Neurocomputing 000 (2017) 1–11

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Instant coherent group motion filtering by group motion representations☆

Na Li[a], Yun Zhang[a,*], Wenhan Luo[b,c], Ning Guo[a]

[a] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, PR China
[b] Department of Electrical and Electronic Engineering, Imperial College London, UK
[c] Tencent AI Lab, Shenzhen, PR China

ARTICLE INFO

Article history:
Received 2 April 2016
Revised 3 March 2017
Accepted 19 May 2017
Available online xxx

Communicated by Marco Cristani

Keywords:
Group refinement
Motion representation
Motion consistency
Motion orientation

ABSTRACT

Group motion direction, which is composed of instant consistent motions, is an interesting property in group dynamics. Understanding the instant group motion direction is a key component of crowd analysis and has a wide range of applications, especially for improving group detection in crowd. In this study, independent from group detection methods, we propose an instant group motion refining framework based on group motion direction. We show that instant group motion direction can be systematically quantified by intra-group motion consistency. This is achieved by proposing two instant group motion representations in terms of velocity consistency and angle consistency. These group motion representations provide a new way to derive the instant group motion direction, which plays an important role in improving the performance of state-of-the-art group detection methods. Besides, to improve traditional group detection methods, a novel group detection method is proposed as well. Extensive experiments on a variety of public scene video clips demonstrate that both group motion consistency representations are not only useful but also necessary for instant group coherent motion filtering.

1. Introduction

Group motion analysis has been extensively studied because of its wide applications in crowd behavior understanding and group detection. Although the definition of group varies in different applications, instant group consistent motions are the primary entities that compose a time-series group evolution. Suppose a set of members with collective behaviors and a common goal are considered as a group [1], the actual motion direction of group at each time is the elementary motion features to understand group behaviors [2] in the crowded scene. Thus, instant group consistent motion direction is an interesting property of group motion, which is academically important and practically useful in a wide range of applications, including pedestrian counting, group behavior analysis and crowd behavior analysis. However, the group detection results [1,3,4] for group motion analysis are usually collected from time-series data, which have to overcome tracking errors to obtain stable groups over time.

Recent studies in socio-psychological and biological suggest that individuals tend to continuously adjust their locations to facilitate verbal exchange [5]. People are more easily influenced by others nearby, especially members in their vicinity with a common destination. However, when pedestrians in the crowd form a group, they exhibit some interesting properties in dynamics. In crossroad, pedestrians walk across the pavement tend to show higher consistency in motion direction. Whilst, in railway station, pedestrians tend to exhibit less consistency in velocity magnitude and motion orientation. Thus, different intra-group motion properties fit to different inter-group motion distinguish requirements. To the best of our knowledge, there has not been a study of using intra-group motion properties to refine group clustering results further.

Thus, one direct application of the instant intra-group consistent motion direction is to distinguish coherent motion among groups, so as to refine group detection results. In the computer vision field, group detection improvement haven't been fully studied.

☆ This work was supported in part by the National High Technology Research and Development Program of China under Grant 2014AA01A302, in part by the PhD Start-up Fund of Natural Science Foundation of Guangdong Province under Grant 2015A030310262, in part by the National Natural Science Foundation of China under Grant 61471348, in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2016A030306022, in part by the Project for Shenzhen Science and Technology Development under Grant JSGG20160229202345378 and in part by the Special Funding for Science and Technology Development in Guangdong Province with Grant 2016A050503035

* Corresponding author.
E-mail addresses: na.li1@siat.ac.cn (N. Li), yun.zhang@siat.ac.cn, zhangyun_8851@163.com (Y. Zhang), w.luo12@imperial.ac.uk (W. Luo), ning.guo@siat.ac.cn (N. Guo).

http://dx.doi.org/10.1016/j.neucom.2017.05.045
0925-2312/

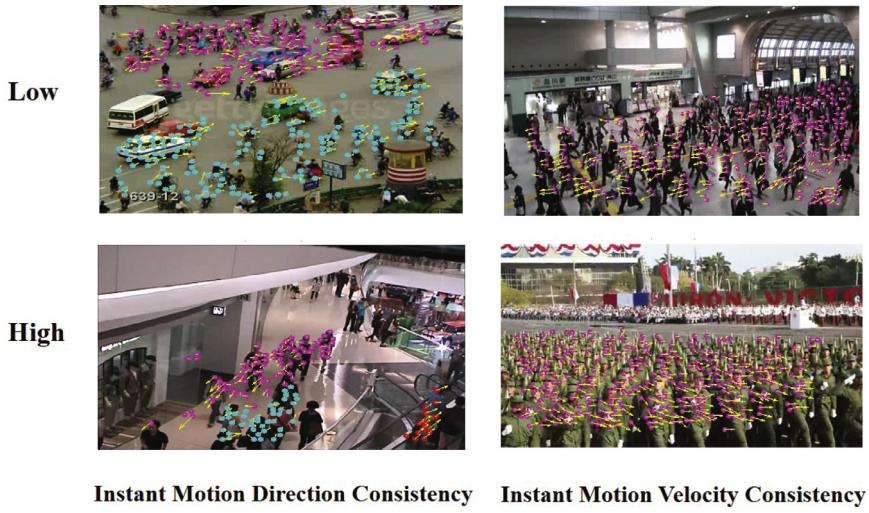

Fig. 1. Instant inter-group motion can be better distinguished by instant individual motion direction (displayed by arrow) consistency and velocity (displayed by both arrow orientation and length) consistency. Across crowded scenes, the consistency level in instant motion direction and instant motion velocity varies from groups to groups. Groups are distinguished with colors.

From the perspective of quantitation, groups can be distinguished by group motion properties, such as instant motion direction and velocity. As shown in Fig. 1, after using Kanade Lucas Tomasi (KLT) tracker in different videos, tracklet clusters of different groups in the same video are marked with different colors. The velocity magnitude and direction are indicated by the arrow orientation and length. For explanation convenience, we make an analogy between a point in image and a member in group in the rest of the paper.

The above observations motivate us to improve group detection not only during the tracklet clustering stage, but also after the tracklet clustering stage. We aim to optimize group detection by proposing a framework to refine group detection results of any given detection methods. We refine the initial tracklet clustering result by studying the instant group motion representation based on characterizing and quantifying group motion consistency, so as to avoid group detection failure caused by either tracking errors or limitations of group detection methods. This study is the first attempt to systematically investigate the instant group motion representations and apply these representations to improve group detection performance. Primarily, we make the following contributions:

(1) *A group motion refining framework.* We introduce instant group motion representations to derive the consistent motion direction as group motion direction for initial groups clustered by the state-of-the-art group detection methods. Then, the group motion direction is adopted to improve intra-group consistency. Extensive experiments show that the proposed framework can improve group detection performance of various state-of-the-art time-series group detection methods.

(2) *A spatial distance robust group detection method.* We propose a group detection method by substituting the velocity correlation with the collectiveness descriptor proposed in [6]. The proposed method outperforms the state-of-the-art methods [1,7] in experimental evaluation. It also enriches the initial tracklet clustering varieties for the evaluation of the proposed instant group motion refining framework.

(3) *Instant group motion representations.* We build up two types of instant group motion representations based on group motion consistencies. They are independent with the initial trackelet clusterings and can avoid tracking errors by identifying group motion consistencies at sub-group level, which are hierarchically overlapping. Two types of group motion consistencies are derived from motion velocity and motion angle of group members respectively. The effectiveness of the proposed instant group motion representations is verified through recognizing intra-group motion orientation for refining initial group detection results.

2. Related works

Crowd analysis granularity can be categorized into three levels: individuals [8–10], groups [1,11,12] and the whole crowd [4,13,14]. Zitouni et al. [15] present a systematic survey and evaluation of crowd modeling techniques. In the computer vision community, researchers have focused mainly on autonomous crowd analysis and significant progress has been achieved. In the field of detecting groups in images, considering the background of the social sciences, Setti et al. [16] present an automatic group detection method for groups specified with rigorous definitions. Choi et al. [17] investigate the intermediate representations based on the socio-psychological concept for modeling groups in images. However, motion features [18] can be more important than static visual features in areas related to the analysis of behaviors and activities in crowd videos. In [18], Li et al. category existing crowd motion features into flow-based features, local spatio-temporal features and trajectory/tracklet. Spatio-temporal gradients [19,20] and histogram functions [21,22] are generally used as basic representations for motion modeling. Vascon et al. [22] propose a game-theoretic framework to model the uncertainty associated with the position and orientation of the engaged persons in sequences. A descriptor of collectiveness is proposed in [7] to measure how individuals act as a union.

Recent works emerging in crowd behavior analysis begin to focus on finer-level group analysis. Therefore, the crowd level motion feature cannot be directly utilized for group motion, which requires finer group segregation. In the following, we briefly discuss some works on finer-level group detection and motion representation.

2.1. Group detection

In the field of finer-level group segregation, state-of-the-art methods [1,6,7,13,23] treat a group either as a collection of individuals or an integrated whole. Zhou et al. [7] segment the coherent motion by developing two coherent neighbor invariances as group coherent motion priors, whilst the group collective transition prior is learned through Markov chains in [1]. In [1],

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

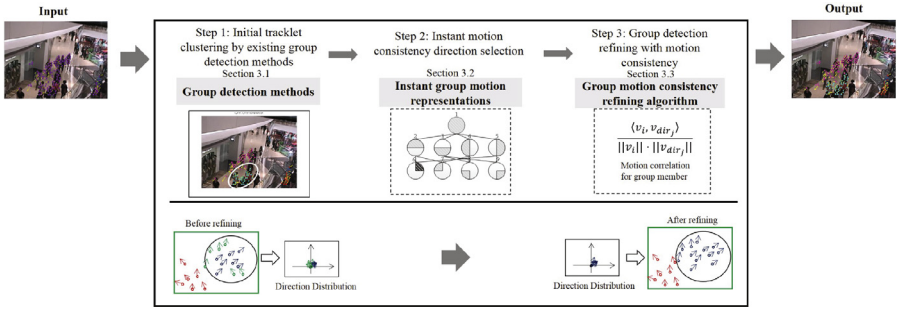

Fig. 2. Framework for group motion consistency refining.

the group collective transition prior is discovered from the initial group clustering results of [7]. Group motion pattern modeled with the spatio-temporal driving force is proposed to address the event analysis and activity recognition [24]. Driven by the typical group structures generated from social interactions among group members, Ge et al. [3] discover small groups hierarchically based on velocity and distance of pair-wise individuals. Hausdorff distance defined with respect to pair-wise proximity and velocity is used to discover groups [3]. Based on the social force model, Mazzon et al. [25] propose a method for detecting and tracking interacting groups of people in crowds.

Most of the above methods segment groups in the crowd by considering intra-group structure, such as velocity and distance coherency of pair-wise individuals. The proposed method utilizes the inter-group motion specificities, such as motion consistency of groups, to refine group detection results. The motivation is that group members tend to move coherently along the group motion direction, which can be characterized by group motion consistency.

2.2. Group motion representation

The motion property is different from groups to groups in the crowd, which depends on group motion patterns. The concept of group motion property is introduced by Shao et al. [1,23] for analyzing and understanding the crowd behavior at the group-level. Both the intra- and inter-group properties proposed in [1,23] are validated for crowd scene identification, which are specifically designed for distinguishing crowd behaviors.

Different from [1,23], this paper studies the group motion property for group motion filtering. Researchers in fields of social-psychological [5] and computer vision [13] begin to study individual motion directions as elemental group dynamics, in terms of pedestrian group properties. As people are more easily influenced by others in their vicinity, spatial positions of neighboring tracklets are used for group destination prediction in social affinity feature learning [13]. Cheng et al. [26] represent motion trajectories from a probabilistic perspective to handle the variability of movements within the group. Zhao et al. [27] establish an effort on building the crowd behavior entropy model using the individual velocity. They provide a method to adopt Shannon entropy to express the crowd macro state through calculating the probability of each microstate. Velocity direction is employed to obtain the particle entropy in [28].

Nevertheless, the above-mentioned methods are short of studying the group motion representation for group motion refining. To refine initial group detection results, we first build up the instant intra-group motion consistency to distinguish members who do not follow the group motion. As group members exhibit relatively high affinity and consistency along instant group motion orientations, it is of advantage in applying instant group motion to characterize motion consistency over tracking errors.

3. Group motion refining framework

Considering a group as a set of members with a common goal, instant motion of group members should be of higher consistency around a certain direction. In this section, we propose a group motion refining framework based on instant group motion direction. This refining framework is developed to refine initial groups detected by existing group detection methods.

As shown in Fig. 2, the input of the framework is a sequence of images. Then initial group detection results are generated by existing group detection methods or a new method (Section 3.1) proposed by us. To describe the instant group motion, we develop the instant group motion representation from two types of instant group motion consistency (Section 3.2). The instant group motion representation facilitate the selection of the instant group motion direction for further group motion refining (Section 3.3). The instant group motion direction is determined by electing the angle window with relatively higher consistency over other angle windows. Then, the group motion direction is calculated as the average motion orientation of group members moving in one concerned angle window. At last, an instant group motion consistency filtering method is developed to search for group members with highly consistent motion, so as to filter out inconsistent tracklets and regain missing tracklets for initial tracklet clustering refinement. Note that, the framework of the group motion refining is independent from specific group detection methods, such as Collective Transition (CT) [1] and Coherent Filtering (CF) [7].

3.1. Group detection: coherent collectiveness filtering (CCF)

In general, our group motion refining framework is independent from specific group detection methods. The flexibility of this framework enables the improvement for a series of group detection methods. Thus, in addition to existing group detection methods such as CT [1] and CF [7], we develop a new method in Algorithm 1 as a complement. This so-called CCF group detection method is based on CF [7] and the collectiveness descriptor [6].

The state-of-the-art group detection methods CT and CF, which highlight the group members, tend to coordinate their behaviors in their neighborhood. Whereas the collectiveness proposed in [6] is designed to cover spatially coherent crowd structure. In the same crowd system, collectiveness [6] is more robust and general to spatial distance than velocity correlation in the matter of describing the similarity between group member motions. Thus, by substituting velocity correlation with collectiveness in CF, we develop the group detection method called CCF, as shown in Algorithm 1. We first obtain the invariant neighbor set $N^i_{t\to t+d}$ by examining the K nearest neighbor set of each tracklet z_i from time t to $t+d$, which is similar to [6]. However, it does not guarantee that all the invariant neighbors acting in collective motions are collected especially when d is small and K is large. Hence, according to the collectiveness descriptor [6], we set a threshold λ on

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

Algorithm 1 Coherent collectiveness filtering.

Input: Tracklets included in the current frame t of a video clip, $\mathbb{Z}$, $R = \phi$;
Output: Group clusters $\{G^{\circ}\}_{j=1}^{M}$;
  for each tracklet $z_i \in \mathbb{Z}$
    search the K nearest neighbor set as $N_{t\to t+d}^{i}$.
    for each $n_i^k \in N_{t\to t+d}$ [7]
      compute the averaged collectiveness [6], $c_{t\to t+d}^{n_i^k}$
      put (i, n_i^k) in R if $c_{t\to t+d}^{n_i^k} > \lambda$, λ is a threshold
      build a graph from R, identify coherent motion $\{G^{\circ}\}_{j=1}^{M}$ as the connected components of the graph.

the average collectiveness to filter inconsistent group members and obtain the pairwise connection set R. λ is a threshold set on the value collectiveness, which is set based on the collectiveness bound in [6]. Finally, a connectivity graph is built, where coherent tracklet clusters are identified as the connected components of the graph. It can not only enrich the variety of initial tracklet clustering for group motion refining framework evaluation in Section 4.3, but also outperform the state-of-the-art methods in [1,7].

3.2. Instant group motion representation

In the crowd, group members are more likely to be influenced by other group members nearby, especially in the case of a common destination. Thus, the group motion direction is one imperative group motion property for group motion analysis. In the spatio-temporal space, group is mainly represented as a set of tracklets $\mathbb{Z} = \{z\}$ detected by the KLT feature tracker but not limited by it. There is motion vector [29,30] that inherent correlated with optical flow [31], which is beyond the scope of this paper. Group members tend to exhibit high motion consistency along group motion orientation. However, the long term group motion consistency tends to be sensitive to noises of tracking. As group motion is composed of instant consistent motions toward a specific destination, the instant group motion consistency direction is more robust for group motion detection.

In this section, to identify consistent motion direction of groups, we first propose two types of general motion consistency at the group level based on group member motion properties, including the velocity magnitude and orientation. Afterwards, based on the group motion consistency at different orientations, we develop histograms of group motion consistency as representations for group orientation selection.

3.2.1. Group motion consistency

Based on the instant velocity of group members, we characterize group motion consistency through describing group member motion properties. Similar to [32], which refers to the entropy model and proposes the crowd entropy utilizing the foreground object, we calculate the group motion consistency entropy as

$$F = H(Q) = -\sum_{i=1}^{m} q_i \log_2 q_i, \tag{1}$$

where H denotes the Shannon entropy of a discrete probability distribution, and $Q = (q_1, q_2, ..., q_m)$ is a discrete probability distribution with $\sum_{i=1}^{m} q_i = 1$. The more inconsistent the group member distribution is, the bigger the entropy is. The discrete probability q_i is calculated by

$$q_i = \beta_i / \sum_{i=1}^{m} \beta_i, \tag{2}$$

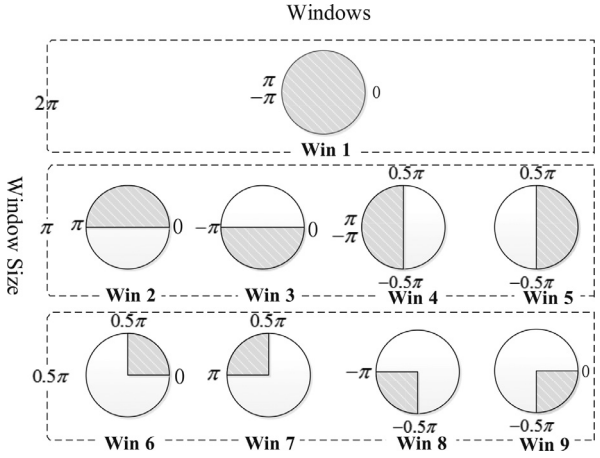

Fig. 3. Overlapping angle windows for generating instant group motion coherency histograms. The first window (top) covers the entire motion direction of group mobility. The windows below cover progressively smaller regions of the motion.

where the frequency of group members with the specific motion property value is defined as β_i, such as the number of group members move toward one specific orientation. By designing β_i with different motion properties, i.e., motion orientation and velocity, we can derive different types of group motion consistency from Eq. (1).

As pedestrian images in the surveillance videos are usually of different scales due to perspective distortion, the individual location affinity have different scales according to pixel-wise tracking. Nevertheless, individual motion orientation is less sensitive to perspective distortion. So we generate two types of group motion consistency in respect of two group motion properties, including the velocity angle θ_i and the magnitude γ in the polar coordinates.

Angle consistency: The group motion consistency is highly related to individual motion orientations in the group. Angle consistency can be employed to avoid particular inherent perspective distortions, when members away from the camera exhibit smaller velocity magnitude. In the angle consistency, $\theta_i (i = 1, 2, ..., m)$ denotes the instant group motion direction, m denotes the number of group members out of the total N group members moving at the orientation windows $W_j (j = 1, 2, ..., 9)$. The set of orientation windows we adopt is shown in Fig. 3. As to calculate Eq. (2), β_i is set with θ_i.

$$\beta_i^A = \theta_i. \tag{3}$$

Velocity consistency: The velocity consistency is computed by introducing a polar coordinate system to denote the distribution of individuals resembling the crowd behavior entropy proposed in [27]. As to calculate Eq. (2), β_i is set based on group member motion velocity by

$$\beta_i^V = \mathrm{N}(\gamma_i, \theta_i), \tag{4}$$

where γ_i and θ_i are the polar angle and polar radius in the log-polar coordinates, they make the motion representation more sensitive to positions of nearby individuals than those of individuals farther away. N is computed as the number of group members whose velocity falls within the bin matrix, where 8 equally spaced θ bins and 5 equally spaced γ bins are used [1]. $\gamma_i \in [\gamma_{min}, \gamma_{max}]$ and $\theta_i \in [-\pi, \pi]$. γ_{min} and γ_{max} are the local minimum and maximum value of γ_i in one group.

3.2.2. Instant group motion representation

Pedestrians with different destinations, e.g., when crossing the zebra-crossing face to face, exhibit relatively different motion orientations. The instant pedestrian group motion orientation can be

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

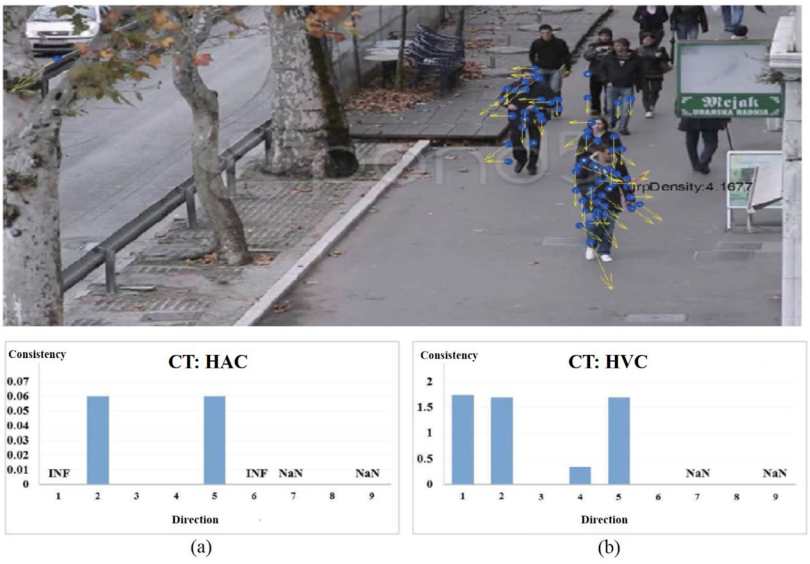

Fig. 4. HAC and HVC of the initial tracklet clustering generated by CT [1]. The more inconsistent the motion is, the bigger the entropy is. According to both HAC and HVC, the sub-group related with Win 3 in Fig. 3 exhibits the highest motion consistency among all orientation windows. The bucket values NaN and INF are scores indicating there is no members move along specific direction or the consistency is lower enough to ignore. This also applies in Figs. 5 and 6.

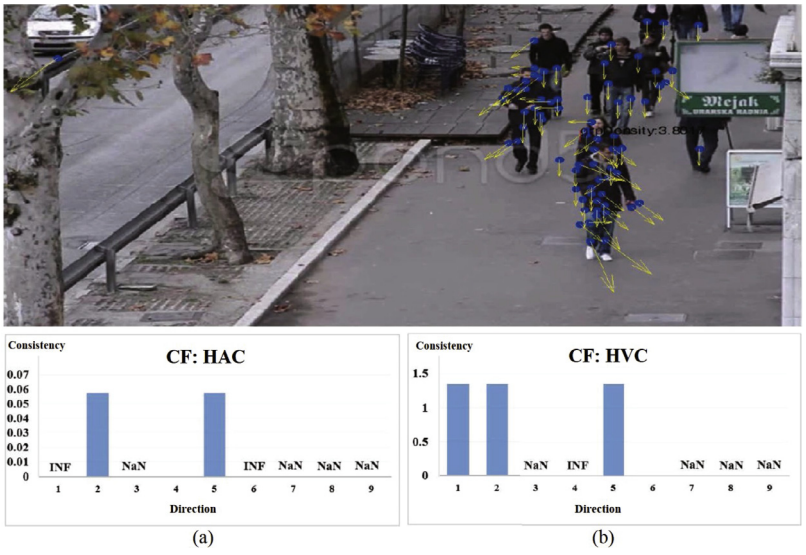

Fig. 5. HAC and HVC of the initial tracklet clustering of method CF [7]. The more inconsistent the motion is, the bigger the entropy is. According to HAC and HVC, the sub-group related with Win 4 and 6 in Fig. 3 separately display the highest motion consistency among all orientation windows.

determined by the actual movement orientation of most pedestrians, who exhibit comparatively high consistency at each time instant. Inter-group motion direction therefore can be distinguished by intra-group pedestrian motion consistency orientation, given the initial group tracklet clustering results. However, the exceptional motion margin or clustering noises prevent group motion orientation from correctly being calculated by averaging all individual motions in the group directly.

We investigate the group motion consistency at different ranges of motion orientation and represent the group motion consistency as histograms. Specifically, we divide group individuals into various sub-groups (or windows) based on motion orientations and then make a quantification for each sub-group motion consistency. The grouping granularity of group motion orientation is highly related to group distributions. The set of motion orientation windows we adopt is shown in Fig. 3. This set of nine overlapping windows organize individuals into groups whose motion orientations are correspondingly located at a range of $[-\pi, \pi]$, $[0, \pi]$, $[-\pi, 0]$, $[-\pi, -\pi/2] \cup [\pi/2, \pi]$, $[-\pi/2, \pi/2]$, $[0, \pi/2]$, $[\pi/2, \pi]$, $[-\pi/2, -\pi]$, $[0, -\pi/2]$. Thus, instead of considering group members in all directions as a whole, we adopt a range of overlapping windows to characterize the group motion coherency hierarchically, resulting in a histogram of 9 bins. The shadowed areas are included in that angle windows, while white areas are ignored. The first window (top) covers the all-around direction range of group motion. The windows below cover progressively smaller regions of the motion.

Instances of the group motion representation are obtained by characterizing motion consistency for sub-groups in the hierarchical range of windows. In this work, as a result of applying the angle consistency and velocity consistency in each motion orientation window, two instances of the group motion representation are noted as the following.

Histogram of angle consistency (HAC): HAC is obtained by collecting the angle consistency of sub groups organized by each orientation window. The proposed representation HAC is employed

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

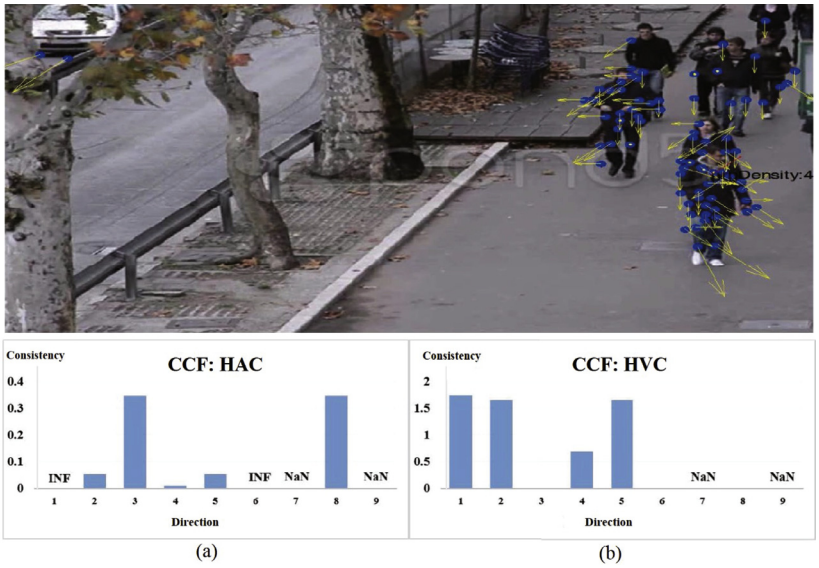

Fig. 6. HAC and HVC of the initial tracklet clustering of method CCF proposed in this work. The more inconsistency the motion is, the bigger the entropy is. According to HAC and HVC, the sub-group related with Win 4 and 3 in Fig. 3 separately display the highest motion consistency among all orientation windows.

to various initial tracklet clustering results generated from various group detection methods, including CT, CF and CCF. The resulting histograms are shown correspondingly in Figs. 4(a), 5(a) and 6(a). NaN and INF indicate there are no members moving along specific direction or the consistency is low enough to be ignored.

Histogram of velocity consistency (HVC): HVC is obtained by collecting the velocity consistency of sub-groups organized by each orientation window. The resulting histograms of applying the proposed representation HVC to various initial tracklet clustering results generated from three different group detection methods are shown in Figs. 4(b), 5(b) and 6(b). In the case that group clusters are generated from noisy tracking results, velocity consistency tends to be robust for the motion analysis task. Thus, HVC outperforms HAC.

3.2.3. Instant group motion consistency direction

We distinguish the motion orientation of groups according to the motion consistency of group members. Note that the total number of group individuals should be greater than a threshold to be effective in representing the orientation of the group motion. In this paper, we set the threshold as N/2, where N is the total number of group members in a specific group. Thus, the sub-group, with relatively higher motion consistency among sub-groups of at least half number of the whole group, is selected for group motion orientation computation. We illustrate the group motion orientation of tracklet clusters through group motion representations. As to the tracklet clustering of CT [1], Win 3 in Fig. 3, which displays the highest velocity consistency and angle consistency, is selected for group motion orientation computation by both HAC and HVC. In terms of CF [7], Win 6 and 4 in Fig. 3 are separately selected by HAC and HVC. Regarding CCF, Win 6 and 3 in Fig. 3 are separately selected by HAC and HVC. These results imply that instant group motion representations perform differently across initial tracklet clusters. Therefore, different instant group motion representations tend to pick up different group motion directions, which is obtained by averaging motion orientations of group members in the selected sub-group.

3.3. Group motion refining

Based on instant group motion representations in Section 3.2, an instant group motion consistency filtering method is proposed to refine the group detection results. The motion consistency refining is complemented to the initial tracklet clusters derived from the time-series images. The key idea is to search for group members with highly consistent motion. The method helps filtering out inconsistent tracklets to obtain stable instant tracklet clustering. The missing tracklets can be regained by increasing the upper bound of velocity correlation, and the incorrectly labeled individuals (tracklets) can be eliminated by rising lower bound. Following this strategy, we define velocity correlation boundary as $\varepsilon_p^{corr} \notin [\min + \alpha \cdot \sigma, \max + \alpha \cdot \sigma]$, where min and max are the minimum and maximum velocity correlation between group members and the computed averaged velocity v_{dir_j}. For each group, the refinement will not stop until the Euclidean distance between coherent motion velocities of two latest iterations is smaller than a threshold ϕ.

To give a clear view, key steps of refining initial tracklet clusters $\{G^\circ\}_{j=1}^M$ are summarized in Algorithm 2.

Algorithm 2 Instant group motion consistency refining.

Input: Initial clusters $\{G^\circ\}_{j=1}^M$, cluster index j, coherent motion velocity variation threshold ϕ ;
Output: Refined tracklet clusters $\{G_j\}_{j=1}^M$;
for each cluster G_j°
  While { v_{dir_j} isn't convergent to ϕ } **do**
    compute instant motion consistency histograms E_j,
    select group motion consistency direction $d = \min_{dir} E_j^{dir}$,
    compute averaged velocity v_{dir_j} at direction d for cluster G_j°,
    for each node $\{i\}$ in cluster G_j°
      compute the velocity correlation $\varepsilon_i^{corr} = \frac{\langle v_i, v_{dir_j} \rangle}{||v_i|| \cdot ||v_{dir_j}||}$,
    compute motion consistency boundary $[\min + \alpha \cdot \sigma \max, +\alpha \cdot \sigma]$ where $\min = \min_i \varepsilon_i^{corr}$, $\max = \max_i \varepsilon_i^{corr}$, σ is the bias of all ε_i^{corr}, α is a scaling parameter,
    eliminate nodes $\mathbb{N} = \{p\}$ from G_j°, if $\varepsilon_p^{corr} \notin [\min + \alpha \cdot \sigma, \max + \alpha \cdot \sigma]$,
    collect isolated nodes $\{q\}$ within the max radius around the center of G_j, namely, $\varepsilon_q^{corr} > \min + \alpha \cdot \sigma$.

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

Fig. 7. Qualitative evaluation over 8 crowd videos in the CUHK crowd dataset, noted as (a)–(h). First column: ground truth of group detection, displayed as tracklet clustering with different colors; Second column: group tracklet clustering produced by CT [1]; Third column: group tracklet clustering produced by CF [7]; Fourth column: group tracklet clustering produced by CCF (proposed). Differences among group detection results of the three group detection methods are marked with white circles. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article).

4. Experimental results

In this section, we demonstrate the effectiveness of the proposed group motion representations for group motion refinement. The quantitative result is measured by 7 measure metrics [33–35], including Normalized Mutual Information (NMI), Purity, Adjusted Rand (AR), unadjusted Rand Index (RI), Mirkin's Index (MI), Hubert's Index (HI) and Accuracy. Smaller HI indicates better clustering performance, the other metrics are vice verse. As far as we know, only NMI, Purity, RI and Accuracy have been selectively evaluated over group detection methods in existing works [1,6,7]. It is the first time that all these 7 metrics are employed to examine group detection performance.

4.1. Data sets

We evaluate the proposed group motion refining framework with both HVC and HAC on the CUHK Crowd Dataset [1]. The dataset includes crowd videos with various densities and perspective scales, collected from various kinds of scenes. The ground truth of the CUHK data set is given in the form of one single frame annotation of each video. Note that, there are noises brought by KLT tracker. We choose a subset of this dataset to provide quantitative evaluation of our instant group motion refining method. In CUHK data set, there are groups annotated in the single frame belonging to fragmented tracklets that fail to sustain over the whole clip [1]. However, the proposed instant group motion refining method, which is designed to refine group members sustain collectiveness in the tracklet fragments cover the given single frame. Thus, we choose a subset of dataset where group members sustain collectiveness cross the given frame to provide quantitative

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

Table 1
Evaluation of group detection with and without HAC and HVC. NMI, Purity, AR, RI, MI, HI, Accuracy are used as quantitative evaluation metrics. The average quantitative performance related to group detection methods CT [1], CF [7] and CCF are presented. For each method, we present at the first column the metric values on the initial clustering without refining, followed by the evaluation of group motion refining with both HAC (second column, with light gray shadow) and HVC (third column, with dark gray shadow).

Average	CT [1]			CF [7]			CCF(proposed)		
	Init.	HAC	HVC	Init.	HAC	HVC	Init.	HAC	HVC
NMI ↑	30.69	44.81	47.27	23.39	35.47	38.43	**32.98**	**47.11**	**47.59**
Purity ↑	88.42	91.79	93.36	88.18	90.61	92.82	**92.66**	**96.78**	**96.87**
AR ↑	29.40	47.23	49.51	20.06	36.65	39.64	**31.75**	**50.25**	**50.61**
RI ↑	80.07	87.48	89.75	73.97	81.61	84.14	**81.60**	**90.31**	**90.49**
HI ↓	19.93	12.52	10.25	26.03	18.39	15.86	**18.40**	**9.69**	**9.51**
MI ↑	60.15	74.95	79.50	47.94	63.22	68.27	**63.20**	**80.62**	**80.98**
Accuracy ↑	41.09	57.72	58.00	**42.68**	55.70	56.12	**44.52**	**60.13**	**60.18**

Table 2
The variance of quantitative performance related to group detection methods CT [1], CF [7] and CCF are presented. For each method, we present at the first column the metric values on the initial clustering without refining, followed by the evaluation of group motion refining with both HAC (second column, with light gray shadow) and HVC (third column, with dark gray shadow).

Variance	CT [1]			CF [7]			CCF(proposed)		
	Init.	HAC	HVC	Init.	HAC	HVC	Init.	HAC	HVC
NMI ↑	0.1229	0.1742	0.1742	0.1229	0.1229	0.1232	**0.1390**	**0.1707**	**0.1742**
Purity ↑	0.0191	0.0125	0.0125	0.0191	0.0058	0.0058	**0.0100**	**0.0020**	**0.0020**
AR ↑	0.1571	0.1896	0.1896	0.1571	0.1368	0.1376	**0.1684**	**0.1851**	**0.1879**
RI ↑	0.0403	0.0215	0.0215	0.0403	0.0205	0.0204	**0.0369**	**0.0166**	**0.0168**
HI ↓	0.0610	0.0861	0.0861	0.1610	0.0818	0.0816	**0.1475**	**0.0665**	**0.0671**
MI ↑	0.0403	0.0215	0.0215	0.0403	0.0205	0.0204	**0.0369**	**0.0166**	**0.0168**
Accuracy ↑	0.0780	0.0906	0.0906	0.1780	0.0941	0.0942	**0.1819**	**0.0866**	**0.0864**

evaluation of our instant group motion refining method. In the selected subset of video clips, we also consider groups from both inside and outside sceneries, different viewpoints of the same scene, different numbers of members, different densities, various uniformities, and different numbers of groups in the same scene.

4.2. Evaluation on initial tracklet clustering for CCF

To verify the independency of the proposed group motion refining framework from initial tracklet clustering result, we adopt results from CT, CF and CCF for evaluation. This enriches the variety of initial tracklet clusters. We carry out the comparison evaluation with the state-of-the-art methods on the same subset of CUHK Crowd Dataset. During the experimental evaluation, parameter configurations for the competing methods are the same as in publications referred to [1,7], where λ=0.6, d=3, and K=10. CCF is verified in selected videos shown in Fig. 7. The quantitative evaluation are listed as three sections in Table 1 which are corresponding to CT, CF and CCF respectively. The variance of the quantitative evaluation are listed in Table 2 in the same way as Table 1. We first apply CT, CF and CCF to selected video clips.[1] Then we evaluate the proposed group motion refining framework on the tracklet clustering of all these three group detection methods.

The quantitative results of initial tracklet clusterings of CT, CF and CCF are shown as the first column in each section of Tables 1 and 2, which are the value of 7 measurements derived from clustering evaluation averaged over video frames listed in Fig. 7. Table 1 indicates that CCF performs better than both CT and CCF in terms of 7 measurements before and after the refinement. According to Table 2, the variances of the proposed method are higher than the competing method on Purity, RI and MI, which are highly related to the purity of a clustering method. After applying the refinement, the stability of the proposed method is improved over the other methods in the light of the second and third columns in each section of Table 2. The ability of generalization of the proposed method could be improved in the future. The ground truth is shown in Fig. 7 followed by the quality result of CT, CF and CCF. Differences among tracklet clustering results of different group detection methods are marked as white circles. As an integration of CF and collectiveness descriptor [6], CCF measures the similarity more accurately when two members are at a distance. Specific individuals in the white circle are detected correctly as the member of the group marked as rose red. CCF performs better in terms of all 7 metrics than other two methods as shown in Table 1. There are unavoidable noises in ground truth of crowd videos in CUHK, as shown in Fig. 7(e), where members grouped by white circle at the bottom right are moving along the opposite direction as an independent group in the crowd. The noise brings no bias to the quantitative evaluation performance comparison of the proposed CCF to the state-of-the-art method CF. However, as to the evaluation of the proposed group motion refining framework, the unavoidable noises in ground truth will affect the accuracy and hence the effectiveness of the instant coherent motion direction, which is the key to group motion refining framework. Therefore, the noisy annotation is eliminated for group motion refining framework evaluation in the following sections without bringing comparison biases.

4.3. Evaluation on representations for group motion refining

The improvement of our group motion refining algorithm is validated at group level across different initial tracklet clusterings in videos of various types of scenes.

In Table 1, the instant group refinement by HAC and HVC are evaluated regarding all the 7 metrics. The comparative improvement ratio of group refining implemented upon two types of group motion representations are listed in Table 3. Through extensive experiments on different combinations of parameter α and ϕ, we set α and ϕ with 0.1 and 0.01, which produce the top two highest values of all 7 metrics over these combinations.

[1] The video results can be downloaded in link https://pan.baidu.com/s/1hsugHF2.

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

Table 3
Relative improvement (percentage) of group motion refining with different instant group motion representations is evaluated, which is computed using Eq. (5). NMI, Purity, AR, RI,MI, HI, Accuracy are quantitatively evaluated in the CUHK Crowd Dataset [1] at each column. The relative improvement of group motion refining with both HAC and HVC are evaluated compared with initial group detection results generated from CT [1], CF [7], CCF(proposed) and the combination of the three.

Ratio(%)	CT [1]		CF [7]		CCF(proposed)		CT + CF + CCF	
	HAC	HVC	HAC	HVC	HAC	HVC	HAC	HVC
NMI ↑	1.13	1.03	0.54	0.49	0.22	0.20	0.63	0.57
Purity ↑	0.20	0.18	0.03	0.02	0.03	0.03	0.08	0.07
AR ↑	0.89	0.81	0.13	0.11	0.13	0.12	0.38	0.35
RI ↑	0.38	0.35	0.10	0.09	0.06	0.05	0.18	0.16
HI ↓	3.45	3.16	1.07	0.98	1.41	1.29	1.97	1.81
MI ↑	0.87	0.80	0.05	0.04	0.12	0.11	0.35	0.32
Accuracy ↑	1.45	1.33	2.06	1.88	1.44	1.32	1.65	1.51

Fig. 8. Qualitative evaluations of instant group motion refining framework on 8 crowd videos in the CUHK crowd dataset, noted as (a)–(h). Evaluation on both HAC and HVC are performed over three methods, including CT, CF and CCF, through group motion refining. Both true and false annotations are illustrated with white circle and white dotted line circle. Note that, the proposed group motion refining framework fails in Fig. 8(b), where initial tracklet clustering at the top row displays relatively less consistency in group member moving orientation.

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

The effectiveness of the proposed group motion representations HAC and HVC to refine groups in crowd videos is demonstrated. Groups detected by different methods are aligned to the ground truth with high match account. The refinement ratio is computed by

$$Ratio = \frac{P_i(G_C{}^{M^*}) - P_i(G^{M^*})}{P_i(G^{M^*})}, \quad (5)$$

where $P_i(G^{M^*})$ is the evaluation on the Metric i for tracklet clustering results G^{M^*} of method M^*, which includes CT, CF and CCF. Table 3 shows the evaluation of group refinement to individual tracklet clustering results of three group detection methods. We also test the refining performance of group motion representations by combing CT, CF and CCF and show the results in Table 3. $P_i(G_C{}^{M^*})$ is the evaluation of metric i for group motion refining of method M^* with group motion consistency representation C, such as velocity consistency and angle consistency.

CCF performs relatively better in terms of all 7 metrics than both CT and CF, no matter with or without the group refinement. With group member motion direction and amplitude complementing each other, HVC outperforms HAC in group motion refining. The performance of refining CCF detection results with HVC is higher than others. The average accuracy improvement ratio of CCF by using HVC is 1.51%. There are groups where HAC can obtain relatively greater improvement. Thus HVC has overwhelm compatibility for coherent motion filtering than HAC when there are relative noisy in initial tracklet clusters.

Relative qualitative improvements of using HAC and HVC for group motion refining are marked as white circles in Fig. 8(a)–(h). Group detection results of CT, CF and CCF are used as the initial tracklet clusters adopted in Section 3.3. The proposed group motion refining framework can refine most of tracklet clusters as well as maintaining the motion consistency for groups otherwise. As shown in Fig. 8(c),(f),(h), the proposed refining framework displays no negative affection on the initial group clustering results. We also observe cases where the proposed refining framework fails. In Fig. 8(b), group members clustered to wrong groups are marked with white dotted line circles, where initial groups at the top row display relative less consistency on group member moving orientations. It proves that group motion representations HAC and HVC in this paper are much more suitable for refining initial tracklet clusters with better consistency in moving orientation.

5. Conclusions

In this paper, we have studied representations of instant group motion for characterizing intra-group motion. We observe that the group motion consistency at each direction can capture the inter-group motion consistency differences. Two group motion representations based on group motion consistency are proposed as HAC and HVC. By the proposed group motion representations, the group motion consistency orientation is distinguished for the instant group motion consistency filtering strategy. An instant group motion refining framework is built upon the group motion consistency filtering strategy. Evaluation is conducted based on group detection results generated by both existing methods (CT,CF) and a newly proposed group detection method (CCF) by us, which proves the universality of our framework. Experimental results show that CCF is better than CT and CF, and refining results show the superiority of two group motion representations for group motion refining.

In the future work, we will explore the application of the group motion consistency on group level crowd event detection and modeling.

References

[1] J. Shao, C.C. Loy, X. Wang, Scene-independent group profiling in crowd, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 431, 2014, pp. 2227–2234.
[2] S.P. Hoogendoorn, P. Bovy, Pedestrian route-choice and activity scheduling theory and models, Transp. Res. Part B: Methodol. 38 (2) (2004) 169–190.
[3] W. Ge, R.T. Collins, R.B. Ruback, Vision-based analysis of small groups in pedestrian crowds, IEEE Trans. Pattern Anal. Mach. Intell. 34 (5) (2012) 1003–1016.
[4] A.B. Chan, N. Vasconcelos, Modeling, clustering, and segmenting video with mixtures of dynamic textures, IEEE Trans. Pattern Anal. Mach. Intell. 30 (5) (2008) 909–926.
[5] M. Mehdi, N. Perozo, S. Garnier, D. Helbing, G. Theraulaz, The walking behaviour of pedestrian social groups and its impact on crowd dynamics, PloS one 5 (4) (2010) e10047.
[6] B. Zhou, X. Tang, H. Zhang, X. Wang, Measuring crowd collectiveness, IEEE Trans. Pattern Anal. Mach. Intell. 36 (8) (2014) 1586–1599.
[7] B. Zhou, X. Tang, X. Wang, Coherent filtering: detecting coherent motions from crowd clutters, in: Proceedings of the IEEE Conference on European Conference on Computer Vision(ECCV), 2012, pp. 857–871.
[8] W. Choi, S. Savarese, A unified framework for multi-target tracking and collective activity recognition, in: Proceedings of the IEEE Conference on European Conference on Computer Vision (ECCV), 2012, pp. 215–230.
[9] Y. Liu, J. Chen, Z. Su, Z. Luo, N. Luo, L. Liu, K. Zhang, Robust head pose estimation using Dirichlet-tree distribution enhanced random forests, Neurocomputing 173 (2016) 42–53.
[10] S. Pellegrini, A. Ess, K. Schindler, L.V. Gool, You'll never walk alone: Modeling social behavior for multi-target tracking, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2009, pp. 261–268.
[11] A.K. Chandran, L.A. Poh, P. Vadakkepat, Identifying social groups in pedestrian crowd videos, in: Proceedings of International Conference on Advances in Pattern Recognition (ICAPR), 2015, pp. 1–6.
[12] L. Leal-Taixe, G. Pons-Moll, B. Rosenhahn, Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2011, pp. 120–127.
[13] A. Alahi, R. Vignesh, F.-F. Li, Socially-aware large-scale crowd forecasting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (1), 2014, pp. 2211–2218.
[14] D. Kuettel, M.D. Breitenstein, L.V. Gool, V. Ferrari, What's going on? discovering spatio-temporal dependencies in dynamic scenes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 1951–1958.
[15] M.S. Zitouni, H. Bhaskar, J. Dias, M. Al-Mualla, Advances and trends in visual crowd analysis: a systematic survey and evaluation of crowd modelling techniques, Neurocomputing 186 (2016) 139–159.
[16] F. Setti, C. Russell, C. Bassetti, M. Cristani, F-formation detection: individuating free-standing conversational groups in images, Plos One 10 (5) (2014) 507–514.
[17] W. Choi, Y. Chao, C. Pantofaru, S. Savarese, Discovering groups of people in images, in: Proceedings of the European Conference on Computer Vision (ECCV), 8692, 2014, pp. 417–433.
[18] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, S. Yan, Crowded scene analysis: a survey, Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) 25 (3) (2015) 367–386.
[19] P.M. Jodoin, Y. Benezeth, Y. Wang, Meta-tracking for video scene understanding, in: Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, 2013, pp. 1–6.
[20] Y. Cong, J. Yuan, J. Liu, Abnormal event detection in crowded scenes using sparse representation, Pattern Recognit. 46 (7) (2013) 1851–1864.
[21] L. Kratz, K. Nishino, Tracking pedestrians using local spatiotemporal motion patterns in extremely crowded scenes, Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 34 (5) (2012) 987–1002.
[22] S. Vascon, E. Mequanint, M. Cristani, H. Hung, M. Pelillo, V. Murino, Detecting conversational groups in images and sequences: a robust game-theoretic approach, Comput. Vis. Image Underst. 143 (2015) 11–24.
[23] J. Shao, C.C. Loy, X. Wang, Learning scene-independent group descriptors for crowd understanding, Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2016). 1–1.
[24] R. Li, R. Chellappa, Group motion segmentation using a spatio-temporal driving force model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 23(3), 2010, pp. 2038–2045.
[25] R. Mazzon, F. Poiesi, A. Cavallaro, Detection and tracking of groups in crowd, in: Proceedings of the 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 13(4), 2013, pp. 202–207.
[26] Z. Cheng, L. Qin, Q. Huang, S. Yan, Q. Tian, Recognizing human group action by layered model with multiple cues, Neurocomputing 136 (2014) 124–135.
[27] Y. Zhao, M. Yuan, G. Su, T. Chen, Crowd macro state detection using entropy model, Physica A: Stat. Mech Appl. 431 (2015) 84–93.
[28] X. Gu, J. Cui, Q. Zhu, Abnormal crowd behavior detection by using the particle entropy, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 125(14), 2014, pp. 3428–3433.
[29] Z. Pan, J. Lei, Y. Zhang, X. Sun, S. Kwong, Fast motion estimation based on content property for low-complexity H.265/HEVC encoder, IEEE Trans. Broadcast. 62 (3) (2016) 675–684.

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045

[30] Z. Pan, Y. Zhang, S. Kwong, Efficient motion and disparity estimation optimization for low complexity multiview video coding, IEEE Trans. Broadcast. 61 (2) (2015) 166–176.
[31] B. Zhang, L. Wang, Z. Wang, Y. Qiao, H. Wang, Real-time action recognition with enhanced motion vector CNNS, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2718–2726.
[32] S. Ali, M. Shah, A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–6.
[33] L. Hubert, P. Arabie, Comparing partitions, J. Classif. 2 (1) (1985) 193–218.
[34] M. Wu, B. Schlkopf, A local learning approach for clustering, Adv. Neural Inf. Process. Syst. 2 (1) (2006) 1529–1536.
[35] C. Aggarwal, A human-computer interactive method for projected clustering, IEEE Trans. Knowl. Data Eng. 16 (4) (2004) 448–460.

Na Li received the B.S. degree in computer science and technology from Hunan University, Changsha, China, in 2009, and the Ph.D. degree in computer application and technology from the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, in 2014. From 2014 to 2016, she was a Post-Doctoral Researcher with the Center for High Performance Computing, Shenzhen Institute of Advanced Technology, CAS, Shenzhen, China. In 2016, she became an Assistant Professor with the Shenzhen Institutes of Advanced Technology, CAS. Her research interests are machine learning, crowded scene analysis, video analysis and video processing.

Yun Zhang received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and received the Ph.D. degree in computer science from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2014, he was a Visiting Scholar with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. In 2010, he joined in the Shenzhen Institutes of Advanced Technology (SIAT), CAS, as an Assistant Professor. Since 2012, he serves as Associate Professor. His research interests are 3D video coding, high efficiency video coding and video signal processing.

Wenhan Luo is currently as a research scientist with the AI Lab in Tencent. His research interests include several topics in computer vision and machine learning, such as motion analysis (especially object tracking), image quality restoration, object detection and recognition. Before he joined Tencent, he received the Ph.D degree from Imperial College London, UK, in 2016, the Master degree from the Institute of Automation, Chinese Academy of Sciences, China, in 2012, and the Bachelor degree from Huazhong University of Science & Technology, China, in 2009.

Ning Guo received her master degree in Signal processing system from GUILIN University of electronic and technology, China, in 2011. She joined Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences as Research Assistant in 2011. Her research interest focus in machine learning, especially in classification and clustering.

Please cite this article as: N. Li et al., Instant coherent group motion filtering by group motion representations, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.05.045