Unsupervised Motion Pattern Mining for Crowded Scenes Analysis

Chongjing Wang, Xu Zhao, Yi Zou and Yuncai Liu
Department of Automation and Key Laboratory of China MOE for System Control and Information Processing
Shanghai Jiao Tong University
Shanghai, China
[e-mail: {ccjj,zhaoxu,jbyiiiii,whomliu}@sjtu.edu.cn]

Abstract

Crowded scenes analysis is a challenging topic in computer vision field. How to detect diverse motion patterns in crowded scenarios from videos is the critical yet hard part of this problem. In this paper, we propose a novel approach to mining motion patterns by utilizing motion information during both long-term period and short interval simultaneously. To capture long-term motions effectively, we introduce Motion History Image (MHI) representation to access to the global perspective about the crowd motion. The combination of MHI and optical flow, which is used to get instant motion information, gives rise to discriminative spatial-temporal motion features. Benefitting from the robustness and efficiency of the novel motion representation, the following motion pattern mining is implemented in a completely unsupervised way. The motion vectors are clustered hierarchically through automatic hierarchical clustering algorithm building on the basis of graphic model. This method overcomes the instability of optical flow in dealing with time continuity in crowded scenes. The results of clustering reveal the situations of motion pattern distribution in current crowded videos. To validate the performance of the proposed approach, we conduct experimental evaluations on some challenging videos including vehicles and pedestrians. The reliable detection results demonstrate the effectiveness of our approach.

1. Introduction

Visually analyzing crowded scenes is an important research topic in computer vision field. It's becoming active in recent years because of its valuable potential applications. As shown in **Fig.1**, a crowded scene may include hundreds even thousands of objects, such as crowd, fauna, vehicles and so on. Crowded scenes commonly appear in our daily lives, such as subway station, supermarket, large gathering, etc. Due to the large number of people and complex situation, these scenes have become accident-prone areas. Many serious accidents had occurred in crowded areas around the world, for example, casualties in Mecca stampede in years of 1990, 1998 and 2004, pedestrian crowd accident in Beijing in 2004, terrorist incidents in public places in Bombay in 2011, and so forth. These accidents have caused huge losses. Therefore, capturing crowd dynamic is very important and meaningful to public security and emergency management.

However, in sharp contrast with the demands of real application, current techniques of visual surveillance on crowded level is still immature, due to the complexity and diversity of the crowded scenarios. With the increase of density and complexity of objects and scenes, clearly exploring the situations of crowded scene become more and more challenging. In summary, the main difficulties about crowded scenes lie in: 1) The effective features from single object are very hard to extracted because its size and resolution in images are too small to recognize clearly. 2) It's hard to track a single object due to severe occlusions and similar appearance in crowded scene. 3) The behaviors of single object are complicated because of the mutual influence and mutual restraint between objects and surrounding environment. 4) The structured or unstructured complicated scenes make the problem diverse. To overcome these difficulties, researchers are seeking new methods according to the specific properties of crowded scenes. Recently, many researchers pay attention to the topic of trajectory tracking, motion analysis, anomaly detection and scene understanding about crowded scenes. Motion pattern detection and analysis play central role in these studies. In the context of crowded analysis, motion pattern refers to a spatial segment of the scene, within which a high degree of local similarity in speed as well as flow direction exist but otherwise outside [24]. Motion patterns not only describe the segmentation in the spatial space, but also reflect the motion tendency in a period. It can present the tendency of the crowd motion at a semantic level.
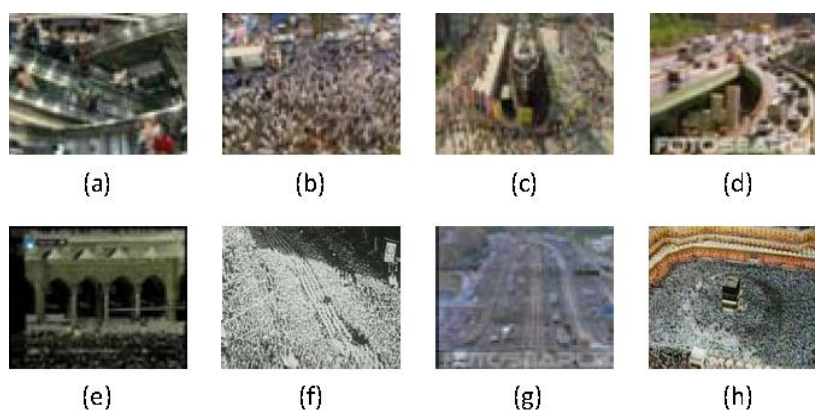

Fig. 1. Some samples of crowded scenarios.

In this paper, we propose a novel approach to detect motion patterns in dynamical crowded scenes. According to the theory of *Conformity Effect* in psychology and *Energy Minimization* in kinetics, the individuals in crowd keep homoplasy in motion field, and they perform with same or similar properties in the same motion pattern. Based on this theory and the visual observation that the information of a single object is difficult to obtain, we focus on detecting and mining motion patterns from global perspective of the crowd. We ignore the behaviors and activities of individuals which are unnecessary to explore in crowded scenarios. Accordingly, we make the following contributions on motion pattern mining for crowded analysis: 1) We introduce Motion History Image (MHI) for the first time into the research of crowd analysis. Motion estimation based on MHI presents cumulative motion of the objects in a period. It is effective and robust for dynamical crowd video processing. The combination of MHI and optical flow, which is used to get instant motion information, gives rise to discriminative spatial-temporal motion features. 2) In an unsupervised manner, we utilize the latest hierarchical automatic clustering approach to detect motion patterns successfully. The algorithm is based on graphic model. We apply it into our research and design reasonable clustering criteria for this problem. 3) We conduct experiment evaluations on the proposed approach and achieve satisfactory performance. The experiments show reliable and robust mining results. Comparisons with other state-of-the-art approaches demonstrate the effectiveness of our proposed approach.

The rest of our paper is organized as follows. In section 2, we introduce the related work on crowd analysis and motion pattern detection. The details of our approach are discussed in section 3 and 4. Section 5 presents the experimental evaluations on diverse crowded scenes. Finally in section 6, we make the conclusions.

2. Related Work

Crowd dynamic analysis is attracting increasing attention recently. A large amount of methods for crowd analysis have been proposed recently. T. Zhao et al. [1] perform people tracking in crowded scenes by modeling the human shape and appearance as articulated ellipsoids and color histograms respectively. This approach is one of the first algorithms for tracking in crowded scenes. N. Vaswani et. al [2] analyze the motion of all the moving objects by learning the temporal deformation through connecting the locations of the objects in consequent frames. Z. Khan et al. [3] use Markov chain Monte Carlo based particle filter to handle the interactions between targets in a crowded scene. They propose a notion that the behaviors of objects are influenced by the neighborhood of the objects. X. Wang et al. [4] propose a similarity measure for trajectories cluster, and then learn the scene models from the cluster. Brostow et al. [5] describe an unsupervised Bayesian clustering method to detect individuals in crowd, the detection is obtained separately for each frame, ignoring the relationship between frames. An adaptive active tracking system with sector based scanning for a single PTZ camera is proposed by Shung Han Cho et al. [33]. All the methods proposed above require that the scenes where objects are not moving densely and the tracking results of objects are available. However, these traditional methods seem to fail in dealing with high density crowded scenes.

To overcome the limitations of traditional methods, researchers are studying new methods according to specific properties of crowd scenes. For this complicated problem of crowded analysis, feature selection and extraction are particularly important. Reasonable features greatly enhance the effectiveness of the research. Generally, these methods can be divided into two categories: physical features based methods and vision features based methods.

Some physical features in other disciplines are introduced to solve our problem. The steakline representation and potential functions in fluid dynamics are discussed to illustrate the crowd movement by R. Mehran et al. [6]. This representation can quickly recognize temporal changes in a sequence, and make a balance between recognition of local spatial changes and filling spatial gaps in the flow. Chaotic invariant features are introduced by Shandong Wu et al. [7] into the crowd context to characterize complicated crowd motion. The method succeed in detecting and localizing anomalies. The force flow in socio-psychological studies is brought to investigate the crowd movement by R. Mehran [8]. The social force model capture the impact of environment on pedestrian in crowd scenes. These approaches should seek the relationship between the vision features and physical features, and integrate the physical features to the framework of visually analyzing crowded scenes.

Vision features based methods try to extract the information about appearance and motion. Antoni B. Chan et al. [9][10][11] present dynamic texture as a suitable representation for both the appearance and dynamics of visual processing to model, cluster, and segment videos. The dynamic texture proposes a generative model which presents a sequence as a collection of spatio-temporal stochastic layers with appearance and dynamics. Kratz et al. [12][13] propose a novel probabilistic method by extracting the local spatial-temporal gradient features to exploit the inherent varying structured motion pattern.This method is ineffective for significant appearance variations and severe occlusion.

Most approaches with motion features stem from the optical flow features. Ali et al. [14] segment coherent crowd flows in video segmentation by using a mathematically exacting framework based on Lagrangian Particle Dynamics. This method is incapability to segment incoherent motions. Ali et al. [15][16] also track persons in high density of crowd scenes by analyzing floor fields of the scene. This method is suitable for structured scenes, heavily dependents on the physical properties of the scene. The features they extracted are extension of optical flow features. Shobhit Saxena et al. [17] model different crowd events for specific end-user scenes. Min Hu et al. [18] propose a method to learn motion patterns by clustering vectors in motion flow field which is a collection of optical flow vectors. And Min Hu et al.[19] also detect the sink of representative modes to video matching by constructing super tracks based on optical flow vectors advection. Imran Saleemi et al. [20] propose to learn dense pixel to pixel transition distributions using tracking trajectories. It is used to detect abnormal events and segment motion foreground from background. X.Wang et al. [21][22] propose a unsupervised learning framework with hierarchical Bayesian models to model activities and interactions in crowded traffic scenes and train station scenes. A Random Field Topic model is proposed for semantic region analysis in crowded scenes [23], the method analyze the tracklet instead of optical flow or trajectories for learning semantic regions. C. wang et al. [25] extract motion features based on Motion History Image, and then detect motion patterns in dynamic crowd scenes. Imran Saleemi et al. [24] propos a mixture Gaussian model representation of salient pattern of optical flow, and learn the patterns through a hierarchical and unsupervised method.

3. Motion Pattern Representation

According to the Gestalt theory of human visual perception, the main factors used in grouping are proximity, similarity, closure, simplicity and common fate (elements with same moving direction are seen as a unit) [18]. In terms of this definition, " Motion pattern " in our research means the spatial-temporal segmentation in a video. Within the segmentation, the local speed

is similar and the direction of motion is proximal or changed smoothly. Motion patterns not only describe the segmentation in the spatial space, but also reflect the movement tendency in a period. In crowded scene, traditional optical flow feature does not capture the long-range temporal dependencies, since it is based on instantaneous information, and it can not describe spatial and temporal features of a flow by itself. The Motion History Image (MHI) was used in human action analysis a few years ago. It captures the the nature of the movement in a period. In this paper, we employ the MHI for the first time into the research about crowd analysis. We consider MHI combined optical flow features as spatial-temporal features to represent the motion patterns. This feature reflects the continuity of motion with time and keep the speed information meanwhile. In this section, we present the representation of motion pattern. We review the MHI in subsection 3.1 and the optical flow algorithm we have employed in subsection 3.2, and then present the detailed algorithm for our method in subsection 3.3.

3.1 Motion History Image

Motion History Image is a real-time motion template that temporally layers consecutive image differences into a static image template [26][27][28]. MHI is calculated as a scalar-valued image, the pixel intensity is a function of the temporal motion history at that point, where more recently moving pixels have brighter values. Traditional method such as Gaussian mixture model is unable to meet the requirements to obtain the moving crowd, because the individuals in foreground are integrated into the background over time. In the research about crowded scenes, the persistent and stable motion is the focus of our attention. The influence among individuals, such as self-occlusion, similar appearance, and so on, have little impact on global motion. MHI can encode a wide range of movement, and represent how or where the image is moving. In MHI, previous movement is weakened over time, while the noise caused by the interaction between individuals has also been faded away. It describes the current movement robustly, and reflects more object moving temporal information over a video sequence efficiently. As MHI can highlight long-term motion, the feature perform well in describing the persistent motion. However, MHI is insensitive to short-duration motion, the feature may not work well in the research of abnormal detection where a quick response is required.

A MHI is calculated by adding simple replacement and decay term as in [28]. At the time instant t and the location (x, y), the MHI is generated as following:

$$MHI_\tau(x,y,t) = \begin{cases} \max(0, MHI_\tau(x,y,t-1)-d) & \text{if } D(x,y,t) < T \\ \tau & \text{if } D(x,y,t) \geq T \end{cases} \qquad (1)$$

where $D(x,y,t)$ is the absolute frame difference with different distance, d is the decay term and T is the threshold. The MHI at time t is determined by the previous MHI and current motion image. Because the gray value interval calculated here is within [0, 255], so we also define the decay term d within [0, 255]. The constant motion generated by moving objects is highlighted while the old objects motion and noise fade away due to the decay term. **Fig. 2** displays the MHI of some videos, of which one sample frame are presented in **Fig. 1(a) - (d).** The time instants in each row are every 10 frames from **Fig. 2(a) - (e)**. We can see that the persistent crowd motion on marathon is highlighted with the increase of time. Meanwhile, the noise and tiny movement are wakened. MHI helps to reduce the disturing factors, and improve the performance of the algorithm about motion estimation.
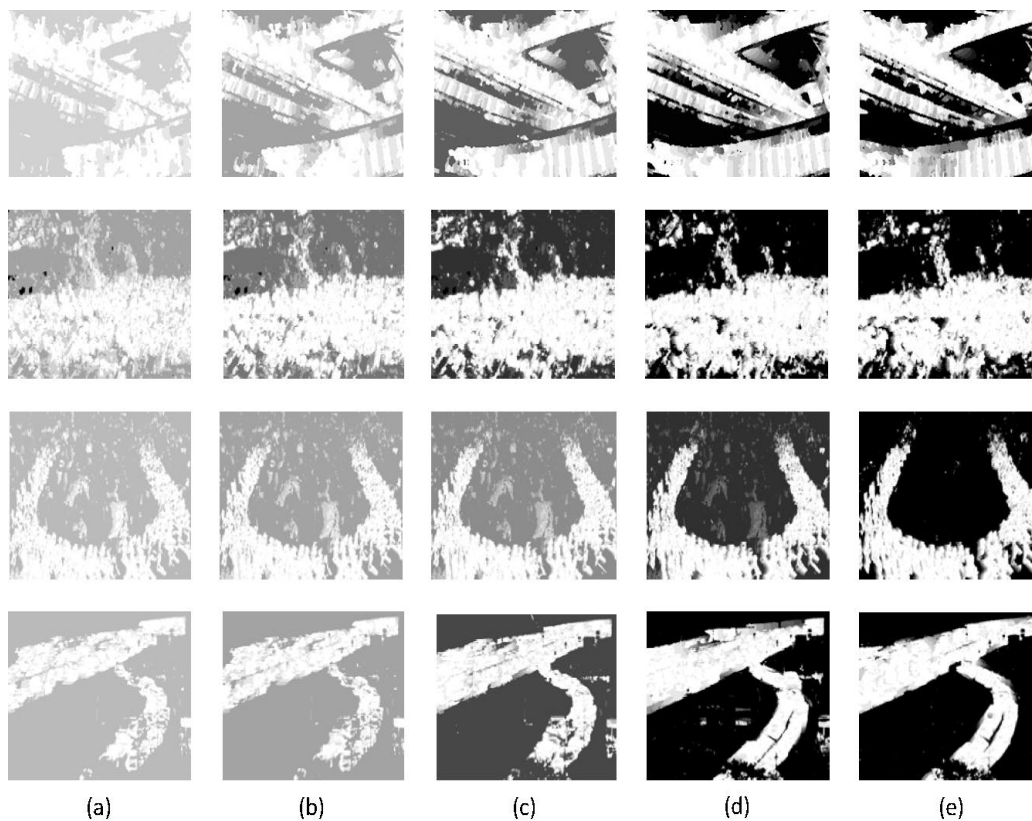
<center>(a)   (b)   (c)   (d)   (e)</center>

Fig. 2. MHI at different time instant.

3.2 Optical Flow Estimation

In the videos with single object and simple background, the direction information of movement can be measured by intensity gradients of the MHI. However, in crowded scene, the intensity gradients of persistent movement do not make sense. We need a method to acquire the directions of the crowd movement. Inspired by the classical work on motion feature extraction, we use the latest optical flow algorithm to estimate the motion directions.

The field of optical flow estimation is making steady progress as evidenced by the increasing accuracy of current methods on the Middlebury optical flow benchmark [29]. D. Sun et al. [30,31] developed a method which ranks at the top of the Middlebury benchmark.

The classical method of optical flow in spatially discrete describes as following:

$$
\begin{aligned}
E(u,v) = \sum_{i,j} \big\{ \; & \rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\
& + \lambda \, [\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) \\
& + \rho_S(v_{i,j} - v_{i+1,j}) + \rho_S(v_{i,j} - v_{i,j+1})] \; \big\}
\end{aligned}
\tag{2}
$$

where $I(i,j)$ is the pixel in the image, u and v are the horizontal and vertical speed in optical flow field from image I_1 and I_2, λ is scale factor, ρ_D and ρ_S are the penalty functions of data space and spatial field respectively. Base on the classical method, modern optimization

and implementations are incorporated into the flow fields to improve the accuracy.

$$
\begin{aligned}
E_A(u,v,\hat{u},\hat{v}) = & \sum_{i,j} \{ \ \rho D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\
& + \lambda \ [\rho S(u_{i,j} - u_{i+1,j}) + \rho S(u_{i,j} - u_{i,j+1}) \\
& + \rho S(v_{i,j} - v_{i+1,j}) + \rho S(v_{i,j} - v_{i,j+1})] \ \} \\
& + \lambda_2 (\|u-\hat{u}\|^2 + \|v-\hat{v}\|^2) \\
& + \sum_{i,j} \sum_{(i',j') \in N_{i,j}} \lambda_3 (|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|)
\end{aligned}
\tag{3}
$$

where $\hat{u}$ and $\hat{v}$ indicate an auxiliary flow field, $N_{i,j}$ is the neighbors of pixel (i,j), and λ_2 and λ_3 denote the scalar weights.

The method described above is employed in our research. Contrast to the tradition method of optical flow, the accuracy has been largely improved，because the performance of the optical flow in a large neighborhood is improved by incorporating median filtering of intermediate flow field and alternating optimization at every pyramid level. An adaptive weight is introduced into the non-local term. The weight in this paper is defined according to spatial distance, color-value distance and occlusion state.

As shown in Fig. 3, the optical flow vectors calculated by traditional method present multiple and scattered directions. However, the directions of optical flow vectors indicated by this method are in similar orientation, the performance therefore is highly promoted.
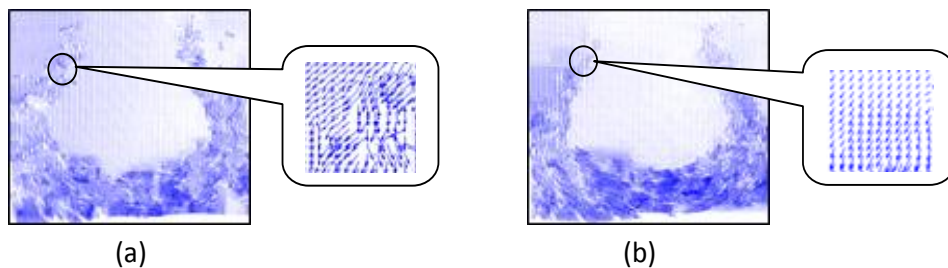

(a)   (b)

Fig. 3. Different methods of optical flow. (a) Traditional optical flow vectors.
(b) Sun's Optical flow vectors.

3.3 Motion Estimation Based on MHI

In crowded scenes, as optical flow algorithm is sensitive to the interaction between small individuals, much noise is produced. The optical flow vectors are chaotic and messed, the algorithm is not robust by directly employing optical flow algorithm. In order to overcome this deficiency, we propose the method of estimating motion by optical flow based on MHI, which has reduced the noise efficiently. The global spatial-temporal features of the crowd are extracted to describe the motion pattern.

MHI represents the temporal information of the moving objects. It integrates the individual movement into global movement and can be applied to the complex background. Optical flow vectors reflect the spatial information of the movement. The combination of the both reflects the spatial-temporal properties. The motion vectors are computed based on MHI, not based on original images. The method avoids much noise caused by individuals in crowd and reflects the global crowd motion property. It also helps to improve the performance of the following detecting algorithm.
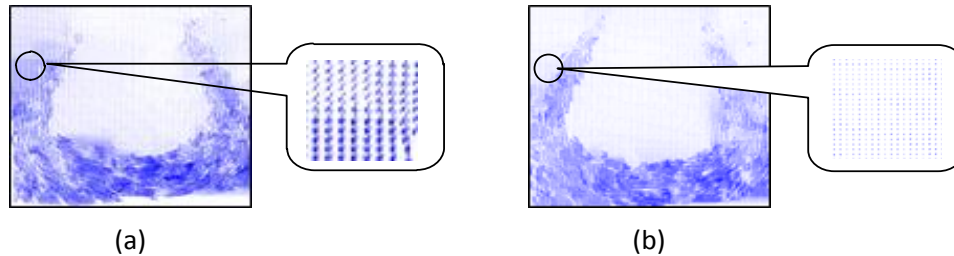
Fig. 4   Noise in optical flow field. (a) Optical flow vectors generated from the original image. (b) Motion vectors based on MHI.

In **Fig. 4(a)**, much noise is generated from the original images because of the complex scene and interferes among individuals. And motion vectors based on MHI in **Fig. 4(b)** reflect the global information of the movement, not dependent on the complex scenes. **Fig. 5(a)** shows a more complex and high density crowded scene in a large market. Thousands of people and several buses move in the market. Individuals can't be discriminated due to the similar appearance, large amount, and small size. It is difficult and unnecessary to distinguish individuals. **Fig. 5(b)** and **Fig. 5(c)** show the result of MHI and motion estimation based on MHI. The constant and obvious motions are highly lightened and the still or tiny motions are faded away in **Fig. 5(b)**. Two main motions with opposite directions are presented by motion estimation in **Fig. 5(c)**.
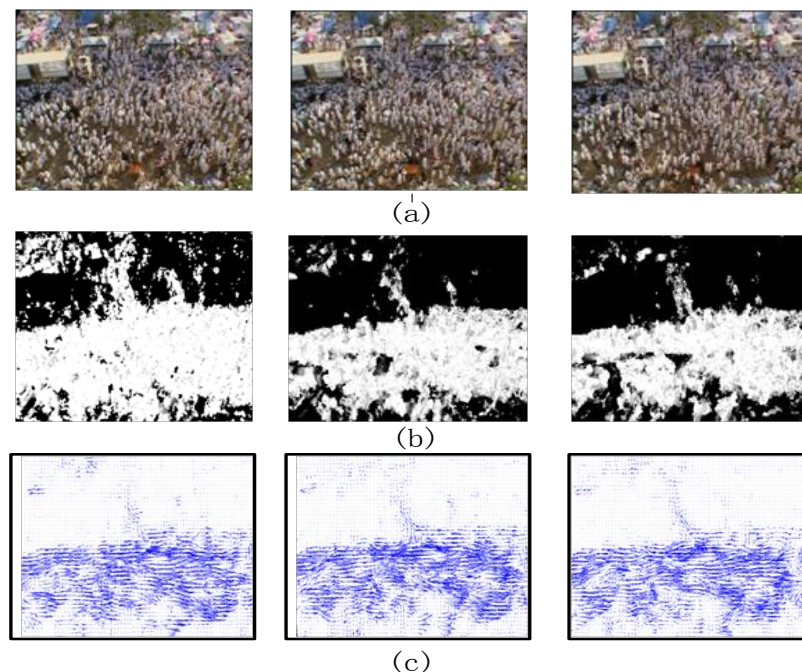

Fig. 5. Results of market video. (a) Original images. (b) MHI. (c) Motion estimation based on MHI.

4. Motion Pattern Mining

Within the same motion pattern, the local speeds are similar and the directions of motions are proximal or changed smoothly. Motion vectors calculated by the method in section 3.3 are four-dimensional including the location and velocity in horizontal and vertical directions. Mining motion pattern means to cluster the motion vectors with similar properties. In this section, we present the mining algorithm of motion pattern. Benefitting from the robustness and efficiency of the novel motion representation, the following motion pattern mining is implemented in a completely unsupervised way. We extend a novel hierarchical clustering algorithm to the vector field building on the basis of graphic model, and the reliable similarity measure makes the clustering method effective and robust.

4.1 Automatic Hierarchical Clustering Method

Because the number of motion patterns is unknown, and the data of motion vectors is very large, the traditional supervised clustering methods are unable to satisfy our requirement. Recently, a novel hierarchical clustering method base on graphic model techniques is proposed by Minsu Cho et al. [32]. The method computes the clusters automatically without pre-defined the number of clusters, and it is effective for large amount of data. Although the method originally aims at the point synthetic, we try to extend the idea to vector field to solve our problem.

The method considers the cluster problem as a node-labeling processing on graphs. The procedure of clustering is seen as the authority seeking on graph, which traversing on each node iteratively by transition matrix until searching high authority scores.

A relational graph $G = (v, \xi, w)$ with nodes v, edges ξ and weights w is given based on a pairwise relation between two components. Weight $\omega \in w$ is related to the edge between the node i to node j, reflecting the similarity between the nodes. PageRank vectors $PPR(i)$ are generated, which make a probabilistic landscape of the authority score around the node i. The hierarchical scheme is performed to recursively aggregate nodes in each cluster. The authority-shift procedure is performed iteratively for each PPR propagation until the n_{th} order PPR converges.

PPR creates the authority scores with respect to the specific nodes, and has been used in topic-sensitive web search based on the user personalization[32]. The PageRank vector r satisfies the following equation:

$$r^T = \alpha r^T P + (1-\alpha) v^T \tag{4}$$

where p refers to the transition matrix that follows the structure of the relational graph, and v expresses the probability distribution that the walker occasionally jumps from one node to a node. PPR implies the authority score vector that specific nodes weighted by the vector v, thus measuring the importance of each node is in relation with the other nodes. PPR propagation is defined as:

$$PPR_{n+1}(i) = PPR(PPR_n(i)) \tag{5}$$

where the PPR vector is employed recursively for high-order personalization. Based on high-order PPR, the authority node to node i for each order is assigned by:

$$Auth_n(i) = \arg\max PPR_n(i) \tag{6}$$

Through the authority nodes, self-authority cluster can be described as the set of nodes, which has its authority node as a member of the set. If it could not be further divided into smaller self-authoritative clusters, a minimal self-authoritative cluster is derived.

Fig. 6 give an example of authority score graph. Each motion vector is defined as a node. The links between nodes reflect the relationship between motion vectors in the graph. The weight w discussed in section 4.2 reflects the similarity or proximity between the motion vectors. The graph G is presented by the weight matrix. The PageRank authority score can be computed by equation (4) in linear system formulation, and the minimal authoritative is obtained by shifting each node toward its authority score until it reaches the convergent node. This process is similar to mode-seeking in the algorithm of mean-shift. But different from the algorithm of mean-shift, the authority-shift only needs to be formulated once per node and no special stopping rule is required.
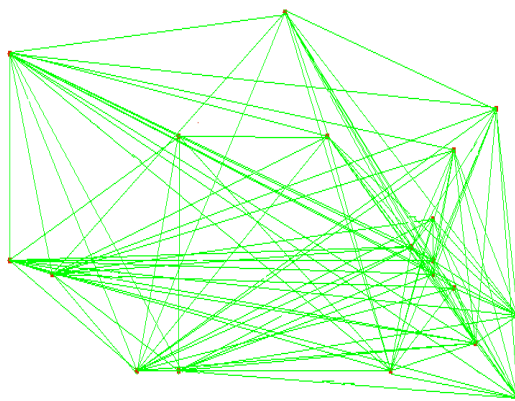

Fig. 6. An example of authority score graph. Each motion vector is defined as a node. The links between nodes reflect the relationship between motion vectors.

4.2 Similarity Measure

Similarity measure is the critical section in clustering. A good measure for clustering should satisfy the condition of intra-cluster similarity and inter-cluster dissimilarity. In the method introduced above, weight $\omega \in w$ denotes the similarity or proximity between node i and node j. In [32], Euclidian distance between point i and point j and RGB distance between pixel i and pixel j are treated as the weight function to cluster the point synthetic set respectively. However, the Euclidian distance or RGB distance are not suitable for crowded motion pattern detection. The similarity measure designed in our research should be related to location and velocity between motion vectors. Motion patterns constitute a highly nonlinear manifold. The result of clustering is not convinced by using the Euclidean distance measure on manifold space. This measure in our paper is more appropriate for the manifold space. And our similarity based on exponential function makes the clustering be in line with the condition of intra-cluster similarity and inter-cluster dissimilarity. The weight matrix W is designed as following:

$$W_{i,j} = \exp(\frac{-D(m,n)^2}{\sigma^2}) \ \ \text{for } m \neq n \text{ and } W_{m,m} = 0 .$$

$D(m,n)$ is the distance between motion vectors, which is calculated by the method in section 3.3. The motion vector is a four-dimensional vector $f = (x,y,u,v) = (P,V)$, where $P = (x,y)$ denotes the location and $V = (u,v)$ denotes the speed in horizontal and vertical direction. The cluster criterion is defined according to location and direction. Different from the method in

[18], the number of cluster in our method doesn't need to be pre-set. Our method can automatically cluster the motion vectors by considering the motion vectors as the nodes in hierarchical clustering. Reasonable similarity measure is given in [18].

The distance $D(m,n)$ between any two motion vectors m and n is designed as following:

$$D(m,n) = (d_P(m,n)d_D(m,n)) \tag{7}$$

where $d_P(m,n)$ is the position distance and $d_D(m,n)$ is the direction distance. The vectors m and n present the case of parallel or intersecting in optical flow field. The two cases are computed as following:

1) The vector m and n are parallel:

$$d_P(m,n) = \|P_m - P_n\| \tag{8}$$

$$d_D(m,n) = (\frac{2}{1+\xi+\bar{V}_m \cdot \bar{V}_n})^2 \tag{9}$$

where $\xi = 10^{-6}$ and $\bar{V} = \frac{V}{\|V\|}$.

2) The vector m and n are intersecting:

$$d_P(m,n) = \|P_m + V_m - P_n\| \tag{10}$$

$$d_D(m,n) = \frac{2}{1+\xi+\cos\theta_m} \cdot \frac{2}{1+\xi+\cos\theta_n} \tag{11}$$

$$\cos\theta_m = \bar{V}_m \cdot \overline{P_m - P_n} \tag{12}$$

$$\cos\theta_n = \bar{V}_n \cdot \overline{P_m - P_n} \tag{13}$$

where θ denotes the angle between the direction of motion vector and horizontal direction.

The final distance $D(m,n)$ is determined by the minimum of the distance in the above two cases so as to seek the maximum similarity between motion vectors. The algorithm of the motion pattern mining is described in Algorithm 1.

Algorithm 1: Motion Pattern Mining Algorithm

Input: Motion vectors based on MHI
Output: Clusters of motion patterns
1.   Compute the distance matrix D according to equation (7)
2.   Compute the weight matrix w
3.   Construct PPR matrix R
4.   Initialize the layers l
5.   **Loop**
6. $l \leftarrow l+1$
7.   Initialize the orders n
8.   **Loop**
9. $n \leftarrow n+1$
10.   Compute authority nodes
11.   Determine the clusters by authority node traversals
12.   Propagate PPRs

13.   **if** any authority node is shifted **then**
14.   Return
15.   **end if**
16.   **end loop**
17.   **If** all nodes lie in a single cluster **then**
18.   Return
19.   **end if**
20. **end loop**

---

5. Experiments

To test our method on mining motion patterns in dynamic crowded scenes, we conduct experiments on some challenging video clips in UCF_Crowds dataset and our SJTU_Crowds dataset. These videos include groups of moving people and vehicles with small size or low resolution of individuals. The backgrounds of the videos generally are very complex and the objects are in severe occlusions. We compared our method with the hierarchical clustering method based on classical optical flow algorithm. We also compare our approach with the method based on the motion flow field in [18] and the state-of-the-art approach in [6].

5.1 Motion Pattern Mining on UCF_Crowds Dataset

We evaluate our algorithm on UCF_Crowds dataset. This dataset contains videos of crowds，high density of moving vehicles, bio-cells under microscopes. These videos are collected mainly from the BBC Motion Gallery and Getty Images website. We conduct experiments on eight typical videos, which are usually used to evaluate the performance of the algorithm about crowded scenes analysis.

The experimental results are shown in **Fig. 7** and **Fig. 8**, in which the first row describes the original test videos, including different scenes in real world, such as, the indoor and outdoor scenes, structured and unstructured scenarios, people and vehicles crowds. The results of MHI are given in the second row, persistent and global motion is reflected over time. The third row shows the motion estimation based on MHI. The reliable motion information is extracted, and much noise is reduced. We made comparison of our method with three typical algorithms as shown in figures of 7 to 10. The results in the fourth row of the figures are conducted directly with the classical optical flow algorithm. The fifth row shows the results of the Min Hu's method. The motion flow field is a union of independent flow vectors computed in different frames [18]. The sixth row shows the results of the streakline method in [6]. The streakline gives a novel representation of the flow that recognizes spatial and temporal changes in the scene. Crowd segmentation is applied based on the similarity of the neighboring streaklines. The seventh row is the results of our method. From the figures, we can see our method can successfully detect the motion patterns that other methods can't. We can see that our method is in accordance with the ground truth in most cases. The results show that our method outperforms the state-of-the-art approach in mining motion patterns.
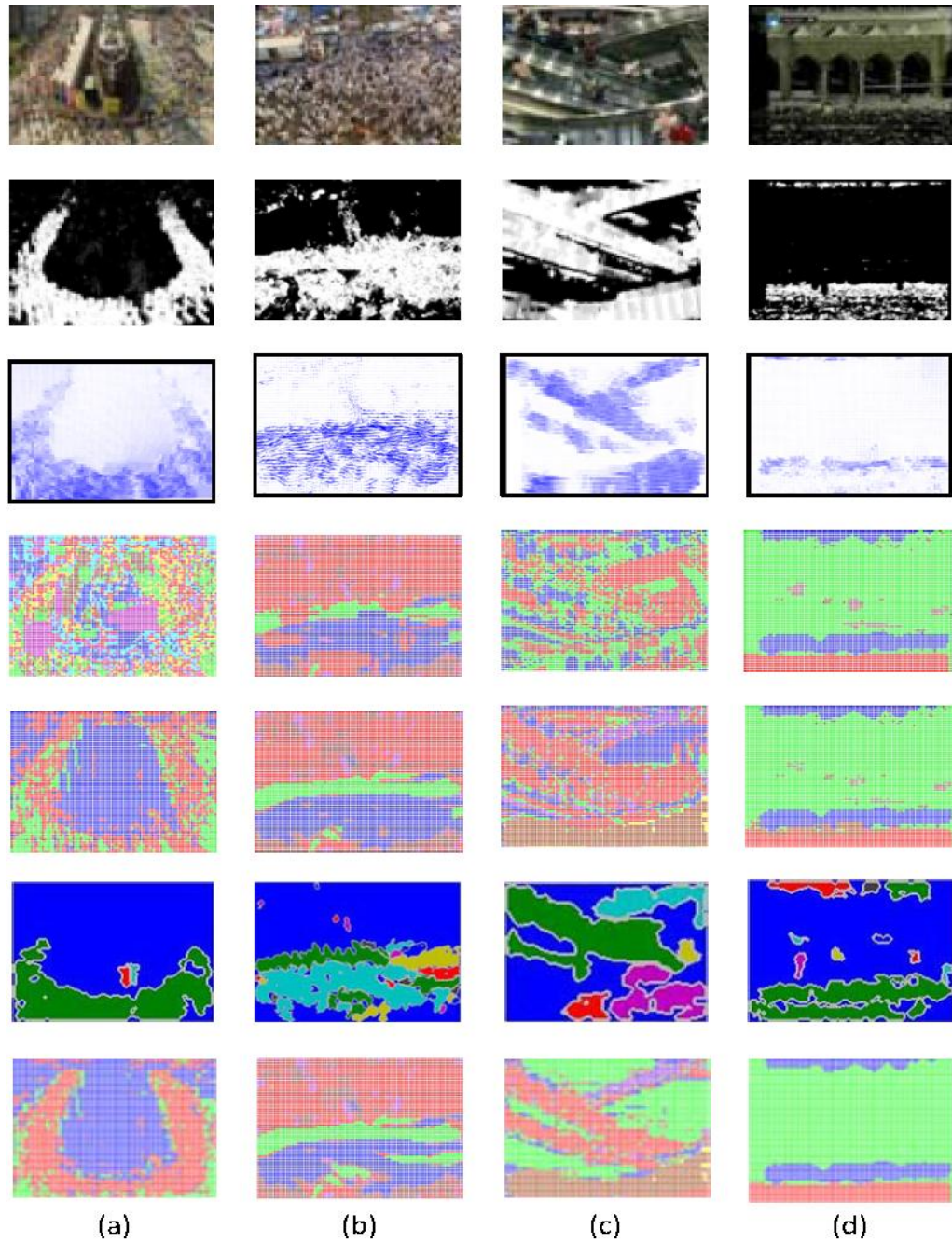
Fig. 7. Results on UCF_Crowds dataset. (a) The video about marathon. (b) The video about market. (c) The video about elevator. (d) The video about theater.
The first row: Original videos. The second row: MHI. The third row: Motion estimation based on MHI. The fourth row: The results based on classical optical flow algorithm. The fifth row: The results based on motion flow field. The sixth row: The results based on streakline. The seventh row: The results of our method.

Fig. 8. Results on UCF_Crowds dataset. (a) The video about traffic flow. (b) The video about high way. (c) The video about parade. (d) The video about hajj.

The first row: Original videos. The second row: MHI. The third row: Motion estimation based on MHI. The fourth row: The results based on classical optical flow algorithm. The fifth row: The results based on motion flow field. The sixth row: The results based on streakline. The seventh row: The results of
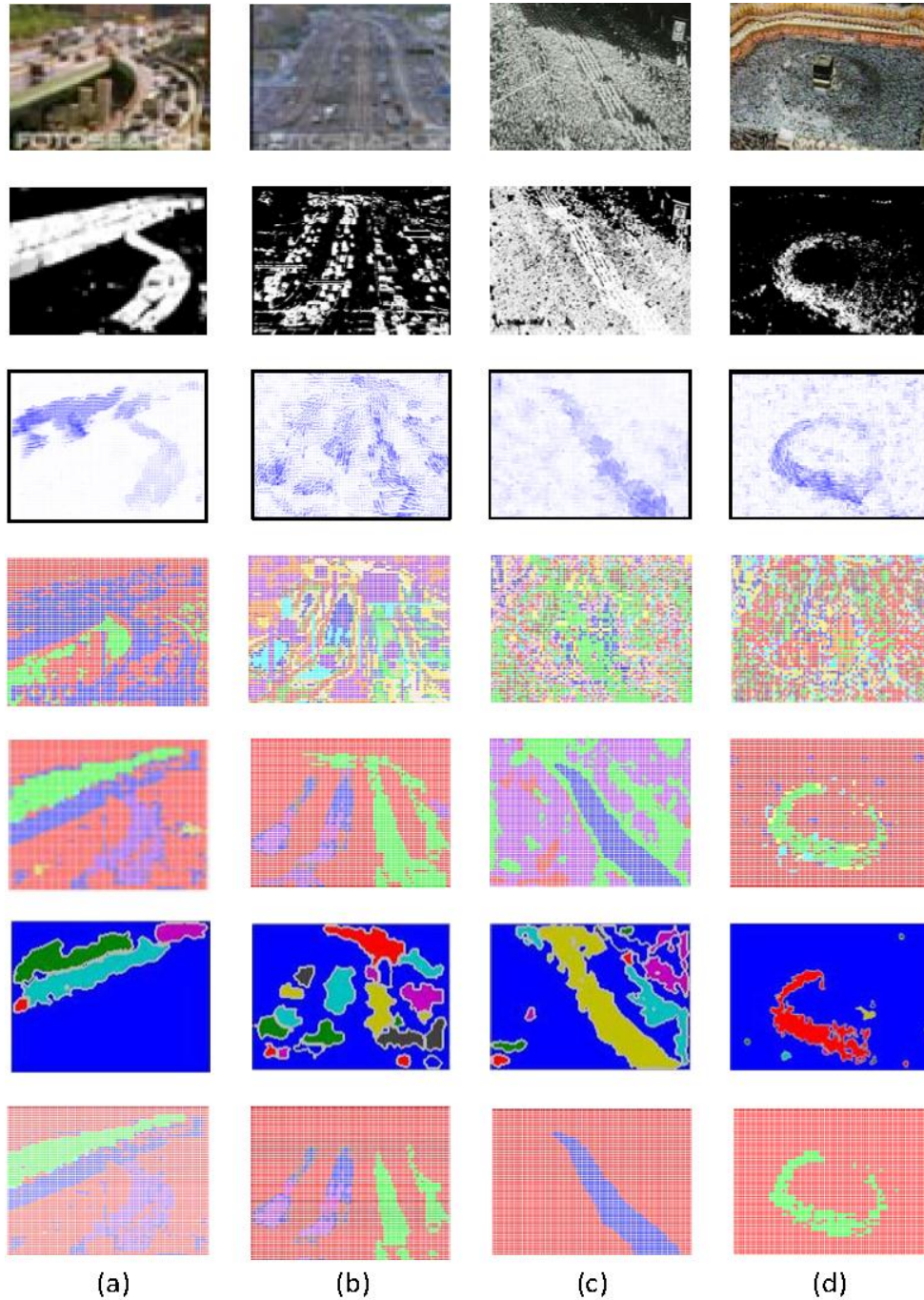
our method.

A marathon video in **Fig. 7(a)** describes the crowded scene which thousands of people are running in a "U-turn" road. In this scene, it is difficult to detect the motions due to the great similarity, intricate interaction among the individuals in the team and the variational directions at the position of "U-turn". The constant movement of the marathon team is highlighted and the noise is waded gradually in the second row of **Fig. 7(a)**. And then the motion estimation algorithm based on MHI in the third row of **Fig. 7(a)** reflects the available information of movement with less noise. It improves the performance tremendously of subsequent processing. The directions of individuals at the position of "U-turn" road change smoothly, and it should belong to the same motion pattern according to our definition of "motion pattern". The seventh row of **Fig. 7(a)** demonstrates the strength of our method, the global motion is obtained successfully. In this crowded scene, individuals at the boundary of the team have more freedom than those at the inside of the team, so scattered directions which present at the boundary result in some small motion patterns. The fourth row of **Fig. 7(a)** shows the result of the hierarchical clustering method based on classical optical flow algorithm, too many disheveled patterns caused by much noise are detected, and the main motion patterns are undetected. The fifth row of **Fig. 7(a)** shows the result of the hierarchical clustering method based on motion flow field, the main motion patterns become blurred. The movement in edges of the team is considered to be another pattern, not been designated as the dominate pattern. The sixth row provides the results from streakline method. It only detects and segments the motion patterns at the bottom of the image. The seventh row shows our results, all the motion patterns are successfully detected. It overcomes the difficulty of detecting uniform motion patterns at the position of "U-turn" with the mutative directions.

Fig. 7(b) shows a large market of complex and high density crowded scene. Thousands of people and several buses are moving in the market. Individuals can't be discriminated due to the similar appearance, extremely jam and small size. In this kind of scenarios, it is difficult also unnecessary to distinguish individuals. Two main motion patterns with opposite directions are detected by our method in the seventh row of **Fig. 7(b)**. To our excitement, the small movement which is surrounded by another large movement in right-bottom part of the figure is detected accurately. And the motion of bus in top part of the figure is also detected. The method based on classical optical flow or motion flow field and streakline method detect only one main motion pattern and some discontinuous movement.

A crowded scene in supermarket where crowds of shoppers take four elevators up and down is shown in **Fig. 7(c)**. It's a challenging scene because of severe occlusion and complicated background. The results of MHI in the second row of **Fig. 7(b)**, present that the temporal information of the persistent moving of elevators is lightened and the static objects, such as several static people sitting around a tea table and the shadows from the background, fade away with time. And also the optical flow estimation algorithm based on MHI in the third row of **Fig. 7(c)** reflects the spatial information, the noise is greatly reduced. The four motion patterns: two groups going up and two groups going down, are detected correctly and corresponded to colors in red, purple and brown in the seventh row of **Fig. 7(c)**. From the results, since the speed and direction are similar on the middle two adjacent lifts, the motion patterns merge together at one end of the elevator. However, it is impossible and necessary to discriminate them. Because of the severe occlusion and complicated background, the method based on classical optical flow algorithm does not detect the motion pattern correctly. Some parts of patterns are detected by the method based on motion flow field. And many unnecessary patterns are over segmented by streakline method.

A grand theater video is described in **Fig. 7(d)**, fans are moving slowly along the fence

upstairs and down stairs. The colors of moving objects and background are very similar, so the extraction of motion feature is significant for this scene. The motion features obtained by the motion estimation based on MHI are shown in the second row of **Fig. 7(d)**. The motion could be described clearly with a little noise. Three motion patterns are detected by the four approaches. We can see that some tiny movement produced by noise also detected by the method based on classical optical flow algorithm, motion flow field and the streakline method. Our method can detect all the motion patterns without noise.

As shown in **Fig. 8(a)**, three statuses of traffic flows which are formed by high density vehicles are detected correctly. In this video, three traffic flows with high speed run in the highway. It is hard to distinguish the two traffic flows with low resolution and opposite directions while the two lanes are very close and parallel. Our method detects the three motions marked in green, blue and purple colors successfully, and the method based on classical optical flow can't distinguish the two traffic flows. One traffic flow is merged into background by the method of streakline method.

A four-lane highway is shown in **Fig. 8(b).** It is difficult to discriminate the motion because the speed and direction are similar in the north of the highway and the street lamps block the movement. The main motion information is extracted by MHI and motion estimation with less noise, and the four motion patterns are detected by our method. The method based on the classical optical flow algorithm seems fail to detect motion patterns in this scene, and the methods based on motion flow field and streakline detect more tiny patterns in the north of the highway.

A parade video is described in **Fig. 8(c)**, thousands of citizens are watching the parade go past. **Fig. 8(d)** shows the scene of pilgerreise nach Mecca. Thousands of followers worship around the tower. In these scenes, it is difficult to detect the dominant motions because much noise is caused by the most people with small irregular movement. The method based on classical optical flow algorithm seems fail to detect in these scene because of much disordered motions. The methods based on motion flow field and streakline detect more tiny patterns. Our approach can separate the parade pattern and the hajj from the chaotic and irregular movement successfully.

5.2 Motion Pattern Mining on SJTU_Crowds Dataset

The SJTU_Crowds Dataset is designed to facilitate the research about crowd analysis. While the research about crowd analysis has become active in recent years, more data need to analyze and evaluate the state-of-art approaches. The UCF_Crowds dataset provides a good choice for the research. And SJTU_Crowds dataset is another choice which is different from the UCF_Crowds dataset in densities, motions, scenes, and so on. We collect the dataset in a square of Shanghai Jiao Tong University. The crowd videos are captured by a calibrated camera with a resolution of 1024 *768. The frame rates of the video system are 30 Hz. This database includes 40 sequences of dynamic crowded scenes. Each scene describes different motions of crowded people. The crowded people present three kinds of densities: 30, 50 and 100 persons, and they move in different velocities. These scenes include various motion patterns of crowded people, such as splitting, merging, intersecting, crossing, linear motion, curvilinear motion, circular motion, emergency collection, evacuation, and so forth. We will public this SJTU_Crowds dataset later.

We test our algorithm on our SJTU_Crowds Dataset. And we also compared our method with the hierarchical clustering method based on classical optical flow, motion flow field and streakline method. The results are shown in **Fig. 9** and **Fig.10**.
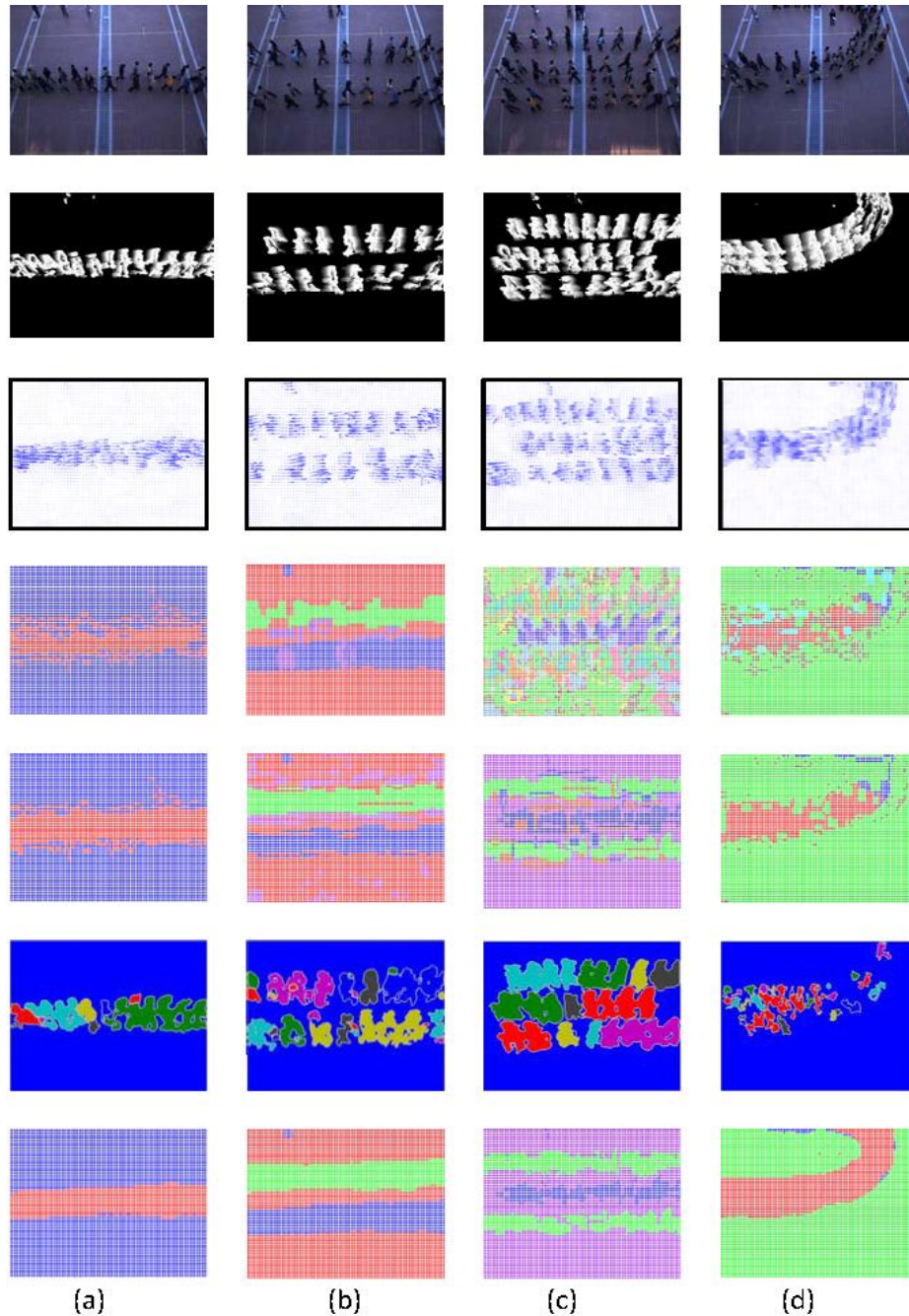
Fig. 9. Results on SJTU_Crowds dataset. (a) The video about one queue. (b) The video about two queues. (c) The video about three queues. (d) The video about one curved queue. The first row: Original videos. The second row: MHI. The third row: Motion estimation based on MHI. The fourth row: The results based on classical optical flow algorithm. The fifth row: The results based on motion flow field. The sixth row: The results based on streakline. The seventh row: The results of our method.

Fig. 10. Results on SJTU_Crowds dataset. (a) The video about one split queue. (b) The video about two curved queues. (c) The video about one circular queue. (d) The video about two circular queues. The first row: Original videos. The second row: MHI. The third row: Motion estimation based on MHI. The fourth row: The results based on classical optical flow algorithm. The fifth row: The results based on motion flow field. The sixth row: The results based on streakline. The seventh row: The results of
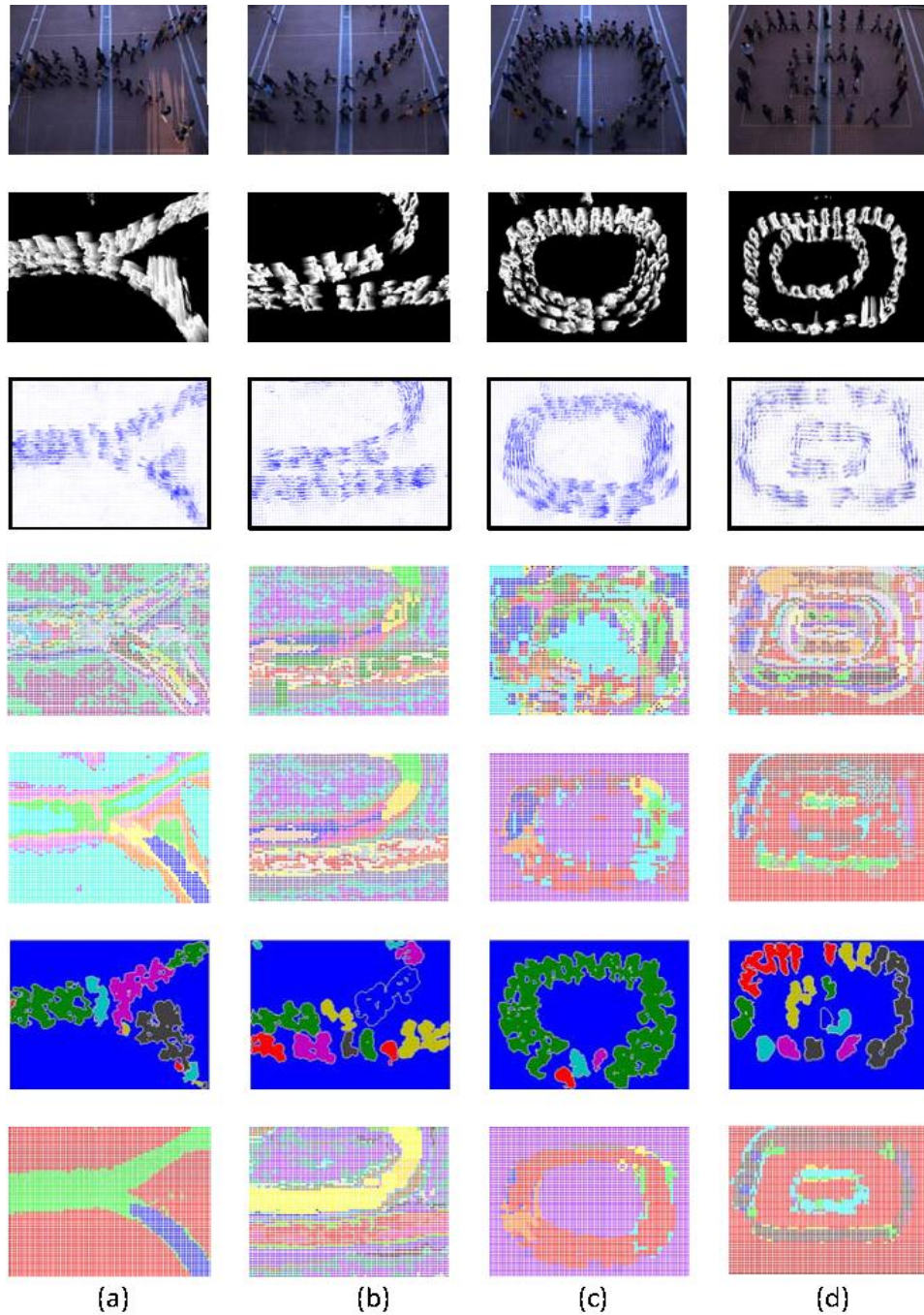
our method.

The videos of one queue, two queues, there queues and curved queue moving in the square are shown in **Fig. 9(a)-(d)**. As can be seen from the results, the MHI presents the information about persistent movement, and the noise is gradually waken. The motion estimation based on MHI reflects the main motion information, and insensitive to the individual movement and interaction between individuals. The motion patterns detected successfully by our method are shown in the seventh row of **Fig. 9**. It seems difficult to detect motion patterns while the individuals in different queues move close to others by the hierarchical clustering method based on classical optical flow algorithm. Much tiny unnecessary motion patterns are also detected by the method based on motion flow field. The streakline representation over segments the crowd scene.

Videos of **Fig.10** describe the curved motions, in which (a) shows a queue splits into two queues, (b) shows two queues moving in opposite directions with a queue moves in line and the another moves curvely. The information of main consistent motions is extracted by motion estimation based on MHI, and the motion patterns are detected correctly, in consistent with the situation of ground truth. However, the patterns are not detected by the classical optical flow algorithm. The motion patterns are over detected with much noise, and the persistent motion is blur by the method of motion flow. The contour of individuals is detected by the streakline method, forming the different and discontinous motion patterns.

The video of crowds moving in one circular queue is shown in **Fig. 10(c)**, and **Fig. 10(d)** describes crowds moving in two circular queues with opposite directions. Our method detects and distinguishes the two motion patterns clearly, but the method based on classical optical flow algorithm fail to detect the main motion patterns. The methods based on motion flow field and streakline method over detect the patterns with much noise.

Different methods present different performance in different crowded scenarios. Due to the complex motion and interaction in crowd, traditional optical flow features does not capture the long-range temporal dependencies. Much noise is produced, the directions of the motion vectors present chaotic and unordered case. Finding effective features is one of the key components in this motion pattern mining algorithm. In simple case where individuals move in one direction in line, the motion patterns detected by the hierarchical clustering method based on classical optical flow algorithm are noisy. In much more complex situation where the individuals move along curves, the method based on classical optical flow algorithm seems to fail in detecting the main motion patterns. As the redundant information and noise are reduced to some extent by the hierarchical clustering method based on motion flow field, the performance of mining motion patterns is prior to the method based on classical optical flow algorithm. In spite of this, there is still much noise to be detected as motion patterns in complex situation. Over-segment problems occur by using the method based on the streakline representation. Many whole motions are dispersed into a number of small motions. In general, our method mines the global motion information effectively, and our method consists with the ground truth in most cases. So our method is more robust in complex scenes.

To evaluate the performance of our approach, we compare our results with the ground truth and the method based on the classical optical flow, motion flow field and streakline method. The ground truth is manually marked from the videos. **Table 1** presents the comparison results on the number of main motion patterns.

Table1. The performance evaluation of our method

Video	GT	Optical Flow			Motion Flow Field			Streakline			Ours		
		#MP	P	R	#MP	P	R	#MP	P	R	#MP	P	R
marathon	1	0	0	0	4	0	0	2	0	0	**1**	**1**	**1**
market	6	5	0.2	0.17	3	0.67	0.33	7	0.43	0.5	**5**	**1**	**0.83**
elevator	4	0	0	0	6	0.33	0.5	5	0.4	0.5	**4**	**1**	**1**
theater	3	4	0.75	1	3	1	1	5	0.4	0.67	**3**	**1**	**1**
Traffic flow	3	3	0.33	0.33	3	1	1	3	0.67	0.67	**3**	**1**	**1**
high way	4	0	0	0	7	0.57	1	7	0	0	**4**	**1**	**1**
parade	1	0	0	0	3	0.33	1	3	0.33	1	**1**	**1**	**1**
hajj	1	0	0	0	2	0.5	0.5	1	1	1	**1**	**1**	**1**
one queue	1	1	1	1	1	1	1	2	0	0	**1**	**1**	**1**
two queues	2	5	0.2	0.5	2	1	1	4	0	0	**2**	**1**	**1**
three queues	3	0	0	0	5	0.4	0.67	6	0	0	**3**	**1**	**1**
one curved	1	0	0	0	2	0.5	1	0	0	0	**1**	**1**	**1**
one split	2	0	0	0	3	0.3	0.5	4	0	0	**2**	**0.5**	**0.5**
two curved	2	0	0	0	7	0.14	0.5	5	0	0	**3**	**0.3**	**0.5**
one circular	1	0	0	0	5	0	0	3	0.33	1	**1**	**1**	**1**
two circular	2	0	0	0	6	0	0	6	0	0	**3**	**0.33**	**0.5**

GT: Ground Truth.
#MP: the number of motion patterns which have been detected.
P: Precision.
R: Recall.

For quantitative evaluation of the mining, two quality measures: precision and recall are computed in our research. From the table, we can see that the results of our method are accordant with the ground truth in most cases. The higher precision and recall are obtained by our method. The great challenge exists in how to mine the motion patterns in the complicated crowded scenes, in which most individuals behave with random small motion, different motion patterns move close to each other, and even the directions vary in the same pattern. The adavantages of our method are significant. One major reason is that the spatial-temporal features we proposed can extract the global and reliable crowd motion information during the motion estimation process. Meanwhile, the noise and the tiny temporary motion can be weakened by using the spatial-temporal features. This helps to provide reliable features for the detecting process. Also, the similar motion is classified as the same motion pattern under the reasonable clustering criterion.

6. Conclusion

In this paper, we propose a novel unsupervised method to mine the motion patterns in dynamic crowded scenes. We introduce the MHI into the research of crowd analysis for the first time. The spatial-temporal features extracted by the motion estimation algorithm based on MHI reflect the global and persistent motion information, and then the motion patterns are mined through the automatic hierarchical clustering algorithm with the reliable cluster criteria. The experiments are conducted on some challenging videos including crowds of vehicles and pedestrians in UCF_Crowds Dataset and our SJTU_Crowds Dataset, and the results demonstrate the effectiveness of our method. We plan to search for more global properties and some special local information to improve the accuracy of the motion pattern mining, and further use the research for behavior analysis and scene understanding for crowded scenes.

References

[1]   T. Zhao, R. Nevatia, "Tracking multiple humans in crowded environment," in: *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, v*ol. 2, pp. II–406. Article (CrossRef Link).

[2]   N. Vaswani, A. Roy Chowdhury, R. Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects," In *Proc. of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. CVPR 2003*. Vol. 2, IEEE, 2003, pp. II–633. Article (CrossRef Link).

[3]   Z. Khan, T. Balch, F. Dellaert, "An mcmc-based particle filter for tracking multiple interacting targets," *Computer Vision-ECCV*, pp. 279–290, 2004. Article (CrossRef Link).

[4]   X. Wang, K. Tieu, E. Grimson, "Learning semantic scene models by trajectory analysis," *Computer Vision–ECCV* , pp.110–123, 2006. Article (CrossRef Link).

[5]   G. Brostow, R. Cipolla, "Unsupervised bayesian detection of independent motion in crowds," in: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.CVPR2006*. Vol. 1, pp. 594–601, 2006. Article (CrossRef Link).

[6]   R. Mehran, B. Moore, M. Shah, "A streakline representation of flow in crowded scenes," *Computer Vision–ECCV*, pp.439–452, 2006. Article (CrossRef Link).

[7]   S. Wu, B. Moore, M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in: *2010 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2054–2060,2010. Article (CrossRef Link).

[8]   R. Mehran, A. Oyama, M. Shah, "Abnormal crowd behavior detection using social force model," in: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 935–942, 2009. Article (CrossRef Link).

[9]   A. Chan, N. Vasconcelos, et al., "Layered dynamic textures," *Advances in Neural Information Processing Systems, 18 (2006)* 203. Article (CrossRef Link).

[10] A. Chan, N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *Pattern Analysis and Machine Intelligence*, vol.30, no.5, pp. 909–926, 2008. Article (CrossRef Link).

[11] A. Chan, N. Vasconcelos, "Layered dynamic textures," *Pattern Analysis and Machine Intelligence*, vol 31, no. 10, pp.1862–1879, 2009. Article (CrossRef Link).

[12] L. Kratz, K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1446–1453, 2009. Article (CrossRef Link).

[13] L. Kratz, K. Nishino, "Tracking with local spatio-temporal motion patterns in extremely crowded scenes," in: *IEEE Conference on Computer Vision and Pattern Recognition, 2010.*, pp. 693–700. Article (CrossRef Link).

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

[15] S. Ali, M. Shah, "Floor fields for tracking in high density crowd scenes," in: *European conference on computer vision*, Vol. 2, 2008, pp. 1–14. Article (CrossRef Link).

[16] S. Ali, Taming crowded visual scenes, Ph.D. thesis, University of Central Florida (2010). Article (CrossRef Link).

[17] S. Saxena, F. Br´emond, M. Thonnat, R. Ma, "Crowd behavior recognition for video surveillance," in: *Advanced Concepts for Intelligent Vision Systems, Springer*, 2008, pp. 970–981. Article (CrossRef Link).

[18] M. Hu, S. Ali, M. Shah, "Learning motion patterns in crowded scenes using motion flow field," in: *Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08)*, Citeseer, 2008. Article (CrossRef Link).

[19] M. Hu, S. Ali, M. Shah, "Detecting global motion patterns in complex videos," in: *19th International Conference on Pattern Recognition, 2008. ICPR 2008*. IEEE, 2008, pp. 1–5. Article (CrossRef Link).

[20] I. Saleemi, K. Shafique, M. Shah, "Probabilistic modeling of scene dynamics for applications in visual surveillance," *IEEE transactions on pattern analysis and machine intelligence (2008)* 1472–1485. Article (CrossRef Link)

[21] X. Wang, X. Ma, E. Grimson, "Unsupervised activity perception by hierarchical bayesian models," in: *IEEE Conference on Computer Vision and Pattern Recognition 2007. CVPR'07. IEEE,* 2007, pp. 1–8. Article (CrossRef Link).

[22] X. Wang, X. Ma, W. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on 31 (3) (2009)* 539–555. Article (CrossRef Link).

[23] B. Zhou, X. Wang, X. Tang, "Random field topic model for semantic region analysis in crowded scenes from tracklets," in: *IEEE Conference on Computer Vision and Pattern Recognition 2011. CVPR'11*. IEEE, 2011, pp. 3441–3448. Article (CrossRef Link).

[24] I. Saleemi, L. Hartung, M. Shah, "Scene understanding by statistical modeling of motion patterns," in: *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE,* 2010, pp. 2069–2076. Article (CrossRef Link).

[25] C. Wang, X. Zhao, Y. Zou, Y. Liu, "Detecting motion patterns in dynamic crowd scenes," in: *2011 Sixth International Conference on Image and Graphics, IEEE,* 2011, pp. 434–439. Article (CrossRef Link).

[26] A. Bobick, J. Davis, "The recognition of human movement using temporal templates," *Pattern Analysis and Machine Intelligence*, vol. *23,* no.3, 257–267, 2001. Article (CrossRef Link).

[27] J. Davis, "Hierarchical motion history images for recognizing human motion," in: *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on, IEEE*, 2001, pp. 39–46. Article (CrossRef Link).

[28] Z. Yin, R. Collins, "Moving object localization in thermal imagery by forward-backward MHI," in: *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, IEEE,* 2006, pp. 133–133. Article (CrossRef Link).

[29] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, R. Szeliski, "A database and evaluation methodology for optical flow," in: *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007*. IEEE, 2007, pp. 1–8. Article (CrossRef Link).

[30] D. Sun, S. Roth, J. Lewis, M. Black, "Learning optical flow," *Computer Vision–ECCV 2008* (2008) 83–97. Article (CrossRef Link).

[31] D. Sun, S. Roth, M. Black, "Secrets of optical flow estimation and their principles," in: *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,* 2010, pp. 2432–2439. Article (CrossRef Link)

[32] M. Cho, K. MuLee, "Authority-shift clustering: Hierarchical clustering by authority seeking on graphs," in: *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*, 2010, pp. 3193–3200. Article (CrossRef Link).

[33] Shung Han Cho, Yunyoung Nam, Sangjin Hong and Weduke Cho, "Sector Based Scanning and Adaptive Active Tracking of Multiple Objects," in: *KSII Transactions on Internet and Information Systems,* Vol. 5, Issue 6, 2011, pp. 1166–1191. Article (CrossRef Link).

Chongjing Wang received a B.S. degree from Xi'an University of technology, China, in 2005, an M.S. degree from Kunming University of Science and Technology, China, in 2008. She is currently a Ph.D. candidate in the Department of Electronic and Electrical Engineering in Shanghai Jiao Tong University. Her research interests include visual analysis of crowded scenes, machine learning, and image/video processing.

Xu Zhao received the Ph.D. degree in pattern recognition and intelligence system from Shanghai Jiao Tong University. He is currently an assistant Professor of Shanghai Jiao Tong University, China. He was a visiting scholar at the Beckman Institute for Advanced Science and Technology at University of Illinois at Urbana-Champaign from 2007 to 2008. His research interests include visual analysis of human motion, machine learning, and image/video processing.
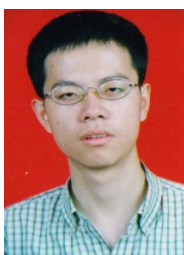
Yi Zou received a B.S. degree and an M.S. degree from Xiamen University, China, in 2006 and 2009.He is currently a Ph.D. candidate in the Department of Electronic and Electrical Engineering in Shanghai Jiao Tong University. His research interests include crowd analysis, computer vision and pattern recognition.

Yuncai Liu received the Ph.D. degree in electrical and computer science engineering from the University of Illinois at Urbana-Champaign in 1990. He worked as an associate researcher at the Beckman Institute of Science and Technology from 1990 to 1991. Since 1991, he had been a system consultant and then a chief consultant of research in Sumitomo Electric Industries Ltd., Japan. In October 2000, he joined the Shanghai Jiao Tong University, China, as a Distinguished Professor. His research interests are in image processing and computer vision, especially in motion estimation, feature detection and matching, and image registration. He also made many progresses in the research of intelligent transportation systems.