a scene and each group might exhibit a specific activity. Most of the existing approaches treat group activity recognition as a singular activity performed by most people visible in a scene. This is not true especially in crowded environments typical of surveillance scenes. There might exist people in the scene that are not part of the group or groups that are engaged in the dominant activity in the scene. For example, from Fig. 1 we can see the dominant activity in the top row is *Crossing* but there are people *Waiting* in the scene. Similarly, the frames in the middle row are associated with *Talking* activity but there exist people marked in the red boxes that are not engaged in *Talking*. Frames in the third row, show the dominant activity to be *Jogging* but some people are *Talking*. If all the people in the scene are used to analyze the group activity it may create misleading recognition of activity due to the underlying noisy or contaminated data. In order to improve the granularity of analyzing group activities, it is important to be able to detect the groups performing the dominant action.

Perspectives from sociology, psychology and computer vision suggest that group activities can be understood by investigating a subject in the context of social signaling constraints (Smith

* Corresponding author. Tel.: +1 713 743 3360.
E-mail addresses: khaitran@cs.uh.edu (K.N. Tran), avbedagk@central.uh.edu (A. Gala), ikakadia@central.uh.edu (I.A. Kakadiaris), sshah@central.uh.edu, shah@cs.uh.edu (S.K. Shah).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.09.015

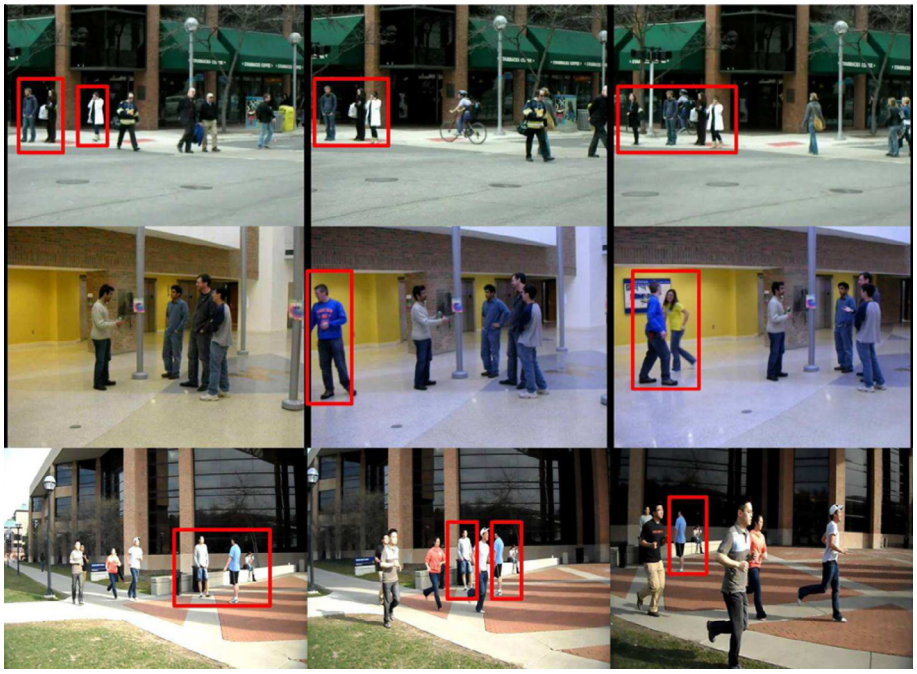

Fig. 1. Example frames of noisy or contaminated data in Collective Activity dataset (Choi et al., 2009). The red boxes depict the people not engaged the dominant activity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

et al., 2008; Helbing and Molnár, 1995; Cristani et al., 2011; Farenzena et al., 2009b). Exploring the spatial and directional relationships between people can facilitate the detection of social interactions in a group. Leveraging the notion of social signaling cues, we develop a two-step top-down process for group activity analysis: first we discover the interacting groups based on the spatial and orientational relationships between individuals, and in the next step, we analyze the local interactions in each group to recognize their group activity. This approach serves two purposes, first it helps to eliminate the clutter in scenes that can mislead the group activity descriptor and the second is to localize the interacting groups in crowded scenes in order to simplify the activity inference process.

In this paper, we propose a graph representation of human interactions to discover interacting groups in the scene. The proposed representation incorporates the social distance (Was et al., 2006)) cue and Visual Focus of Attention (VFoA) (Farenzena et al., 2009b) cue in modeling social interactions and is generative so many robust graph algorithms can be applied to detect the groups efficiently. Our representation is motivated by the recent success of social signal processing (Cristani et al., 2011; Farenzena et al., 2009b) and our clustering algorithm is inspired by the fundamentals of dominant set for clustering (Pavan and Pelillo, 2007). Further, using the detected groups we propose a novel group activity representation along with an efficient recognition algorithm to learn and classify group activities in top-down mechanism.

The contributions of our work are:

1. *A graph representation for human interactions along with dominant set based clustering algorithm to discover interacting groups.* We propose a new social interaction cue based representation using graph theory where each vertex represents one person and weighted edges describe the interaction between any two people in a group. We use the dominant set based clustering algorithm to discover the interacting groups in the scene.
2. *A group activity descriptor along with bag of words recognition framework.* We propose a novel group activity descriptor that encodes social interaction cues and motion information of people in particular interacting groups that are discovered by our first contribution.

This paper is an extension of our work in Tran et al. (2013) and the extensions are: (1) introducing the use of another social cue approach (VFoA) along with graph based clustering algorithm to discover interacting group, (2) presenting an extended baseline top-down model integrating with social cues for group activity recognition. Those extensions provide more comprehensive understanding of different approaches in using valuable social information for crowded analysis. The rest of the paper is organized as follows. We review related work on group activity analysis in Section 2. Section 3 describes the discovery of interacting groups in the scene and its use in representing group activity along with the classification algorithm used to address the activity recognition task. Experimental results and evaluations are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related work

Group activity analysis can often be considered a multi-step process, one that involves individual person activity, individuals forming meaningful groups, interaction between individuals and interactions between groups. In general, the approaches to group activity analysis can be classified into two categories: bottom-up and top-down. The bottom-up approaches rely on recognizing activity of each individual in a group context. Vice versa, top-down approaches recognize group activity by analyzing at the group level rather than at the individual level. Bottom-up approaches show the understanding of activities at the individual or local level, however they are more sensitive to group or global level of activities in crowded scenes. Top-down approaches show the understanding of activities at global level but they are less sensitive to recognizing activities at the individual level.

In bottom-up approach, group context is used to differentiate ambiguous activities e.g. standing and talking, which are normally