Analyzing Motion Patterns in Crowded Scenes via Automatic Tracklets Clustering

WANG Chongjing, ZHAO Xu, ZOU Yi, LIU Yuncai

Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China
Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

Abstract: Crowded scene analysis is currently a hot and challenging topic in computer vision field. The ability to analyze motion patterns from videos is a difficult, but critical part of this problem. In this paper, we propose a novel approach for the analysis of motion patterns by clustering the tracklets using an unsupervised hierarchical clustering algorithm, where the similarity between tracklets is measured by the Longest Common Subsequences. The tracklets are obtained by tracking dense points under three effective rules, therefore enabling it to capture the motion patterns in crowded scenes. The analysis of motion patterns is implemented in a completely unsupervised way, and the tracklets are clustered automatically through hierarchical clustering algorithm based on a graphic model. To validate the performance of our approach, we conducted experimental evaluations on two datasets. The results reveal the precise distributions of motion patterns in current crowded videos and demonstrate the effectiveness of our approach.

Key words: crowded scene analysis; motion pattern; tracklet; automatic clustering

I. INTRODUCTION

Visually analyzing crowded scenes is recently attracting much more attention from research community of computer vision field. It has been an important research topic because of its valuable potential applications. As shown in Figure 1, a crowded scene may include hundreds even thousands of objects, such as crowds, faunas, vehicles and so on. Crowded scenes arise commonly in our daily life, such as supermarkets, downtown streets, public celebrations, etc. Due to the large number of people and the complex situation, public safety in crowded places had been common concerns. Many accidents occurred in the past with regard to mass evacuations, political parades, mobs, or natural disasters have caused huge losses. Capturing crowd dynamic is becoming increasingly important and meaningful to public security and emergency management.

Conventional tracking methods that typically acquire static backgrounds or moving objects are limited to scenes with a few objects. However, with the increase in density and complexity of objects and scenes, clearly exploring the situations of crowded scenes becomes more and more challenging. Due to the complexity and diversity of the crowded scenarios, current techniques of visual surveillance on the crowded level are still immature. In summary, there exist several main difficulties in the research about crowded scenes. First, the effective features from single object are very hard to extract because of its small size and low resolution. Second, a single object is difficult to track due to the severe occlusion and similar appearance in crowded scenes. Third, the mutual influences and restraints between objects and the surrounding environment make the problem diverse. To

Received: 2012-12-15
Revised: 2013-01-30
Editor: YUAN Baozong

overcome these difficulties, researchers are exploring effective methods according to the specific properties of crowded scenes.

Recently, many researchers take great interest in the topic of crowd tracking, motion analysis, anomaly detection and scene understanding about crowded scenes. Zhao et al. [1] perform human tracking in crowded scenes by modeling the human shape and appearance as articulated ellipsoids and color histograms respectively. This approach is one of the algorithms firstly applied in tracking in crowded scenes. Khan et al. [2] 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. Hue et al. [3] detect interest points in each frame to describe the objects. Tracking is performed in all the objects by establishing correspondences among points between frames. Brostow et al. [4] describe an unsupervised Bayesian clustering method to detect individuals in a crowd, and the detection is obtained separately for each frame, ignoring the relationship between frames. Sugimura et al. [5] detect individual moving entities by assuming that the subjects move in distinct directions. 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 are in absence of ability 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. Ali et al. [6] segment coherent crowd flows in video segmentation by using a mathematical exacting framework based on Lagrange Particle Dynamics. Ali et al. [7-8] also track pedestrians in high density of crowd scenes by analyzing floor fields that describe how a pedestrian should move based on scene- wide constraints. This method is suitable for structured scenes, heavily depending on the physical properties of the scene. Rodriguez et al. [9] represent the local motion in different directions on each spatial location with a topical model. Saleemi et al. [10] propose to learn dense pixel to pixel transition distributions using tracking trajectories. It is used to detect abnormal events and the segment motion foreground from the background. The streakline representation and potential functions in fluid dynamics are discussed to illustrate the crowd movement by Mehran et al. [11]. 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. Wang et al. [12] propose an unsupervised learning framework with hierarchical Bayesian models of model activities and interactions in crowded traffic scenes and train station scenes. In Ref. [13], a Random Field Topic model is proposed for semantic region analysis in crowded scenes. The method analyzes the tracklet instead of optical flow or trajectories for learning semantic regions. Wang et al. [14] extract motion features based on Motion History Image, and then detect motion patterns in dynamic crowd scenes. Saleemi et al. [15] propose a mixture Gaussian model representation of salient pattern of optical flow, and learn the patterns through a hierarchical and unsupervised method. Zhou Bolei et al. [16] propose a new Mixture model of Dynamic Pedestrian-Agents to learn the collective behavior patterns of pedestrians in crowded scenes. They also characterize the local spatio-temporal relationship of individuals by coherent filtering to detect coherent motion patterns from noisy

We propose a method to analyze motion patterns in crowded scenes in a completely unsupervised way. The tracklets are captured by tracking dense points according to the characteristics of crowded scenes. Motion pattern analyzing is implemented by clustering tracklets automatically through a hierarchical clustering algorithm building on the basis of graphic model.

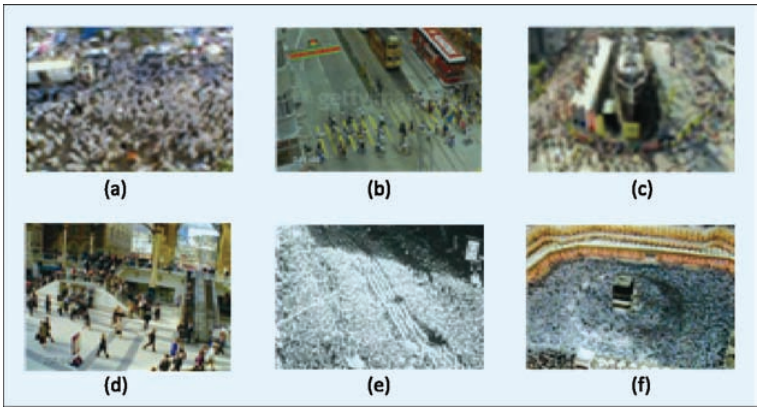

Fig.1 *Some examples of crowded scenes*

time-series data [17].

In above related works, motion pattern detection and analysis play an important and basic role. Under the context of crowded analysis, the 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 [15]. 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.

In this paper, we propose a novel approach to analyze motion patterns from tracklets 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, we focus on detecting and analyzing motion patterns from a global perspective of the crowd. Accordingly, we make the following contributions on motion pattern analyzing for crowded analysis: 1) we collect tracklets by tracking dense feature points from the video of crowded scenes. 2) Motion patterns are learned by clustering the Longest Common Subsequences (LCSS) of tracklets and the similarity between tracklets is measured thereby. 3) We conduct experiment evaluations on the proposed approach and achieve satisfactory performance. The experiments show reliable and robust results in analyzing motion patterns. 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 II and III, the details of our approach are discussed. The section IV presents the experimental evaluations on diverse crowded videos. Finally in section V, we make the conclusion.

II. Dense Point Tracking

Accurate and dense tracking from videos is an important requirement for crowded video surveillance. The most popular point tracker is Kanade-Lacas-Tomasi Tracker (KLT) [18]. The method generates an image pyramid, detects points which have obvious structural features, and tracks these points across frames. Generally this method is fast and accurate, but it is unsuitable in crowded scenes. Only a few points could be tracked due to the obscure structural features and severe occlusions.

It's hard to track an individual for a long period in crowded scenes as the inter and intra-object occlusions in crowded scene make the problem suffering from tracking drift and the situation become more serious over time. To deal with this case, we propose an efficient approach to extract tracklets in crowded scenes. Inspired by Sundaram and Wang [19-20], we propose three constraints to track the dense points according to the specific properties of crowded scenes. We collect the tracklets by tracking densely sampled points using optical flow fields. A tracklet is a fragment of a trajectory during a short period. Tracklets terminate when ambiguities caused by occlusion and scene mass arise.

The optical flow field $w_t = (u_t, v_t)$ is calculated by the classical LK optical flow algorithm, where u and v are the horizontal and vertical speed in optical flow field is from two adjacent images, and $w_{t+1} = (u_{t+1}, v_{t+1})$ denotes the flow from frame t forward to $t+1$.

A set of dense points is initialized at the beginning of the video. We initialize the points at every pixel because the flow field is dense in crowd video. Each point (u_t, v_t) at frame t is tracked forward to the frame $t+1$ by median filtering in a dense optical flow field.

$$(x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * w_t) \mid (\overline{x}_t, \overline{y}_t) \quad (1)$$

where M denotes the median kernel, and $(\overline{x}_t, \overline{y}_t)$ is the rounded position of (x_t, y_t). Global smoothness constraints M are employed among the points propagation, which is more robust than bilinear interpolation in Ref. [19], especially for points near motion boundaries.

In crowded scenes, severe occlusion occurs

frequently. Tracking has to be interrupted instantly as a point gets occluded. This is very necessary so as to avoid the point merging into another motion pattern. N. Sundaram et al. [19] propose two reliable rules to check the consistency of tracking by GPU-accelerated large displacement optical flow.

In a non-occlusion situation, the backward flow vector should keep in accordance with the inverse direction of the forward flow vectors. If this consistency constraint is not fullfilled, the tracked point may be get occluded at next frame or the optical flow vectors may not be estimated correctly. Both possibilities could construct good reasons to stop tracking this point at t.

$$\left|w+\hat{w}\right|^2 < \alpha_1(\left|w\right|^2+\left|\hat{w}\right|^2)+\beta_1 \qquad (2)$$

where $\hat{w}=(\hat{u},\hat{v})$ denotes the flow from frame $t+1$ back to t. As some small errors always occur in optical flow algorithm, a tolerance interval $\alpha_1(|w|^2+|\hat{w}|^2)+\beta_1$ is allowed to increase linearly with the magnitude of flow vector.

In crowded scenes, ambiguities arise at the boundary of the motion. The exact position of the motion boundary estimated by optical flow algorithm typically drifts a little. This phenomenon results in the same consequences as the occlusion: a point drifts to the side of another motion boundary and mixes in different motions. To avoid this defect, the third rule is proposed to stop the tracking:

$$\left|\nabla u\right|^2+\left|\nabla v\right|^2 > \alpha_2\left|w\right|^2+\beta_2 \qquad (3)$$

where ∇u and ∇v describe the divergence of the optical flow vector. Crowd motion is similar to the motion of fluid, so divergence is employed to limit the dispersion of the crowd.

The above three rules are established to track the dense points in crowded scenes. The dense tracklets have better quality than KLT tracker. The length of the tracklet is always short, but more reliable to reflect the ground truth of the crowd motion. This helps to reduce the disturbing factors in tracking, and improve the performance of the algorithm about motion pattern analyzing. Figure 2 shows the tracklets in crowded airport. Original images are described in Figure 2 (a-d), and tracklets obtained by KLT algorithm and our method respectively are shown in Figure 2 (e-h) and Figure 2 (i-l). Obviously, the tracklets obtained by

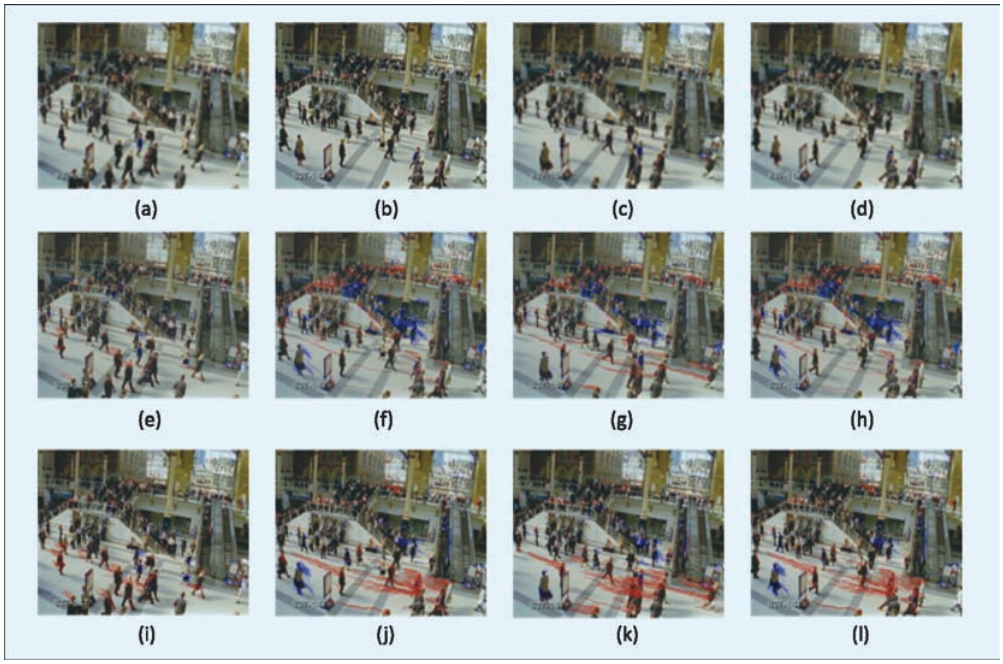

Fig.2 *Tracklets in the video of airport scene* **(a-d)** *original images* **(e-h)** *tracklets by KLT tracker* **(i-l)** *tracklets by our method*

KLT tracker are sparse, and the noise caused by tiny disturbance accumulates gradually at the top of the image. Benefitting from our rules, the dense points can be tracked robustly, and the noise is greatly reduced.

III. Motion Pattern Analyzing

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 the same moving direction are seen as a unit) [21]. According to this definition, "Motion pattern" in our research means the spatial-temporal segmentation in a video. Within the same motion pattern, the tracklets are close to each other and have similar or proximal motion direction. Motion patterns not only describe the segmentation in the spatial space, but also reflect the movement tendency in a period. In this section, our goal is to cluster the tracklets to analyze the motion patterns. We extend a novel unsupervised hierarchical clustering algorithm to tracklets clustering based on graphic model, and the reliable similarity measure derived by LCSS makes the clustering method effective and robust.

3.1 Longest common subsequence

Previous approaches to model the similarity between time-series include the algorithm of the Euclidean and Dynamic Time Warping (DTW) distance [22-23], which however is more sensitive to noise. LCSS is an efficient alignment model for unequal length data, and is more robust to noise and outliers than DTW [24-26].

The basic idea of LCSS is to match two time-series by allowing that not all the points need to be matched. Instead of one-to-one mapping between points, some elements in LCSS model can be unmatched; a point with no good match can be abandoned to prevent unfair biasing.

Let A and B denote two tracklets with the size n and m respectively defined as Eq. (4) and Eq. (5). We also defined the sequence Head (A) and Head (B) as Eq. (6) and Eq. (7). The LCSS is defined as follows:

$$A = \{(a_{x,1}, a_{y,1}), \cdots, (a_{x,n}, a_{y,n})\} \quad (4)$$

$$B = \{(b_{x,1}, b_{y,1}), \cdots, (b_{x,m}, b_{y,m})\} \quad (5)$$

$$\text{Head}(A) = \{(a_{x,1}, a_{y,1}), \cdots, (a_{x,n-1}, a_{y,n-1})\} \quad (6)$$

$$\text{Head}(B) = \{(b_{x,1}, b_{y,1}), \cdots, (b_{x,m-1}, b_{y,m-1})\} \quad (7)$$

$$\text{LCSS}(A,B) = \begin{cases} 0, & \text{if } A \text{ or } B \text{ is empty} \\ 1 + \text{LCSS}(\text{Head}(A), \text{Head}(B)), & \\ & \text{if } |a_{x,n} - b_{x,m}| < \varepsilon \\ & \text{and } |a_{x,n} - b_{x,m}| < \varepsilon \\ & \text{and } |n - m| \leqslant \delta \\ \max(\text{LCSS}(\text{Head}(A), B), \text{LCSS}(A, \text{Head}(B))), & \\ & \text{otherwise} \end{cases} \quad (8)$$

The constant ε controls the spatial matching threshold, and the constant δ controls how far in time to match a given point from one tracklet to a point in another tracklet. The $\text{LCSS}(A,B)$ specifies the matching cost between two tracklets, and can be efficiently computed using dynamic programming.

The similarity measure between tractlet A and B is defined as:

$$s(A,B) = \frac{\text{LCSS}(A,B)}{\min(n,m)} \quad (9)$$

This similarity function based on LCSS model allows time stretching. This satisfies the specific properties of crowded scenes. The tracklets in the same motion pattern which are close in space but different at time can be matched efficiently.

3.2 Automatic hierarchical clustering method

The classical supervised clustering algorithms are unable to satisfy our requirement as the number of motion patterns is unknown, and the data of tracklets is very large. A novel hierarchical clustering method base on graphic model techniques is proposed by Minsu Cho et al. [27]. The algorithm makes the clustering automatically without pre-defined number of clusters, and is effective for large size of data set. The method originally aims at the point

synthetic, and we extend the algorithm to automatic tracklets clustering.

This algorithm considers the cluster problem as a node-labeling processing on graphs. The procedure of clustering is constructed as the authority seeking on graph.

A relational graph $G = (v, \xi, w)$ with vertices v, edges ξ and weights w is given based on a pairwise relation, reflecting the similarity between two components. The hierarchical scheme is executed to recursively aggregate nodes in each cluster. The authority-shift procedure is performed iteratively for each Personalized Page Rank (PPR) propagation until the n_{th} order PPR converges.

PPR score creates the authority scores with respect to the specific nodes, and has been used in topic-sensitive web search based on the user personalization [27]. The PageRank vector r satisfies the following equation:

$$r^T = \alpha r^T P + (1-a)v^T \quad (10)$$

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 another. PPR implies the authority score that specific nodes weight by the vector v, thus measuring the importance of each node is in relation with the other nodes. PPR propagation is defined as:

$$\mathrm{PPR}_{n+1}(i) = \mathrm{PPR}(\mathrm{PPR}_n(i)) \quad (11)$$

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

$$\mathrm{Auth}_n(i) = \arg\max \mathrm{PPR}_n(i) \quad (12)$$

In our paper, a tracklet is constructed as a vertex. The links between vertices reflect the relationship between tracklets. The weight reflects the similarity between the tracklets discussed in Section 3.1. The graph G is presented by the weight matrix. The PageRank authority score can be computed by Eq. (10) in linear system formulation, and the minimal authoritative is obtained by shifting each node toward its authority score until it reaches the convergent node.

The process of seeking authority scores on graph is similar to mode-seeking in the mean-shift algorithm. But the difference is that the authority-shift only needs to be formulated once per node without special stopping rule.

IV. Experiment

To test our approach on analyzing motion patterns in dynamic crowded scenes, we implement experiments on some challenging video clips in UCF_Crowds dataset and our SJTU_Crowds dataset.

4.1 Dataset

The UCF_Crowds dataset is constructed by The University of Central Florida. The dataset contains videos of crowds, high density of moving vehicles and bio-cells under microscopes. These videos are collected mainly from the BBC Motion Gallery and Getty Images website.

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, few available large and public datasets about crowd in community can be obtained. It is partially the reason why we collect our SJTU_Crowds dataset. This dataset is different from the UCF_Crowds dataset in densities, motions, scenes, and so on.

The UCF_Crowds dataset is collected in a square of Shanghai Jiao Tong University campus. The crowd videos are captured by a calibrated camera with a resolution of 1 024*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. These scenes include various motion patterns of crowded people, such as splitting, merging, intersecting, crossing, linear motion, curvilinear motion, circular motion, emergency collection, evacuation, etc. We will public this SJTU_Crowds dataset later.

4.2 Experiment

We conduct experiments on typical videos in

UCF_Crowds dataset, which are usually used to evaluate the performance of the algorithm about crowded scenes analysis. And we also implement experiments in our SJTU_Crowds dataset. We compared our method with the hierarchical clustering method based on classical KLT algorithm. We also compare our approach with the state-of-the-art proposed in Ref. [11].

The values we used for $\alpha, \beta, \varepsilon, \delta$ are clearly dependent on the application and the dataset. For most datasets, we discover that setting α_1 between 0.05 and 0.07, β_1 between 0.01 and 0.02, α_2 between 0.1 and 0.2, β_2 between 0.001 and 0.005 is better.

The experimental results are shown in Figures 3 and 4. A video in a large market with complex and high density crowded is described in Figure 3 (a-d). Thousands of people and several buses are moving in the market. Individuals cannot be discriminated due to the similar appearance, especially occlusion with small size. In this case, it is difficult and also unnecessary to distinguish individuals. The tracklets become denser with the increase of time, and the noise is gradually wakened. This help to capture the persistent motion pattern of crowd. Four main motion patterns are detected, which are marked in different color in Figure 3 (i-l). We are encouraged that the small movement colored in green which is surrounded by another large movement marked in blue is detected accurately.

In unstructured scene, individuals move according to the social psychology. The dominant path which individuals move along can be analyzed through the motion patterns. We infer the four main paths in Figure 5 (b) by the motion patterns in Figure 5 (a).

The videos of five queues are shown in Figure 4 (a-e). As can be seen from the results, the tracklets reflect the reliable movement information of the queues in Figure 4 (f-j). It seems difficult to detect motion patterns while

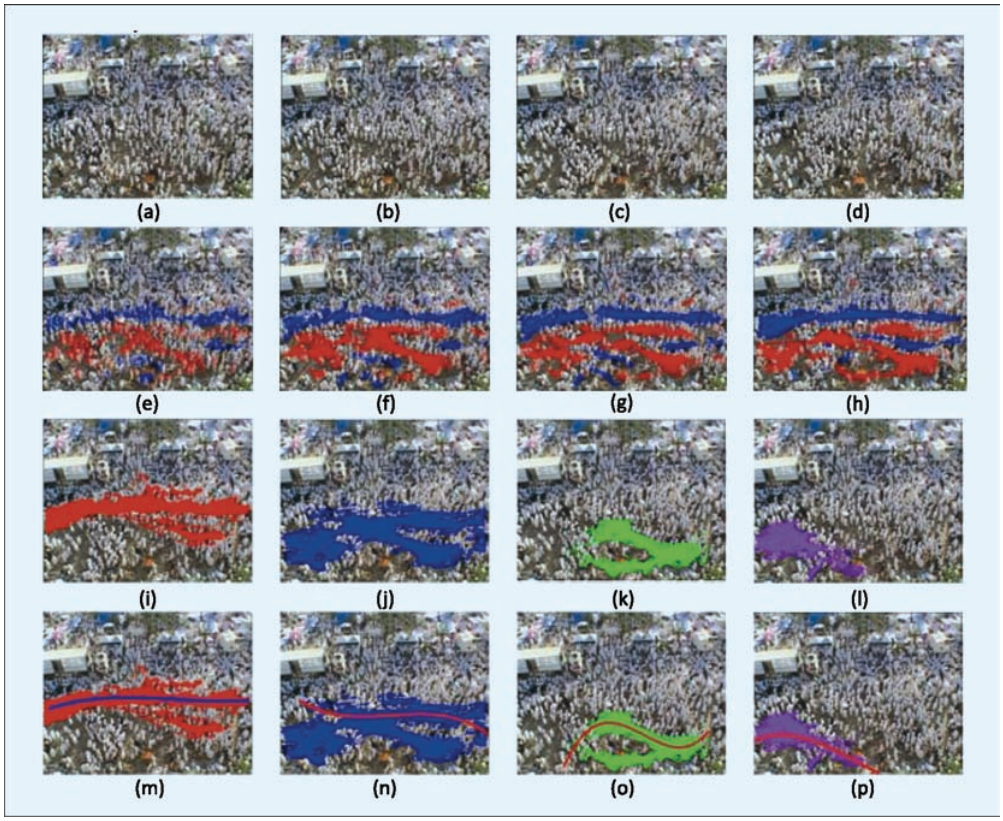

Fig.3 *The results of market video* **(a-d)** *original images* **(e-f)** *tracklets* **(i-l)** *motion patterns* **(m-p)** *main paths*

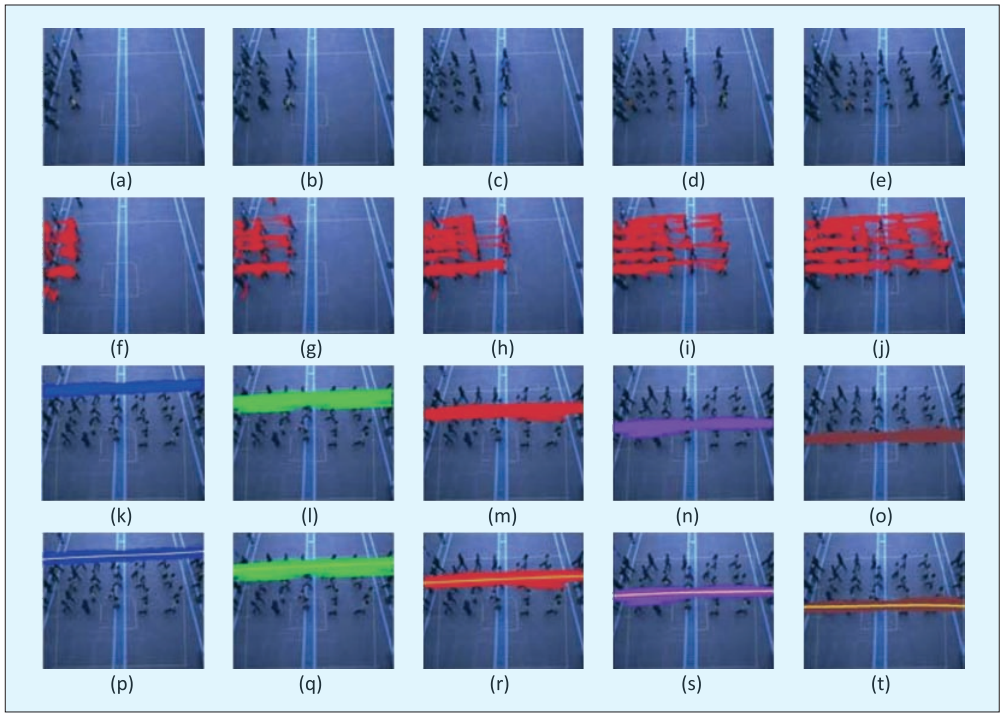

Fig.4 *The results of five queues video* **(a-e)** *original images* **(f-j)** *tracklets* **(k-o)** *motion patterns* **(p-t)** *main paths*

the individuals in different queues move close to others. However, the motion patterns are detected precisely in Figure 4 (k-o). The main paths are generated by the motion patterns in Figure 6.

In order to evaluate the performance of our algorithm, we select six typical videos on both datasets. We compare our results with the ground truth and other methods. The ground truth is manually marked from the videos. The comparisons about the number of motion pattern detected by different methods are shown in Figures 7 and 8.

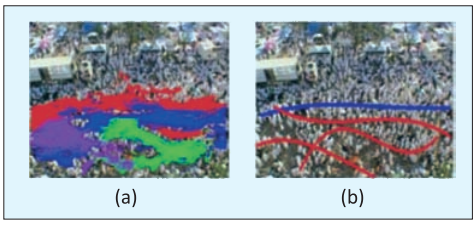

Fig.5 *Main paths of market video* **(a)** *motion patterns* **(b)** *main paths*

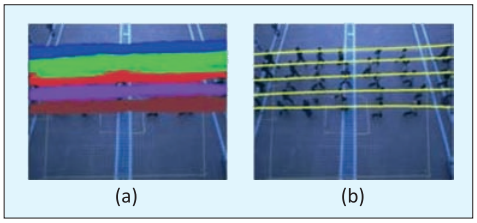

Fig.6 *Main paths of five queues video* **(a)** *motion patterns* **(b)** *main paths*

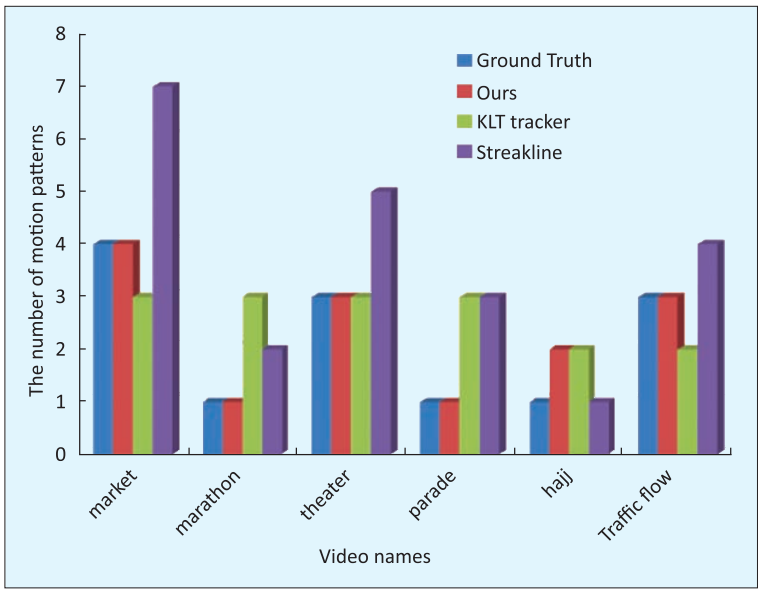

Fig.7 *The performance evaluation on UCF_Crowds dataset*

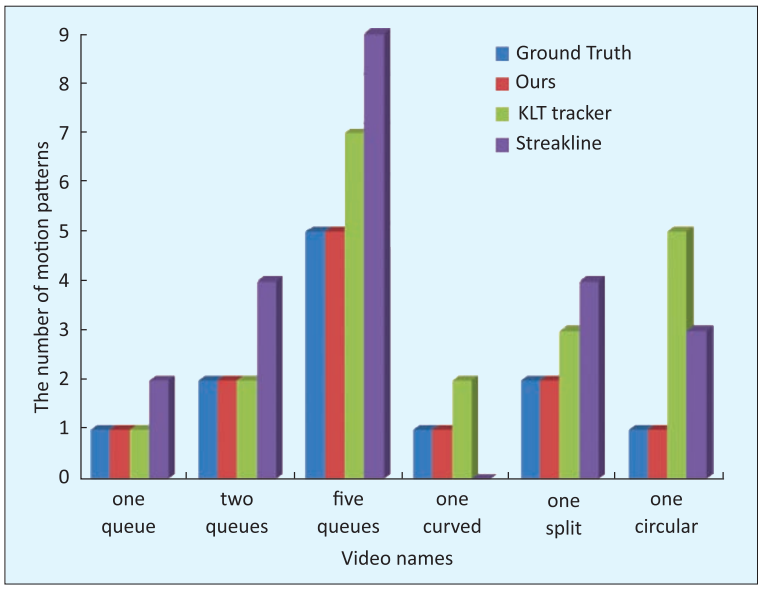

Fig.8 *The performance evaluation on SJTU_Crowds dataset*

From the comparisons, we can see that the results of our method are accordant with the ground truth in most cases. The advantages of our method are significant. One major reason is that the tracklets we proposed can extract the valuable crowd motion information. Meanwhile, the noise can be weakened with the increase of time. This helps to provide reliable features for the detecting process. Also, the similar measure based on LCSS provides a good distance measurement for unequal length data.

V. Conclusion

In this paper, we propose an unsupervised method to analyze motion patterns in dynamic crowded scenes. We track dense points under three rules through LK optical flow algorithm. And then the motion patterns are analyzed by the automatic hierarchical clustering algorithm with LCSS criteria. The experiments are conducted on some challenging videos in UCF_Crowds dataset and our SJTU_Crowds dataset, and the results demonstrate the effectiveness of our method.

We plan to investigate more effective properties about the crowd, and further use this research for motion analyzing and scene understanding in crowded scenes.

Acknowledgement

This work was supported in part by National Basic Research Program of China (973 Program) under Grant No. 2011CB302203 and the National Natural Science Foundation of China under Grant No. 61273285.

References

[1] ZHAO Tao, NEVATIA R. Tracking Multiple Humans in Crowded Environment[C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004): June 27-July 2, 2004. San Diego, CA, USA. IEEE Press, 2004, 2: II-406-II 413.

[2] KHAN Z, BALCH T R, DELLAERT F. An MCMC-Based Particle Filter for Tracking Multiple Interacting Targets[C]// Proceedings of the 8th European Conference on Computer Vision: May 11-14, 2004. Prague, Czech Republic. Springer, 2004: 279-290.

[3] HUE C, LE C J, PÉREZ P. Posterior Cramer-Rao Bounds for Multi-Target Tracking[J]. IEEE Transactions on Aerospace and Electronic Systems, 2006, 42(1): 37-49.

[4] BROSTOW G J, CIPOLLA R. Unsupervised Bayesian Detection of Independent Motion in Crowds[C]// Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition: June 17-22, 2006. New York, NY, USA. IEEE Press, 2006: 594-601.

[5] SUGIMURA D, KITANI K M, OKABE T, *et al.* Using Individuality to Track Individuals: Clustering Individual Trajectories in Crowds Using Local Appearance and Frequency Trait[C]// Proceedings of IEEE 12th International Conference on Computer Vision: September 27-October 4, 2009. Kyoto, Japan. IEEE Press, 2009: 1467-1474.

[6] ALI S, SHAH M. A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis[C]// Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE: June 18-23, 2007. Minneapolis, Minnesota, USA. IEEE Computer Society, 2007: 1-6.

[7] ALI S, SHAH M. Floor Fields for Tracking in High Density Crowd Scenes[C]// Proceedings of the 10th European Conference on Computer Vision: October 12-18, 2008. Marseille, France. Springer, 2008: 1-14.

[8] ALI S. Taming Crowded Visual Scenes[D]. Or-

lando, Florida, USA: Ph.D. thesis, University of Central Florida, 2010.

[9] RODRIGUEZ M, ALI S, KANADE T. Tracking in Unstructured Crowded Scenes[C]// Proceedings of IEEE 12th International Conference on Computer Vision: September 27-October 4, 2009. Kyoto, Japan. IEEE Press, 2009: 1389-1396.

[10] SALEEMI I, SHAFIQUE K, SHSH M. Probabilistic Modeling of Scene Dynamics for Applications in Visual Surveillance[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(8): 1472-1485.

[11] MEHRAN R, MOORE B E, SHAH M. A Streakline Representation of Flow in Crowded Scenes[C]// Proceedings of 11th European Conference on Computer Vision (ECCV): September 5-11, 2010. Heraklion, Crete, Greece, 2010: 439-452.

[12] WANG Xiaogang, MA Xiaoxu, GRIMSON E E. Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 539-555.

[13] ZHOU Bolei, WANG Xiaogang, TANG Xiao'ou. Random Field Topic Model for Semantic Region Analysis in Crowded Scenes from Tracklets[C]// Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition: June 20-25, 2011. Colorado Springs, Colorado, USA. IEEE Press, 2011: 3441-3448.

[14] WANG Chongjing, ZHAO Xu, ZOU Yi, *et al*. Detecting Motion Patterns in Dynamic Crowd Scenes[C]// Proceedings of 2011 Sixth International Conference on Image and Graphics (ICIG): August 12-15, 2011. Hefei, Anhui, China. IEEE Press, 2011: 434-439.

[15] SALEEMI I, HARTUNG L, SHAH M. Scene Understanding by Statistical Modeling of Motion Patterns[C]// Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): June 13-18, 2010. San Francisco, CA, USA. IEEE Press, 2010: 2069-2076.

[16] ZHOU Bolei, WANG Xiaogang, Tang Xiao'ou. Understanding Collective Crowd Behaviors: Learning A Mixture Model of Dynamic Pedestrian-Agents[C]// Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): June 16-21, 2012. Providence, RI, USA. IEEE Press, 2012: 2871-2878.

[17] ZHOU Bolei, WANG Xiaogang, Tang Xiaoou. Coherence Filtering: Detection Coherent Motions from Crowd Clutters[C]// Proceedings of 12th IEEE European Conference on Computer Vision (ECCV): October 7-13, 2012. Florence, Italy, 2012: 857-871.

[18] SHI Jianbo, TOMASI C. Good Features to Track [C]// Proceedings of 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition: June 21-23, 1994. Seattle, Washington, USA. IEEE Press, 1994: 593-600.

[19] SUNDARAM N, BROX T, KEUTZER K. Dense Point Trajectories by GPU-Accelerated Large Displacement Optical Flow[C]// Proceedings of 2010 11th IEEE European Conference on Computer Vision (ECCV): September 5-11, 2010. Heraklion, Crete, Greece. Springer, 2010: 438-451.

[20] WANG Heng, KLÄSER A, SCHMID C W, *et al*. Action Recognition by Dense Trajectories[C]// Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR): June 20-25, 2011. Colorado Springs, CO, USA. IEEE Press, 2011: 3169-3176.

[21] HU Min, ALI S, SHAH M. Learning Motion Patterns in Crowded Scenes Using Motion Flow Field[C]// Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08): December 8-11, 2008. Tampa, Florida, USA. IEEE Press, 2008: 1-5.

[22] PICIARELLI C, FORESTI G L. On-line Trajectory Clustering for Anomalous Events Detection[J]. Pattern Recognition Letters, 2006, 27(15), 1835-1842.

[23] MORRIS B T, TRIVEDI M M. Learning and Classification of Trajectories in Dynamic Scenes: A General Framework for Live Video Analysis[C]// Proceedings of IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance: September 1-3, 2008. Santa Fe, NM, USA. IEEE Press, 2008: 154-161.

[24] VLACHOS M, KOLLIOS G, GUNOPULOS D. Discovering Similar Multidimensional Trajectories[C]// Proceedings of 18th International Conference on Data Engineering: February 26-March 1, 2002. San Jose, CA, USA. IEEE Computer Society, 2002: 673-684.

[25] CHERIYADAT A M, RADKE R J. Automatically Determining Dominant Motions in Crowded Scenes by Clustering Partial Feature Trajectories[C]// Proceedings of First ACM/IEEE International Conference on Distributed Smart Cameras: September 25-28, 2007. Vienna, Austria, 2007: 52-58.

[26] CHERIYADAT A M, RADKE R J. Detecting Dominant Motions in Dense Crowds[J]. IEEE Journal of Selected Topics in Signal Processing, 2008, 2(4): 568-581.

[27] CHO M, LEE K M. Authority-shift Clustering: Hierarchical Clustering by Authority Seeking

on Graphs[C]// Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): June 13-18, 2010. San Francisco, CA, USA. IEEE Press, 2010: 3193-3200.

Biographies

WANG Chongjing, received the B.S. degree from Xi'an University of Technology, China in 2005, and 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. Email: ccjj@sjtu.edu.cn

ZHAO Xu, received the Ph.D. degree in pattern recognition and intelligence system from Shanghai Jiao Tong University, China. He is currently an Assistant Professor of Shanghai Jiao Tong University. He was a visiting scholar at the Beckman Institute for Advanced Science and Technology at University of Illinois at Urbana-Champaign, USA from 2007 to 2008. His research interests include visual analysis of human motion, machine learning, and image/video processing. Email: zhaoxu@sjtu.edu.cn

ZOU Yi, received his B.S. and M.S. degrees from Xiamen University, China in 2006 and 2009, respectively. He is currently a Ph.D. candidate in the Department of Electronic and Electrical Engineering in Shanghai Jiao Tong University, China. His research interests include crowd analysis, computer vision and pattern recognition. Email: jbyiiiii@sjtu.edu.cn

LIU Yuncai, received the Ph.D. degree in electrical and computer science engineering from the University of Illinois at Urbana-Champaign, USA 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 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. Email: whomliu@sjtu.edu.cn