ANCHOR-BASED GROUP DETECTION IN CROWD SCENES

Mulin Chen[1] *Qi Wang*[1*] *Xuelong Li*[2]

[1]School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL),
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
[2]Center for OPTical IMagery Analysis and Learning (OPTIMAL),
Xi'an Institute of Optics and Precision Mechanics,
Chinese Academy of Sciences, Xi'an, 710119, Shaanxi, P. R. China

ABSTRACT

Group detection aims to classify pedestrians into categories according to their motion dynamics. It's fundamental for analyzing crowd behaviors and involves a wide range of applications. In this paper, we propose a Anchor-based Manifold Ranking (AMR) method to detect groups in crowd scenes. Our main contributions are threefold: (1) the topological relationship of individuals are effectively investigated with a manifold ranking method; (2) global consistency in crowds are accurately recognized by a coherent merging strategy; (3) the number of groups is decided automatically based on the similarity graph of individuals. Experimental results show that the proposed framework is competitive against the state-of-the-art methods.

Index Terms— Crowd Motion, Group Detection, Manifold Structure, Clustering

1. INTRODUCTION

In recent years, the analysis of crowd behavior has been a active research area in the realm of computer vision. As the primary component that make up a crowd, coherent groups have received plenty of attentions and involve many practical applications, such as crowd tracking [1–4], anomaly detection [5–7] and semantic scene segmentation [8]. However, the detection of groups remains to be a challenging issue due to the complexity of crowd behaviors.

A major difficulty in group detection comes from the complicate structures of crowd motions. As shown in Figure 1 (a), people in crowd scenes tend to form manifold structures, where individuals only keep consistency with their surroundings. In these cases, pedestrians in one group tend to show big behavioral differences, which represents a serious impediment in the detection of groups. Another barrier is the recognition of global consistency. As Figure 1 (b) visualizes, individuals in crowd scene may exhibit global behavior consistency, which is neglected by traditional local clustering methods. Thus, it's necessary to take the global coherency into account when detecting groups.

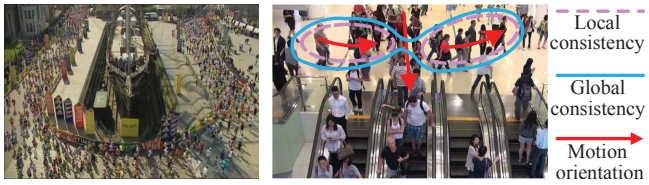


Fig. 1. (a) Crowd motion with manifold structure. (b) Illustration of local and global consistency in crowds.

Many efforts have been made on group detection. Ali and Shah [9] and Lin et al. [8] segmented coherent motions by transferring the flow field, which is time-consuming and limited to handle various crowd motions. Due to the serious occlusion in crowd scenes, many approaches [10–17] treat feature point as study object. Zhou et al. [10] and Shao et al. [11] detected groups by identifying the invariant neighbors of each feature point. Zhou et al. [12] introduced a collectiveness descriptor to quantifying coherent motions. Wu et al. [13] detected coherent motions by collective merging. The above techniques are either limited to exploit the topological relationship of individuals or unable to discover global consistency.

In this paper, we propose an Anchor-based Manifold Ranking method (AMR) to detect coherent groups. Firstly, the anchor individuals are identified to reveal the distinct motion patterns in crowd scenes. Then, a manifold ranking method is employed to cluster individuals into local coherent motions based on their topological relationship with the anchors. Finally, a coherent merging strategy is developed to combine the local motions and recognize global consistency. We compare the proposed method with four state-of-the-art methods and provide quantitative experimental results.

*Corresponding author. This work is supported by the National Natural Science Foundation of China under Grant 61379094 and Natural Science Foundation Research Project of Shaanxi Province under Grant 2015JM6264.

2. DECISION OF ANCHORS

In crowd scenes, we assume that there exist an anchor individual in each group, which connects to others and has the ability to represent the shared motion pattern.

First, the individuals need to be extracted. Since the detection and tracking in crowd scene are still not solved, we use feature points to represent pedestrians, which can be effectively obtained by a generalized KLT (gKLT) tracker [12]. Then, we begin to detect groups on each frame separately. A similarity graph is built to perceive the correlations of points on the current frame. Denoting the spatial position and motion orientation of a point i as $\overrightarrow{p_i} = (p_i^x, p_i^y)$ and $\overrightarrow{ori_i} = (ori_i^x, ori_i^y)$ respectively, the similarity of point i and j can be defined as

$$G_{ij} = \begin{cases} max(\frac{\overrightarrow{ori_i} \cdot \overrightarrow{ori_j}}{|\overrightarrow{ori_i}| \times |\overrightarrow{ori_j}|}, 0), & \text{if } d(i,j) < r \\ 0, & \text{else} \end{cases}, \quad (1)$$

where $d(i,j) = \sqrt{(p_i^x - p_j^x)^2 + (p_i^y - p_j^y)^2}$ is the spatial distance between i and j. Supposing there are N points, the distance threshold r is empirically set as the N-th smallest element in all pairs of the distance d. Thus, the similarity will be high if two points reside close to each other and share similar motion orientation.

Given the graph G, the desired group number c can be roughly estimated by calculating the number of strongly connected components in G, which can be solved by the depth-first search method [18]. Then we aim to find c anchor individuals. As mentioned before, an anchor should interact closely with other individuals. So we define the interaction intensity of point i as

$$\rho_i = \sum_j G_{ij}. \quad (2)$$

Then a larger ρ indicates a closer connection with surroundings. As Rodriguez and Laio [19] pointed out, an anchor point, which can be considered as the cluster center of a group, should have higher interaction intensity than its neighbors, and keep far away from those of higher intensity (cluster center of other groups). Thus, similar to [19], a quantity δ_i is introduced to measure the minimum distance between i and points with higher intensities,

$$\delta_i = \min_{j:\rho_j > \rho_i} d(i,j). \quad (3)$$

If point i has the highest intensity, we simply set its δ_i as $\max_j d(i,j)$. Then δ will be large for points with local or global maxima interaction intensity. So the anchor individuals can be obtained by finding the points with the top c largest δ, as shown in Figure 2 (b).

(a) (b)

Fig. 2. (a) Original video frame. (b) Tracked feature points (yellow color) and detected anchors (red color). Arrows indicate motion orientations.

3. CLUSTERING BY MANIFOLD RANKING

Regarding the anchors as labelled data, the group detection procedure can be viewed as a multi-class semi-supervised classification task. According to the relationship with anchors, points can be classified into different clusters.

For the purpose of precisely detecting groups, it's necessary to exploring the structures of crowds. To this end, a manifold ranking method [20, 21] is utilized. Given c anchors, the objective is to classify the points into c categories. First, let $Y \in \mathbb{R}^{N \times c}$ denotes the initial label of points, where $Y_{ij} = 1$ if point i is the j-th anchor and $Y_{ij} = 0$ otherwise. And define $F \in \mathbb{R}^{N \times c}$ as a relationship matrix, where F_{ij} indicates point i's topological relevance to anchor j. Thus, given the similarity graph G, F can be obtained by solving the following problem

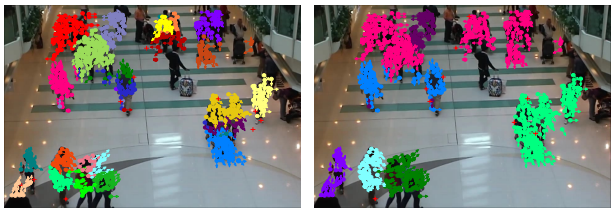
$$\arg\min_F \frac{1}{2}(\sum_{i,j=1}^{N} G_{ij} || \frac{F_i}{\sqrt{D_{ii}}} - \frac{F_j}{\sqrt{D_{jj}}} ||^2 + \alpha \sum_{i=1}^{N} ||F_i - Y_i||^2), \quad (4)$$

where D is the degree matrix of G, F_i and Y_i are the i-th row of F and Y respectively. In the above problem, the smoothness constraint (first term) ensures that the clustering result don't change too much between neighbors, and the fitting constraint (second term) prevents all the elements of F to be equal. The parameter α (α is set to be 0.1) balances the two terms. Setting the derivative of the above equation to be 0, the optimal solution can be computed as

$$F^* = (D - \frac{1}{1+\alpha} G)^{-1} Y, \quad (5)$$

where D is the degree matrix of G. For a point i, a larger F_{ij} indicates a higher probability that i belongs to the j-th cluster. Furthermore, for outliers that don't belonging to any group, they keep low relevance to the anchors. Then point i is denoted as an outlier if $\max F_i \leq \epsilon$ (ϵ is set as 0.2), and its cluster index is 0. If point i is not an outlier, its cluster label $cluster_i$ is defined as

$$cluster_i = \arg\max_{j \leq c} F_{ij}^*. \quad (6)$$



(a) (b)

Fig. 3. (a) Detected local clusters. (b) Final groups. Scatters with different colors indicate different clusters/groups, and arrows indicate motion orientations. The plus sign indicates outliers. It can be seen that the coherent merging method successfully combines the local clusters into groups.

After assigning a label to each point, the local clusters are obtained. However, due to the global consistency in crowd motions, the local clusters can't represent the actual groups, as shown in Figure 3 (a). So a coherent merging refinement is followed to combine the local motions into final groups.

4. COHERENT MERGING

To detect global consistency, a coherent merging strategy is designed to combine the obtained local clusters.

Since anchors have the ability to reveal the motion dynamic of groups, the coherency of groups can be measured according to the anchors' motion orientations. Supposing point i and j are the anchors of cluster C_i and C_j respectively, the coherency of C_i and C_j is defined as

$$Coh(C_i, C_j) = \frac{\overrightarrow{ori_i} \cdot \overrightarrow{ori_j}}{|\overrightarrow{ori_i}| \times |\overrightarrow{ori_j}|}. \quad (7)$$

In addition, the spatial distance should also be taken into consideration. If two clusters are close to each other, they are like to belong to the same global group. To measure the spatial distance of two clusters, an intuitive way is to calculate their center positions' distance. However, if the clusters contain a great many of points, their centers may be far away even when the clusters are adjacent. So we alternatively calculate two clusters' distance according to their nearest points

$$Dist(C_i, C_j) = \min_{i \in C_i, j \in C_j} \sqrt{(p_i^x - p_j^x)^2 + (p_i^y - p_j^y)^2}. \quad (8)$$

Afterwards, C_i and C_j are considered to be continuous if $Coh(C_i, C_j) > \theta$ and $Dist(C_i, C_j) < r$, where threshold θ is set to be 0.6 and r has been introduced in Section 2. Thus, global consistency can be recognized by merging the continuous clusters iteratively. In each iteration, only the clusters with the highest motion coherency are combined, so the final result will not affected by the merging order. As shown in Figure 3 (b), after coherent merging, local clusters are successfully combined into global coherent groups. The whole procedure of the proposed method is shown in Algorithm 1.

Algorithm 1 The proposed framework

Input: Input frame with N tracked feature points, parameters α, thresholds ϵ and θ.
Output: Detected groups.

Stage: Decision of anchors
Calculate points' spatial distance d and the threshold r.
Build similarity graph G.
Predict desired group number c.
Compute ρ and δ for all points with Equation 2 and 3.
Identify anchors by finding the maximum δ.
Stage: Clustering by manifold ranking
Set the initial label matrix Y.
Calculate relevance matrix F with Equation 5.
Threshold F with ϵ to find outliers.
Obtain local clusters with Equation 6.
Stage: Coherent merging
repeat
  Compute the coherency of clusters with Equation 7.
  Calculate the distance between clusters with Equation 8.
  Threshold coherency and distance with θ and r to find continuous clusters.
  Combine the continuous clusters with the highest coherency.
until no continuous clusters

5. EXPERIMENTS

In this section, extensive experiments are conducted to validate the effectiveness of the proposed method. CUHK Crowd Dataset [11] is employed to evaluate the performance, and four state-of-the-art group detection techniques are used for comparison.

Dataset: CUHK Crowd Dataset contains 474 crowd videos with various densities and shapes. It provides the locations and velocities of tracked feature points, and annotates the group label of each point.

Competitors: To quantitatively evaluate the performance, four state-of-the-art methods are used as competitors, including Coherent Filtering (CF) [10], Collective Transition (CT) [11], Measuring Crowd Collectiveness (MCC) [12] and Collective Density Clustering (CDC) [13]. We let all the competitors use their respective optimal parameters

Performance: We detect groups on each video clip and employ two widely used metrics, the accuracy (ACC) [22] and F-score [23] as measurements. Higher values of ACC and F-score indicate better results. The quantitative comparison of different methods is shown in Table 1. It's manifest that the

Fig. 4. Representative comparison results of group detection. Scatters with different colors indicate different detected groups, and the plus sign indicates outliers. The result of AMR is closer to the ground truth than the competitors.

	CF	CT	CDC	MCC	AMR
ACC	0.70	0.75	0.67	0.69	**0.78**
F-score	0.67	0.74	0.68	0.67	**0.76**

Table 1. Quantitative comparison on group detection. Best results are in bold faces

	CF	CT	CDC	MCC	AMR
AD	2.45	1.63	1.59	2.02	**1.47**
VAR	3.01	1.83	1.84	2.56	**1.56**

Table 2. Quantitative comparison on group number estimation. Best results are in bold face

proposed AMR obtains the highest ACC and F-score, which means that AMR is more effective than other methods. CF detects groups by combine the invariant points into a group. CT introduces a transition prior and refines the results of CF. Both of the two methods need long term information and can't deal with the immediate motion change of individuals. MCC measures the collectiveness of points and utilizes it to detect groups, which neglects the global consistency in crowd motions. CDC detects coherent motions with a two-stage merging procedure. However, it's limited to handle crowds with manifold structures. The proposed AMR investigates the topological relationship between points and reasonably combines the local motions into global coherent groups. So it doesn't share the above deficiencies and achieves promising performance. Some representative examples are visualized in Figure 4, and it can be seen that the results of AMR are consistent with the ground truth.

The proposed method decides group number automatically, so we further verify whether the estimation is accurate. The widely used Average Difference (AD) and Variance (VAR) [13] are taken as measurements. The lower AD corresponds to the less deviation from real group number, and the lower VAR indicates a higher stability of group detection. Table 2 denotes the performance of each method. The AD and VAR of the proposed method are the lowest. CDC also obtains relatively good results, which is benefit from its global clustering procedure. The performance of CF is unsatisfactory because it can't distinguish groups with subtle difference. The proposed AMR utilizes a coherent merging processing, so it performs well. Besides, the accuracy of group number estimation also reflects the ability to detect global consistency. Thus, we can conclude that AMR is capable of recognizing the global consistency in crowds.

6. CONCLUSIONS

In this paper, a new group detection method has been put forward. In this procedure, the interaction intensities of points are first utilized to find the anchors, which reveal the motion dynamics of groups. Points' topological relevance to the anchors is deeply exploited with a manifold ranking method, based on which the points are classified into local clusters. A coherent merging strategy is developed to combine the continuous local clusters and recognize global consistency. Experimental results and comparison to the state-of-the-art methods validate the proposed method's capability to detect groups and estimate group number accurately. In the future, we will investigate how to apply our method into some specific applications, such as event recognition and crowd counting.

References

[1] Y. Wang, X. Luo, and S. Hu, "Visual tracking via multi-task non-negative matrix factorization," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 1516–1520.

[2] Q. Liu, T. Campos, W. Wang, and A. Hilton, "Identity association using PHD filters in multiple head tracking with depth sensors," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 1506–1510.

[3] F. Zhu, X Wang, and N. Yu, "Crowd tracking with dynamic evolution of group structures," in *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI*, 2014, pp. 139–154.

[4] Q. Wang, J. Fang, and Y. Yuan, "Multi-cue based tracking," *Neurocomputing*, vol. 131, pp. 227–236, 2014.

[5] A. Li, Z. Miao, Y. Cen, and Q. Liang, "Abnormal event detection based on sparse reconstruction in crowded scenes," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 1786–1790.

[6] D. Ma, Q. Wang, and Y. Yuan, "Anomaly detection in crowd scene via online learning," in *International Conference on Internet Multimedia Computing and Service*, 2014, p. 158.

[7] Y. Yuan, J. Fang, and Q. Wang, "Online anomaly detection in crowd scenes via structure analysis," *IEEE Trans. Cybernetics*, vol. 45, no. 3, pp. 562–575, 2015.

[8] W. Lin, Y. M, W. Wang, J. Wu, J. Wang, and T. Mei, "A diffusion and clustering-based approach for finding coherent motions and understanding crowd scenes," *IEEE Trans. Image Processing*, vol. 25, no. 4, pp. 1674–1687, 2016.

[9] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.

[10] B. Zhou, X. Tang, and X. Wang, "Coherent filtering: Detecting coherent motions from crowd clutters," in *European Conference on Computer Vision*, 2012, pp. 857–871.

[11] J. Shao, C. Loy, and X. Wang, "Scene-independent group profiling in crowd," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2227–2234.

[12] B. Zhou, X. Tang, H. Zhang, and X. Wang, "Measuring crowd collectiveness," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1586–1599, 2014.

[13] Y. Wu, Y. Ye, and C. Zhao, "Coherent motion detection with collective density clustering," in *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, 2015, pp. 361–370.

[14] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, 2000.

[15] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2871–2878.

[16] B. Zhou, X. Tang, and X. Wang, "Learning collective crowd behaviors with dynamic pedestrian-agents," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 50–68, 2015.

[17] X. Li, M. Chen, and Q. Wang, "Measuring collectiveness via refined topological similarity," *TOMCCAP*, vol. 12, no. 2, pp. 34, 2016.

[18] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM J. Comput.*, vol. 1, no. 2, pp. 146–160, 1972.

[19] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[20] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Advances in Neural Information Processing Systems*, 2003, pp. 169–176.

[21] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 6, pp. 1279–1289, 2016.

[22] F. Nie, X. Wang, M. Jordan, and H. Huang, "The constrained laplacian rank algorithm for graph-based clustering," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 1969–1976.

[23] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 40, no. 6, pp. 1438–1446, 2010.