



# Video Anomaly Detection Based on Hierarchical Clustering

Chunyue Zhao, Beichen Li, Qing Wang, and Zhipeng Wang<sup>(✉)</sup>

School of Electrical Automation and Information Engineering,  
Tianjin University, Tianjin 300072, China  
zpwang@tju.edu.cn

**Abstract.** Recently the video surveillance market has developed rapidly, but judging whether there is abnormal behavior in the video relying on manpower is too expensive. Therefore, a method is needed to identify the abnormal behavior automatically. Scholars at home and abroad have done in-depth research on video abnormal events detection in different scenarios. However, the current detection technology still needs improvement in the speed of the algorithm. From this point of view, this paper proposes a video abnormal event detection method based on hierarchical clustering.

In order to construct the sparse coefficient matrix more accurately and quickly, the hierarchical clustering is introduced into sparse coding in this paper. And the structure information of the sparse coefficient matrix is used as the clustering criteria, which improves the standard group sparse coding method. In addition, the BK-SVD algorithm is used to train the dictionary so that we can further improve the speed of the algorithm through dictionary division. In the experimental part, we prove that the proposed algorithm has great performance in frame level and pixel level in MATLAB environment.

**Keywords:** Video abnormal event detection · Dictionary learning · Sparse coding

## 1 Introduction

Nowadays, video anomaly detection technology attracts more and more attention because video surveillance has a wide range of applications in various fields such as production and life. It is too expensive to judge whether there is abnormal behavior in the surveillance video relying solely on manpower. Therefore, we urgently need a technology that can automatically recognize abnormal behavior.

The difficulty of detecting anomalies is diverse in different scenes. And the research on dense scenes is the most challenging but also the most practical work. This paper proposes a video anomaly detection algorithm based on hierarchical clustering for video anomaly detection in dense scenes. The algorithm can automatically detect abnormal behaviors and meet the needs of practical applications.

In the work of abnormal event detection, there are mainly three parts that need to be solved, namely feature extraction, dictionary learning and sparse coding. At present, previous researches have focus on feature extraction methods. However, dictionary

learning and sparse coding algorithms still need more in-depth research. Therefore, the content of this paper mainly focuses on these two parts.

In the field of dictionary learning and sparse coding, a relatively new and effective idea is to consider the structural characteristics of the dictionary or the sparse matrix, and add some constraints during dictionary learning and coefficient matrix construction to make the trained dictionary reconstruct the original information more accurately. However, there is always a general deficiency in this kind of ideas. This is because the artificially defined constraints always have certain limitations, which leads to some deviations from the restoration of the original features. Therefore, this paper supposes to improve the way in which artificial constraints are formulated. In addition, the idea of dictionary division also plays a dominant role in dictionary learning, because the division of sub-dictionary is beneficial to reduce the computation time and improve the robustness of the algorithm. Therefore, this paper uses the idea of dictionary partitioning and hierarchical clustering to automatically fuse atoms in the dictionary, which greatly improves the accuracy and speed of abnormal event detection.

The contributions of this paper: (1) On the basis of making full use of the structure of the sparse coefficient matrix, hierarchical clustering is used in the sparse representation. (2) BK-SVD is used in the dictionary learning process. Blocking the dictionary atom helps to improve the speed of dictionary training. (3) Deformation of the reconstruction error by the least squares method makes it easier to judge the abnormality in experiments.

## 2 Related Work

### 2.1 Trajectory-Based Video Anomaly Detection

In the sparse scene, the complexity of the feature is not high (this complexity generally refers to the dimension of the feature) because there is no occlusion. Therefore, detecting anomalies by analyzing motion trajectories is more robust. The process of obtaining the trajectory is actually the process of dividing the input video into video blocks of a specified size. Since the video block cannot be directly analyzed quantitatively, the trajectory need to be described by the model. And then the algorithm is used to analyze the obtained trajectory.

Scholars of Northwestern University proposed a dynamic hierarchical clustering anomaly detection method based on trajectory [1]. The author first described the target trajectory with HMM model. The distance between different trajectories calculated by the Bayesian information criterion is used as the criterion for the similarity. The cluster of the target is then obtained by the 2-depth greedy algorithm. After each clustering, all remaining trajectories are retrained and classified to achieve the effect of dynamic clustering. The algorithm has a good effect on the accuracy of detection and the speed of the algorithm, so it is of great significance to promote the development of video anomaly detection technology.

## 2.2 Population-Based Video Anomaly Detection

In dense scenes, it is not easy to analyze the video directly because each frame contains numerous complex information. We often selectively extract features of each frame in the video, and then describe them by the dictionary and sparse coefficient matrix. In different algorithms, the way to train the dictionary is diverse. But all algorithms are for strong generalization ability of the dictionary, and the trained dictionary can meet the speed requirements of practical application.

Scholars have proposed many excellent algorithms for the crowd-based video anomaly detection. These algorithms have achieved good detection results in practical applications.

The scholars from Aalborg University [2] proposed an unsupervised dictionary learning method. The authors paid attention to the distribution problem of sparse coefficient matrix. They not only used reconstruction error as the criterion for judging anomalies, but also believed that features corresponding to dense non-zero coefficients are basically normal. Therefore, dictionary corresponding to normal behavior was trained and then linked into a complete dictionary corresponding to all normal behaviors. The dictionary was supplemented with atoms corresponding to dense non-zero coefficients in order to obtain a higher recovery dictionary. However, this algorithm does not focus on the speed problem, so it is not suitable for scenarios with lots of operations.

Scholars from the Chinese Academy of Sciences have proposed a structural dictionary learning method for complex scenes [3]. The authors believe that the sparse coefficient matrix corresponding to normal features are similar in many ways, so the normal features will have resemble sparse representations. And the authors also believe that features that are close to the normal feature space are basically normal, so the sparse matrices of features with smaller spatial distances are similar. The authors use these two spatial structure features as constraints for dictionary learning and coefficient matrix construction, which obtain relatively high precision in the detection of abnormal events.

Professors of Tianjin Normal University proposed a method for detecting abnormal events by compact and low rank sparse learning (CLSR) [4]. The authors believe that the characteristics corresponding to normal behavior are similar, so the dictionary trained has low rank. The author also believes that the sparse matrices corresponding to normal behavior are similar, so the tightness rule is proposed to constrain the construction of sparse coefficient matrices. Compared with many algorithms that ignore structural information, the experimental part of the paper verifies that the method of dictionary learning and sparse coefficient construction under the low rank rule and the tightness rule improves the detection accuracy.

Scholars from the Chinese University of Hong Kong have proposed a fast dictionary learning method [5], which can achieve a detection rate of 140–150 frames per second in MATLAB software. This paper automatically combines the dictionary atoms (columns in the dictionary) but it need promise the sufficient number of combinations and minimum reconstruction error. Finally, a target dictionary composed of many sub-dictionaries is formed. Since the fusion process of the sub-dictionary is automatic and the dictionary is sequentially updated by sub-dictionary, the speed of the algorithm is

guaranteed. At present, the idea of dictionary division plays an important role in dictionary learning, because it may have a positive effect on algorithm speed, algorithm robustness or algorithm accuracy.

Zelnik-Manor scholars proposed a method of dictionary optimization using block sparse representation [6]. The author proposes a BK-SVD (block K-singular value decomposition) dictionary training method, which is a breakthrough in the research of dictionary learning. We are more familiar with the K-SVD dictionary construction method, in which the dictionary atom is updated column by column. Comparatively, the block updating strategy in the BK-SVD algorithm can help to improve the update speed. This paper also takes the speed of updating into consideration, so the BK-SVD algorithm is adopted.

Based on references [7–13], we can understand that there are two main types of algorithms with good performance in the field of sparse coding representation, namely greedy algorithm and L1 norm approximation, respectively. Greedy algorithms include MP, OMP, StOMP and there are mainly BP and FOCUSS in L1 norm approximation algorithms. According to the literature, we can discover that the greedy algorithm is faster and more suitable for larger computing needs. Since research in this paper is aimed at dense scenes with crowds which means a lot of operations, we adopt the greedy algorithm in the sparse coding part. In addition, this paper takes advantage of the structural characteristics of the sparse coefficient matrix itself and utilizes it as the clustering criterion for hierarchical clustering.

This paper studies crowded and complex scenarios, so we detect the anomalies based on the crowd.

In summary, it can be clearly recognized that there are many excellent algorithms in dictionary learning and sparse matrix construction, but basically no algorithm can meet the requirements of practical application for both calculation speed and accuracy. It means that there is still a long way to go. We not only need to construct a sparse coefficient matrix and a dictionary which can describe the training samples, but also require the dictionary to have generalization ability. Otherwise, it will increase the missed detection and misjudgment of abnormal events.

The paper is organized as follows: Section 1 presents an overview of the background of the topic. Section 2 introduces the related work. Section 3 discusses the sparse coding algorithm for video anomaly detection. In Sect. 4, this paper refers to the K-SVD algorithm and the BK-SVD algorithm in details, followed by comparison of experimental results and conclusion in Sect. 5.

### 3 Sparse Coding Based on Hierarchical Clustering

#### 3.1 Several Basic Clustering Methods

Clustering is to divide a data set into several disjoint subsets according to the specified criteria. There are great similarities among data in the same subset and little similarities between different subsets.

**Prototype Clustering.** The K-means algorithm [14] is a typical prototype clustering algorithm. The clustering process is the process of minimizing the square error. At this time, each object closely surrounds the mean of all objects within the cluster. The process of clustering is not easy, because there is no prior knowledge about this dataset and there is not any tag information. Therefore, it needs to find the cluster corresponding to the least square error which is actually an NP-hard problem. Applying the greedy algorithm to solve such clustering problems and find the optimal clustering results by continuous iterative optimization. The learning vector quantization algorithm [15, 16] is also a type of prototype clustering, but the difference is that it assumes that the data samples have category markers.

**Hierarchical Clustering.** The [17–20] is divided into two categories: “top-down” and “bottom-up” [21]. The “top-down” algorithm, also known as the divisive method, treats the collection to be classified as a large class. And after each iteration, smaller classes will be generated. The “bottom-up” algorithm can also be known as the agglomerative method. In this algorithm, each element in the set to be classified is regarded as a class, and the number of classes decreases every iteration.

The amount of data in this paper is relatively large, and the final number of clusters cannot be determined in advance. Therefore, from the perspective of computational complexity and algorithm speed, hierarchical clustering is more suitable for sparse coding. In this paper, the similarity of sparsity degree is taken as the criterion of clustering, and the maximum size of each clustering set is defined. It is hoped that the objects with the closest sparsity degree are gathered together. In summary, hierarchical agglomerative clustering is more suitable for sparse coding in this paper.

Applying hierarchical clustering to sparse coding can also achieve the purpose of dictionary partitioning. This is because the dictionary atoms corresponding to the sparse coefficient sub-matrix are automatically aggregated into one piece, and finally the purpose of dictionary division is achieved.

### 3.2 Group Sparse Dictionary Learning

Group sparse representation is also an important research direction in machine learning, and its application is very extensive. The combination of group sparse coding and dictionary learning has been widely used in the research of image processing. Applying group sparse and graph rules to medical image denoising and medical image fusion [22] greatly reduces the distortion probability of the image. This paper also introduces group sparse idea into the dictionary encoding.

If all the pixels are used to redisplay an image, the calculation speed cannot meet the requirements of the actual application. This is because there are too many pixels in each frame. Therefore, a part of the pixels must be selected to construct a dictionary. And the original image is obtained by multiplying the dictionary with a sparse coefficient matrix. Assuming that the extracted training features are  $Y = [y_1, \dots, y_n]$ , the corresponding dictionary  $D = [d_1, \dots, d_n] = [D_1, \dots, D_g]$ ,  $d_i (i = 1, \dots, n)$  is the column vector of dictionary  $D$ ,  $D_i (i = 1, \dots, g)$  is the sub-dictionary, and the sparse coefficient matrix  $X = [X_1^T, \dots, X_L^T]^T$ , then  $Y$  can be represented by the dictionary  $D$  and the sparse coefficient matrix  $X$  in the formula (1).

$$Y = DX = (D_1, \dots, D_g) \begin{pmatrix} X_1^T \\ \vdots \\ X_g^T \end{pmatrix}^T \tag{1}$$

Among them,  $X_i(i = 1 \dots g)$  can be solved by the optimization problem shown by Eq. (2), and its physical meaning is that each sub-matrix in the sparse coefficient matrix is desirably as sparse as possible.

$$\min_X \|X\|_{0,i} \quad \text{subject to} \quad Y = DX, \quad i = 1, \dots, g \tag{2}$$

In addition, the L0 norm’s calculation is difficult, so the L0 norm will be relaxed to the L1 norm. And the solution problem of  $X_i$  will be solved by the Eq. (3). Where  $k$  is an arbitrary small positive real, and the smaller the value of  $k$ , the smaller the reconstruction error.

$$\min_X \|X\|_{1,g} \quad \text{subject to} \quad \|Y - DX\|_2^2 \leq k \tag{3}$$

### 3.3 Sparse Coding Based on Hierarchical Clustering

In the sparse coding process, this paper uses the BOMP algorithm to obtain the sparse coefficient matrix. The BOMP algorithm [23, 24] is a variant of the OMP algorithm [25, 26], which integrates the grouping idea into the OMP algorithm. In the BOMP algorithm, the value of a sparse coefficient sub-matrix is updated, which improve the algorithm speed.

Different from the general block sparse dictionary learning, this paper considers the structural information of the sparse coefficient matrix itself when dividing the block. We know that normal behavior is common and the distribution is relatively regular, so their corresponding sparse coefficient matrices are similar. We can consider the row vector with “short distance” as the clustering object in the sparse coefficient matrix. This paper stipulates that the maximum number of matrix rows is  $s$  after clustering, so that the  $s$  rows with the most similar sparsity are gathered together.

Since there is no prior knowledge about the sparse coefficient matrix, we first need to initialize the clustering block and get the initial sparse coefficient matrix by OMP algorithm. Then, under the condition that the sparse coefficient sub-matrix is required to be as sparse as possible, sub-matrixes just obtained are clustered again to obtain a new block partitioning result. Finally, the obtained block is used to perform iterative updating by using the BOMP algorithm to obtain a sparse coefficient matrix.

Suppose the training feature is  $Y = [y_1, \dots, y_n]$ , the corresponding dictionary  $D = [d_1, \dots, d_n] = [D_1, \dots, D_L]$ ,  $d_i (i = 1, \dots, n)$  is the column vector of dictionary  $D$ ,  $D_i (i = 1, \dots, L)$  is the sub-dictionary after dividing the block, and  $L$  is the number of blocks in the dictionary. The sparse coefficient matrix  $X = [X_1^T, \dots, X_L^T]^T$ ,  $P_j (j = 1, \dots, L)$  is the number of row vectors included in each sparse coefficient sub-matrix, where  $L$  is the number of blocks of the sparse coefficient matrix  $X$ .  $l$  is the sparsity of each sparse coefficient sub-matrix, which means that the non-zero element’s number of each column in the sparse coefficient sub-matrix does not exceed  $l$ . Therefore, the

construction of the dictionary  $D$  and the sparse coefficient matrix  $X$  in this paper needs to satisfy the formula (4).

$$\min_{D,X,L} \|Y - DX\|_2^2 \quad \text{subject to} \quad \|X_m\|_{0,n} \leq l, P_j \leq s \quad (4)$$

Where  $m = 1, \dots, G$  ( $G$  is the number of columns in the sparse coefficient sub-matrix  $X_n$  ( $n = 1, \dots, L$ )),  $\|X\|_{0,n} \leq l$  means that the sum of the number of non-zero elements in each column in each sparse coefficient sub-matrix need to be less than  $l$  which is specified in this paper. In the clustering process, each sparse coefficient sub-matrix should not include more than  $s$  rows and the non-zero element's number of each column does not exceed  $l$ .

The following article describes in detail how the clustering process is implemented in the algorithm. Assume that the set of column indices of non-zero element positions in each row is  $R_i$  ( $i = 1, \dots, L$ ), where  $L$  is the number of blocks of the sparse coefficient matrix  $X$ . In this paper, we follow Eq. (5) to find out the two rows (or two block) with the closest sparsity.

$$[u, v] = \arg \max_{u \neq v} (R_u \cap R_v) \quad \text{subject to} \quad P^* \leq s \quad (5)$$

Where  $u, v$  are the two blocks or two rows with the closest sparsity, and  $P^*$  is the number of row vectors included in the new set to be found. If two blocks or two rows found satisfy the requirements of Eqs. (3–5), the two blocks or two rows are combined into one block and the corresponding row index will be removed from the sparse coefficient matrix.

In this paper, the original intention of clustering the sparse coefficient matrix lies in the division of the corresponding dictionary. The rows in the sparse coefficient matrix are divided into blocks as cluster objects, so that the columns in the corresponding dictionary can automatically generate sub-dictionaries. The clustering process of the sparse coefficient matrix is shown in Fig. 1.

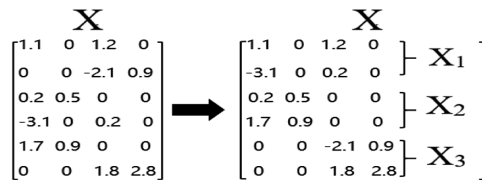


Fig. 1. Clustering process of sparse coefficient matrix

In fact, the difference from the general group sparse algorithm is that this paper not only considers the structural characteristics of the sparse coefficient matrix, but also limits the sparsity of each column in the sparse coefficient sub-matrix in the clustering process.

## 4 Dictionary Learning Based on Block K-Singular Value Decomposition

### 4.1 BK-SVD Algorithm

The BK-SVD algorithm (block K-singular value decomposition algorithm) is a variant of the K-SVD algorithm. This algorithm updates the dictionary atom block by block. The hierarchical clustering has divided the sparse coefficient matrix into blocks, which corresponds to the sub-block of the dictionary. So the sub-dictionary is also obtained. When updating each block, the remaining blocks are guaranteed to be fixed, and the singular value analysis is performed on the current error term.

### 4.2 The K-SVD and BK-SVD Algorithm

The update of the sparse coefficient matrix and the dictionary is actually performed crosswise, which means that after the dictionary initialization, the BOMP algorithm is used to find the sparse coefficient matrix at the time when keeping D unchanged. After that, we keep the updated sparse coefficient matrix unchanged and update the dictionary D with the BK-SVD algorithm. It is assumed that the nth dictionary learning is performed, and the sparse coefficient matrix is the one obtained by updating the n-1th step. After using the sparse coefficient matrix obtained, the dictionary in the iterative process can be obtained.

$$X^{(n)} = \arg \min_x \|Y - D^{(n-1)}X\|_2^2 \quad \text{subject to} \quad \|X_i\|_{0,j} \leq l, P^* \leq s \quad (6)$$

Where  $i = 1, \dots, G$  (the number of columns of the sparse coefficient matrix),  $j = 1, \dots, L$  (the number of blocks in the dictionary). This update step needs to be done with the help of the BOMP algorithm.

$$D^{(n)} = \arg \min_D \|Y - DX^{(n)}\|_2^2 \quad (7)$$

Equation (7) shows that after the sparse coefficient matrix is obtained in the nth iteration, the target dictionary is obtained with the minimum reconstruction error. In this experiment, after 50 iterations, the dictionary with good performance and the sparse coefficient matrix are obtained. Assume that when the  $j$  ( $j \in [1, L]$ ) block is updated, the influence of the  $j$ th block is removed. The difference between the reconstructed feature and the training feature is  $W_j = Y - \sum_{i \neq j} D_i X_i^T$ , and the reconstruction error of the training feature is  $E = \|W_j - D_j X_j^T\|_2^2$ . The singular value decomposition is performed on the reconstruction error, then  $W_j = U \Delta V'$ . At this time, the general form of the dictionary D and the sparse coefficient matrix is obtained, as shown in Eq. (8).



$$\begin{aligned} D_j &= [U_1, \dots, U_t] \\ X_j^T &= [\Delta_1^1 V_1, \dots, \Delta_t^t V_t]^T \end{aligned} \quad (8)$$

Equations (4–3) is the result of applying the BK-SVD algorithm to dictionary learning, where  $t$  is the number of columns corresponding to the current block in the dictionary. The singular value decomposition of the corresponding error of each block is better than the singular value decomposition of the corresponding error of a single column, because the final result in former case is more likely to achieve global optimality. So this is part of the reason why the BK-SVD algorithm is better than the K-SVD algorithm.

### 4.3 Judgment of Abnormal Events

In general, the method of judging anomalies is mainly through reconstruction error. If the reconstruction error is bigger than a given threshold, the test feature is judged to be abnormal; otherwise, the test feature is judged to be normal. The reconstruction error  $E$  is represented by Eq. (9), which is basically the most general representation of the reconstruction error. But different algorithms will deform it to more easily and accurately detect abnormal behavior.

$$E = \|Y - DX\|_2^2 \quad (9)$$

Since it is desirable to minimize the reconstruction error (even close to zero), the corresponding sparse coefficient matrix can be found by the least squares method. The general representation of the sparse coefficient matrix solved by the least squares method is given by Eq. (10).

$$X = (D^T D)^{-1} D^T Y \quad (10)$$

After the Eq. (10) is obtained, the reconstruction error  $E$  can be rewritten into the form of the Eq. (11). We can see that the reconstruction error at this time is only related to the dictionary  $D$  and the test feature  $Y$ . Although this kind of solution brings some errors to some extent, it provides great convenience for the anomaly detection in this paper. From the perspective of the simplicity of the algorithm, this paper chooses this method.

$$E = \|Y - DX\|_2^2 = \left\| \left( D(D^T D)^{-1} D^T - I \right) Y \right\|_2^2 \quad (11)$$

Where,  $I$  is the identity matrix. We define an auxiliary variable  $F$  to facilitate the operation.

$$F = \left( D(D^T D)^{-1} D^T - I \right) \quad (12)$$

Therefore, when the reconstruction error  $E = \|FY\|_2^2$  is bigger than a given threshold, it is judged as an abnormal event; otherwise, it is judged as a normal event.

The above is the main idea of this algorithm. This algorithm has made some improvements to the more advanced algorithms to some extent. The following paper will prove the advancement of this algorithm in video anomaly detection through experiments.

## 5 Experiments

### 5.1 Feature Extraction

The data set used in this paper is the Avenue data set [5], which contains 16 training videos and 21 test videos. Each frame has 120\*160 pixels. For each video, this paper first divides it into 10\*10\*5 pixel blocks, which means that each frame of 5 consecutive frames consists of 10\*10 small pixel blocks. It need to be judged whether there is motion behavior in each pixel block. If there is motion behavior, it means that this part of the feature is relatively significant, and the feature will be extracted. The process of extracting features is actually a compression process as for each frame in the original video. In this process, this paper hopes that the finally obtained features can ensure the accurate restoration of the original video as much as possible. The features extracted in this way have a high advantage for the retention of information. However, the number of such features is still a considerable amount. So after the preliminary feature extraction, 500\*16000 training features can be obtained. The PCA algorithm is used to reduce the dimension. Finally 10\*16000 training features are obtained, which will be used to train the dictionary in this paper.

### 5.2 Parameter Settings

In the end, this paper hopes to build a 10\*96 dictionary. Although the size of the dictionary will have certain deficiencies as for the accuracy of the restoration training features, it is very beneficial to the speed of the algorithm. This article sets the value of  $s$  to 3, that means each block in the dictionary does not include more than three columns. The setting of this value is affected by the size of the dictionary.

In this paper, the number of iterations of the dictionary is set to 50. In each iteration process, this paper runs sparse coding based on hierarchical clustering and dictionary learning algorithm based on BK-SVD. In the experiment, it was found that even if iterating 50 times, the speed of the algorithm is still relatively fast. After 50 iterations, we can get a dictionary with better performance.

In this experiment, the value of  $l$  is set to 2, which means that the number of non-zeros of each column in each sparse coefficient sub-matrix does not exceed 2. The size of this value is actually affected by the size of the  $s$  value. In this experiment, the value of  $s$  is 3, which means that each block in the sparse coefficient matrix has only 3 rows, so the value of  $l$  must be less than the value of  $s$ . When the value of  $l$  is 1, the requirement for the sparse condition is too high, because it may cause a large loss of the original information, so the

value of  $\lambda$  is set to 2. It will ensure that the coefficient matrix is as sparse as possible and the original information is not lost as much as possible.

### 5.3 Frame-Level and Pixel-Level ROC Curves

At the frame level, the abnormal frame is defined as follows: If one of the pixels in the frame is abnormal, the frame is abnormal. The frame-level ROC curve in this paper is shown in Fig. 2. The AUC value corresponding to this ROC curve is 0.75218.

At the pixel level, the criterion for determining a frame as anomalies is as follows: When the frame in Groundtruth is abnormal, if more than 40% of the pixels in the testing frame are abnormal, then the frame is judged as Abnormal; when the frame in Groundtruth is normal, if one of the pixels of the testing frame is abnormal, the frame is judged to be abnormal. The pixel-level ROC curve in this paper is shown in Fig. 3. The AUC value corresponding to this ROC curve is 0.5701.

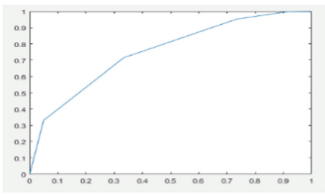


Fig. 2. Frame-level ROC curve

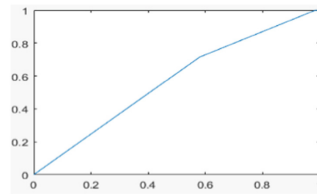


Fig. 3. Pixel-level ROC curve

Table 1. Frame-level and pixel-level AUC values

Algorithm	Frame-level AUC(%)	Pixel-level AUC(%)
Lu's algorithm [5]	81.75	63.3
Ours	75.2	57.0

The AUC values of the present algorithm and Lu's algorithm at the frame level as well as at the pixel level are listed in Table 1.

It can be seen from Table 1 that although the detection accuracy of the algorithm at the frame level and the pixel level is lower than that of the Lu's algorithm, the

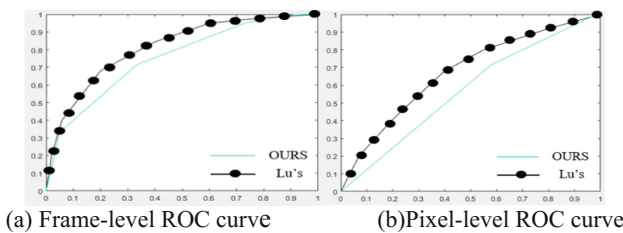


Fig. 4. ROC curve of two algorithms at the frame level and pixel level

algorithm still has a satisfactory detection result. In the future, we can further improve the detection accuracy of the algorithm by other methods. Figure 4 shows the ROC curves of the two algorithms at the frame and pixel levels.

**Acknowledgments.** This work was supported by National Natural Science Foundation of China under Grant 61731003, the projects of International Cooperation and Exchanges NFSC under Grant 61520106002.

## References

1. Jiang, F., Wu, Y., Katsaggelos, A.K.: A dynamic hierarchical clustering method for trajectory-based unusual video event detection. *IEEE Trans. Image Process.* **18**(4), 907–13 (2009). Publication of the IEEE Signal Processing Society
2. Ren, H., Liu, W., Olsen, S.I., et al.: Unsupervised behavior-specific dictionary learning for abnormal event detection. *IEEE Trans. Signal Process.* **17**(2), 99–111 (2015)
3. Yuan, Y., Feng, Y., Lu, X.: Structured dictionary learning for abnormal event detection in crowded scenes. *Pattern Recogn.* **62**(4), 129–138 (2018)
4. Zhang, Z., Mei, X., Xiao, B.: Abnormal event detection via compact low-rank sparse learning. *IEEE Intell. Syst.* **31**(2), 29–36 (2016)
5. Lu, C., Shi, J., Jia, J.: Abnormal event detection at 150 FPS in MATLAB. In: *IEEE International Conference on Computer Vision*, pp. 2720–2727. IEEE (2014)
6. Zelnik-Manor, L., Rosenblum, K., Eldar, Y.C.: *Dictionary Optimization for Block-Sparse Representations*, pp. 34–47. IEEE Press, Piscataway (2012)
7. Ren, H., Pan, H., Olsen, S.I., et al.: An in-depth study of sparse codes on abnormality detection. *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 66–72. IEEE Computer Society (2016)
8. Aharon, M., Elad, M., Bruckstein, A.: The K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006)
9. Li, S., Yin, H., Fang, L.: Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **45**(23), 531–541 (2015)
10. Olshausen, B.A., Field, D.J.: Natural image statistics and efficient coding. *Network: Comput. Neural Syst.* **2**(7), 333–339 (1996)
11. Lesage, S., Gribonval, R., Bimbot, F., et al.: Learning unions of orthonormal bases with thresholded singular value decomposition. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 293–296. IEEE (2005)
12. Duarte-Carvajalino, J.M., Sapiro, G.: Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Trans. Image Process.* **18**(7), 1395 (2009)
13. Kreutz-Delgado, K., Murray, J.F., Rao, B.D., et al.: Dictionary learning algorithms for sparse representation. *Neural Comput.* **15**(2), 349–396 (2003)
14. Hartigan, J.A.: Means clustering algorithm. *Appl. Stat.* **28**(1), 100–108 (1979)
15. Asikainen, A., Kolehmainen, M., Ruuskanen, J., et al.: Structure-based classification of active and inactive estrogenic compounds by decision tree, LVQ and kNN methods. *Chemosphere* **62**(4), 658–673 (2006)
16. Ganesh Murthy, C.N.S.: Classification of encoded patterns using constructive learning algorithms based on learning vector quantization (LVQ). *Tech. Rep.* **65**(11), 245–257 (1996)

17. Liu, Y., Xu, H., Yi, H., et al.: Network anomaly detection based on dynamic hierarchical clustering of cross domain data. In: IEEE International Conference on Software Quality, Reliability and Security Companion, pp. 200–204. IEEE (2017)
18. Ren, W.W., Liang, H., Zhao, K., et al.: An efficient parallel anomaly detection algorithm based on hierarchical clustering. *J. Networks* **8**(3), 672–679 (2013)
19. Hu, L., Ren, W.W., Ren, F.: An adaptive anomaly detection based on hierarchical clustering. In: International Conference on Information Science & Engineering, pp. 1626–1629. IEEE (2009)
20. Chen, W., Liu, X., Li, T., et al.: A negative selection algorithm based on hierarchical clustering of self set and its application in anomaly detection. *Int. J. Comput. Intell. Syst.* **4**(4), 410–419 (2011)
21. 周爱武, 潘勇, 崔丹丹等.: AGNES算法在K-means算法中的应用. *微型机与应用* **30**(23), 79–81 (2011)
22. Li, S., Yin, H., Fang, L.: Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **59**(12), 3450–3459 (2012)
23. Eldar, Y.C., Kuppinger, P., Bölcskei, H.: Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.* **58**(6), 3042–3054 (2010)
24. Eldar, Y.C., Bölcskei, H.: Block-sparsity: coherence and efficient recovery. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2885–2888. IEEE (2009)
25. Goklani, H.S.: A review on image reconstruction using compressed sensing algorithms: OMP, CoSaMP and NIHT. *Int. J. Image Graphics Sig. Process.* **9**(8), 30–41 (2017)
26. Xu, Y., Sun, G., Geng, T., et al.: An improved method for OMP-based algorithm; using fusing strategy. In: IEEE, International Colloquium on Signal Processing & ITS Applications, pp. 202–207. IEEE (2017)