Online Bayesian Nonparametrics for Group Detection

Matteo Zanotto
matteo.zanotto@iit.it
Loris Bazzani
loris.bazzani@iit.it
Marco Cristani
marco.cristani@iit.it
Vittorio Murino
vittorio.murino@iit.it

Pattern Analysis & Computer Vision
Istituto Italiano di Tecnologia
Via Morego 30 - 16163
Genova, Italy

Social interactions are essential in our daily activities and automatic modelling of interactional exchanges has become an active research topic over the last few years. An important step towards the analysis of social behaviour and the understanding of social interactions involves the identification of groups of people, which is the goal of this work.
We propose to perform group detection from videos in real surveillance scenarios through a Dirichlet Process Mixture Model (DPMM) [1]. Within this model, groups are represented as components of an infinite mixture model, and individuals are seen as observations generated from them. Each individual is represented by its position (x and y coordinates) and velocity (decomposed in heading direction and module) and groups are detected in an unsupervised way by performing clustering in the feature space defined above. The introduction of a "social" constraint based on proxemics rules proposed by Hall [3] allows to only maintain components associated to groups satisfying theories from social psychology.
In order to keep computation fast and compatible with real-time video analysis, we propose a sequential variational inference performing single parameters updates rather then iterating to convergence for each single frame (see Figure 1). This approach builds upon ideas by Neal and Hinton [5] and is possible because grouping configurations evolve smoothly, allowing to exploit the temporal correlation across consecutive frames to refine the detection of groups taking advantage of the evolution of the observations. The posterior distribution estimated at one time step can then act as prior knowledge for the following one, distributing inference over time. This sequential approach also allows to take advantage of the dynamics of observations evolution without explicitly modelling it, which is a valuable feature in cases where no prior knowledge on such dynamics is available to be coded in the model.

The method has been tested on two public benchmark datasets for group detection: the *BIWI* dataset [6], containing the `eth` and `hotel` sequences, and the *Crowd by Example* dataset [4], containing the `zara01`, `zara02`, `students003` sequences. In order to capture all the relevant aspects of group detection, comparison with two benchmark methods has been performed. The proposed method outperforms that by Yamaguchi *et al.* [7] on 4 out of the 5 considered sequences, and that by Bazzani *et al.* [2] on the chosen test sequence. Examples of the qualitative results obtained by our method are reported in Figure 2 along with the ground truth.

Summarising in this paper we propose a model for group detection which

- Does not require to fix a priori the number of groups to find.
- Can dynamically adapt the number of groups from frame to frame to effectively match the observed data, also coping with split and merge of groups.
- Can take advantage of the dynamics of the data without explicitly modelling it.
- Produces results through online inference.
- Can perform real-time processing up to 42 fps (bounded by the pedestrian detector performance).

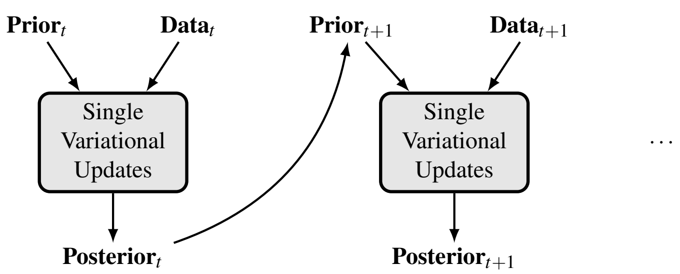

Figure 1: Pictorial representation of the proposed single-step variational inference scheme.

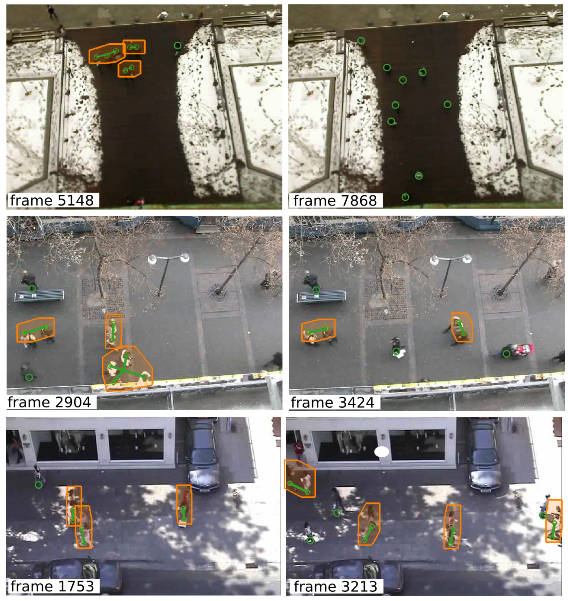

Figure 2: Qualitative results on `eth` (first row), `hotel` (second row), and `zara01` (last row). The ground truth position of individuals and groups are shown with green circles and segments, respectively. The estimated groups are depicted as orange convex hulls (best viewed in colors).

Experimental results show that our method outperforms state-of-the-art methods on evaluation metrics capturing different aspects of group detections, suggesting a better overall performance.

[1] C.E. Antoniak. Mixtures of Dirichlet Processes with Applicatioins to Bayesian Nonparametric Problems. *The Annals of Statistics*, 2(6): 1152–1174, 1974.

[2] L. Bazzani, V. Murino, and M. Cristani. Decentralized particle filter for joint individual-group tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012.

[3] E.T. Hall. *The Hidden Dimension*. 1966.

[4] A. Lerner, Y. Chrysanthou, and D. Lischinski. Crowds by example. volume 26, pages 655–664, Sep 2007.

[5] R.M. Neal and G.E. Hinton. A new view of the EM algorithm that justifies incremental and other variants. In Michael I. Jordan, editor, *Learning in Graphical Models*. MIT Press, 1998.

[6] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *International Conference on Computer Vision*, 2009.

[7] K. Yamaguchi, A.C. Berg, L.E. Ortiz, and T.L. Berg. Who are you with and where are you going? In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011.