Crowd Motion Analysis for Group Detection

Neha Bhargava
Indian Institute of Technology Bombay
Mumbai
India
neha@ee.iitb.ac.in

Subhasis Chaudhuri
Indian Institute of Technology Bombay
Mumbai
India
sc@ee.iitb.ac.in

ABSTRACT

Understanding crowd dynamics is an interesting problem in computer vision owing to its various applications. We propose a dynamical system to model the dynamics of collective motion of the crowd. The model learns the spatio-temporal interaction pattern of the crowd from the track data captured over a time period. The model is trained under a least square formulation with spatial and temporal constraints. The spatial constraint allows the model to consider only the neighbors of a particular agent and the temporal constraint enforces temporal smoothness in the model. We also propose an effective group detection algorithm that utilizes the eigenvectors of the interaction matrix of the model. The group detection is cast as a spectral clustering problem. Extensive experimentation demonstrates a superlative performance of our group detection algorithm over state-of-the-art methods.

CCS Concepts

•**Computing methodologies** → *Spectral methods; Modeling methodologies;* Motion capture;

Keywords

crowd dynamics, group detection

1. INTRODUCTION

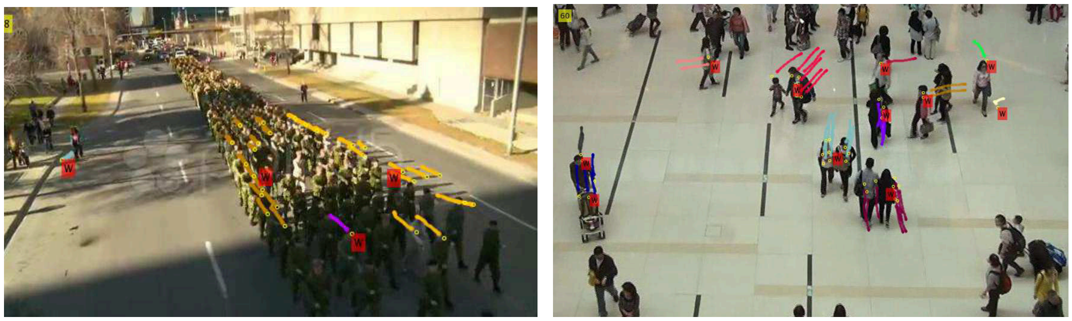
Understanding human behavior in different scenarios has always attracted the researchers. The variability and complexity in the behavior make it a highly challenging task. However, this decade is witnessing a huge interest of researchers in the area of crowd motion analysis due to its various applications in surveillance, safety, public place management, hazards prevention, and virtual environments. This interest has resulted in many interesting papers in the area. We are aware of at least four survey papers on the subject of crowd analysis that indicate the amount of attention, it has drawn in this and the previous decade [8],[18],[7],[20]. The latest survey paper [8] by Chang *et al.* encapsulates the recent works published after 2009, covering topics of motion pattern segmentation, crowd behavior and anomaly detection. Thida *et al.* [18]provide a review on macroscopic and microscopic modeling methods. They also present a critical survey on crowd event detection. Julio *et al.* cover various vision techniques applicable to crowd analysis such as tracking, density estimation, and computer simulation [7]. Zhan *et al.* discuss various vision based techniques used in crowd analysis. They also discuss crowd analysis from the perspective of different disciplines − psychology, sociology and computer graphics [20]. At the top level, the techniques used in crowd motion analysis can be divided into two major classes − holistic and particle based. The holistic methods consider crowd as a single entity and analyze the overall behavior. These methods fail to provide much insight at an individual or intermediate level. On the other hand, particle based methods consider crowd as a collection of individuals. But their performance degrades with the increase in crowd density due to occlusion and tracking problems. Hence the group level analysis and consequently group detection algorithm becomes important.

We believe that a moderately dense crowd consists of groups. We define a group as a set of individuals having some sort of interaction to achieve a common goal, e.g. walking together to reach the same destination. Spatial proximity is required to form a group; if there are agents with a similar motion pattern but are far away from each other, they do not form a group as per our definition. Each group has its own set of goals that leads to various interaction patterns among the members of the group and together form a crowd. The crowd behavior can vary from a highly structured to a totally unstructured pattern. In case of a structured crowd, for example − marching of soldiers, all groups are in coordination and share the same goal (see Figure.1a); whereas in an unstructured crowd, for example − at railway station or at a shopping complex, there are multiple groups with different goals (see Figure.1b). We are interested in modeling such crowd dynamics and identify the groups. The paper has following contributions:

1. A framework is proposed to model the collective motion of the crowd by a first order dynamical system. The model captures the interaction patterns among the individuals. Although the proposed model does not capture the possible non-linear relations, its usefulness for short-term analysis has been verified experimentally.

2. We also provide an optimization formulation for the

ICVGIP, December 18-22, 2016, Guwahati, India
 ISBN 978-1-4503-4753-2/16/12. . . $15.00
DOI: http://dx.doi.org/10.1145/3009977.3010071



(a) Uniform crowd (b) Mixed crowd

Figure 1: Output of the our proposed group detection algorithm: (a) and (b) give examples of structured and unstructured crowd. Tracklets for some of the agents over past few frames are also shown. Each color represents a group (Best viewed in color). The videos are from CUHK [14] datasets

estimation of the interaction matrix under the constraints of temporal continuity, spatial restriction and sparsity of inter-agent relationship.

3. Since the interaction matrix is learned from the trajectory data, it captures the spatio-temporal patterns among the agents. We observe that the eigenvectors reflect the spatio-temporal patterns captured by the matrix. Thus, we propose a spectral clustering [10] based algorithm to identify the groups present in the scene. Extensive experimentation demonstrates the effectiveness of the algorithm.

The remaining part of the paper is organized as follows. Next section reviews the related literature. Section 3 explains the proposed mathematical formulation followed by group detection algorithm in Section 4. The experimental results are presented in Section 5 followed by conclusions in Section 6.

2. RELATED WORK

There are numerous research papers in the challenging and interesting area of crowd behavior analysis and group detection. There are many holistic approaches (e.g. [9], [16], [2]) as well as particle based algorithms (e.g. [6], [13], [5], [21]) in the literature. Holistic methods analyze crowd as a single entity and ignore individuals or groups. In many papers, a dense crowd is considered analogous to fluid and hence concepts from fluid mechanics are applied for analysis. Mehran *et al.* in [9] present streakline representation of crowd flow for behavior analysis. Solmaz *et al.* recognize crowd behaviors such as bottlenecks, fountainheads, lanes, arches and blocks through stability analysis of a dynamical system [16]. Benabbas *et al.* detect motion patterns and events in the crowded scenes by modeling motion and velocity at each spatial location [2].

The agent based approaches analyze each individual or group to discover the global behavior. Shao *et al.* introduce a collective transition prior in [14] and represent each group by a Markov chain. They define interesting group descriptors which proved to be useful in group state analysis and crowd classification. In [13], Sethi and Chowdhury propose phase space algorithm to identify pairwise correlation between the motion patterns. Ge *et al.* find groups by hierarchical clustering based on pairwise velocities and distance [4], [5]. Zhou *et al.* find groups by using coherent filtering [21]. They propose a coherent neighbor invariance which characterizes coherent moving individuals. Sochman *et al.* [15] infer groups based on social force model [6]. They define a pairwise group activity confidence to identify groups. Srikrishnan and Chaudhuri in [17] define a linear cyclic pursuit based framework for collective motion modeling with the goal of short-term prediction. But they do not explore group detection and there is no analysis of crowd behavior.

Most of the particle based algorithms compute pairwise velocity and spatial cues to find the groups hierarchically. They do not model spatio-temporal patterns of the agents collectively which might capture more complex interactions. In this work, we model motion trajectories collectively. Also instead of relying on spatio-temporal information (which is prone to noise) directly for group detection, we use spectral clustering to identify groups.

3. MATHEMATICAL FORMULATION

We define a group as a set of agents having spatial proximity and some sort of interaction. In general, such interactions are complex and non-linear in nature. We approximate these interactions locally in time by a first order dynamical model. Note that we refer by agent an individual entity (represented by a point to be tracked) in the crowd.

3.1 Proposed Interaction Model

We model the collective relationship among the agents by a first order affine system. Our hypothesis is based on the intuition that each agent takes into consideration (a) the movement of other agents present nearby and (b) her/his desired goal, while taking the next step. The model relates the next positions of the agents to the current positions. Let $\mathbf{x}(k) = [x_1(k), x_2(k), ..., x_N(k)]^T$, then

$$\mathbf{x}(k+1) = [\mathbf{A}_k|\mathbf{a}_k]\begin{bmatrix}\mathbf{x}(k)\\ 1\end{bmatrix} = \mathbf{A}'_k\mathbf{x}'(k) \qquad (1)$$

where N is the total number of agents, $\mathbf{A}_k \in \mathbb{R}^{N\times N}$, $\mathbf{A}'_k \in \mathbb{R}^{(N+1)\times(N+1)}$, $\mathbf{a}_k \in \mathbb{R}^{N\times 1}$, $\mathbf{x}'(k) \in \mathbb{R}^{(N+1)\times 1}$ and $x_i(k) \in \mathbb{R}$ is the location of i^{th} agent at time instant k along x-axis. We call $\mathbf{A}_k$ as the interaction matrix which captures the evolution of an agent as a function of all agents present in the scene. Note that $\mathbf{A}_k$ has no assumption on its form and entries. It need not be symmetric i.e. agent i may not depend on agent j in the same way as agent j depends on agent. For example, consider a case where agent i is stationary and agent j approaches him/her. Since their behaviors are not symmetric with respect to each other, we assume that it implies $a_k(i,j) \neq a_k(j,i)$.

In this paper, it is assumed that the motion along x and y directions are independent and hence can be analyzed independently. However, we experimented with concatenated x and y but that did not improve the results. On the contrary, it increased the size of interaction matrix leading to requirement of more past data to learn and leads to chances of violation of short-term assumption. Hence to avoid this violation, we analyze the motion in x and y direction separately. The corresponding model along y direction is $\mathbf{y}(k+1) = \mathbf{B}_k\mathbf{y}(k) + \mathbf{b}_k$. In the rest of the paper, we discuss the solution for matrix $\mathbf{A}_k$ noting this fact that the same process is also carried out for $\mathbf{B}_k$. We expect matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ to be dependent on crowd motion. Since crowd behavior might change with time, the interaction matrix is time varying in nature, which we represent as $\mathbf{A}_k$ where k is a time instant. Assuming $\mathbf{A}'_k$ has $N+1$ independent eigenvectors, the general solution to Eq.(1) is given as

$$\mathbf{x}(k) = \sum_{\substack{i=1\\ \lambda_i \neq 1}}^{N+1} \{c_i\lambda_i^k\mathbf{v}_i + d_i\frac{(\lambda_i^k - 1)}{\lambda_i - 1}\mathbf{v}_i\} + \sum_{\substack{i=1\\ \lambda_i = 1}}^{N+1} (c_i + kd_i)\mathbf{v}_i, \quad (2)$$

where λ_i is the i^{th} eigenvalue, $\mathbf{v}_i$ is the corresponding normalized eigenvector, c_i and d_i are the corresponding constant coefficients that depend on the initial condition and $\mathbf{a}$ respectively. Different values of λ_i and $\mathbf{v}_i$ generate various motion patterns for an agent. These patterns can be associated to different motion tracks generated by an agent while walking, approaching, splitting or stationary.

3.2 Estimation of Interaction Matrix

The matrix $\mathbf{A}'_k$ at any time instant is learned from the immediate past trajectory data of all the agents in a least squares framework. We update $\mathbf{A}'_\mathbf{k}$ with each incoming frame as interaction patterns may change over the time. In addition, sudden changes in these interactions are unlikely. Therefore it is desired that the entries of $\mathbf{A}'_k$ do not change drastically in consecutive time instants − we assume them to be varying smoothly over time. We incorporate this constraint by minimizing l_2 norm of the difference between current matrix $\mathbf{A}'_k$ and previous estimate at $(k-1)^{th}$ instant. Furthermore for crowded scenes, it is unlikely that an agent's motion depends on all the agents present in the scene. We capture sparsity in $\mathbf{A}'_k$ by minimizing l_1 norm of $\mathbf{A}'_k$. Adding these constraints to the cost function, the final formulation at k^{th} time instant becomes:

$$\begin{aligned}\hat{\mathbf{A}}'_k = \quad & \arg \min_{\mathbf{A}'_k \in \mathbf{R}^{N\times(N+1)}} \Big\{ ||\mathbf{A}'_k\mathbf{X}^{k-1}_{k-L} - \mathbf{X}^{k}_{k-L+1}||_2^2 \\ & + r_1||\mathbf{A}'_k - \mathbf{A}'_{k-1}||_2^2 + r_2||\mathbf{A}'_\mathbf{k}||_1 \Big\}, \qquad (3)\end{aligned}$$

where $\mathbf{X}^j_i \in \mathbb{R}^{N\times L}$ contains the positions of all N agents from i^{th} to j^{th} frames concatenated together, $\mathbf{A}'_{k-1}$ is the estimate at the previous frame and r_1 and r_2 are appropriate regularization parameters. Note that we will use $\mathbf{A}'$ instead of $\mathbf{A}'_k$ for notation convenience.

One requires at least $L \geq (N+1)$ past positions to solve this least squares. Therefore the interaction pattern is assumed to remain constant over L frames. However, a large N leads to two major problems: (i) longer trajectories (i.e. higher L) are required to learn the interaction matrix $\mathbf{A}'$ which may not be available and (ii) the interaction may not remain constant over L past positions. To address these problems, we identify spatial neighbors of each agent separately and learn only the corresponding entries in the matrix (one row at a time), others are kept as zero. The neighborhood is defined as follows − the agent $\mathbf{p}$ is a neighbor to the agent $\mathbf{q}$ if $dist(\mathbf{p},\mathbf{q}) < R_\mathbf{p}$. The assumption is that it is unlikely that far away agents influence the motion of an agent. The advantage is that the shorter trajectories are now sufficient as the number of entries of $\mathbf{A}'$ to be learned are lesser. Note that we estimate matrix $\mathbf{A}'$ in a row-wise manner where i^{th} row has number of entries to be estimated as equal to one plus the number of the neighbors of agent i due to neighborhood constraint. Further, there could be an agent within the spatial proximity of another agent but there may not be any interaction between them. Hence it is required that the corresponding entry in the matrix $\mathbf{A}'$ should be zero. This is enforced by adding sparsity constraint in Eq. 3. We use L1General package developed by Schmidt [12] for solving L1-regularization problems.

For an illustration, see Figure.2. There are a total of $N = 20$ agents present in the scene. Estimation of the row of matrix $\mathbf{A}$ corresponding to agent $\mathbf{p}$ requires 50 previous frames (assuming $L = 2.5N$) whereas the neighborhood based estimation reduces this to 23. Also consider a case where agents $\mathbf{p}$ and $\mathbf{r}$ interact with each other but are not within the spatial proximity owing to neighborhood constraint. The interaction is captured when intersection of neighborhoods of $\mathbf{p}$ and $\mathbf{r}$ has at least one interacting agent, in this case its $\mathbf{q}$ who is in the spatial proximity of both.

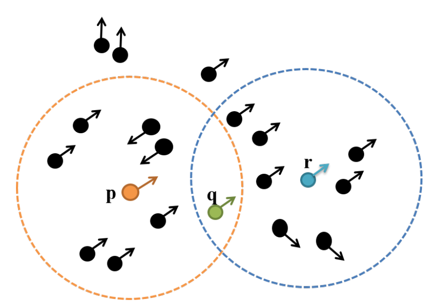


Figure 2: **Neighborhood criteria**: Spatial neighborhoods around agents $\mathbf{p}$ and $\mathbf{r}$ are represented as circles around them. There are a total of 20 agents in the scene out of which only 8 are neighbors of $\mathbf{p}$. Estimation of elements of row of $\mathbf{A}$ corresponding to agent $\mathbf{p}$, considering all agents present in the scene requires $2.5 \times 20 = 50$ previous video frames (assuming $L = 2.5N$). While the use of neighborhood constraint reduces this to $2.5 * 9 \approx 23$ frames.

3.3 Validation of the Model

We use an average k-step prediction error as a measure to test the validity of the proposed model on real videos. Figure. 3 shows average errors for different step size prediction on videos from BEHAVE and CUHK datasets, each curve corresponding to a different video. The k-step prediction

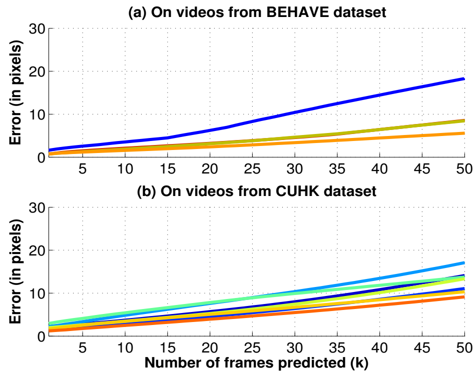


Figure 3: **Illustration of suitability of the proposed model**: Average k-step prediction error for sample videos from BEHAVE and CUHK datasets, each curve corresponds to different video

error at any time instant n is calculated as follows:

$$E_n(k) = \frac{1}{kN} \sum_{i=1}^{k} \sum_{j=1}^{N} |x_j^{actual}(n+i) - x_j^{pred}(n+i)|, \quad (4)$$

It may be noted that matrix $\mathbf{A}$ is estimated from the latest video frames upto n and then Eq. 1 is used to obtain x_j^{pred}. The k-step prediction error for the video is obtained by averaging $E_n(k)$ over all the frames of the video. As expected, error increases with k but with a marginal increment. We observe that, for both the databases, prediction is quite valid up to 1-1.5 seconds (about 40 frames). Since the model assumes that the interaction remains same for some time, it may not capture the changes occurred during the longer duration which may lead to significant error at higher k. These error plots show that the proposed model is suitable for short-term analysis, which is the underlying theme of the proposed algorithm.

4. GROUP DETECTION ALGORITHM

In this section, we discuss the algorithm for identifying the groups present in the scene by analyzing the interaction matrix $\mathbf{A}$. Let eigenvector matrix contains all the eigenvectors column-wise. From Eq. 2, notice that if any two rows of eigenvector matrix are similar, the corresponding agents belong to same group. Hence, we define a mapping for i^{th} agent as

$$f(x_i) : x_i \in \mathbb{R} \;\rightarrow\; \mathbf{z}_i = (v_{1i}, v_{2i}, \ldots, v_{ri})^T \in \mathbb{R}^{r\times 1}$$

where v_{ji} is the i^{th} entry of j^{th} eigenvector of interaction matrix $\mathbf{A}$ and r is the number of significant eigenvalues. A clustering algorithm is applied on the points $\{\mathbf{z}_i\}, \forall i = 1, 2, \ldots, N$ to identify the groups. The clustering algorithm runs on the components of eigenvectors, therefore this algorithm falls in the category of spectral clustering [10]. Since the number of groups is unknown, we apply a threshold based clustering. The adaptive threshold used for i^{th} point is $c|\mathbf{z}_i|$, where $|\mathbf{z}_i|$ is its magnitude and c is found empirically. For example, all the agents within the distance of $c|\mathbf{z}_1|$ from $\mathbf{z}_1$ will form a group with agent 1. In this way, all the groups are obtained. Also we consider only significant eigenvectors with $|\lambda| \geq 0.90$, of $\mathbf{A}$ for group detection since the response from the eigenvectors with $|\lambda| < 0.9$ dies down to an insignificant level within the period of L frames (about 10% level for $N = 8$ and $L = 2.5N$).

It may be noted that this group detection algorithm remains same in the case where $\mathbf{A}$ does not have N independent eigenvectors. In such a case, the clustering algorithm runs on generalized eigenvectors.

Note that the group detection algorithm runs in x and y directions independently and results need to be combined together. For group detection, a group is formed only if it is formed in both the directions. For example, if $Z_x = [1, 1, 2, 1]$ and $Z_y = [2, 1, 2, 2]$ are the label vectors (indicating assigned group number for all the four agents) obtained in x and y directions respectively, the final label vector would be $Z = [1, 2, 3, 1]$. That is, out of 4 agents, 1 and 4 are grouped together while agents 2 and 3 are separate groups.

5. EXPERIMENTS AND RESULTS

We tested our algorithms on BEHAVE [3] and CUHK datasets [14] which are quite common among the researchers for crowd analysis and group activity detection . CUHK dataset is a comprehensive crowd video dataset containing 474 video clips covering various crowd behaviors with varying crowd density. BEHAVE dataset has video clips covering various types of group activities. We tested group detection algorithm on all the 474 videos from CUHK dataset and on 2 video clips (having duration of more than 10 minutes) from BEHAVE dataset. We compared the proposed algorithm with other methods on these selected agents. The ground truth for CUHK dataset was obtained manually.

Table 1: Performance comparison of different group detection algorithms

	CF [21]	CT [14]	*Proposed*
NMI	0.66	0.69	**0.86**
Purity	0.71	0.72	**0.90**
RI	0.67	0.69	**0.85**

We compare the proposed algorithm for group detection with state-of-the-art methods by Shao *et al.* [14] and Zhou *et al.* [21]. Comparison with other major recent works is already mentioned in [14]. Hence we restrict our comparison to these two for brevity. Furthermore, for quantitative analysis on CUHK videos, we randomly select two time instants for each video where we compare the proposed algorithm with other methods and ground truth instead of manually deciding on the instants when the performance has to be evaluated. We use Normalized Mutual Information (NMI) [19], Purity [1] and Rand Index (RI) [11] which are widely used for evaluation of clustering algorithms. NMI is inspired by information theory concepts while Rand Index penalizes both false positives and false negatives during clustering. Table 1 shows the comparison on these measures. It is quite evident from the table that the performance of the proposed algorithm far surpasses those of [14] and [21].

Figure. 4 demonstrates a visual comparison for different scenarios. Since Zhou *et al.* in [21] find coherent motion patterns at one time and then update them over time, hence it is sensitive to tracking errors and has the possibility of accumulation of errors if any frame has tracking error. Shao *et al.* [14] assign every agent to a collective transition prior. They have spatial proximity constraint only at the initial time instant which might not be effective as time progresses, hence their algorithm groups all the agents moving in the same direction giving less importance to their spatial relationships. This can be observed from the output figures in column (b) of Figure. 4. Further in 4^{th}row, a person with red hat is moving faster than the group behind him but CT and CF fail to capture this difference in velocity while the proposed algorithm could capture it. The groups in last row have small changes in their directions of movement which is again not captured by these two methods while proposed method detects such small changes.

Our algorithm outperforms these state-of-the-art methods because it is more robust to tracking errors since we extract groups from the eigenvectors rather than directly using the tracklets. It is quite evident from the Figure. 4 where the tracklets for various agents are marked with different colors to indicate the group they belong to, that the proposed algorithm is able to detect agents in a group much better than the other existing methods. Also the proposed algorithm yields $NMI = 0.92$, $Purity = 0.94$ and $RI = 0.93$ on video clips from BEHAVE dataset whereas the corresponding measures for [14] and [21] have very low values (*e.g.* Purity for CF is 0.35). It shows that these methods do not perform well in videos of sparse crowd whereas the proposed method can also handle a sparse crowd effectively.

6. CONCLUSIONS

In this work, we presented a framework for analysis of medium dense crowd videos. We proposed a first order dynamical system to model agent trajectories collectively and subsequently demonstrated the effectiveness of this interaction model for group detection. We show that eigenvector based clustering for group detection is effective. As a next venture, we are interested in exploring the proposed model to analyze crowd and group behaviors. Also our algorithm assumes the availability of tracks which itself is a challenge in many crowded videos due to occlusion and other tracking problems. We also aspire to define a unified framework where the proposed model and a tracker work together to improve each other's performance in crowded videos by incorporating group interaction cues.

7. REFERENCES

[1] C. C. Aggarwal. A human-computer interactive method for projected clustering. *IEEE Transactions on Knowledge and Data Engineering*, 16(4):448–460, 2004.

[2] Y. Benabbas, N. Ihaddadene, and C. Djeraba. Motion pattern extraction and event detection for automatic visual surveillance. *Journal on Image and Video Processing*, 2011:7, 2011.

[3] S. Blunsden and R. Fisher. The behave video dataset: ground truthed video for multi-person behavior classification. *Annals of the BMVA*, 4(1-12):4, 2010.

[4] W. Ge, R. T. Collins, and B. Ruback. Automatically detecting the small group structure of a crowd. In *Workshop on Applications of Computer Vision (WACV), 2009*, pages 1–8. IEEE, 2009.

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

[6] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[7] J. S. J. Junior, S. Musse, and C. Jung. Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine*, 5(27):66–77, 2010.

[8] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan. Crowded scene analysis: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3):367–386, 2015.

[9] R. Mehran, B. E. Moore, and M. Shah. A streakline representation of flow in crowded scenes. In *Computer Vision–ECCV 2010*, pages 439–452. Springer, 2010.

[10] A. Y. Ng, M. I. Jordan, Y. Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[11] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

[12] M. Schmidt, G. Fung, and R. Rosales. Optimization methods for l1-regularization. *University of British Columbia, Technical Report TR-2009*, 19, 2009.

[13] R. J. Sethi and A. K. Roy-Chowdhury. Individuals, groups, and crowds: Modelling complex, multi-object behaviour in phase space. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011*, pages 1502–1509. IEEE, 2011.

[14] J. Shao, C. C. Loy, and X. Wang. Scene-independent group profiling in crowd. In *CVPR, 2014*, pages 2227–2234. IEEE, 2014.

[15] J. Šochman and D. C. Hogg. Who knows who-inverting the social force model for finding groups. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011*, pages 830–837. IEEE, 2011.

[16] B. Solmaz, B. E. Moore, and M. Shah. Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):2064–2070, 2012.

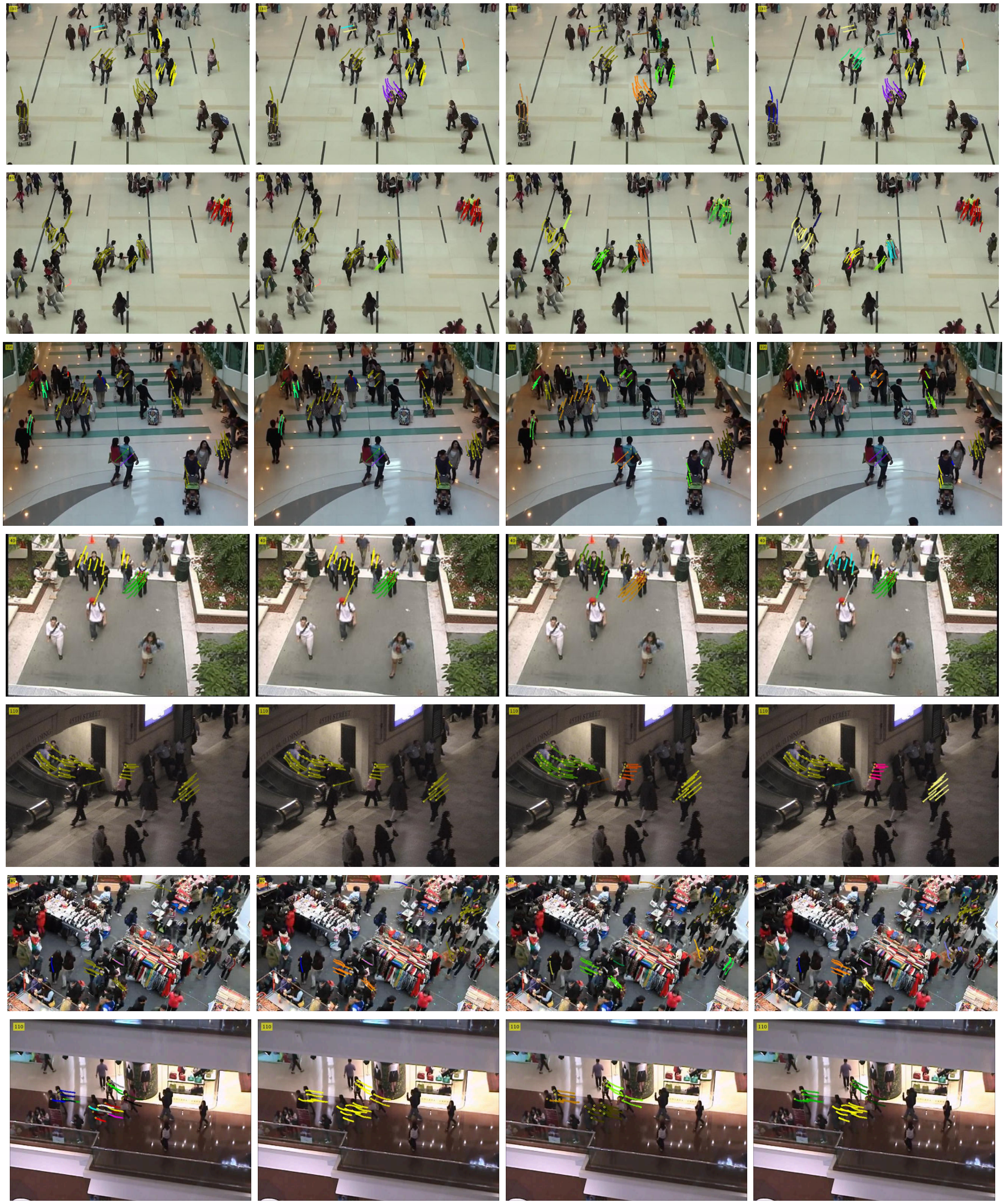
[17] V. Srikrishnan and S. Chaudhuri. Crowd motion analysis using linear cyclic pursuit. In *International Conference on Pattern Recognition (ICPR), 2010*, pages 3340–3343. IEEE, 2010.

[18] M. Thida, Y. L. Yong, P. Climent-Pérez, H.-l. Eng, and P. Remagnino. A literature review on video analytics of crowded scenes. In *Intelligent Multimedia Surveillance*, pages 17–36. Springer, 2013.

[19] M. Wu and B. Schölkopf. A local learning approach for clustering. In *Advances in neural information processing systems*, pages 1529–1536, 2006.

[20] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu. Crowd analysis: a survey. *Machine Vision and Applications*, 19(5-6):345–357, 2008.

[21] B. Zhou, X. Tang, and X. Wang. Coherent filtering: detecting coherent motions from crowd clutters. In *Computer Vision–ECCV 2012*, pages 857–871. Springer, 2012.



(a) CF (b) CT (c) P (d) GT

Figure 4: Comparison of group detection results from Coherent Filtering [21] in column (a), Collective Transition [14] in column (b), our proposed method in column (c) with the ground truth in column (d) for different types of scenes. Each group is represented by a different color. Best viewed in color and when zoomed.