Social Group Discovery from Surveillance Videos: A Data-Driven Approach with Attention-Based Cues

Isarun Chamveha[1]
isarunc@iis.u-tokyo.ac.jp
Yusuke Sugano[1]
sugano@iis.u-tokyo.ac.jp
Yoichi Sato[1]
ysato@iis.u-tokyo.ac.jp
Akihiro Sugimoto[2]
sugimoto@nii.ac.jp

[1] The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[2] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

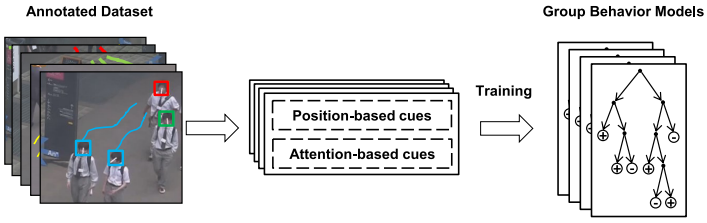

Figure 1: The framework of our proposed approach. Given annotated dataset of pedestrian states, set of measurements based on attention-based and position-based cues is created for each pair of pedestrians. Estimator is then applied to construct set of decision trees representing group behavior models for social group discovery task.

Table 1: Accuracy of estimates with of our dataset. Accuracy is measured as average accuracy between two classes to avoid bias problems in the test samples. **[6]** presented results using the approach of Yamaguchi *et al*. [6]. **Proposed** indicates results with our proposed approach. Note that the frame rate in the datasets is 30 fps.

Dataset	Approach	N_{past}					
		0	30	60	120	240	∞
UT Surveillance	[6]	74.9	75.8	76.5	75.8	76.0	76.4
	Proposed	**76.4**	**76.6**	**77.9**	**77.9**	**80.3**	**81.2**
Town Centre	[6]	68.2	67.5	69.8	69.3	68.5	68.5
	Proposed	**68.3**	**73.9**	**75.4**	**75.2**	**81.4**	**81.8**

Abstract

This paper presents an approach to discover social groups in surveillance videos by incorporating attention-based cues to model group behaviors of pedestrians in videos. Group behaviors are modeled as a set of decision trees with the decisions being basic measurements based on position-based and attention-based cues. Rather than enforcing explicit models, we apply tree-based learning algorithms to implicitly construct the decision tree models. The experimental results demonstrate that incorporating attention-based cues significantly increased the estimation accuracy compared to the conventional approaches that used position-based cues alone.

1 Introduction

Group discovery is an important problem in computer vision as group information can be used to aid pedestrian tracking in low frame-rate videos [5, 6] and to analyze human behavior [1]. Position-based cues, such as the relative position and velocity of pedestrians has been proved to be useful in solving the problem of social group discovery. However, attention-based cues, *i.e*., how people pay attention to one another, have not yet been taken into account there. We aim to combine attention-based cues and position-based cues to discover social groups in this work. We propose a data-driven approach to discover social groups using statistics of both attention-based and position-based cues over pedestrian trajectories.

2 Proposed Framework

We define social group discovery as a pairwise problem to determine whether two people belong to the same group given information on their past states. The framework for our approach is outlined in Figure 1. We model the social group behaviors of pedestrians in a scene from two types of cues: *attention-based cues* and *position-based cues*. Given a set of pair-wise pedestrian states, several measurements of both cues are calculated at each time step, and are aggregated into histograms to evaluate the frequencies of each behavior over the entire trajectory. Using an annotated dataset of pedestrian states, a tree-based learning algorithm is then applied to construct the decision trees representing group behavior models for the social group discovery task.

Two types of attention-based cues are exploited in this work. The first cue is the *gaze exchange* between pedestrians. In order to perform group events, *e.g*., conversation events, pedestrians in the same group often exchanged their gazes and fixed their attentions at one another. The second cue is the *mutual attention* of pedestrians in the same group. This is based on the observation that pedestrians often pay attention to the same object of interest. Measurements of position-based cues are derived from trajectories of two pedestrians based on the approach by Yamaguchi *et al*. [6].

To train decision trees, we calculate a set of measurements for every pair of pedestrians in the training set with overlapping existent. These measurements are aggregated into histograms. The decision trees could be implicitly constructed by using tree-based learning algorithms with the combination of histogram bins as decisions. A random trees classifier [3] was used in our approach to construct the trees. The histograms for each measurement were concatenated to represent the feature vector for that sample, and these vectors were used to train the random trees.

3 Experimental Results

We conducted experiments using two sequences: the UT-Surveillance sequence used by [4] and the Town Centre sequence used by [2] with varying number of available information of past frames N_{past}. We compared our method with the one by Yamaguchi *et al*. [6], who proposed to solve the same problem in the similar setting to our approach The results are listed in Table 1. It can be seen that our approach improves the accuracy of social group discovery tasks in every case in both datasets.

[1] L. Bazzani, M. Cristani, G. Paggetti, D. Tosato, G. Menegaz, and V. Murino. Analyzing groups: A social signaling perspective. In *Video Analytics for Business Intelligence*, pages 271–305. 2012.

[2] B. Benfold and I. Reid. Unsupervised learning of a scene-specific coarse gaze estimator. In *ICCV*, 2011.

[3] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[4] I. Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Okabe, Y. Sato, and A. Sugimoto. Appearance-based head pose estimation with scene-specific adaptation. In *ICCV Workshops*, 2011.

[5] S. Pellegrini, A. Ess, and L. Van Gool. Improving data association by joint modeling of pedestrian trajectories and groupings. In *ECCV*, 2010.

[6] K. Yamaguchi, A.C. Berg, L.E. Ortiz, and T.L. Berg. Who are you with and where are you going? In *CVPR*, 2011.