

# Coherent Filtering: Detecting Coherent Motions from Crowd Clutters

Bolei Zhou<sup>1</sup>, Xiaoou Tang<sup>1,3</sup>, and Xiaogang Wang<sup>2,3</sup>

<sup>1</sup> Department of Information Engineering, The Chinese University of Hong Kong

<sup>2</sup> Department of Electronic Engineering, The Chinese University of Hong Kong

<sup>3</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

**Abstract.** Coherent motions, which describe the collective movements of individuals in crowd, widely exist in physical and biological systems. Understanding their underlying priors and detecting various coherent motion patterns from background clutters have both scientific values and a wide range of practical applications, especially for crowd motion analysis. In this paper, we propose and study a prior of coherent motion called *Coherent Neighbor Invariance*, which characterizes the local spatiotemporal relationships of individuals in coherent motion. Based on the coherent neighbor invariance, a general technique of detecting coherent motion patterns from noisy time-series data called *Coherent Filtering* is proposed. It can be effectively applied to data with different distributions at different scales in various real-world problems, where the environments could be sparse or extremely crowded with heavy noise. Experimental evaluation and comparison on synthetic and real data show the existence of Coherence Neighbor Invariance and the effectiveness of our Coherent Filtering.<sup>1</sup>

## 1 Introduction

Coherent motion is a universal phenomenon in nature and widely exists in many physical and biological systems. For example, tornadoes, storms, and atmospheric circulation are all caused by the coherent movements of physical particles in the atmosphere. Meanwhile, the collective behaviors of organisms such as swarming ants and schooling fishes have long captured the interests of social and natural scientists [1, 2]. Detecting coherent motions and understanding their underlying principles are related to many important scientific research topics such as self-organization of biological systems [3]. There is also a wide range of practical applications. For example, in video surveillance, detecting coherent motion patterns of pedestrian groups in crowd has important applications to object counting [4, 5], crowd tracking [6], and crowd management [7]. Furthermore, clusters of coherent motions provide a mid-level representation of crowd dynamics, and could be used for high-level semantic analysis such as scene understanding and crowd activity recognition [8–11].

Generally speaking, coherent motion detection can be formulated as finding clusters of dynamic particles with coherent motion from their time-series observation and removing background noise as outliers. Under different scene context, the detected coherent motions may be interpreted as different semantic behaviors. As shown in Figure

---

<sup>1</sup> Source codes and videos are available from  
<http://mmlab.ie.cuhk.edu.hk/coherentfiltering/>

9, moving keypoints tracked in the scenes exhibit a wide variety of coherent motion patterns, corresponding to individual and group movements, traffic mode, crowd flow *etc.* These examples show that detecting coherent motions from noisy observations is of great importance to crowd behavior analysis and scene understanding.

The goal of this work is to explore the underlying prior in the dynamics of coherent motions and to leverage it for coherent motion detection. We propose a prior called *Coherent Neighbor Invariance*, which exists in the local neighborhoods of individuals in coherent motions, and show that it well distinguishes coherent and incoherent motions. Then we develop a general coherent motion detection technique called *Coherent Filtering* based on such a prior. Experiments show that it can work robustly in online mode for various applications.

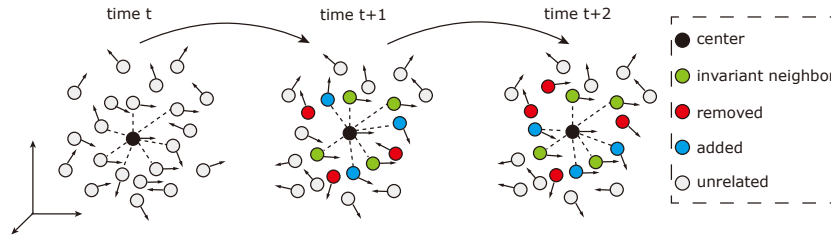
### 1.1 Related Works

Crowd motion analysis has been an active research topic in computer vision. It is closely related to many low-level and high-level computer vision problems in crowded or cluttered environments. For example, Rabaud *et al.* [4] and Brostow *et al.* [5] proposed approaches to detect independent motions in order to count moving objects. Lin *et al.* [12] used Lie algebra of affine transform to learn the global motion patterns of crowds. Ali *et al.* [6] used floor fields from fluid mechanics for the segmentation of crowd flows. Hu *et al.* [13] clustered the single-frame optical flows to learn the motion patterns. Brox *et al.* extended spectral clustering to group long-term dense trajectories for the segmentation of moving objects in video. Meanwhile, the high-level semantic analysis in crowded scenes focuses on modeling scene structures and recognizing crowd behaviors. Wang *et al.* [8, 14] and Zhou *et al.* [10] used hierarchical topic models to learn the models of semantic regions from the co-occurrence of optical flow and tracks. Zhou *et al.* [9] proposed a mixture model of dynamic pedestrian-agents to learn the collective crowd behavior patterns in crowded scenes. In 3D motion segmentation [15], under the assumption of affine transform there are several subspace approaches proposed, such as Generalized Principal Component Analysis (GPCA) [16] and RANSAC [15].

Although these problems are closely related to coherent motions, they were studied separately and had different limited assumptions and solutions in the past literature. For example, the approaches in [4, 5] are specifically designed for object counting and had assumptions on the shape and distribution of coherent motions. The models in [12] and [6] assume the geometric transformation of motion. The clustering methods in [13, 17] are easily affected by noises. GPCA [16] and RANSAC [18] consider the motion trajectories lies on motion subspace and are projected under the affine camera model. Therefore, these approaches cannot be well generalized to other problems related to coherent motions. We develop a general coherent motion technique which can be well applied to the various problems discussed above. It achieves comparable results or even outperforms the state-of-the-arts in different application domains, especially when the background noise is heavy.

## 2 The Prior: Coherent Neighbor Invariance

Although coherent motions are the macroscopic observations of collective movements of individuals, recent studies [2, 19] show that it actually can be characterized by the



**Fig. 1.** Illustration of coherent neighbor invariance. The green dots are the invariant  $K$  nearest neighbors of the central black dot over time (here  $K = 7$ ). The invariant neighbors have a higher probability to be the dots moving coherently with the central dot, since their local spatiotemporal relationships and velocity correlations with the central dot are inclined to remain invariant over time. The red and blue dots change their neighborhood over time (removed or added), so that they have a small probability to move coherently with the central dot.

interactions among individuals within local neighborhoods. Inspired by these observations and results, we propose a prior underlying the dynamics of coherent motion as *Coherent Neighbor Invariance*. There are two key properties of Coherent Neighbor Invariance, which distinguish coherent motions from random motions:

- Invariance of spatiotemporal relationships: the neighborhood of individuals with coherent motions tends to remain invariant over time.
- Invariance of velocity correlations: the velocity correlations of neighboring individuals with coherent motions remain high when being averaged over time.

On the contrary, incoherently moving individuals do not have such properties because of the mutual independence of their movements. Coherent neighbor invariance exists in the dynamics of  $K$  nearest neighbors of individuals for consecutive time. It reflects the self-organization of coherent motions in neighborhood, and explains the formation of global coherent motion patterns from local coordinations of individuals. An illustration of the prior is shown in Figure 1.

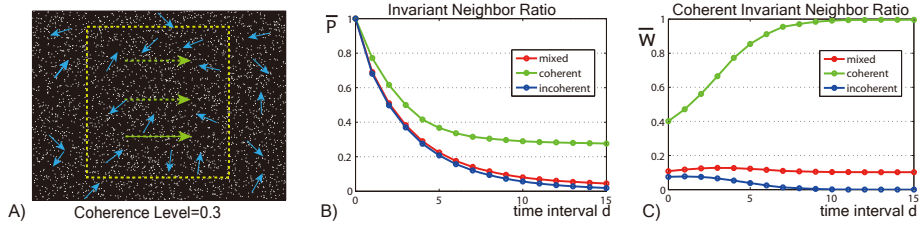
To quantitatively analyze this prior, we first define two measurements of Coherent Neighbor Invariance: the coherent neighbor invariance of spatiotemporal relationships (in Section 2.2), and the coherent neighbor invariance of velocity correlations (in Section 2.3). From the following experiments, we show that this prior not only helps reveal the mechanism of coherent motions, but also can be effectively leveraged to separate various coherent motion patterns from background noise using a technique called Coherent Filtering.

### 2.1 Random Dot Kinematogram

We take Random Dot Kinematogram (RDK) as an example to analyze the coherent neighbor invariance, because it is easy to understand and can be well generalized. RDK is a classic psychophysical stimulus, and is often used for investigating coherent motion perception of visual systems [20]. The stimulus consists of a completely random array of thousands of tiny dots that move either coherently or randomly. An illustration is shown in Figure 2A. Incoherently moving dots (referred as incoherent dots) are

**Table 1.** Notations used in the paper

$ \cdot $	The cardinality of a set.
$\cap$	The intersection of two sets.
$\mathcal{F}$	$\mathcal{F} = \{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ . The set of coherent dots.
$\mathcal{B}$	The set of incoherent dots.
$\mathcal{N}_t^i$	$\mathcal{N}_t^i = \{i_t^1, \dots, i_t^K\}$ . The set of the $K$ -NNs of dot $i$ at time $t$ .
$\mathcal{M}_{t \rightarrow d}^i$	$\mathcal{M}_{t \rightarrow d}^i = \mathcal{N}_t^i \cap \mathcal{N}_{t+1}^i \cap \dots \cap \mathcal{N}_{t+d}^i$ . The $d^{th}$ order invariant neighbor set of $i$ .
$\mathcal{C}_{t \rightarrow d}^i$	$\mathcal{C}_{t \rightarrow d}^i = \mathcal{M}_{t \rightarrow d}^i \cap \mathcal{F}$ . The coherent invariant neighbor set of $i$ .
$g_{t \rightarrow d}^{i_k}$	Averaged velocity correlation between $i$ and $i_k$ from $t$ to $t+d$ .
$\mathcal{R}$	$\mathcal{R} = \{(i, i_k)   g_{t \rightarrow d}^{i_k} > \lambda, i_k \in \mathcal{A}\}$ . The set of pairwise connections.
$\mathcal{S}$	$\mathcal{S} = \{s_i   i \in \mathcal{I}\}$ . The set of the cluster index $s_i$ for each dot $i$ .



**Fig. 2.** A) Illustration of random dot kinematogram. Here the number of coherent motion pattern  $N = 1$ . The cyan arrows indicate the random moving directions of some noisy dots with incoherent motions. The green arrows indicate the direction of coherent motion. B) Averaged invariant neighbor ratios  $\bar{P}$  with time interval  $d$ . C) Averaged coherent invariant neighbor ratios  $\bar{W}$  with time interval  $d$ . All these measurements are computed and averaged for coherent dots, incoherent dots and all the dots respectively for comparison.

randomly placed over the whole scene and serve as background noise. In the central rectangular area, a group of coherently moving dots (referred as coherent dots) are also randomly placed. The proportion of coherent dots to all the dots in the rectangular area (mixing of coherent dots and incoherent dots) is called the coherence level. In psychophysical study, human subjects are required to identify coherent motions of dots from the background noise. The coherence level determines the difficulty level of the identification task.

Formally, we denote  $\mathcal{I}$  as the set of all the dots (mixed dots) in the central rectangular area,  $\mathcal{F}$  as the set of coherent dots, and  $\mathcal{B}$  as the set of incoherent dots. There could be  $N$  different coherent motion patterns which divide  $\mathcal{F}$  into subsets  $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ . Thus the problem of detecting coherent motions from noisy observations of dot movements is formulated as estimating the separation  $\mathcal{I} = \{\mathcal{F}, \mathcal{B}\}$ , and the sub-separation  $\mathcal{F} = \{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ . In Sections 2.2 and 2.3, we will analyze two coherent neighbor invariance measurements and study their dynamic behaviors. Table 1 shows the notations used in the paper.

## 2.2 Invariance of Spatiotemporal Relationships

This subsection analyzes the invariance of spatiotemporal relationships in coherent motions. We first define some related concepts. At time  $t$ , the set  $\mathcal{N}_t^i$  contains the  $K$  near-

est neighbors of dot  $i$  under Euclidean distance. It evolves into  $\mathcal{N}_{t+1}^i$  at time  $t + 1$ <sup>2</sup>. We denote  $\mathcal{M}_{t \rightarrow 1}^i$  as the 1<sup>st</sup> order invariant neighbor set, which contains the invariant neighbors among the  $K$  nearest neighbors of dot  $i$  from time  $t$  to  $t + 1$ . Similarly,  $\mathcal{M}_{t \rightarrow d}^i$  is denoted as the  $d^{\text{th}}$  order invariant neighbor set which contains the invariant neighbors from time  $t$  to  $t + d$ . We denote  $\mathcal{C}_{t \rightarrow d}^i$  as the intersection of  $\mathcal{M}_{t \rightarrow d}^i$  and  $\mathcal{F}$ , so that it only contains the invariant neighbors with coherent motion. It is called the coherent invariant neighbor set of dot  $i$  at  $d^{\text{th}}$  order.

Two ratios are defined. The first is the invariant neighbor ratio  $P_{t \rightarrow d}^i$ , measuring the proportion of invariant neighbors among the  $K$  nearest neighbors during time  $t$  to  $t + d$ . The second is the coherent invariant neighbor ratio  $W_{t \rightarrow d}^i$ , measuring the proportion of coherent dots among the invariant neighbors during time  $t$  to  $t + d$ . Specifically,

$$P_{t \rightarrow d}^i = \frac{|\mathcal{M}_{t \rightarrow d}^i|}{K}, \quad W_{t \rightarrow d}^i = \frac{|\mathcal{C}_{t \rightarrow d}^i|}{|\mathcal{M}_{t \rightarrow d}^i|},$$

where  $P_{t \rightarrow d}^i$  and  $W_{t \rightarrow d}^i \in [0, 1]$ . Obviously  $P_{t \rightarrow d}^i$  and  $W_{t \rightarrow d}^i$  will change with the value of  $d$  and they describe the temporal behaviors of dots with coherent and incoherent motions. For an incoherent dot  $i$ , since most of the dots in its neighborhood move independently with dot  $i$ , its  $K$  nearest neighbors would vary greatly over time. Thus  $P_{t \rightarrow d}^i$  is expected to decrease quickly when  $d$  increase, and gradually to 0. On the contrary, for a coherent dot  $i$ , dots in its neighborhood, which move coherently with  $i$ , would tend to remain in the neighborhood during the whole time interval  $d$  because of their consistent movements with  $i$ , while other incoherent dots in its neighborhood will easily move out of the neighborhood. Thus  $P_{t \rightarrow d}^i$  is expected to decrease slower than that of incoherent dots and then remains as a constant when  $d$  increases further. On the other hand,  $W_{t \rightarrow d}^i$  measures the proportion of coherent dots in the invariant neighbor set  $\mathcal{M}_{t \rightarrow d}^i$ . Obviously only the dots which move coherently with dot  $i$  tend to remain in the neighborhood of  $i$  from time  $t$  to  $t + d$ . For an incoherent dot  $i$ , because all dots in  $\mathcal{M}_{t \rightarrow d}^i$  are moving independently with dot  $i$ ,  $W_{t \rightarrow d}^i$  is expected to be low over time. For a coherent dot  $i$ , since majority of the remaining dots in  $\mathcal{M}_{t \rightarrow d}^i$  move coherently with  $i$  as  $d$  increases,  $W_{t \rightarrow d}^i$  is expected to increase with  $d$ .

Since the groundtruth of dots as  $\mathcal{F}$  or  $\mathcal{B}$  are known in RDK, we can compute and analyze the two ratios for coherent dots, incoherent dots, and mixed dots respectively. We set the coherence level as 0.3, which means there are 30% dots ( $\sim 800$ ) moving coherently in the central rectangular area of Figure 2A. As shown in Figures 2B and 2C, the experimental results of the two ratios in RDK verify our analysis. We can see that as  $d$  increases the averaged invariant neighbor ratios  $\overline{P}$  for coherent dots and incoherent dots are clearly separated. Meanwhile, the averaged coherent invariant neighbor ratio  $\overline{W}$  for coherent dots increases almost to 1, and that for incoherent dots decreases to 0.

Our analysis and illustrative results in RDK show the invariant neighbor ratio and the coherent invariant neighbor ratio have good discriminability for coherent and incoherent motions. We call this property of coherent motion as *coherent neighbor invariance of spatiotemporal relationships*.

<sup>2</sup> The correspondence of dots over time is assumed.

### 2.3 Invariance of Velocity Correlations

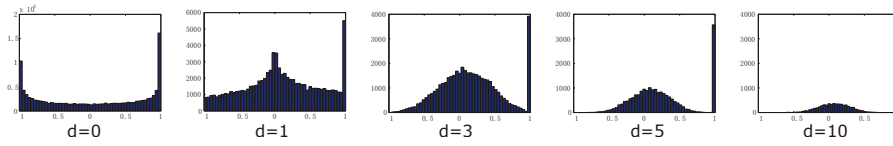
The other property of coherent motion is the invariance of velocity correlations between neighboring dots. Suppose that dot  $i_k$  belongs to the invariant neighbor set of dot  $i$ . Their velocity correlation averaged from time  $t$  to  $t + d$  is

$$g_{t \rightarrow d}^{i_k} = \frac{1}{d+1} \sum_{\tau=t}^{t+d} \frac{\mathbf{v}_\tau^i \cdot \mathbf{v}_\tau^{i_k}}{\|\mathbf{v}_\tau^i\| \cdot \|\mathbf{v}_\tau^{i_k}\|},$$

where  $\mathbf{v}_\tau^i$  is the velocity of  $i$  at time  $\tau$ . If dot  $i_k$  moves incoherently with dot  $i$ ,  $g_{t \rightarrow d}^{i_k}$  would be low as  $d$  increases. Otherwise,  $g_{t \rightarrow d}^{i_k}$  remains high. Therefore, the velocity correlations of coherently moving dots and incoherently moving dots in local regions can be well separated as  $d$  increases.

Figure 3 shows the histograms of  $g_{t \rightarrow d}^{i_k}$  generated from the invariant neighbor set in RDK, with  $d = 0, 1, 3, 5, 10$  respectively. The experimental results verify our analysis above. We can see that as  $d$  increases, the histogram gradually separates into two modes: one near 0 and the other near 1. This property of coherent motions is called *coherent neighbor invariance of velocity correlations*.

Because of this property, it is simple to remove the incoherent dots  $i_k$  from the invariant neighbor set  $\mathcal{M}_{t \rightarrow d}^i$ : setting a threshold  $\lambda$  on the value of  $g_{t \rightarrow d}^{i_k}$  and then removing  $i_k$  from  $\mathcal{M}_{t \rightarrow d}^i$  if  $g_{t \rightarrow d}^{i_k} < \lambda$ . After thresholding, we can create a set  $\mathcal{R}$  containing the thresholded pairwise connections, in which  $(i, i_k)$  are connected if  $i_k$  still remains in  $\mathcal{M}_{t \rightarrow d}^i$ . Then coherent motions can be easily detected according to  $\mathcal{R}$ , using the algorithm proposed in Section 3.1.



**Fig. 3.** Histograms of  $g_{t \rightarrow d}^{i_k}$  generated from the invariant neighbor set in RDK, with  $d = 0, 1, 3, 5, 10$  respectively. As  $d$  increases,  $g_{t \rightarrow d}^{i_k}$  of coherently moving dots and incoherently moving dots are well separated. The bar near 1 is the histogram of  $g_{t \rightarrow d}^{i_k}$  of coherently moving dots, and the hump near 0 is the histogram of  $g_{t \rightarrow d}^{i_k}$  of incoherently moving dots.

## 3 Coherent Filtering Using Coherent Neighbor Invariance

Based on Coherent Neighbor Invariance, a clustering technique Coherent Filtering is proposed for coherent motion detection from time-series data. Coherent Filtering consists of two algorithms. The first is to detect coherent motion patterns at one time, the second is to associate detected coherent motion and update existing coherent motion over time. Algorithms are listed in Table 2 and Table 3 respectively.

Coherent Filtering has some important merits. 1) It stems from the coherent neighbor invariance, which is a prior widely observed in coherent motions. Therefore it is a general technique for clustering time-series data and detecting coherent motions in various

real-world problems, such as object counting, group movement detection, traffic mode recognition, and crowd flow segmentation. 2) It only relies on local spatiotemporal relationships and velocity correlations without any assumption on the global shape of coherent motion patterns and the distribution of background noise. Therefore it can be robustly applied to data at different scales and distributions without substantial change. 3) In practical applications, it might be difficult to obtain the correspondence of key-points over a long time, especially in crowded environments. Experiments show that our algorithm only requires correspondence over a short period (normally 4 or 5 frames). This means that it can work robustly in crowded scenes and in an online mode. 4) The cluster number  $N$  is automatically decided from data without knowing *a priori*.

### 3.1 Algorithm for Detecting Coherent Motions

In Algorithm **CoheFilterDet** we first obtain the invariant neighbor set  $\mathcal{M}_{t \rightarrow d}^i$  by examining the neighborhood in  $\mathcal{N}_\tau^i$  from  $t$  to  $t + d$  for each dot  $i \in \mathcal{I}$ . According to *coherent neighbor invariance of spatiotemporal relationships*, most dots in  $\mathcal{M}_{t \rightarrow d}^i$  should be coherent dots. However, it does not guarantee that *all* the dots in  $\mathcal{A}$  are coherent dots especially when  $d$  is small. Then, according to *the coherent neighbor invariance of velocity correlations*, we set a threshold  $\lambda$  on the averaged velocity correlations to remove incoherent dots and obtain the pairwise connection set  $\mathcal{R}$ . Finally a connectivity graph is built, where nodes are dots and edges are defined by connection relationships in  $\mathcal{R}$ . With this graph, incoherent dots  $\mathcal{B}$  are identified as isolated nodes and coherent motion clusters  $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$  are identified as the connected components of the graph.

**Table 2.** Algorithm **CoheFilterDet** for detecting coherent motion patterns

---

**FUNCTION**  $(\mathcal{F}_1, \dots, \mathcal{F}_N) = \text{CoheFilterDet}(\mathcal{I})$

---

01: **for**  $\tau = t$  to  $t + d$   
 02:   search the  $K$  nearest neighbor set as  $\mathcal{N}_\tau^i$  for each dot  $i \in \mathcal{I}$   
 03: **for** each dot  $i \in \mathcal{I}$   
 04:   search the invariant neighbor set as  $\mathcal{M}_{t \rightarrow d}^i$   
 05:   **for** each  $i_k \in \mathcal{M}_{t \rightarrow d}^i$   
 06:     compute the averaged velocity correlations  $g_{t \rightarrow d}^{i_k}$   
 07:     include  $(i, i_k)$  in  $\mathcal{R}$  if  $g_{t \rightarrow d}^{i_k} > \lambda$ .  
 08: Build a graph from  $\mathcal{R}$ . Remove incoherently moving individuals as the isolated nodes and identify coherent motion  $\{\mathcal{F}_1, \dots, \mathcal{F}_N\}$  as the connected components of the graph.

---

### 3.2 Algorithm for Associating Continuous Coherent Motion

Coherent motion clusters will continue and evolve, and new cluster of coherent motion will emerge over time. Based on the temporal overlaps of trajectories we develop another algorithm **CoheFilterAssoci** to associate and update the clusters of coherent motion over consecutive frames.

To associate the clusters of coherent motion over time, we define a variable  $s_i$  as the cluster index for each trajectory  $i$ . As illustrated in Figure 4A, the algorithm will update

**Table 3.** Algorithm **CoheFilterAssoci** for associating coherent motions

---

**FUNCTION** ( $\mathcal{S}^{t+1}$ ) = **CoheFilterAssoci**( $\mathcal{S}^t, \mathcal{I}^{t+1}$ )

---

01:  $(\mathcal{F}_1, \dots, \mathcal{F}_{N_{t+1}}) = \mathbf{CoheFilterDet}(\mathcal{I}^{t+1})$  /\*detect new coherent motion at frame t+1\*/

02: **for** each  $i \in \mathcal{I}^t \cap \mathcal{I}^{t+1}$

03:  $s_i^{t+1} = s_i^t$  /\*firstly assume there is no cluster index changing for dot  $i$ \*/

04:  $M = \max(\mathcal{S}^t)$  /\*maximum cluster index value in  $\mathcal{S}^t$ \*/

05: **for** each  $\mathcal{F}_{n_{t+1}}$

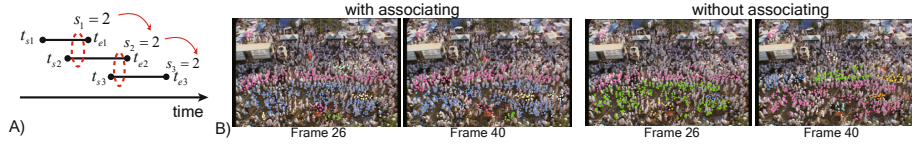
06:  $S = \mathbf{mode}(\mathcal{H})$ , where  $\mathcal{H} = \{s_i^{t+1} | i \in \mathcal{F}_{n_{t+1}}\}$  /\*get the most frequent value in  $\mathcal{H}$ \*/

07: **if**  $S=0$  **then**  $S=M+1$  /\*add a new cluster index\*/

08: **for** each  $i \in \mathcal{F}_{n_{t+1}}$  let  $s_i^{t+1}=S$

09: identify dot  $i \in \mathcal{I}^{t+1}$  as foreground and its cluster index as  $s_i^{t+1}$  if  $s_i^{t+1} > 0$

---



**Fig. 4.** Illustration of associating continuous coherent motions. A) There are temporal overlaps between trajectory 1 and 2, trajectory 2 and 3. If trajectory 1 and 2 are detected into one coherent motion cluster at one time, and trajectory 2 and 3 are detected into one coherent motion cluster at next time, the index  $s_1 = 2$  of trajectory 1 will be transferred to the other two trajectories. Red circles indicate trajectories are detected into one coherent motion cluster,  $t_{s_i}$  and  $t_{e_i}$  denote the starting and ending time of trajectory  $i$ . B) Two representative frames of coherent motion detection result with associating and without associating respectively. Dots in the same color belong to one coherent motion cluster over time and space.

the cluster indice of trajectories by majority voting and keep on detecting new coherent motion cluster over time. The detail of algorithm is listed in Table 3.

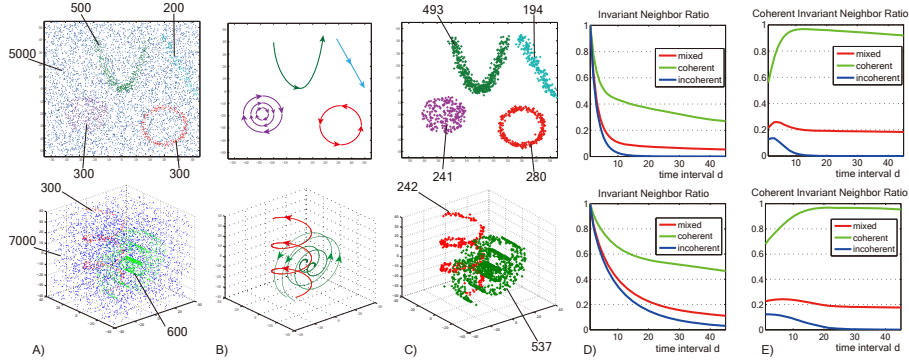
## 4 Experimental Results

In this section, we will evaluate the robustness and effectiveness of Coherent Filtering on complex synthetic data, real 3D motion segmentation database, and crowd videos. On the synthetic data, we test the technique by detecting coherent motion patterns with different dynamics from high-density Brownian motion noise. Then we evaluate the technique on the 3D affine motion segmentation Hopkins155 database [15], and compare it to several baseline methods on the database in the presence of outliers. Lastly we test Coherent Filtering on videos by detecting coherent motion patterns in real scenes with a variety of scales and crowdedness.

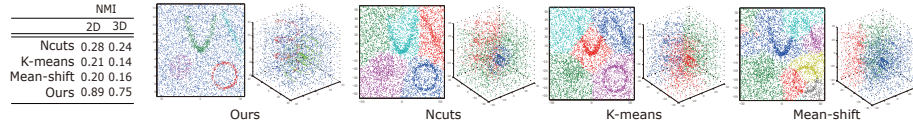
### 4.1 Coherent Motion in Synthetic Data

Figure 5A shows the two synthetic datasets for evaluation. The coherent motion patterns emerging in the datasets vary greatly in scales, shapes, and dynamics. For the 2D





**Fig. 5.** Coherent motion detection on synthetic 2D and 3D datasets. A) The shapes and the numbers of coherently moving dots (colors indicate different coherent motion patterns) and noisy dots (in blue). B) The traces of each coherent motion patterns. C) The detected coherent motion patterns by Coherent Filtering. D) Invariant neighbor ratios for different types of dots over time. E) Coherent invariant neighbor ratios for different types of dots over time. The algorithm parameters are  $K = 15$ ,  $d = 4$  and  $\lambda = 0.6$ .



**Fig. 6.** The qualitative and quantitative results of the four methods for comparison. Colors indicate different clusters. NMI is used to quantitatively evaluate the clustering results. Our technique achieves the best performance.

dataset, the parametric forms of parabola, circle, line, and disk are used to generate four 2D coherent motion patterns. For the 3D dataset, the parametric forms of helix and spiral surface are used to generate two 3D coherent motion patterns. Figure 5B illustrates the traces of the synthetic coherent motion patterns. Initial positions of the coherent dots are randomly sampled from Gaussian distribution along the traces of the coherent motion patterns.

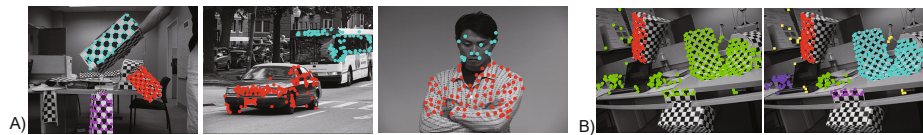
As the detection results shown in Figure 5C, our technique detects well these coherent motion patterns from rather noisy data. The good performance of detecting various coherent motion patterns shows the robustness and generalization ability of our technique. Figures 5D and 5E show that the invariant neighbor ratios and the coherent invariant neighbor ratios for the two datasets. We can see they have good discriminability between coherent dots and incoherent noisy dots. It also verifies the existence of coherent neighbor invariance in the synthetic data.

Three representative clustering methods, *i.e.*, Normalized Cuts (Ncuts) [21], K-means and Mean-shift [22], are selected for comparison. Ncuts, K-means, and Mean-shift are often extended for time-series data clustering [23] and trajectory clustering [24]. By convention, we treat the trajectory of each dot  $i$  from time  $t$  to  $t + d$  as a feature vector

$(x_t^i, y_t^i, v_{x,t}^i, v_{y,t}^i, \dots, x_{t+d}^i, y_{t+d}^i, v_{x,t+d}^i, v_{y,t+d}^i)$ , where  $(x_\tau^i, y_\tau^i)$  and  $(v_{x,\tau}^i, v_{y,\tau}^i)$  are the location and velocity of dot  $i$  at time  $\tau$ . Dots are clustered based on the feature vectors. For Ncuts and K-means, the cluster numbers are chosen as 5 and 3 for the two datasets. Mean-shift automatically determines the cluster number. The clustering results are shown in Figure 6. The quantitative result is measured by the Normalized Mutual Information (NMI) [25]. Larger NMI indicates better clustering performance. The time interval  $d$  is set as 4, which means that the trajectory of each dot has six samples as the inputs of all the algorithms. Our algorithm achieves the best performance.

## 4.2 3D Motion Segmentation

There are many potential applications of our Coherent Filtering algorithm in real-world problems. We test it on 3D affine motion segmentation on the Hopkins155 Database [15]. We choose this application because its ground truth is available and it can provide quantitative evaluation of our technique. This database consists of 120 video se-

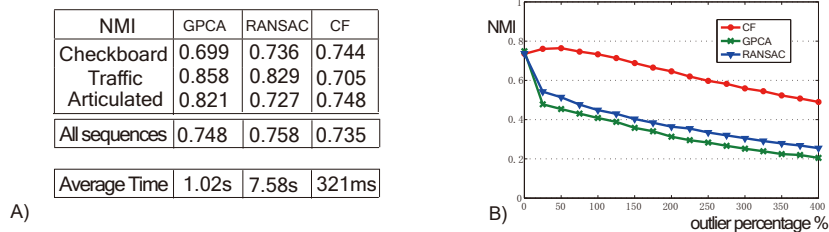


**Fig. 7.** A) Representative sequences from Hopkins155 database and the segmentation results of Coherent Filtering. B) The first image is one representative frame with groundtruth. There are 2 clusters, one cluster is on the moving object, the other cluster is on the static background objects, which results from moving camera. The second image is the segmentation result of our method. Since Coherent Filtering has no assumption on the number of clusters, it tends to segment some dispersed background cluster into several clusters of separated objects. Yellow + dots are the detected noises.

quences with two motions and 35 video sequences with three motions. Trajectories of feature points on each motion are clustered as ground truth and input for each testing method. The sequences can be categorized into three main groups, checkboard, traffic, and articulated, which contain a variety of motions, such as degenerate and non-degenerate motions, independent motions, and motions from camera movement. Figure 7A shows the representative frames of sequences in the database. For comparison, typical subspace motion segmentation methods, Generalized Principal Component Analysis (GPCA) [16] and RANSAC [15] are taken as the baseline. These approaches utilize the fact that object movements in this database are rigid and under affine transform. However, our method does not need the assumption since it is used to detect motions in more general form. Figure 8A shows the segmentation performance in terms of Normalized Mutual Information and average computation time.<sup>3</sup> Coherent Filtering achieves com-

<sup>3</sup> Codes of comparison methods are downloaded from authors' websites. Average computation time is tested on a computer with 3GHz Core Quad CPU and 4GB RAM with Matlab implementation. Note that since the number of clusters detected by our method may not accurately correspond to the ground truth cluster number, NMI is a more suitable measurement than the misclassification rate reported in [15].

parative performance to these subspace segmentation methods, though it is not specifically designed for 3D affine motion segmentation. The major error of our method comes from that our method has no assumption on the number of motion clusters in the sequence, so that it would tend to segment some dispersed background cluster of ground truth into several small clusters instead of one background cluster, as shown in Figure 7B. This problem is hard to be solved using our method alone without assuming the affine transform since these small clusters are far away to each other in space. Strictly speaking, they do not belong to a single coherent motion, but the database ground truth treats them as one pattern. To further evaluate the robustness of different methods in the

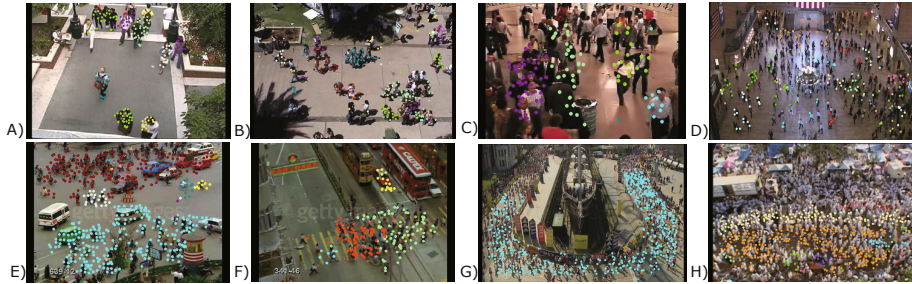


**Fig. 8.** A) NMI of different methods on the Hopkins155 Database, along with the average computation time. Though Coherent Filtering is not specifically designed for 3D motion segmentation, it achieves comparative performance to other subspace segmentation methods with a better computational efficiency. B) NMI of different methods as the function of the outlier percentage (from 0% to 400%).

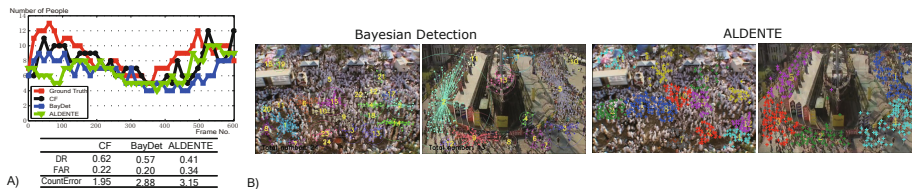
presence of outliers, we add outlier trajectories from 0% to 400% into the groundtruth trajectories of every 155 sequences, then test the performances of these methods. Outlier trajectories are generated by randomly choosing initial points in the first frame and then extending the trajectories with random walk. Figure 8B shows the NMI of these methods with different outlier percentages. Coherent Filtering works more robust with heavy noise.

### 4.3 Coherent Motion Detection in Crowd

Experiments are conducted on 8 video clips with coherent motion emerging in different scales and distributions. The first video clip is captured on the USC campus [4]. The scene is relatively sparse and the scales of pedestrians are high. The second one is from [26] with higher crowd density, and it contains both individual and groups of pedestrians. In the third one, many middle-sized people are coming in and out. The fourth is acquired from a far-view railway station. The resolutions of pedestrian are very small. And the last four clips are selected from Getty Image website, containing high-density crowd running and traversing. Some of them have been used in [7]. In initialization, KLT keypoint tracker [27] is used to automatically detect keypoints and extract their trajectories as the observation  $\mathcal{I}$  for the input of algorithm. Tracking terminates when severe clutters or occlusions arise, and new tracks will be initialized later. For all videos, the parameters of Coherent Filtering are  $\lambda = 0.6$ ,  $d = 4$ , and  $K = 10$ . We further discuss the influence of parameters on the clustering results in Section 4.4 .



**Fig. 9.** Representative frames and the coherent motion clusters detected by Coherent Filtering. Moving keypoints from videos exhibit a variety of coherent motion patterns at different scales in different scene context. A) The majority of detected coherent motion clusters result from the independent walking pedestrians, since the scale of pedestrians is rather big. B) Coherent motions from both individual pedestrians and groups of pedestrians walking together are detected. C) Different queues of walking-in-and-out people are detected. D) From the far view to the railway station, there are merely one or two keypoints tracked on each pedestrian in the scene. Thus the emergent coherent motions of keypoints represent the clusters of nearby pedestrians heading in the same directions, and they are related to different traffic modes. E) Two major lanes of vehicles on the road are detected, among them several small clusters representing jaywalkers are also detected because of their difference in motion directions to the major lanes. F) Two groups of pedestrians are detected to pass each other on the crosswalk. G) There is one circular coherent motion cluster detected as athletes running. H) The population density in the scene is extremely high, the detected coherent motion patterns characterize the dominant crowd flows. The crowd is separated into several bidirectional flows: the yellow flow is moving to the left, the orange flow is moving to the right, and the blue flow is moving against the orange flow dividing the orange flow of people.



**Fig. 10.** A) The number of pedestrians detected at each key frame with respect to Frame No., and Detection Rate(DR), False Alarm Rate(FAR), and counting error(CountError) for Coherent Filter(CF), BayDet[5], and ALDENTE[4]. B) BayDet and ALDENTE fail to detect the coherent motions when the crowdedness and the level of noise arise.

Figure 9 shows the representative frames and coherent motion clusters detected by Coherent Filtering in different scenes. Coherent Filtering detects well the underlying coherent motion patterns from the noisy time-series observations of detected keypoints. From these results, we can see that the detected coherent motion clusters correspond to a variety of semantic behaviors and activities, which are of great importance to further video analysis, surveillance and scene understanding. However, it is difficult to quan-

titatively evaluate these behaviors from detected coherent motions since it is hard to obtain the ground truth.

To quantitatively evaluate the detection performance, we conduct the people counting experiment on the scene shown in Figure 10, and compare with trajectory clustering-based people counting method ALDENTE [4] and Bayesian Detection counting method BayDet [5]. The experimental setting is the same as [5]. We count the number of pedestrian detected at each key frame (every key frame per 15 frames), and use the average people counting error as the evaluation criteria for the three methods. Meanwhile, since at each frame the detected clusters can be either True Positive or False Positive, and the False Positive also can be counted as the False Negative (the undetected one) in the number of pedestrians detected in each key frame. That makes people counting evaluation criteria not so accurate. Thus we further evaluate the Detection Rate ( $DR$ ) and False Alarm Rate ( $FAR$ ) as

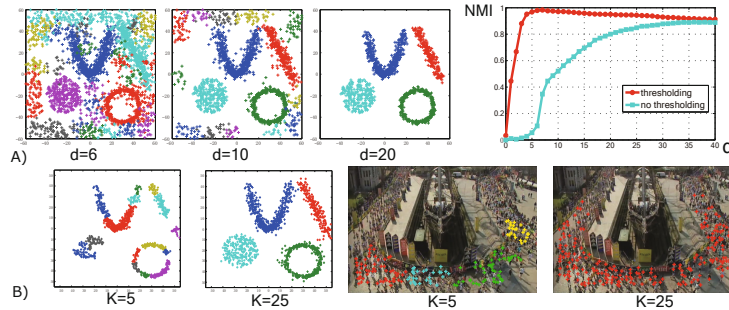
$$DR = \frac{\sum_i TP_i}{\sum_i GT_i}, \quad FAR = \frac{\sum_i FP_i}{\sum_i TP_i + FP_i},$$

where  $TP_i$ ,  $FP_i$ , and  $GT_i$  are the number of True Positive, False Positive, and groundtruth at frame  $i$ . Figure 10A shows the numbers of pedestrian detected by the three methods and the groundtruth at each key frames, and Figure 10B shows the average people counting error,  $DR$ , and  $FAR$  for the three methods respectively. Coherent Filtering achieves the best performance. On the other hand, ALDENTE and BayDet work poorly when the density of the crowd and the level of the noise increase. As shown in Figure 10C, they both fail to detect the coherent motions in crowded scenes.

#### 4.4 Further Analysis of the Algorithm

**Necessity of Two Filtering Steps in the Algorithm.** The algorithm **CoheFilterDet** of Coherent Filtering can be divided into two steps: first removing variant neighbors and then filtering out the neighbors with low averaged velocity correlations. As discussed in Section 2.2, the coherent neighbor invariance of spatiotemporal relationships does not guarantee that *all* the dots in  $\mathcal{A}$  are coherent dots, especially when  $d$  is small. Figure 11A shows the clustering results directly obtained from  $\mathcal{A}$  without thresholding when  $d = 6, 10,$  and  $20$  respectively, and the plot of Normalized Mutual Information (NMI) under different  $d$  with thresholding and without thresholding respectively. No thresholding, when  $d$  is small there remains a significant amount of noise. As  $d$  increases, NMI increases, which means the clustering performance is improved. Then NMI with thresholding and NMI without thresholding gradually converge. In principle, all the incoherent dots can be removed by setting a large  $d$ , such as when  $d = 20$ . However in practice, it is difficult to obtain the correspondence of dots over a long period. Thus filtering with thresholding the averaged velocity correlations on  $\mathcal{A}$  is necessary.

**Influence of  $K$ .**  $K$  decides the size of the neighborhood. Figure 11B shows the clustering results with  $K = 5$  and  $K = 25$  on the 2D synthetic data and the real data. When  $K$  is small, the detected coherent motion patterns are inclined to be divided into parts. However, when  $K$  is too large, some noise might be included. Thus the choice of  $K$  is related to the scale of coherent motion patterns to be detected in specific videos.



**Fig. 11.** A) The clustering results without thresholding when  $d = 6, 10, 20$ , and the plot of NMI under different  $d$  with thresholding and without thresholding. B) The clustering results on the 2D synthetic data and the real data with  $K = 5$  and  $K = 25$  respectively.

## 5 Conclusion

In this paper, we study the *Coherent Neighbor Invariance* as a prior of coherent motions and propose an effective clustering technique *Coherent Filtering*. Experimental evaluation and comparison on synthetic and real datasets show the existence of coherent neighbor invariance in various dynamic systems and the effectiveness of our technique under high-density noise. In the future work, we will study coherence neighbor invariance in a wider range of physical and biological systems and explore more potential applications.

**Acknowledgement.** This work is partially supported by the General Research Fund sponsored by the Research Grants Council of Hong Kong (Project No. CUHK417110 and CUHK417011) and National Natural Science Foundation of China (Project No. 61005057). It is also supported through Introduced Innovative R&D Team of Guangdong Province 201001D0104648280 and Shenzhen Key Lab of Computer Vision and Pattern Recognition.

## References

1. Couzin, I.: Collective cognition in animal groups. *Trends in Cognitive Sciences* (2009)
2. Moussaid, M., Garnier, S., Theraulaz, G., Helbing, D.: Collective information processing and pattern formation in swarms, flocks, and crowds. *Topics in Cognitive Science* (2009)
3. Camazine, S.: *Self-organization in biological systems*. Princeton Univ. Pr. (2003)
4. Rabaud, V., Belongie, S.: Counting crowded moving objects. In: *Proc. CVPR* (2006)
5. Brostow, G., Cipolla, R.: Unsupervised bayesian detection of independent motion in crowds. In: *Proc. CVPR* (2006)
6. Ali, S., Shah, M.: Floor Fields for Tracking in High Density Crowd Scenes. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part II*. LNCS, vol. 5303, pp. 1–14. Springer, Heidelberg (2008)
7. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: *Proc. CVPR* (2007)

8. Wang, X., Ma, X., Grimson, W.: Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Trans. on PAMI* (2008)
9. Zhou, B., Wang, X., Tang, X.: Understanding collective crowd behaviors: learning a mixture model of dynamic pedestrian-agents. In: *Proc. CVPR* (2012)
10. Zhou, B., Wang, X., Tang, X.: Random field topic model for semantic region analysis in crowded scenes from tracklets. In: *Proc. CVPR* (2011)
11. Wang, X., Tieu, K., Grimson, E.: Learning Semantic Scene Models by Trajectory Analysis. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3953, pp. 110–123. Springer, Heidelberg (2006)
12. Lin, D., Grimson, E., Fisher, J.: Learning visual flows: A lie algebraic approach. In: *Proc. CVPR* (2009)
13. Hu, M., Ali, S., Shah, M.: Learning motion patterns in crowded scenes using motion flow field. In: *Proc. ICPR* (2008)
14. Wang, X., Ma, K., Ng, G., Grimson, W.: Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. *Int'l Journal of Computer Vision* (2011)
15. Tron, R., Vidal, R.: A benchmark for the comparison of 3-d motion segmentation algorithms. In: *Proc. CVPR* (2007)
16. Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (gpca). *IEEE Trans. on PAMI* (2005)
17. Brox, T., Malik, J.: Object Segmentation by Long Term Analysis of Point Trajectories. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part V*. LNCS, vol. 6315, pp. 282–295. Springer, Heidelberg (2010)
18. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* (1981)
19. Petit, O., Bon, R.: Decision-making processes: the case of collective movements. *Behavioural Processes* (2010)
20. Lamme, V.: The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience* (1995)
21. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. on PAMI* (2000)
22. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on PAMI* (2002)
23. Liao, W., et al.: Clustering of time series data—a survey. *Pattern Recognition* (2005)
24. Morris, B., Trivedi, M.: A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* (2008)
25. Strehl, A., Ghosh, J.: Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research* (2003)
26. Pellegrini, S., Ess, A., Van Gool, L.: Improving Data Association by Joint Modeling of Pedestrian Trajectories and Groupings. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part I*. LNCS, vol. 6311, pp. 452–465. Springer, Heidelberg (2010)
27. Tomasi, C., Kanade, T.: Detection and Tracking of Point Features. *Int'l Journal of Computer Vision* (1991)