Research Article **Open Access**

Object Detection Algorithm in Underground Mine Based on Sparse Representation

Yan Lu* and Qing-xiang Huang

School of Energy Engineering, Xi'an University of Science and Technology, Xi'an 710054, China

***Corresponding author:** Yan Lu, School of Energy Engineering, Xi'an University of Science and Technology, Xi'an 710054, China, E-mail: tyccly123@126.com

Abstract

This paper puts forward an improved K-SVD object detection algorithm for the problem of multiple noise sources in underground mine video. Firstly, the background modeling is applied in the video; then, the improved non-local mean filtering algorithm is used to enhance the image quality; finally, the improved image is processed by the sparse representation algorithm to further detect the moving object. In order to verify the effectiveness of the proposed algorithm, the algorithm and other algorithms are applied to video object detection in two different scenarios. The experimental results show that, in the underground mine video, the proposed algorithm can increase the accuracy by more than 8% compared with the traditional K-SVD algorithm, and the proportion of error points decreases by about 25%. Better detection of the moving object is achieved by the proposed algorithm.

Keywords: Underground mine; Background modeling; Image denoising; Sparse representation; Object detection; Non-local mean filtering

Introduction

The purpose of moving object detection is to extract the foreground area that people are interested in in a video stream, that is, the moving object, in order to deal with the next step of object recognition, tracking, behaviour analysis and so on [1]. At the same time, the intelligent vision technology was introduced in the underground video surveillance system to further improve its digitalization and intelligence level, thereby reducing the injury and death rate of accident and ensuring the safety of coal mine production [2,3].

In recent years, domestic and foreign researchers have made intensive achievements in object detection. Morioka et al. [4] proposed to use colour histograms as regional features for matching, however, colour features are more affected by light and accordingly it is not robust. Bouzaraa et al. [5] proposed to use a statistical property of reference images to improve the accuracy of histogram matching, but the experiments were performed under outdoor conditions without considering the dim underground mine environment. Cong et al. [6] studied the background difference method of K-SVD (K-Singular Value Decomposition) [7,8] for dictionary learning, and achieved better detection results, but the ability to background model adaptive and anti-counterfeiting prospects is not strong enough, and not be suitable for low-illumination, high-noise underground mine complex environments. These detection algorithms only consider the illumination and noise changes in the outdoor environment and are not suitable for coal mine scenarios [9]. Therefore, for the characteristics of underground mine video, Wang et al. [10] proposed a combination of YOLOv2 (You Only Look Once, YOLO) and FCN (Fully Convolutional Networks) to improve the accuracy of underground mine small pedestrian detection, but there is a correlation between the accuracy of detecting small pedestrians and the speed of pedestrians for this algorithms. According to the characteristics of video underground mine, an algorithm for the linear fusion of colour and edge is proposed by Xian et al. [11]. It overcomes the detection problem of miners and background grey scale and can remove false moving objects caused by illumination changes. However, the detection threshold of the algorithm needs to be manually selected, and it is difficult to have the optimal results.

To sum up, this paper applies the sparse representation algorithm to the specific situation of underground intelligent video surveillance system. Considering the characteristics of the captured videos, an improved K-SVD by considering the non-local mean filtering algorithm is proposed. Major steps are: (1) the background is modelled. (2) The image to be detected is filtered to obtain a high-quality object image, and the sparse representation algorithm is used to detect the moving object. (3) The video in different scenes is detected to verify the effectiveness of being proposed algorithm. Experimental results show that our new algorithm can be used to detect moving objects in underground mine video more accurately.

Method

Image sparse representation

Sparse representation was originally proposed in the field of signal processing [12,13]. The core idea is to select a small number of basis vectors in a large number of base vector sets to represent the object signal and to obtain high compression ratio while ensuring high fidelity of the signal. The image sparse representation is to represent the image in the most concise way, that is, most of the coefficients are zero after expression, leaving only a few non-zero coefficients, and the non-zero coefficients can reveal the inherent structure and essential properties of the image.

Set the original signal $x \in R^M$, data dictionary $D = [d_1, d_2, ..., d_L] \in R^{M \times L}$, where $dd_i(i = 1, 2, ...L)$, is atom of data dictionary. In this case, each signal can be represented as a linear

combination of dictionary atoms, as shown in equation (1):

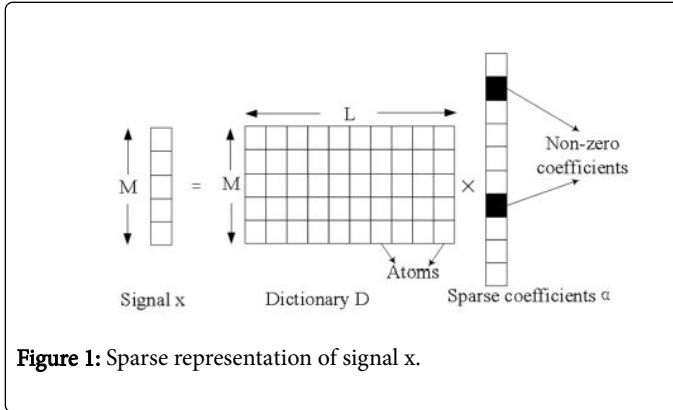$$x = D\alpha = \sum_{i=1}^{L} d_i \alpha_i \qquad (1)$$

Figure 1: Sparse representation of signal x.

Where $\alpha = [\alpha_1, \alpha_2, ... \alpha_L]^T$ is the sparse representation of the signal x, and also known as the sparse representation coefficients. If α has only a finite number of (such as k) non-zero elements, and all other elements are zero, then the sparsity of α is k. The signal $x \in R^M$ is sparsely represented by the dictionary D as shown in Figure 1.

The meaning of image sparse representation (also known as sparse coding) is that the coefficients represented by the dictionary are sparse, that is, only a small number of atoms are needed to uniquely represent the image linearly. Its mathematical model is expressed as:

$$\alpha = \arg\min \| \alpha \|_0, \quad s.t \| x - D\alpha \|_2^2 \leq \varepsilon \qquad (2)$$

Where $\| \alpha \|_0$ represents the norm of l_0 and is the number of zero elements in the α. ε refers to the sparse representation error and the error constraint term $\| x - D\alpha \|_2^2 \leq \varepsilon$ guarantees that the coefficient α is as close as possible to the original image signal x. Due to the non-convexity of the norm of l_0, the solution to this problem is essentially a NP-Hard combinatorial optimization problem [14]. At present, commonly used methods mainly include Basis Pursuit (BP) [15], FOCUSS [16], Matching Pursuit (MP) [17], and Orthogonal Matching Tracking (OMP) [18] etc. Among them, the BP algorithm solves the above problem by changing l_0 into l_1 the formula (2) can be rewritten as: $\alpha = \arg\min \| x - D\alpha \|_2^2 + \mu \| \alpha \|_1 \qquad (3)$

Where $\| \alpha \|_1 = \sum_i \alpha_i$, μ is called the penalty factor.

Dictionary learning

The sparsity of a signal is relative to its expression domain. In terms of image signals, it is usually not sparse in the spatial domain (pixel value), but in a specific dictionary, the expression of the signal become possibly sparse [19]. Therefore, whether or not the image signal can be sparsely represented depends largely on how the dictionary is constructed. There are two ways to construct a sparse representation dictionary [20]. The first type of method is a base dictionary. The dictionary is calculated by a mathematical model and represented by an implicit matrix. These types of dictionaries commonly use wavelet, curve wave, contour wave and other forms, but only when the

characteristics of the signal match the dictionary, it can get better results; the second method is based on the learning method, through which a more specific dictionary can be constructed. Therefore, this paper also uses this method to construct the required dictionary. When constructing a dictionary, the atoms in it also need to meet two requirements: Firstly, in order to linearly represent the information in the M-dimensional space of the original signal, these dictionary atoms are required to form the entire M-dimensional space; secondly, in order to ensure uniqueness of signal sparse representation, it requires that each dictionary atom is linearly independent. The base dictionary method includes a Fourier transform dictionary, a discrete cosine transform dictionary, a wavelet transform dictionary, etc.; the learning class dictionary includes a principal component analysis dictionary, a K-SVD dictionary, an Independent Component Analysis (ICA) dictionary, and so on.

In practical applications, the corresponding dictionary design methods should be selected according to the actual situation. The choice of dictionary is generally divided into three cases: 1. when there is no training data, some transforms are used as a dictionary, that is, a base dictionary method; 2. if there is a large amount of training data, a learning class dictionary can be utilized. In this paper, the K-SVD dictionary learning algorithm is adopted. Firstly, the background is modeled, and then the detected image is processed by removing noise, and the BP algorithm in the sparse representation algorithm is used to solve the sparse coefficient of equation (3). Finally, the K-SVD algorithm is adopted. The error is calculated and the kth atom dk in dictionary D is updated, and the coefficient k α_T^k corresponding to the kth row is also updated.

Image denoising algorithm

Due to the low illumination, multi-noise and high noise in the mine environment, it has brought many difficulties to the detection of moving objects. Therefore, this paper proposes an Improved Non-Local Mean (INLM) filtering algorithm. First, the image is reconstructed by using the low-frequency coefficients of DCT to filter out some noise while protecting the main content of the image [21]. Then, use the Mahalanobis distance to improve the NLM algorithm [22], and the denoising processing was applied to the object, so that the image has better quality before detection. Finally, sparse representation was implemented, so as to improve the accuracy of moving object detection.

Discrete cosine transformation

DCT is an orthogonal transform whose transform kernel is a real cosine function. After discrete cosine transform of an image, much important visual information about the image is concentrated in a small part of the low-frequency coefficients of the DCT transform.

Given two-dimensional signal f of size $N \times N$, its DCT is defined as:

$$F(u,v) = \frac{2}{N} c(u) c(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \times \cos[\frac{(2i+1)u\pi}{2N}]\cos$$

$$[\frac{(2j+1)v\pi}{2N}] \qquad (4)$$

The Inverse Discrete Cosine Transform (IDCT) is defined as:

$$F(i, j) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c(u) c(v) f(u,v) cos[\frac{(2i+1)\mu\pi}{2N}] \times cos$$

$$[\frac{(2j+1)v\pi}{2N}] \qquad (5)$$

In the above formula

$$c(u) = \begin{cases} \frac{1}{\sqrt{2}}, u = o \\ 1, u = 1, 2, ..., N-1 \end{cases}$$

$$c(v) = \begin{cases} \frac{1}{\sqrt{2}}, v = o \\ 1, v = 1, 2, ..., N-1 \end{cases}$$

It is because the discrete cosine transform has a strong "energy concentration" characteristic, the reconstructed image reconstructed by a small number of low-frequency DCT coefficients can be used to replace the noise image to calculate the similarity weight value between the image blocks.

NLM algorithm

Classical NLM filtering, when filtering the additive noise image, the Gaussian weighted Euclidean distance between each pixel in the non-local large-scale searching window and the central pixel under a small-scale similar window was estimated. And the degree of similarity between the pixels and the center pixel was measured. This will be set as the main parameter for determining the size of each pixel weighting coefficient.

Given at position i, the observation value is $X(i)$, and useful signal is $S(i)$, then the additive noise image model is

$$X(i) = S(i) + N(i) \qquad (6)$$

Then, the classical NLM filtering algorithm performs non-local weighted average filtering on the image i position pixels, which can be described as

$$\bar{X}_{NLM}(i) = \sum_{j \in \Omega(i)} \omega(i,j)X(j) \qquad (7)$$

Where $\Omega(i)$ represents a large-scale search window centered on the center of i; $\omega(i,j)$) represents the weighting coefficient of j pixels to i pixels in the search window. The calculation of the weighting factor $\omega(i,j)$ can be expressed as

$$\omega(i,j) = \frac{1}{B(i)}exp(\frac{D_{GWED}(i,j)}{h^2}) \qquad (8)$$

Where h refers to control the decay rate of the exponential function, it determines the degree of filtering; $B(i)$ refers to the regularization factor and it can be expressed as

$$B(i) = \sum_{j \in \Omega(i)} exp(-\frac{D_{GWED}(i,j)}{h^2}) \qquad (9)$$

Where $D_{GWED}(i,j)$ represents the Gaussian weighted Euclidean distance between i pixels and j pixels in the search window of $\Omega(i)$, which can be expressed as

$$D_{GWED}(i,j) = \|Y(\psi(i)) - Y(\psi(j))\|_{2,\sigma}^2 \qquad (10)$$

Where $Y(\psi(i))\psi(i)$ and represent small scale similar windows centered on i, j, respectively. $Y(\psi(i))$ and $Y(\psi(j))$ represent the image

observation value matrix at the similar window $\psi(i)$ and $\psi(j)$ represents the Gaussian weighted norm with the calculated standard deviation of σ^2.

Mahalanobis distance

The Mahalanobis distance proposed by Indian statistician Mahalanobis is a more effective and reliable method of measuring the similarity of sample data than the Euclidean distance. It overcomes the shortcomings of Euclidean distance. The Mahalanobis distance eliminates the interference of correlation between variables by standardizing the sample data, and more effectively measures the similarity between the sample data.

Assuming that the two sets of sample data i, j constitute the column vectors x_i and x_j, the Mahalanobis distance d_{ij} between the two vectors is defined as $d_{ij} = \sqrt{(x_i - x_j)^T S^{-1}(x_i - x_j)}$ (11)

Where S is the covariance matrix of vectors x_i and x_j,

$$S = (x_{ij} - x_m)(x_{ij} - x_m)^T \qquad (12)$$

Where $x_{ij} = (x_i \ x_j), x_m = 1/2((x_i \ x_i) + (x_j \ x_j))$. It can be seen from equation (5) that when S is a unit matrix, the Mahalanobis distance degenerates to the Euclidean distance, so the Mahalanobis distance is an extension of the Euclidean distance.

Generalizing calculation of Mahalanobis distance

When calculating the Mahalanobis distance, the inverse matrix of the covariance matrix S in equation (5) does not exist, so the Mahalanobis distance defined by equation (4) is unstable. To solve this problem, a generalized inverse matrix of S is introduced. By implementing singular value decomposition on equation (5), one obtains

$$S = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^H \qquad (13)$$

Where U, V matrix is unitary matrix. $\Sigma = diag(\lambda_1, \lambda_2, ..., \lambda_r), \lambda_i(i=1,2,....,r)$ is a non-zero singularity of the matrix and $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_r, r$ is the rank of S; and the superscript H indicates that the matrix is conjugate transposed. By solving Moore-Penrose inverse matrix S^+ in equation (6), one gets

$$S^+ = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^H \qquad (14)$$

Where $\Sigma^{-1} = diag(\lambda_1^{-1}, \lambda_2^{-1}, ..., \lambda_r^{-1})$. Equation (7) is a generalization of the feature space pair for S^{-1} by using S^+ to substitute S^{-1} in Equation (4), one gets the Mahalanobis distance in the feature space $d_{ij} = \sqrt{(x_i - x_j)^T S^+ (x_i - x_j)}$ (15)

In the $\Sigma = diag(\lambda_1 \lambda_2 .... \lambda_r)$ value of $\lambda_i(i=1,2,....,r)$ represents the amount of energy of each unrelated data component. For image data, $\lambda_i(i=1,2,....,r)$ represents the variance of each irrelevant component information in the image. It was found that λ_1 is always

much larger than other values, i.e. $\lambda_1 \gg \lambda_i (\lambda_i \neq \lambda_1)$. When calculating, we can just keep λ_1 and ignore $\lambda_i (\lambda_1 \neq \lambda_i)$ and think that other singular values are zero. This not only effectively reduces the data dimension, but also suppresses interference from non-useful signals, i.e. noise disturbance. According to the above analysis, one gets

$$\Sigma \approx \lambda_1 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \qquad (16)$$

Invert the equation (9), one gets $\Sigma^{-1} = \dfrac{1}{\lambda_1}\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$. By substituting Σ^{-1} into equation (7) and deduce it, we can get

$$S^+ = \frac{1}{\lambda_1}V\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}U^H = \frac{1}{\lambda_1^2}S \qquad (17)$$

By substituting equation (10) into equation (8), we get improved Mahalanobis distance:

$$d_{ij} = \frac{1}{\lambda_1}\sqrt{\left(x_i - x_j\right)^T\left(x_i - x_j\right)} \qquad (18)$$

Equation (11) not only avoids the instability of the Mahalanobis distance, but also simplifies the calculation and suppresses the noise interference. Because of the symmetry of the Mahalanobis distance, symmetry can be used to halve the computational complexity without sacrificing denoising performance.

Detection of Moving Objects Based on Improved Sparse Representation

Background modelling

When the camera is fixed, the background area can be regarded as static, and only the pixel value of the position where the moving object is located will change. The successive frames are combined by column vector, and the background values of adjacent columns are equal if the interference of noise was ignored. At this time, the image in the current frame can be represented by the foreground object and the background image. Assuming that the current frame image is X, the background image is X_b, and the foreground image of the motion is X_f, then the image at this time can be expressed as:

$$X = X_b + X_f \qquad (19)$$

However, the actual image in underground mine generally has noise interference. Therefore, in the case of considering noise interference, the current image is composed of three parts. The mathematical expression is as follows:

$$X = X_b + X_f + X_E \qquad (20)$$

Where E_X is the error matrix caused by noise. The background model can be obtained by dictionary learning and updating. The problems of low illumination and multiple noise sources in the underground mine environment have brought many difficulties to the detection of moving objects. Therefore, this paper first uses the Improved Non-Local Means (INLM) filtering mentioned above to remove the noise of the detected object, so that the image has better image quality before detection. Then the sparse representation is

applied to each image, so as to improve the accuracy of detection of the moving object. The background is initially modelled by constructing the minimization constraint function using following equations:

$$\hat{X}_b = \arg\min\left\|X_b - D\alpha\right\|_2^2 + \mu\left\|\alpha\right\|_1 \qquad (21)$$

Where D is the data dictionary to be trained, and each column vector corresponds to a dictionary atom. The initial training dictionary $D(0)$ and the sparse coefficient of the background image Xb decomposed on D are obtained by K-SVD. In equation (21), the first term is to ensure that the image can be approximated by a linear combination of the dictionary D and the coefficient α; the second term is a sparse control to ensure that the sparsest coefficients α are constantly sought during the solution. The product of the optimization object D and α is the initial background sparse representation model $\hat{X}_b$.

The above minimization algorithm requires an initial value of the background X_b, which is obtained by using a multi-frame averaging algorithm is obtained. The mean value of the P-frame training sample image is taken as the b X initial value in equation (21), namely:

$$X^b = \frac{1}{P}\sum_{i-1}^{P} X_i \qquad (22)$$

Where $X_i(i = 1, 2, ..., P)$ represents the P frame training sample image.

Background update

After obtaining the initial background sparse representation model, five adjacent frames in the video sequence are extracted. The iterative update operation of the dictionary atom of dkand the sparse coefficient of α are completed by the K-SVD dictionary learning method. And the optimal data dictionary D is found. The background modelling based on the dictionary is optimally approximated to the observation value of the background modelling of the neighbouring frame image, thereby obtaining an updated optimal background model of $\hat{X}_{new}^b$. That is, $j = t, t-2, t-4, t-6, t-8$. Therefore, the following optimization problem will be solved:

$$\hat{X}_{new}^b = \arg_{D, \alpha}\min\left\|X_j - D\alpha\right\|_2^2 + \mu\left\|\alpha\right\|_1 \qquad (23)$$

In the formula, $X_j(j = t, t-2, t-4, t-6, t-8)$ represents the five adjacent frames of the video sequence at time t.

Background update algorithm

Input: video sequence of X to be detected, initial dictionary of $D^{(0)}$ and sparse coefficient of α.

Steps:

Five adjacent frames $X_j(j = t, t-2, t-4, t-6, t-8)$ of the video sequence to be detected at time t were extracted. Set the iteration counter C, and its initial value is set as 1; Dictionary updating: Based on equation (16), each atom $d_k(k = 1, 2, ..., k)$ in dictionary D is updated step by step using the K-SVD dictionary learning method.

Rewrite equation (23) as

$$\left\| X - D\alpha \right\|_2^2 = \left\| X - \sum_{j=1}^{K} d_j \alpha_T^j \right\|_2^2$$

$$= \left\| \left(X - \sum_{j \neq 1}^{K} d_j \alpha_T^j \right) - d_k \alpha_T^k \right\|_2^2 \qquad (24)$$

$$= \left\| E_k - d_k \alpha_T^k \right\|_2^2$$

Where Ek: the error to the representation of the image caused by removing the atoms. In order to get the minimum value of the above formula, it is necessary to make $d_k \alpha_T^k$ approaching Ek as close as possible. Therefore, the singular value decomposition is performed to Ek, so that $E_k = U \Lambda V^T$. The dictionary atom dk is updated by the first column of U, and the new coefficient α_T^k is obtained by V multiplying $\Lambda(1,1)$. Let $C = C + 1$, If the C is less than the pre-defined cycle number, go to step 2, otherwise exit the loopOutput: Updated optimal background model $\hat{X}_{b-new}$.

Determination of moving object

After the background model is updated, the background modeling difference method of equation (20) is used to perform a differential operation between the current frame image X and the updated background model b new $\hat{X}_{b-new}$ to obtain a foreground moving object.

$$X_f = \begin{cases} 0 \\ \left| X - \hat{X}_{b-new} \right| \end{cases} if \left| X - \hat{X}_{b-new} \right| < T \qquad (25)$$

Experimental Results and Analysis

In this paper, the software MATLAB is used to perform moving object detection experiments for video in two different scenarios. The first video comes from the PETS2001 (Performance Evaluation of Tracking and Surveillance) database. The video captures the pedestrian walking process and takes the common pedestrians in motion as the object to be tested. As can be seen from the video images, the image quality in this video is clear and not affected by the illumination changes. Figure 2 shows the 10th and the 100th frame of the video. It can be found in Figure 2, the illumination is uniform, and the image is not disturbed by excessive noise.
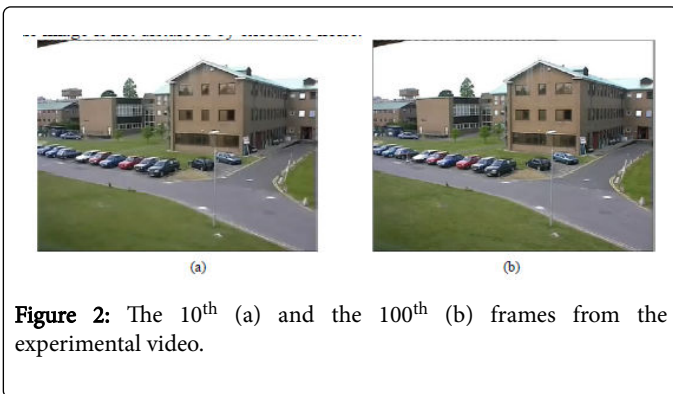
Figure 2: The 10th (a) and the 100th (b) frames from the experimental video.

For the above video, the moving object detection is performed on the video by MoD algorithm, MOG algorithm, traditional K-SVD algorithm and the algorithm in this paper. The obtained results are shown in Figure 3.
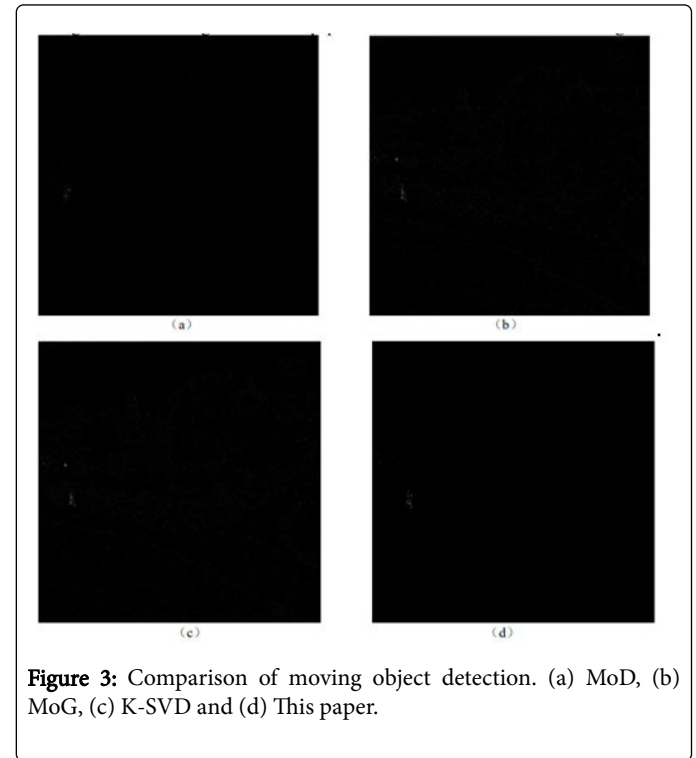
Figure 3: Comparison of moving object detection. (a) MoD, (b) MoG, (c) K-SVD and (d) This paper.

It can be seen from the test results that the object detected by the MoD algorithm is very weak, and the object is hardly seen from the result. The MoG algorithm is more obvious than the MoD algorithm, but there is noise in the picture; for the traditional K-SVD algorithm, the same problem exists. Using the algorithm of this paper, the detected moving object contours and detailed results are clear and there is no noise interference.
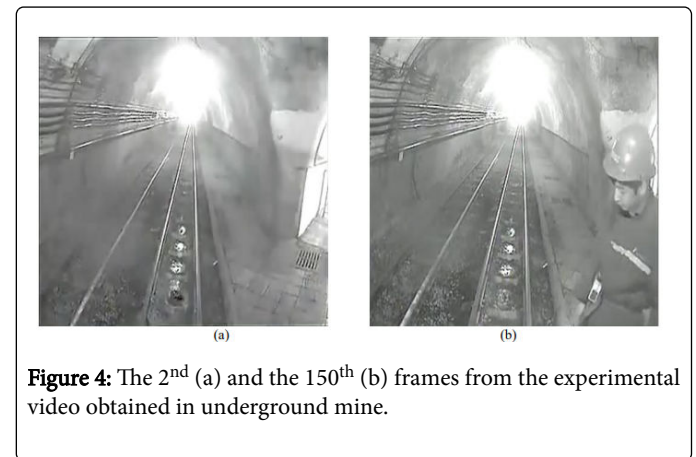
Figure 4: The 2nd (a) and the 150th (b) frames from the experimental video obtained in underground mine.

The video used in experiment 2 was actually taken in the underground mine. The camera is located in the roadway. The object to be detected is also the pedestrian walking in the roadway. The 2nd and the 150th frame image in the video are shown in Figure 4. It can be seen from the video quality used in Experiment 2 is worse than the video quality in experiment 1. The images of video in experiment 2 have several problems such as uneven illumination and many noises.

Due to the uneven illumination and many noises in the underground mine video, the object to be detected will be interfered and affected by these problems. Therefore, the idea of this paper is to use less representative information to reconstruct the detected object. The selected representative information usually has better robust. Even in the case of more interference sources, these information can be well adapted to the complex environment of the underground mine, and can be applied to reconstruct the moving object stably. Since the sparse representation algorithm uses less feature information for recovering the moving object. In this paper, the sparse representation algorithm was introduced to reconstruct the moving object in the underground mine environment.

For the experimental-2 video, the moving object detection is performed by MoD algorithm, MOG algorithm, traditional K-SVD algorithm and the algorithm in this paper. The obtained results are shown in in Figure 5.
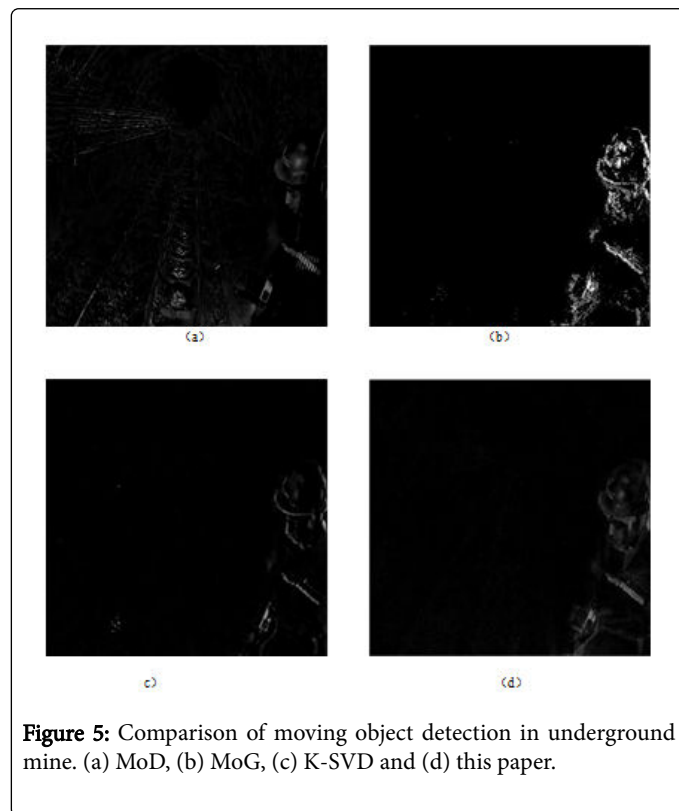

Figure 5: Comparison of moving object detection in underground mine. (a) MoD, (b) MoG, (c) K-SVD and (d) this paper.

It can be seen from the test results in Figure 5, for underground mine video, the change of illumination and the interference of noise will have a corresponding impact on the detection result of the moving object. Figure 5(a) shows the detection results obtained by using the MoD algorithm. It can be seen from the results that due to uneven illumination and multi-noise F interference in the mine video, the detection result contains a lot of background scene information. Such detection results will bring great difficulties for subsequent further processing. Figure 5(b) shows the detection results of the MoG algorithm. It can be seen from the figure that the detection result of the MoG algorithm has been improved to some extent than the MoD algorithm. Most of the detection results are foreground information. Due to the noise interference, there exist some noise information, and the moving object also loses a part of the contour information. Figure 5(c) shows the detection results by using the traditional K-SVD algorithm. It can be seen that in the complex environment in underground mine, the sparse representation algorithm shows better robustness, and the detection results are obviously better than the above two algorithms. However, there is still a problem of noise interference and loss of object local information in the result. Figure 5(d) shows the detection results of the proposed algorithm in this paper. Since the filtering algorithm is applied in advance, the influence of the interference factor has been well eliminated. And the object is well detected with rich details. Results show that the detecting algorithm in this paper is significantly better than the previous three algorithms.

It can be seen from the above two experiments that although the two videos come from different scenes and environments, the algorithm proposed in this paper can detect the moving objects better in both the daily environment and the specific scenes in the underground mine. The detected objects have more detailed information, which provides a good basis for subsequent motion recognition and tracking.

The above results visually illustrate the superiority of the proposed algorithm over the other three algorithms. In order to compare the different detecting algorithms quantitatively, Tables 1 and 2 present the detection error and detection precision, respectively. In this paper, TP is used to indicate the correct number of front attractions, TN is the number of correct background points, FP is the number of wrong front spots, and FN is the number of wrong background points. Therefore, the effectiveness of the proposed can be explained by parameters of PBC and Precision

PBC indicates the proportion of the error points in the object detection. The smaller the value, the higher the accuracy of the segmentation. That is to say, the result of the detection of the moving object is that the wrong pixel is the least. Precision indicates the proportion of the correct number of front attractions in the detected foreground. The higher the value, the more accurate the detected foreground. The calculation results of the PBC and the foreground segmentation precision (Precision) of the four algorithms in the two experiments are shown in Tables 1 and 2.

Video	MoD	MoG	Traditional K-SVD	This paper
Video 1	0.44376	0.42142	0.40751	0.38055
Video 2	0.78722	0.56325	0.55694	0.30471

Table 1: Comparison of PBC for different detecting algorithms.

Video	MoD	MoG	Traditional K-SVD	This paper

Video 1	0.74557	0.80113	0.81353	0.87355
Video 2	0.63674	0.80429	0.81637	0.89751

Table 2: Comparison of detecting precision for different detecting algorithms.

As can be seen from Table 1, in the 1st experiment, since the image quality of the video used is high, the calculation results of the four algorithms in the PBC are relatively similar. However, for the video used in the 2nd experiment, due to the complexity of the underground mine environment, the video has uneven illumination and many noises. From the PBC calculation results for 2nd experiment, the detection algorithm in this paper obviously better than the other three algorithms.

The comparison of object detection precision for different four algorithms is presented in Table 2. For the 1st experiment, because the video quality is high, the four algorithms can detect the foreground object well, and the calculated accuracy is also very similar. However, in the 2nd experiment, the quality of the video image decreased linearly. Therefore, the detection accuracy also decreased. Especially the calculation result of the MoG algorithm decreased by nearly 0.11. However, the accuracy of the proposed algorithm in this paper did not decrease, and the proportion of error points has not increased. Which indicates that the algorithm in this paper has strong adaptability and effectiveness in certain complex dark situations?

Conclusions

The algorithm proposed in this paper is aimed at the specific scene of the underground mine environment. The video in this scene has problems such as uneven illumination and more noise. Therefore, this paper uses the improved non-mean local filtering algorithm to filter the object to improve the image quality. Then the sparse representation is introduced into the underground mine video, and the moving object has been well detected. Results in this paper show that the sparse representation algorithm has good robustness in the underground mine environment, and the proposed algorithm can overcome the problems of illumination and noise interference in the underground mine environment. The object detection algorithm provides a solid foundation for the next step of moving object recognition and tracking.

References

1. Xue Mei, Ling Haibin Robust (2011) Visual tracking and vehicle classification via sparse representation. IEEE Trans Pattern Anal Mach Intell 33: 2259-2272.

2. Jeremy GM (2014) Machine learning classification models for more effective mine safety inspections. Trans ASME J Appl Mech 11:14-20.

3. Zhang Xiehua, Zhang Shen, Fang Shuai (2014) Clearing research on fog and dust images in coal mine intelligent video surveillance. J China Coal Soc 39:198-204.

4. Morioka K, Mao Xuchu, Hashimoto H (2006) Global color model based object matching in the multi-camera environment. International Conference on Intelligent Robots and Systems.

5. Bouzaraa Fahd, Urfalioglu Onay (2015) A Naive Bayes approach to improve histogram matching quality. IEEE International Workshop on Machine Learning for Signal Processing.

6. Cong Z, Wang XG, Wai KC (2011) Background subtraction via robust dictionary learning. EURASIP Journal on Image and Video Processing 972-961.

7. Azimi-Sadjadi Mahmood R, Kopacz Justin, Klausner Nick (2014) K-SVD dictionary learning using a fast OMP with applications. IEEE International Conference on Image Processing.

8. Zhu Rong, Wang Yong (2015) Color image denoising via dictionary learning and sparse representation. J Comput Theor Nanosci 12: 3911-3916.

9. Li Dan, Qian Jiansheng, ChaiI Yanli (2011) The object detecting in dangerous areas for coal mine underground. J China Coal Society 36: 527-532.

10. Wang Lin, Li Weishan, Zhang Yuliang (2017) Pedestrian Detection Based on YOLOv2 with Skip Structure in Underground Coal Mine. Proceedings of 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference.

11. Xian Xiaodong, Jiang Qiqi (2017) Target detection in mine based on fusion of color and edge information. International Conference on Advanced Mechatronic Systems.

12. Candes EJ (2006) Compressive sampling. Proceedings of the international congress of mathematicians.

13. Donoho DL (2006) Compressed sensing. IEEE Trans on International Theory 52: 1289-1306.

14. Amaldi EK (1998) On the approximability of minimizing nonzero variables or unsatisfied relations in linear system. Theor Comput Sci 209: 237-260.

15. Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. Siam J Sci Comput 20: 33 -61.

16. Gorodnitsky IF, Rao BD (1997) Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm. IEEE Trans Signal Process 45:600-616.

17. Mallat S, Zhifeng Z (1993) Matching pursuits with time-frequency dictionaries signal processing. IEEE Trans Signal Process 41: 3397-3415.

18. Pati Y, Rezaiifar R, Krisnaprasad P (1993) Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers.

19. Azghai Masoumeh, Marvasti Farokh (2016) Application of sparse signal processing. IEEE Global Conference on Signal and Information Processing.

20. Yeganli Faezeh, Ozkaramanli Huseyin (2014) Improve online dictionary learning for sparse signal representation. 22nd Signal Processing and Communications Applications Conference.

21. Honglei T (2013) Non-local means denoising algorithm based on dct. Sci Tech Engineering 13: 3123-3126.

22. Yin Panqiang, Lu Dongming, Yuan Yuan (2016) An improved non-local means image denoising algorithm using Mahalanobis distance. J Com Aided Design Com Graph 28: 404-410.