Crowd Scene Understanding from Group Profiling*

Jing Shao[1] Chen Change Loy[2] Xiaogang Wang[1]

[1]Department of Electronic Engineering, The Chinese University of Hong Kong

[2]Department of Information Engineering, The Chinese University of Hong Kong

jshao@ee.cuhk.edu.hk, ccloy@ie.cuhk.edu.hk, xgwang@ee.cuhk.edu.hk

1. Introduction

Groups are the primary entities that make up a crowd. Understanding group-level dynamics and properties is thus scientifically important and practically useful in a wide range of applications, especially for crowd understanding. Socio-psychologists and biologists [10, 11] have extensively studied group dynamics as the primary processes that influence crowd behaviors. In these studies, group dynamics are characterized by both intra- and inter-group properties.

In the context of visual surveillance, groups also primarily make up human crowds. Indeed, a rich body of literature [2, 6] suggest that majority of the pedestrians tend to move in groups with their friends and family members. When pedestrians form groups, they exhibit some interesting properties in their dynamics, which share commonalities with socio-psychological and biological studies (Fig. 1). For instance, collective behavior is observed when pedestrians in a group maneuver towards a common destination. During a crowd disaster, turbulent dynamics in the crowd can be characterized by the stability property. Crowd tends to have non-uniform distribution when its members have different social relationships and walk in less restricted area. Two pedestrian groups with different goals, *e.g.* when crossing roads from different directions, exhibit conflict behavior. Clearly, understanding such properties provides critical mid-representation to crowd motion analysis [1, 7], and could facilitate other high-level semantic analysis such as crowd scene understanding, crowd video classification, and crowd event retrieval.

2. Methodology and Experiment

Our goal is to characterize and quantify these group properties from vision point of view, and study their potentials on crowd behavior analysis and crowd scene understanding. We consider a group beyond just a collection of spatially proximate individuals, but also a dynamic unit that exhibits various fundamental intra- and inter-group properties, which can be used to compare group activities across different crowd systems. The proposed framework is shown in Fig. 2. In detail, we make the following contributions:

Figure 1. Crowd behavior can be better understood through inherent intra- and inter-group properties. In this study, we show the possibility of quantifying such properties with scene-independent visual descriptors. Best viewed in color.

1) *A robust group detector* - We introduce a novel Collective Transition (CT) prior to capture the underlying dynamics of a group. The CT prior of a group refers to the transition matrix describing coherent motion of all group members. Based on this prior we formulate a robust group detector which outperforms state-of-the-art methods [4, 12].

2) *Scene-independent group descriptors* - Based on the CT prior, we devise a set of visual descriptors to quantify four fundamental intra- and inter-group properties, namely collectiveness, stability, uniformity, and conflict. These descriptors convey richer group-level information in comparison to the conventional group size and velocity information [3]. Importantly, these descriptors are scene invariant and robust to public scenes with variety of crowdedness.

Collectiveness: The collectiveness property indicates the degree of individuals acting as a union in collective motion. Differ from [13] using manifold learning, we quantify collectiveness at group level with the CT prior, since it captures the coherent motion of all group members.

Stability: The stability property characterizes whether a group can keep internal topological structure over time. In particular, stable members tend to (1) maintain a similar set of nearest neighbors; (2) keep a consistent topological dis-

*The long version "Scene-Independent Group Profiling in Crowd" is presented in the main conference as oral.

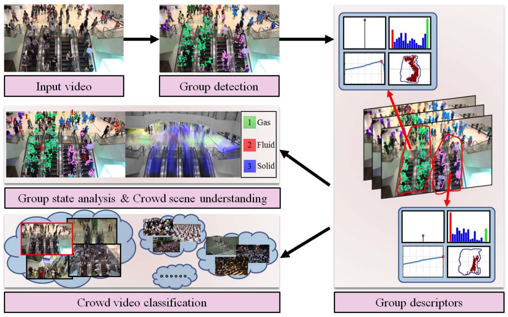

Figure 2. The proposed framework.

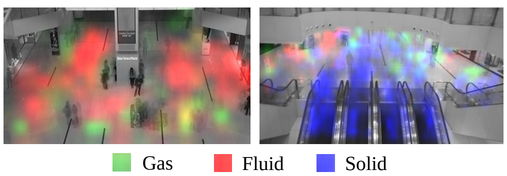

Figure 3. Distributions of different types of groups in a shopping mall and a escalator scene. Colors indicate group states automatically recognized with our descriptors. Best viewed in colour.

tance with its neighbors throughout a clip; and (3) a member is less likely to leave its current nearest neighbor set.

Uniformity: In contrast to the two previous properties that measure temporal aspects, uniformity property characterizes homogeneity of a group in terms of spatial distribution. We quantify it by inferring the optimal number of graph cuts on the K-NN graph.

Conflict: Conflict describes interaction/friction between groups when they approach each other. We quantify this property via both spatial distribution by shape context and level of conflict by CT prior.

3) *Group-driven crowd scene understanding* - We show that the proposed descriptors are effective in identifying the intrinsic group states (gases, fluids, and solid) following the common analogy employed in crowd modeling literature [9, 5]. These states are useful in crowd scene understanding because they are decided by crowd density, goals, interactions and relationships of group members, and scene structures. As examples shown in Fig. 3, in a large open area, pedestrians behave more like gas and fluid, while move as flying solid on an escalator track or in a queue.

We also demonstrate their superiority for crowd video classification over existing activity descriptors [8]. All the 474 video clips in our dataset are manually assigned into 8 classes as shown in Table 1. The 8 classes are commonly seen in crowd videos and some are of special interest in crowd management and traffic control. Leave-one-out evaluation is used. Each time one scene (which may include multiple video clips) is selected for test, and the remaining scenes for training. Thus it tests the cross-scene generalization capability. SVM is used for classification. The confusion matrices are shown in Figure 4. The average accuracy

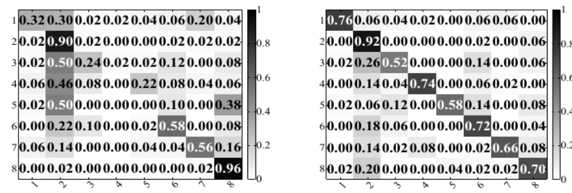

Figure 4. Confusion matrices of crowd video classification. Left: using holistic features in [8]. The average accuracy is 44%. Right: using our descriptors. The average accuracy is 70%.

Table 1. List of crowd video classes.

	Class name
1	Highly mixed pedestrian walking
2	Crowd walking following a mainstream and well organized
3	Crowd walking following a mainstream but poorly organized
4	Crowd merge
5	Crowd split
6	Crowd crossing in opposite directions
7	Intervened escalator traffic
8	Smooth escalator traffic

of our approach is shown 70%, much higher than that of random guess (12.5%) and the result of using the holistic crowd scene descriptor proposed in [8] (44%).

All the experiments are conducted on a new CUHK Crowd Dataset[1] with hundreds of video clips collected from over 200 crowded scenes.

References

[1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *CVPR*, 2007. 1

[2] A. F. Aveni. The not-so-lonely crowd: Friendship groups in collective behavior. *Sociometry*, 40:96–99, 1977. 1

[3] M.-C. Chang, N. Krahnstoever, and W. Ge. Probabilistic group-level motion analysis and scenario recognition. In *ICCV*, 2011. 1

[4] W. Ge, R. T. Collins, and R. B. Ruback. Vision-based analysis of small groups in pedestrian crowds. *TPAMI*, 34(5):1003–1016, 2012. 1

[5] D. Helbing and A. Johansson. Pedestrian, crowd and evacuation dynamics. In *Extreme Environmental Events*, pages 697–716. 2011. 2

[6] I. Karamouzas and M. Overmars. Simulating and evaluating the local behavior of small pedestrian groups. *TVCG*, 18(3):394–406, 2012. 1

[7] S. Kim, S. J. Guy, and D. Manocha. Velocity-based modeling of physical interactions in multi-agent simulations. In *SIGGRAPH*, pages 125–133. ACM, 2013. 1

[8] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *CVPR*, 2009. 2

[9] J. Toner, Y. Tu, and S. Ramaswamy. Hydrodynamics abd phases of flocks. *Annals of Physics*, 318:170–244, 2005. 2

[10] S. A. Wheelan. *The handbook of group research and practice*. Sage, 2005. 1

[11] H.-P. Zhang, A. Beér, E.-L. Florin, and H. L. Swinney. Collective motion and density fluctuations in bacterial colonies. *PNAS*, 107(31):13626–13630, 2010. 1

[12] B. Zhou, X. Tang, and X. Wang. Coherent filtering: detecting coherent motions from crowd clutters. In *ECCV*. 2012. 1

[13] B. Zhou, X. Tang, and X. Wang. Measuring crowd collectiveness. In *CVPR*, 2013. 1

[1] http://www.ee.cuhk.edu.hk/~xgwang/CUHKcrowd.html