Finding Coherent Motions and Understanding Crowd Scenes: A Diffusion and Clustering-based Approach

Weiyao Lin[1], Yang Mi[1], You-Ping Zhong[2], and Weiyue Wang[1]

[1]Department of Electronic Engineering, Shanghai Jiao Tong University, China
[2]National Grid Jiangxi Electric Power Co. Maintenance Branch, China

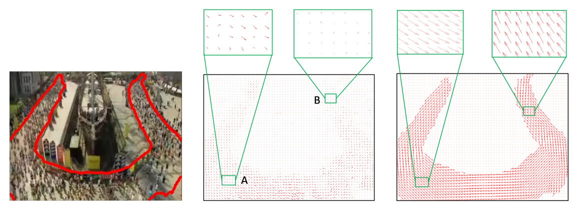
Abstract

Coherent motions, which represent coherent movements of massive individual particles, are pervasive in natural and social scenarios. Examples include traffic flows and parades of people (cf. Figs 1a and 2a). Since coherent motions can effectively decompose scenes into meaningful semantic parts and facilitate the analysis of complex crowd scenes, they are of increasing importance in crowd-scene understanding and activity recognition.

In this paper, we address the problem of detecting coherent motions in crowd scenes, and subsequently using them to understand input scenes. More specifically, we focus on 1) constructing an accurate coherent motion field to find coherent motions, 2) finding stable semantic regions based on the detected coherent motions and using them to recognize pre-defined activities (i.e., activities with labeled training data) in a crowd scene, and 3) automatically mining recurrent activities in a crowd scene based on the detected coherent motions and semantic regions.

First, constructing an accurate coherent motion field is crucial in detecting reliable coherent motions. In Fig. 1, (b) is the input motion field and (c) is the coherent motion field which is constructed from (b) using the proposed approach. In (b), the motion vectors of particles at the beginning of the Marathon queue are far different from those at the end, and there are many inaccurate optical flow vectors. Due to such variations and input errors, it is difficult to achieve satisfying coherent motion detection results directly from (b). However, by transferring (b) into a coherent motion field where the coherent motions among particles are suitably highlighted in (c), coherent motion detection is greatly facilitated. Although many algorithms have been proposed for coherent motion detection [1, 9, 10, 12, 3], this problem is not yet effectively addressed. We argue that a good coherent motion field should effectively be able to *1)* encode motion correlation among particles, *such that particles with high correlations can be grouped into the same coherent region; and, 2)* maintain motion information of individual particles, *such that activities in crowd scenes can be effectively parsed by the extracted coherent motion field. Based on these intuitions, we propose a thermal-diffusion-based approach, which can extract accurate coherent motion fields.*

Second, constructing meaningful semantic regions to de-



(a) (b) (c)

Figure 1. (a) Example frame of a Marathon video sequence, the red curve is the coherent motion region; (b) Input motion vector field of (a); (c) Coherent motion field from (b) using the proposed approach (Best viewed in color).

(a) (b) (c)

Figure 2. (a) Example time-varying coherent motions in a scene, where different coherent motions are circled by curves with different color; (b) Constructed semantic regions for the scene in (a); (c) Recurrent activities for the scene in (a), where the arrows represent the major motion flows in each recurrent activity (Best viewed in color).

scribe activity patterns in a scene is also essential. Coherent motions at different times may vary widely. In Fig. 2a, changing of traffic lights will lead to different coherent motions. Coherent motions alone may not effectively describe the overall semantic patterns in a scene either. Therefore, semantic regions need to be extracted from these time-varying coherent motions to achieve stable and meaningful semantic patterns, as in Fig. 2b. However, most existing works only focus on the detection of coherent motions at some specific time, while the problem of handling time-varying coherent motions is less studied. We proposed a two-step clustering process for this purpose.

Third, mining recurrent activities is another important issue. Many crowd scenes are composed of recurrent activities [2, 5, 11]. For example, the scene in Fig. 2 is composed of recurrent activities including vertical motion activities and horizontal motion activities, as in Fig. 2c. Automatically mining these recurrent activities is important in understanding scene contents and their dynamics. Although many researches have been done for parsing recurrent activities in low-crowd scenes [8, 4, 6, 7], this issue is not well addressed in crowd scene scenarios where reliable motion trajectories are unavailable. We proposed a cluster-and-merge process, which can effectively extract recurrent activities in crowd scenes.

Our contributions to crowd scene understanding and activity recognition are summarized as follows.

1. *We introduce a coarse-to-fine thermal diffusion process to transfer an input motion field into a thermal energy field (TEF), which is a more accurate coherent motion field. TEF effectively encodes both motion correlation among particles and motion trends of individual particles. To our knowledge, this is the first work that introduces thermal diffusion to detect coherent motions in crowd scenes. We also introduce a triangulation-based scheme to effectively identify coherent motion components from the TEF.*

2. *We present a two-step clustering scheme to find semantic regions according to the correlations among coherent motions. The found semantic regions can effectively catch activity patterns in a scene. Thus good performance can be achieved when recognizing predefined crowd activities based on these semantic regions.*

3. *We propose a cluster-and-merge process to automatically mine recurrent activities by clustering and merging the coherent motions. The obtained recurrent activities can accurately describe recurrent motion patterns in a crowd scene.*

Acknowledgements

This work was supported in part by the National Science Foundation of China grants (61471235, 61472370 and 61202207) and the Chinese national 973 grants (2013CB329603).

References

[1] S. Ali and M. Shah. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 1–6, 2007. 1

[2] R. Emonet, J. Varadarajan, and J.-M. Odobez. Temporal analysis of motif mixtures using Dirichlet processes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 36(1):140–156, 2013. 2

[3] M. Hu, S. Ali, and M. Shah. Learning motion patterns in crowded scenes using motion flow field. *Intl. Conf. Pattern Recognition (ICPR)*, pages 1–5, 2008. 1

[4] W. Hu, X. Li, G. Tian, S. Maybank, and Z. Zhang. An incremental DPMM-based method for trajectory clustering, modeling, and retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(5):1051–1065, 2013. 2

[5] V. Jagannadan, R. Emonet, and J.-M. Odobez. A sequential topic model for mining recurrent activities from long term video logs. *Intl. J. Computer Vision*, 103(1):100–126, 2013. 2

[6] B. Morris and M. Trivedi. Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(11):2287–2301, 2011. 2

[7] J. Nascimento, M. A. T. Figueiredo, and J. S. Marques. Trajectory classification using switched dynamical Hidden Markov Models. *IEEE Trans. Image Processing*, 19(5):1338–1348, 2010. 2

[8] X. Wang, K. T. Ma, G. Ng, and E. Grimson. Trajectory analysis and semantic region modeling using nonparametric Bayesian models. *Intl. J. Computer Vision*, 96:287–321, 2011. 2

[9] S. Wu and H. Wong. Crowd motion partitioning in a scattered motion field. *IEEE Trans. Systems, Man, and Cybernetics*, 42(5):1443–1454, 2012. 1

[10] B. Zhou, X. Tang, and X. Wang. Coherent filtering: detecting coherent motions from crowd clutters. *European Conf. Computer Vision (ECCV)*, pages 857–871, 2012. 1

[11] B. Zhou, X. Wang, and X. Tang. Random field topic model for semantic region analysis in crowded scenes from tracklets. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 3441–3448, 2011. 2

[12] B. Zhou, X. Wang, and X. Tang. Measuring crowd collectiveness. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 3049–3056, 2013. 1