

## High Density Crowd Behaviors Recognition Based on Micro-behavioral and Sparse Representation

Dongping Zhang, Jiao Xu, Yanjie Li and Ye Shen

College of Information Engineering China Jiliang University, Hangzhou310018,  
China,  
[silenttree\\_zju@cjlu.edu.cn](mailto:silenttree_zju@cjlu.edu.cn), [joanne7906@gmail.com](mailto:joanne7906@gmail.com),

### Abstract

*This paper presents a novel method to recognize high density crowd behaviors using micro-behaviors combining with Sparse Representation based on Locally Linear Embedding (named LLE-based Sparse Representation or LLE-SR). We extract micro-behaviors from each frame, respectively named Fountainhead, Bottleneck, Blocking, Lane and Ring/Arch, and construct micro-behaviors histograms to better describe complex high density crowd scenes. As the mid-level semantic, micro-behaviors solve the gap between the high-level semantics and low-level semantic, and the creation of them does not need any information of target track. LLE-SR method fully considers the behavior of local manifold structure of samples. Through adding LLE regularization term in sparse classification model, the unstable property of the manifold structure can be solved, and then recognition rate is improved. Numerous experiments have been conducted in real scenes, the results of which demonstrate the effectiveness and robustness of the proposed method for high density crowd behavior recognition.*

**Keywords:** high density crowds, behavior recognition, micro-behaviors, LLE-based Sparse Representation

### 1. Introduction

The safety of crowds has drawn great attention of researchers in the area of video monitor. Most studies focus on the abnormal events detection, and traditional approaches of surveillance tasks are almost based on motion detecting /tracking individuals [1-6], using trajectories to describe the scene and behaviors of people. However, these methods are not applicable for high density crowds. In real-world situation, we find that high density crowd behaviors in complex scenarios can be composed by multiple micro-behaviors without knowing individual behaviors, which will be used to conduct training examples of Sparse Representation, realizing crowds' behaviors recognition.

Recent methods for crowd's analyses focus on abnormal or unexpected behaviors detection [7-8]. With the development of compressed sensing, the method based on sparse representation is widely used in various fields [9-12]. Conventional sparse representation of the model does not consider the local manifold structure of the test samples, leading to the instability of sparse decomposition. Inspired by sparse representation and manifold learning [13-14] method, this paper presents crowd behavior recognition solution based on local linear embedding [15] sparse representation to distinguish different behaviors, not only making a simple judgment about abnormal behaviors, fully considering local structure of samples. Experimental results show that the sparse representation based on locally linear embedding algorithm can effectively improve the ability to distinguish samples and get good experimental results for behavior recognition of high density crowds.

The aim of our work is to devise a holistic approach that can achieve specific high object-density behavior analysis. It can be expressed by micro-behaviors combined with LLE-based Sparse Representation that we propose in this paper, and does not need to know the individual behavior accurately. The method is not only suitable for low but also high density crowd scenes.

The remainder of this paper is organized as follows: In Section 2 and 3 the methodology is laid out, including micro-behaviors extraction and common crowd abnormal behaviors recognition. Section 4 shows experimental results and Section 5 concludes this paper.

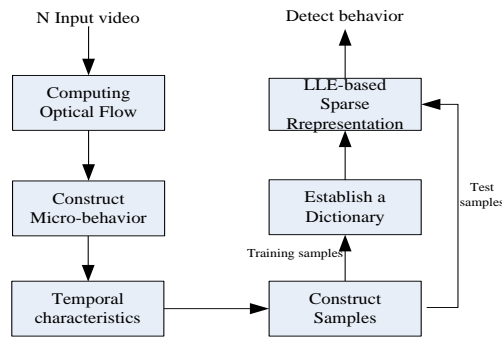


Figure 1. The Framework of the Proposed Method

## 2. Construction of Micro-behaviors

In this section, we will introduce micro-behaviors, and the framework is shown as follows:

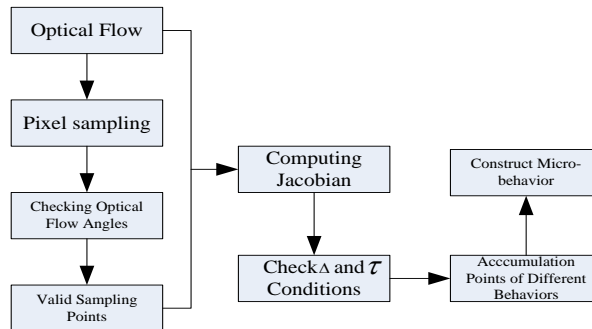


Figure 2. The Construction of Micro-behaviors

In paper [17], a framework has been introduced to identify multiple crowd behaviors through stability analysis for dynamical systems. But our method is different. Firstly, our method does not need any particles trajectories to represent the motion in the scene. Secondly, we do not analyze all of the particles within the ROI. The article emphasizes that a grid of particles is overlaid on the initial frame and advected with the flow. Through particle advection and the resulting accumulation of particles, each ROI is computed. What's more, the algorithm is repeated over sequential clips of a video in order to record changes in eigenvalues, which may imply changes in behavior. Thirdly, we use eigenvalues of valid sampling points to conduct five micro-behaviors, and use short-term characteristics to express different behaviors.

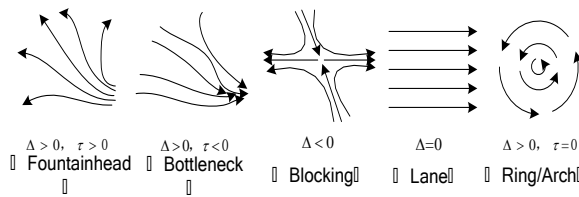
Through the improved spectral flow method, the optical flow in motion area can be calculated, and we can obtain a continuous dynamical system (1):

$$F(w) = [u(w), v(w)], \quad (1)$$

with  $w(t) = [x(t), y(t)]$ , and  $u$  and  $v$  represent pixel velocities in the  $x$  and  $y$  directions, respectively. According to Taylor's theorem, we denote  $J_F$  as follows:

$$J_F = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}. \quad (2)$$

$J_F$  is the Jacobian matrix for  $F(w)$ , and we hope that the important information about the five crowd micro-behaviors flow as shown in Fig.3 can be reflected by the eigenvalues of  $J_F$ . However, in the process of crowd behavior analysis, there are some questions which we must note as follows:



**Figure 3. Illustration of Five Crowd Micro-behaviors Flow**

- (a) A certain noise does exist in the movement velocity field, such as human arms and legs swinging back and forth and so on. This will significantly affect the results of the upcoming analysis of micro-behavior;
- (b) A target's movement information can be expressed by multiple pixels (What we should pay more special attention to is that these pixels might tend to share the same or similar movement information), but in the process of micro-behavior analysis, using these pixels to build Jacobian matrix is not what we want, and unreasonable, even leading to erroneous analysis result directly;
- (c) In the velocity field of crowd scenery, there are a lot of background points (that means the velocities of these pixels are zero). If appropriate filter is not employed for these points, the problems mentioned in (b) can also be encountered because of the invalid computation.

Based on the above reasons, it is necessary to make proper sampling for every frame of video sequences. So, differently from article [17], we make appropriate sample in accordance with the targets size after obtaining  $u$  and  $v$  from optical flow, being maximum extent consistent with the actual movement. The details will be introduced in Section 5.

At the same time, sampling pixel angles  $\theta = \arctan(v / u)$  can be checked. We cluster the sampling pixels by checking  $\theta$ , and the number of sampling pixels in each cluster is a signification measure for that cluster, because candidate points (valid sampling pixels) will be defined as pixels which included in major clusters(Typically, a major cluster contains at least one third of the total sample pixels.). We put each frame image change as a short-term movement, and use these valid sampling point motion information to obtain Jacobian matrix has shown, the eigenvalues of which are the

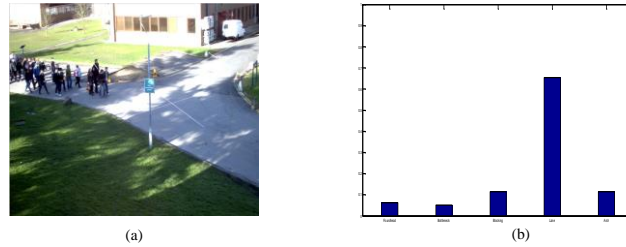
solutions of a characteristic equation  $\lambda^2 - \tau\lambda + \Delta = 0$ , where  $\tau$  is the trace and  $\Delta$  is the determinant of the matrix. We can easily find that:

$$\lambda_{1,2} = \frac{1}{2}(\tau \pm \sqrt{\tau^2 - 4\Delta}), \quad (3)$$

in which  $\Delta = \lambda_1\lambda_2$  and  $\tau = \lambda_1 + \lambda_2$ .  $\lambda_1$  and  $\lambda_2$  are the eigenvalues, which can reflect important information about the flow as described in Fig.3, in connection with specific crowd behaviors. The specific introduction of the analysis can be found in paper [17], and the following five specific flows are simply summarized as:

Blocking:  $\Delta < 0$ ; Ring/Arch:  $\Delta > 0, \tau = 0$ ; Lane:  $\Delta = 0$ ; Bottleneck:  $\Delta > 0, \tau < 0$ ; Fountainhead:  $\Delta > 0, \tau < 0$ .

As motioned above, we only consider the valid sampling pixels in each frame. Then for each frame, we build five kinds behavior flow histograms, as a micro-behavior, which will use describe complex scenes.



**Figure 4. (a) is a Crowd Scene, and (b) is the Corresponding Histogram**

X coordinate is five kinds micro-behaviors, respectively Fountainhead, Bottleneck Blocking, Lane, Ring/Arch, and Y is the ratio.

### 3. LLE-based Sparse Representation

Traditional sparse representation algorithms suppose that there are  $M$  distinct categories training samples making up an over complete dictionary:

$$D = [D^1, \dots, D^M] \in R^{m \times N}, \quad (4)$$

and  $D^i = [d_1^i, \dots, d_{N_i}^i] \in R^{m \times N_i}$  is used to describe the  $i$ th class containing  $N_i$  training samples. For a test sample  $y$ , it can be sparsely represented by the over complete dictionary, which is merged by all these  $M$  classes:

$$y = D \times \alpha \in R^m, \quad (5)$$

in which  $\alpha \in R^N$  is the sparse representation of  $y$  over the dictionary  $D$ . Influenced by noise, we renewal sparse representation  $\hat{\alpha}$  of  $y$  as Eq.(6) show:

$$\hat{\alpha} = \min_{\alpha} \|y - D\alpha\|^2 + \beta \|\alpha\|_1, \quad (6)$$

and  $\beta$  ( $|\beta| < 1$ ) is used to enforce sparsity. In order to identify the class of  $y$ , we calculate the residual error between the test sample and the  $i$ -th reconstruction class:

$$r^i(y) = \|y - D^i \hat{\alpha}^i\|, i = 1, \dots, M, \quad (7)$$

in which  $\hat{\alpha}^i$  refers to the nonzero coefficients corresponding to the indices of  $i$  th class's atoms, and the test sample  $y$  which class belongs to can be determined as:

$$\text{class}(y) = \arg \min_{i=1, \dots, M} r^i(y) . \quad (8)$$

The more attention we should pay is that it is difficult to label training samples in practice. So Eq.(5) is modeled as an over-determined case, i.e.,  $m > N$  . In order to solve the problem, the researchers usually increase the dimension of  $D$  . That is to say, we make  $\tilde{D} = [D \ I] \in R^{m \times (N+m)}$  to replace  $D$  :

$$\tilde{c} = \min_{\alpha} \left\| y - \tilde{D} c \right\|^2 + \beta \|c\|_1 , \quad (9)$$

in which  $I$  is a unit matrix and  $c = [\alpha^T \ \tilde{\alpha}^T]^T \in R^{(N+m)}$  , and  $\tilde{\alpha} \in R^m$  is an error vector. Then the residual error is defined as:

$$r^i(y) = \left\| y - \tilde{D}^i \tilde{c}^i \right\| , i = 1, \dots, M . \quad (10)$$

The class of  $y$  is then determined as the one with the minimal residual:

$$\text{class}(y) = \arg \min_{i=1, \dots, M} r^i(y) . \quad (11)$$

Locally linear embedding (LLE) is a promising manifold learning algorithm, and has attracted a lot of attentions. LLE is based on simple geometric intuitions, We give a test sample set  $Y = [y_1, y_2, \dots, y_n]$  . Inspired by LLE, an assumption generated to alleviate the instability of sparse decompositions is that, if a test sample  $y_i$  can be estimated by the linear combination of its neighbors in the manifold of the test sample set, then  $c_i$  can be also estimated by the linear combination of corresponding neighbors using the same reconstruction weight of the sample  $y_i$  . Based on the assumption above, an LLE-based quadratic constraint is defined as:

$$\sum_{i=1}^n \left\| c_i - \sum_{j \in N(y_i)} w_{ji} c_j \right\|^2 , \quad (12)$$

where  $w_{ji}$  denotes the reconstruction weight of  $c_j$  .  $N(y_i)$  is selected if it is within the first  $y_i$  closest samples with respect to  $y_i$  .  $w_{ji}$  can be obtained as

$$w_{ji} = \arg \min_{v_{ji}} \left\| y_i - \sum_{j \in N(y_i)} w_{ji} y_j \right\|^2 ; s.t. \sum_{j \in N(y_i)} w_{ji} = 1 . \quad (13)$$

Eq.(16) can be transformed as:

$$\sum_{i=1}^n \left\| c_i - \sum_{j \in N(y_i)} w_{ji} c_j \right\|^2 = \|C - CW\|^2 = \|(C - CW)^T\|^2 = Tr(C(I - W)(I - W)^T C^T) = Tr(CMC^T) . \quad (14)$$

where  $I$  is the unit matrix and  $M = (I - W)(I - W)^T$  . And  $w_{ji}$  is defined as:

$$W_{ji} = \begin{cases} w_{ji} , & \text{If } y_j \text{ is within the } k \text{ closest to } y_i \\ 0 , & \text{otherwise} \end{cases} , \quad (15)$$

Then the LLE-based optimization problem named LLESR is formulated as:

$$\min_c \left\| Y - \tilde{D} C \right\|^2 + \beta \|C\|_1 + \beta_1 Tr(CMC^T) , \quad (16)$$

where  $\beta_1 \geq 0$  is the regularization parameter. Our aim is to optimize  $C$ . Eq.(16) can be represented as follows:

$$\min_{c_1, \dots, c_m} \sum_{i=1}^m \left\| y_i - \tilde{D} c_i \right\|^2 + \beta_1 \sum_{i,j=1}^m M_{ij} c_i^T c_j + \sum_{i=1}^m \beta \|c_i\|_1. \quad (17)$$

According to [18, 19], the key to solve Eq.(17) is to search for the sign of coefficient  $c_i^{(j)}$ , which means the  $j$ -th coefficient of  $c_i$ . In our paper, we employ feature-sign search algorithm introduced in [18, 19] to obtain the optimal sparse representation  $c_i^*$ .

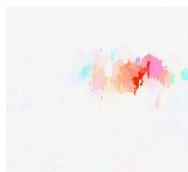
#### 4. Experimental Results

We use the real-world video sequences downloaded from web (such as BBC, Youtube) and sequences from PETS 2009 Dataset, representing crowd scenes.

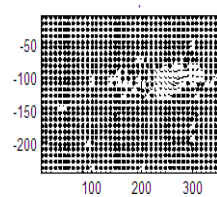
We take a set of scattered video downloaded from PETS 2009 as an example to introduce our method. Fig.5. has described one frame of the scene with the size of  $240 \times 360$ . As described in Fig.6. (b), in order to eliminate the noise, we set velocity threshold  $\kappa$  first, and set  $v$  of the pixel to zero if  $|v| < \kappa$  ( $\kappa = 0.5$ ). At the same time, we downscale the samples by the size of  $12 \times 20$  according to the size of crowd density of the scene.



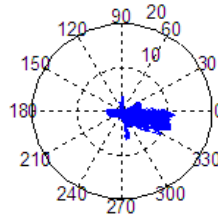
Figure 5. "Panic" Scenery of Crowds



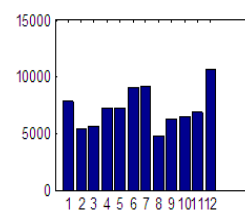
(a) Middlebury color coding

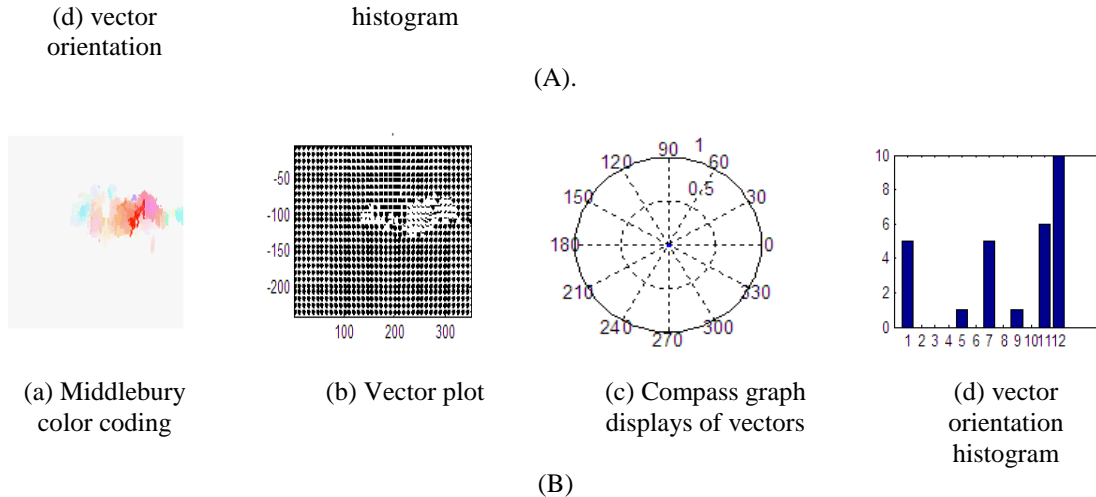


(b) Vector plot



(c) Compass graph displays of vectors





**Figure 6. The Effect Comparison of :(A) Pixels with No-sampling and (B) Pixels with Sampling**

In the video image processing, a kind of behavior cannot be completely expressed only by one frame’s information. Instead, multi-kinds of behaviors are often expressed by a video sequence. Hence, in this paper, we adopt the method of constructing space-time cube to express behaviors using micro behaviors, and the entire crowd behavior is described by these micro behaviors. For example, a “Panic” scene can be judged by video scene which is composed of several micro-behaviors mainly include "Fountainhead" flow in a period of time. There are three abnormal behaviors commonly existing in crowds described as follows. They can be expressed by five micro-behaviors flows vividly.

- (1) Panic: Crowds diverge from one location and persist in many separate directions;
- (2) Cluster: Crowds converge to one location suddenly for some reason, such as fight with each other.
- (3) Congestion: Crowds bounce off of each other in somewhat random directions and surrounding crowd moving in opposite directions block each other as crowd density increase and may lead to gridlock.

For each frame, we can get the micro-behavior, representing as:

$$R = [R_{Fountainhead}, R_{Bottleneck}, R_{Blocking}, R_{Lane}, R_{Arch}] \quad (18)$$

The ratios of one frame just reflect temporary behaviors, and a specific crowd behavior can be expressed by multiple micro-behaviors during a durational time. In this paper, the entire sequence is divided into non-overlapped small fragments uniformly, and the length of each fragment is  $l$ . In our experiment, we choose  $l = 16$  in practice, and the selection of the frame number has been fully taken into consideration of the length of test data. Then a training sample of micro-behavioral semantic features can be described by a  $5 \times 16 = 80$  dimensional vector by sequential clips of the video as follows:

$$d_j^i = [R_1, R_2 \dots R_{16}] \quad (19)$$

where  $d_j^i$  denotes the  $j$ -th sample of the  $i$ -th class.

Our purpose is to use micro-behaviors to describe the common abnormal crowd behaviors. Table1 gives a set of data as a sample for typical scenery “Panic”. From the table we can find that the “Panic” behavior is mainly composed of “Fountainhead”

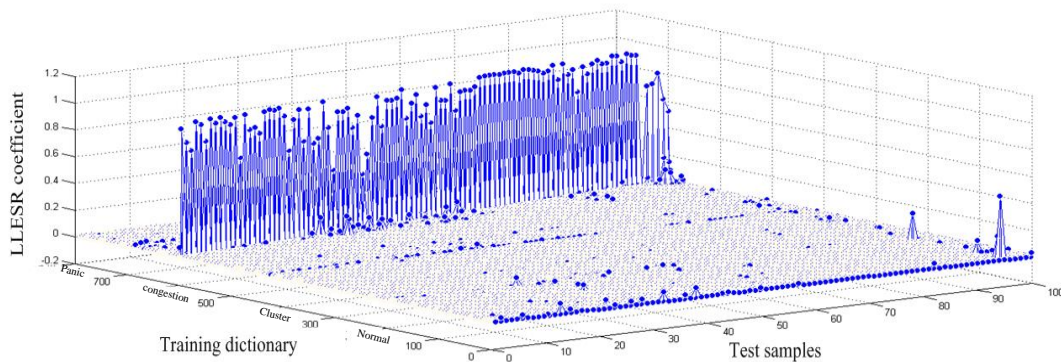
micro-behavior flow and the ratio tendency of effective sample points basically remains a certain correlation in certain duration.

**Table 1. Ratio Conditions Corresponding to Behaviors**

	$R_{Panic}$	$R_{Congestion}$	$R_{Cluster}$	$R_{Lane}$	$R_{Arch}$
1	0.654	0.062	0.051	0.115	0.115
2	0.684	0.053	0.066	0.066	0.131
3	0.679	0.095	0.024	0.059	0.143
4	0.643	0.092	0.020	0.020	0.225
5	0.644	0.022	0.044	0.089	0.200
6	0.588	0.069	0.019	0.059	0.265
7	0.617	0.121	0.020	0.030	0.212
8	0.638	0.088	0.039	0.049	0.186
9	0.691	0.048	0.049	0.049	0.163
10	0.616	0.024	0.056	0.048	0.256
11	0.607	0.063	0.063	0.031	0.236
12	0.624	0.050	0.050	0.050	0.226
13	0.631	0.059	0.050	0.042	0.218
14	0.603	0.059	0.044	0.066	0.228
15	0.642	0.070	0.056	0.063	0.169
16	0.661	0.052	0.026	0.069	0.192

We do not use a Tolerance L of ratio to make a conclusion which behavior it belongs to as in [11], because the choice of L might be uncertain. In our paper, we adopt LLE-SR algorithm ( $\beta=0.1, \beta_1=1, k=5$ ) to distinguish different abnormal behaviors.

After obtaining the sparse representation dictionary and the behavior characteristics of test samples, we can use the proposed LLE-SR method to classify behaviors. In our proposed algorithm, we construct a  $80 \times 800$  dictionary and adopt a set of video sequences, including three abnormal behaviors mentioned above and normal behaviors to verify this algorithm. Fig.6 shows the test result of "Panic" with 100 test samples. From Fig.6 we can see that sparse coefficients of the test samples mainly concentrate in the dictionary's atoms position of "Panic".



**Figure 6. The Video Sequences of "Panic" Detection Result**

In order to further prove the effectiveness of our algorithm, we use the same test samples to detect abnormal behaviors with other methods and the result is shown in Table 2. Optical flow method is often used as a description of the events in complex scenes nowadays [20-21]. It has a good description of the change rate of pixel gray values and indirectly reflects the target's moving speed and direction. So it is widely used in the field of abnormal behavior detection. In Table 2, Method 1 represents our LLE-SR algorithm proposed in this paper; Method 2 represents SR



algorithm combining with micro-behaviors mentioned in this paper; Method 3 represents optical flow feather combining with HMM.

According to the test of different kinds of video sequences, we can see from Table 2 that, our proposed LLE-SR method achieves the highest accuracy.

**Table 2. The Accuracy of Different Methods**

Classes		Samples		Accuracy		
Name		Train	Test	Method 1	Method 2	Method 3
Abnormal behavior	Panic	200	100	92%	87%	89%
	Congestion	200	126	87.6%	83.3%	81.7%
	Cluster	200	88	88.6%	87.5%	85.22%
Normal behavior		200	100	91%	88%	89%

## 5. Conclusion

In this paper, we put forward a method that combines Sparse Representation based on Locally Linear Embedding with semantic feathers of micro-behaviors to recognize behaviors of high density crowd in visual scenes. The main achievements of this work are the micro-behaviors construction using gradient matrix of sampling points and the addition of LLE regularization term for sparse representation. Micro-behaviors combines low-lever motion features with high-lever semantic to express complex scenes better. LLE-based Sparse Representation method fully considers the behavior of local manifold structure of samples. By the way of adding LLE regularization term in sparse classification model, the unstable property of the manifold structure can be overcome and the recognition rate can be improved. Experimental results show that our LLE-SR method improves the accuracy and robustness.

## Acknowledgements

This work was supported by Zhejiang Provincial Science & Technology Research Program (Grant No. 2013C33046, 2014C33075) and Zhejiang Provincial NSF (Grant No. Y1110506, LY13H180011).

## References

- [1] Y. O. Javed and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, (2006), p. 13.
- [2] G. J. Brostow and R. Cipolla, "Unsupervised bayesian detection of independent motion in crowds[C]//Computer Vision and Pattern Recognition", 2006 IEEE Computer Society Conference on. IEEE, vol. 1, (2006), pp. 594-601.
- [3] A. Ess Pellegrini, K. Schindler and L. J. V. Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking[C]//Computer Vision, 2009 IEEE 12th International Conference on. IEEE, (2009), pp. 261-268.
- [4] S. Ali and M. Shah, "Floor fields for tracking in high density crowd scenes [M]", *Computer Vision—ECCV 2008, Springer Berlin Heidelberg*, (2008), pp. 1-14.

- [5] L. Kratz and K. Nishino, "Tracking with local spatio-temporal motion patterns in extremely crowded scenes[C]", Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, (2010), pp. 693-700.
- [6] Z. Lin, L. Davis, D. Doermann, and D. DeMenthon, "Simultaneous appearance modeling and segmentation for matching people under occlusion," in ACCV07, (2007), pp. II: 404-413.
- [7] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models[C]", Computer Vision and Pattern Recognition, 2009, CVPR 2009, IEEE Conference on. IEEE, (2009), pp. 1446-1453.
- [8] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly detection in crowded scenes[C]", Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, (2010), pp. 1975-1981.
- [9] J. Yang, J. Wright, T. S. Huang and Y. Ma, "Image super-resolution via sparse representation [J]", Image Processing, IEEE Transactions on, vol. 19, no. 11, (2010), pp. 2861-2873.
- [10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries [J]", Image Processing, IEEE Transactions on, vol. 15, no. 12, (2006), pp. 3736-3745.
- [11] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation [J]", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 11, pp. (2011), 2259-2272.
- [12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation [J]", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, (2009), pp. 210-227.
- [13] X. Lu, Y. Yuan and P. Yan, "Image Super-resolution via Double Sparsity Regularized Manifold Learning [J]", Circuits and Systems for Video Technology, vol. 99, no. 1, (2013), pp. 1-12.
- [14] X. Zhang, Y. Yang and L. C. Jiao, "Manifold-constrained coding and sparse representation for human action recognition [J]", Pattern Recognition, vol. 46, no. 7, (2013), pp. 1819-1831.
- [15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding [J]", Science, vol. 290, no. 5500, (2000), pp. 2323-2326.
- [16] P. Sand and S. Teller, "Particle video: Long-range motion estimation using point trajectories [C]", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York: IEEE Computer Society, (2006), pp. 2195-2202.
- [17] B. Solmaz, B. E. Moore and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems [J]", Pattern Analysis and Machine Intelligence, IEEE Transactions on, (2012), vol. 34, no. 10, pp. 2064-2070.
- [18] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," IEEE Trans. Image Process., vol. 20, no. 5, (2011) May, pp. 1327-1336.
- [19] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," in Proc. Adv. Neural Inf. Process. Syst., (2007), pp. 801-808.
- [20] S. Wu and H. S. Wong, "Crowd Motion Partitioning in a Scattered Motion Field", IEEE transactions on systems, man, and cybernetics, Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society, (2012).

## Authors



**Dongping Zhang**, he was born in 1970. He received the Ph.D in Information & Communication Engineering from Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2006. Since 2006, He is an associate professor at College of Information Engineering, China Jiliang University in Hangzhou. His research interests include image processing and pattern recognition, computer vision and videos.



**Jiao Xu**, she graduated at the China Jiliang University in Hangzhou in 2012. Currently she is a MSc. student at the College of Information Engineering of China Jiliang University. Her research interests are oriented to image processing, pattern recognition and image compression for security and multimedia applied imaging systems.



**Yanjie Li**, She received the bachelor degree in College of Information Engineering from China Jiliang University, China, in 2014. She is pursuing a master's degree from China Jiliang University, China, at present. Her research interests include machine learning, pattern recognition, data mining, and image processing



**Shen Ye**, he received the Ph.D. degree in Biomedical Engineering from Zhejiang University, China in 2014. He is an associate professor in the department of signal and information processing from China Jiliang University, China at present. His research interests are in machine learning, data mining ,computer-aided diagnosis.

