

MOVING TARGET DETECTION BASED ON AN ADAPTIVE LOW-RANK SPARSE DECOMPOSITION

Jiang CHONG

*School of Information Science and Engineering
Hunan Women's University
410004 Changsha, China
e-mail: jessiejch@qq.com*

Abstract. For the exact detection of moving targets in video processing, an adaptive low-rank sparse decomposition algorithm is proposed in this paper. In the paper's algorithm, the background model and the solved frame vector are first used to construct an augmented matrix, then robust principal component analysis (RPCA) is used to perform a low-rank sparse decomposition on the enhanced augmented matrix. The separated low-rank part and sparse noise correspond to the background and motion foreground of the video frame, respectively, the incremental singular value decomposition method and the current background vector are used to update the background model. The experimental results show that the algorithm can deal with complex scenes such as light changes and background motion better, and the algorithm's delay and memory consumption can be reduced effectively.

Keywords: Detection of moving objects, low-rank, sparse decomposition, adaptive robust, principal component analysis

Mathematics Subject Classification 2010: 68U10

1 INTRODUCTION

The important source of people's knowledge is from image information in the world. In many occasions, the transmitted information in images is richer, truer and more specific than other forms of information. The cooperation between the human eye

and the brain enables people to acquire, process, and understand visual information. Humans use vision to perceive external environmental information efficiently. In fact, according to statistics of some foreign scholars, about 80% of the external information, which is obtained by humans, comes from images which are taken by the eyes. If computers are to be intelligent, as the main carrier for humans to obtain external information, our vision must be able to process image information. Especially in recent years, image data processing with large capacity (such as graphics, images and video) has been widely used in medical, transportation and industrial automation fields.

Most images in nature are constantly changing simulated images. In daily life, the moving targets in these images are often more concerned about us, such as pedestrians, vehicles, and other moving objects. Moving object detection is a popular direction in computer vision and digital image processing, and it is widely used in robot navigation, intelligent video surveillance, industrial detection, aerospace and many other fields. Therefore, moving object detection has become a research hotspot in theory and application in recent years. Moving object detect is an important branch of image processing and computer vision, and it is also a core part of intelligent monitoring systems. The purpose is how to quickly and accurately detect moving targets in surveillance video, that is, moving targets are extracted from sequence images.

The detection of moving targets is the most basic and very important step in video processing. The accurate detection of moving targets is of great significance for tracking, classifying, understanding and analyzing the behavior of subsequent moving targets. In recent years, many scholars have conducted many studies on the detection of moving targets. However, the detection of moving targets faces many challenges, such as multi-mode background interference, lighting changes, and camera shake. They are still research hotspots and difficulties in the field of computer vision.

The representative algorithms include optical flow method, frame difference method and background modeling method in the research of moving target detection. The optical flow method can detect and track moving targets without prior knowledge of the background, but there are problems such as high complexity of the algorithm and sensitivity to illumination changes. In the frame difference method, the moving target contour is obtained by performing differential operations on two adjacent frames in the video image sequence. The algorithm is simple and easy to implement, but it depends on the selected inter-frame time interval. The idea of the background modeling method is to establish a background model, and the video frame is detected with the background model to determine the motion foreground. Compared with the optical flow method and the frame difference method, it has the advantages of less calculation, faster speed, and higher precision. The core of the background modeling method is how to build a model, how to update the model properly to deal with the changes of the background itself.

At present, although there is a large number of moving object detection algorithms, due to the complexity and variability of the actual environment, these

algorithms are not all very robust. The faced problems and difficulties can be summarized as follows:

1. Model initialization problem: During the background initialization training period, because high-quality background models have not been obtained, it often leads to false detection of moving targets;
2. Camouflage phenomenon: Some moving targets may be extremely similar to the background, it may cause the moving targets to be indistinguishable from the background;
3. Illumination change: It is divided into sudden change and gradation of light. The background model should be able to adapt to the gradual change of light in the outdoor environment during the day; correspondingly, the background model can also be adapted to the indoor environment where the lights are suddenly turned on. In short, the change of light will strongly affect the background model, which will most likely cause false detection;
4. Foreground hole phenomenon: When the moving target has a large number of regions with the same color, the internal changes of these regions may lead to inaccurate detection, some internal regions of the foreground are misjudged as the background;
5. Dynamic background: The most common is the leaf shake, of course, water ripples, small target shake;
6. Suddenly stopped moving targets: After some moving objects enter the scene, they stop in the scene. Obviously, the moving target in this case should be identified as the background;
7. Shadow: The shadow of the moving target and the original shadow of the background area can be detected;
8. Noise interference: Noise interference basically belongs to the low data quality, it is caused by the video image which is transmitted or compressed by the webcam;
9. Camera shake: Under some conditions, wind will cause camera shake;
10. Camera self-adjustment: At present, many cameras have automatic control functions, such as lighting control, white balance, and zoom-in and zoom-out functions.

2 RELATED WORKS

Computer vision applications based on videos often require the detection of moving objects in their first steps. Background subtraction is then applied in order to separate the background and the foreground. Background subtraction is surely among the most investigated field in computer vision providing a big amount of publications. Most of them concern the application of mathematical and machine learning models to be more robust to the challenges met in videos. However, the

ultimate goal is that the background subtraction methods developed in research could be employed in real applications like traffic surveillance. There is often a gap between the current methods used in real applications and the current methods in fundamental research. In addition, the videos evaluated in large-scale datasets are not exhaustive in the way that they only covered a part of the complete spectrum of the challenges met in real applications [1].

Background modeling is a technique for extracting moving objects in video sequences. The different foreground segmentation methods are categorized based on the way of obtaining the reference frame. The different approaches can be categorized into basic modeling, statistical modeling, clustering algorithms, methods based on fuzzy modeling, neural and neuro-fuzzy methods, etc.

Neural network methods train networks for a specific amount of video frames, allowing it to update its weights in order to model the background. Recently, the most common background subtraction methods are based on deep neural networks. The convolutional Neural Networks (CNN) was introduced to perform the segmentation task [2]. The benefit of this approach is that foreground objects segmentation can be achieved independently to the temporal characteristics. CNN can achieve the segmentation with only spatial characteristics because the incoming examples are independent. The pertinent features are selected using CNNs and transmitted into a classifier to segment moving objects. Finally, a median filter and or a fully connected framework are applied. Lim et al. introduced an approach based on a triplet CNN for a multistage background feature embedding using an encoder-decoder model [3]. A pre-trained convolutional network, VGG-16 Net is adapted under a triplet framework in the encoder part to incorporate an image in multiple scales into the feature space. Only a few training samples are used. The network takes an RGB image in three different scales and generates a foreground segmentation probability mask for the corresponding image. In the period of 2018–2019, numerous deep learning models either based on auto-encoder [4, 5, 6] and CNNs [7, 37, 9, 10, 11] have been proposed. However, all these methods are supervised and have been trained on ground truth video frames of datasets and tested on the same types of videos.

The representative work of the background modeling method is as follows: Gaussian mixed model (GMM) was proposed by Stauffer et al. [12], multiple Gaussian models were used to fit the multi-peak state of pixel brightness changes, and it has better illumination robustness. The non-parametric kernel density estimation (KDE) was proposed by Elgammal et al. [13], it does not need to assume the type of the probability density function in advance, and multiple background samples were used to estimate the probability density of pixels. The algorithm can effectively suppress shadows, but there exists a large cache size, difficulty in selecting core bandwidth, etc. The codebook model were used proposed by Kim et al. [14], CodeBook of a multiple codeword was established for each pixel, multi-modal background can be detected in real time, there is good global illumination robustness. However, the memory consumption of the algorithm is large. The self-organizing background subtraction (SOBS) was proposed by Maddalena et al. [15, 16], it draws

on the characteristics of neural networks, a pixel in the background model is mapped to multiple locations in the model. The relevance of pixel neighbors' domain space is used in the updating method. The VIBE (visual background extractor) modeling method was proposed by Barnich et al. [17], a random sampling strategy was used to initialize the background model, and the background model is updated with a second random sampling method. This airspace information propagation mechanism enables the algorithm to deal with camera shake, the real-time performance of the algorithm is also high. Pixelmann adaptive segmentation (PBAS) was proposed by Hofmann et al. [18], the idea of cybernetics is introduced. The foreground judgment threshold and background model update rate change adaptively, there is high recognition accuracy in the algorithm, but it takes more time to calculate and set multiple thresholds. In short, these algorithms based on statistical models treat each pixel as an independent, unrelated individual, without excavating the inherent nature of high-dimensional data.

In recent years, compressive sensing has received a great deal of attention as a new signal sampling theory [19], and it has been widely used in data compression, pattern recognition, and wireless communication. Compressed sensing is a signal sparse in a certain transform domain. The signal is sampled at a sampling rate much lower than the Nyquist frequency. The observation matrix is used to linearly project the high-dimensional sparse signal to a low-dimensional subspace. By solving the optimization problem, the original signal with high probability is reconstructed from the subspace. A low-rank representation can be seen as a generalization of compressed sensing in two-dimensional data [20]. The rank of the matrix is used as a sparse measure. Compared with the traditional subspace learning model, sparse representation has better robustness to data containing outliers and strong noise. In sparse representation, the robust principal component analysis (RPCA) is also known as low rank and sparse matrix decomposition [21, 22], it mainly considers how to recover low rank data from observations that are subject to large sparse noise pollution observations. Taking into account the low rank of the background and the sparseness of the motion foreground, Candes et al. first applied the low-rank sparse decomposition to background modeling [23]. In this method, the moving target and the background area can be separated effectively from the monitoring video sequence without an independent training stage. The premise of the algorithm is to assume that the background is almost or completely stationary. Subsequent researchers also launched a series of studies based on this method of moving target detection algorithms. DECOLOR algorithm combines RPCA with motion recognition [24] and it combines Markov random field (MRF) a priori. Noise and small background motion can be eliminated effectively in the smoothness of MRF, but the foreground area will be "Smooth" and result in the loss of details of the moving target. Gao et al. proposed the block sparse RPCA [25]. First the low-rank sparse decomposition of the sampled samples was performed to obtain the outlines block, and then the second decomposition was performed to obtain the moving target. Liu et al. introduced the concept of structured sparse in considering the spatial correlation of outliers, and low-rank sparse decomposition is used with different regularization parameters for

each pixel group [26]. Bouwmans et al. summarized and evaluated the technique of decomposition into Low-rank [27].

When RPCA is applied to moving target recognition, the above algorithms are improved from different perspectives. However, they all have the following problems:

1. All vectors of the video are vectorized and loaded into memory once, when the number of video frames is relatively more, it will cause memory overflow;
2. They do not take into account complex changes such as changes in light, background motion in the actual video surveillance environment, and poor robustness.

3 RPCA APPLIED TO BACKGROUND SUBTRACTION A SHORT OVERVIEW

3.1 Matrix Approaches

In many practical applications, the given data matrix D tends to be low-rank or approximately low-rank, the random amplitude is arbitrarily large, but the error of sparse distribution destroys the low rank of the original data. Therefore, matrix D can be decomposed into the sum of two matrices, that is, $D = A + E$, where matrices A and E are unknown, and A is a low-rank matrix. When the noise obeys an independent and identically distributed Gaussian distribution, the classical principal component analysis (PCA) can be used to obtain the optimal matrix A [28], that is, the following optimization problem is solved:

$$\min \|D - A\| \quad \text{s.t.} \quad \text{rank}(A) \leq k \quad (1)$$

wherein, $\text{rank}(\cdot)$ represents the rank of the matrix, and $\|\cdot\|$ represents the 2 norm. Classical PCA can be effectively solved by singular value decomposition, but it is limited to cases where noise N_0 is small and independent of Gaussian distribution. Candès et al. proposed RPCA to solve the problem of E , it is sparse and large noise [23]. That is, matrix E in the above model is a sparse matrix with arbitrary amplitude. The RPCA model is shown in Equation (2), where $\|\cdot\|_0$ represents 0 norm.

$$\min(\text{rank}(A), \|E\|_0) \quad \text{s.t.} \quad D = A + E. \quad (2)$$

Introducing the regularization parameter λ , this bi-objective optimization problem is transformed into a single-objective optimization problem:

$$\min \text{rank}(A) + \lambda \|E\|_0 \quad \text{s.t.} \quad D = A + E. \quad (3)$$

In Equations (2) and (3), the objective functions includes $\text{rank}(A)$ and $\|E\|_0$, it is non-linear non-convex combined optimization functions, and its solution is an NP-Hard problem. Since the matrix norm $\|\cdot\|_*$ of the matrix is the envelope of its rank,

the 0 norm and the 1 norm of the matrix can be equivalent under certain conditions, so the convex relaxation of the matrix is the following optimization problem:

$$\min \|A\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad D = A + E. \quad (4)$$

3.2 Tensor Approaches

The convex optimization problem shown in Equation (3) is also called principal component pursuit (PCP). Candes et al. prove that as long as the singular vector distribution of the low-rank matrix A is reasonable and the non-zero elements of the sparse matrix are evenly distributed, then the PCP can recover the original low-rank matrix A from any unknown error with a probability close to unity [23].

Robust Principal Component Analysis (RPCA) is a modification of the widely used statistical procedure of principal component analysis (PCA) which works well with respect to grossly corrupted observations. A number of different approaches exist for Robust PCA, including an idealized version of Robust PCA, which aims to recover a low-rank matrix A from highly corrupted measurements $D = A + E$ [23]. RPCA can be applied to video surveillance, face recognition, collaborative filtering and other aspects. In video surveillance, each frame of the video is vectorized and successively arranged into a matrix. The stable background part corresponds to the low-rank part of the matrix, and the motion foreground can be used as a sparse noise part. Therefore, PCP can be used to detect moving targets in a smooth background.

This decomposition in low-rank and sparse matrices can be achieved by techniques such as Principal Component Pursuit method (PCP) [23], Stable PCP [29], Quantized PCP [30], Block based PCP [31], and Local PCP [32]. Then, optimization methods are used such as the Augmented Lagrange Multiplier Method (ALM [33]), Alternating Direction Method (ADM [34]), Fast Alternating Minimization (FAM [35]) or Iteratively Reweighted Least Squares (IRLS [36]), Real-time Robust Principal Components Pursuit [37], Memory-efficient dynamic robust PCA [38], Incremental Principal Component Pursuit [39], Spatio-temporal Sparse Subspace Clustering [40], Double-constrained RPCA [41], ARF and OR-PCA Background Subtraction [42].

In addition to the matrix segmentation technique used in target detection, there are Tensors Decomposition techniques. In mathematics, tensors are geometric objects that describe linear relations between geometric vectors, scalars, and other tensors. Elementary examples of such relations include the dot product, the cross product, and linear maps. Geometric vectors, often used in physics and engineering applications, and scalars themselves are also tensors. A more sophisticated example is the Cauchy stress tensor T , which takes a direction v as input and produces the stress $T(v)$ on the surface normal to this vector for output, thus expressing a relationship between these two vector. With Tensors Decomposition as target detection, people study some effective methods, such as Outlier-Robust Tensor PCA [43], Tensor Based Low-Rank and Saliency Fused-Sparse Decomposition [44], Online Stochas-

tic Tensor Decomposition [45], Stochastic Decomposition into Low Rank and Sparse Tensor [46].

4 ADAPTIVE LOW RANK SPARSE DECOMPOSITION ALGORITHM

Compared with the moving object detection algorithm based on pixel background modeling, the advantage of RPCA is that it can separate the moving foreground area from the background area effectively without a separate training stage. But the premise of the algorithm is to assume that the background is almost or completely static. However, this assumption is difficult to establish in most of the real monitoring scenarios. And Equation (4) is used to accurately separate the foreground and the background, you need to load a larger number of video frames. If the number of frames is too small, slow moving objects are recognized as part of the background. Since the time complexity of the singular value decomposition is $O(m^3)$ ($D \in R^{m \times n}$) in the RPCA solution, the time cost of the algorithm rises in a cubic order with the increase of m , and loading and solving a large number of video frames cause that a low Rank-sparse decomposition has high latency in moving target recognition.

Based on this, an adaptive low-rank sparse decomposition algorithm is proposed in this paper. Through the partial loading and iterative updating of the video frame, the memory occupancy can be reduced while the accuracy of the detection effect can be effectively guaranteed, and the motion target detection delay can be reduced. The algorithm framework is shown in Figure 1.

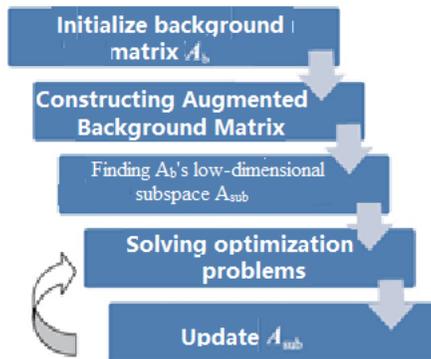


Figure 1. Algorithm framework

Robust PCA considers such a problem: the general data matrix D contains structural information and also contains noise. Then this matrix can be decomposed into two matrices $D = A + E$, A is low rank (due to a certain amount of structural

information inside causing a linear correlation between rows or columns), E is sparse (containing noise, here is the foreground (moving image) , it is sparse). Robust PCA is to decompose a matrix D into a matrix with as low a rank as possible and a matrix as sparse as possible. D consists of original image sequence frames, its each column is an image pixel. A_b is the column subset of A (image background). D consists of original image sequence frames, its each column is an image pixel. A_b is the column subset of A (image background). The implementation of the algorithm in Figure 1 is described below.

4.1 Dynamic Designer Motion Simulation

In order to reduce delay of the algorithm, we consider processing a small number of video frames at a time. However, in order to ensure the accuracy of the motion foreground and background separation, A still needs a large number of columns to guarantee the low rank of $\|\cdot\|_*$. Therefore, an augmented matrix is composed of the calculated background vector and the background vector to be calculated [47].

If the background vector of the first k frames of video has been obtained, assuming that b_j represents the background vector of the j^{th} frame of video, where $j \in [1, k], n$ is the total number of pixels per frame, the following background matrix is obtained: $A_b = [b_1, b_2, \dots, b_k] \in R^{n \times k}$.

The augmented background matrix \hat{A} is composed of A_b and A , where $A \in R^{n \times m}$ is the background matrix of m new frames to be calculated: $\hat{A} = [A_b, A] \in R^{n \times (k+m)}$.

Therefore, $\|A\|_*$ in the model is shown in Equation (4), it can be replaced by $\|\hat{A}\|_*$.

4.2 Low Dimensional Background Modeling and Model Solving

Since the time complexity of optimizing the augmented low-rank matrix \hat{A} is $O(k+m)^3$, as the new frame is continuously processed, its time cost increases in a cubic order. Therefore, it is very necessary to perform dimensionality reduction on A_b , that is to project A_b into a low-dimensional subspace. Let this subspace be $A_{sub} \in R^{n \times p}$ in order, a new augmented matrix $[A_{sub}, A] \in R^{n \times (p+m)}$ ($p \ll k$) is obtained. In other words, the goal is to find an optimal solution A_{sub} , that makes the nuclear norm of $[A_{sub}, A_{new}]$ closest to \hat{A}_{aug} , the optimization problem of Equation (5) is obtained:

$$A_{sub} = \arg \min \|\|\hat{A}\|_* - \|[A_{sub}, A_{new}]\|_*|. \tag{5}$$

First, Singular Value Decomposition (SVD) is performed on A_b , the main features of the high dimensional matrix is extracted to construct A_{sub} :

$$A_b = UDV^T. \tag{6}$$

In Equation (6), $D \in R^{n \times k}$ is a diagonal matrix which is composed of singular values of A_b ; $U \in R^{n \times n}$ is a matrix which is composed of left singular vectors;

$V \in R^{k \times k}$ is a matrix which is composed of right singular vectors. Selecting the first P largest singular values of D , it constitutes a diagonal matrix $D_p \in p \times p$, and selecting the top P column of U , it constitutes a matrix $U_p \in n \times p$, which results in a low-dimensional subspace A_{sub} :

$$A_{sub} = U_p D_p. \tag{7}$$

The variables of the model (4) is replaced to obtain the model (8):

$$\min ||[A_{sub}, A]||_* + \lambda ||E||_1 \quad \text{s.t.} \quad D = A + E. \tag{8}$$

Considering the efficiency of the implementation, the inexact Lagrange multiplier method (inexact ALM) is used to solve the above optimization problem [33, 34]. In Equation (8), the augmented matrix of the kernel norm term consists of known and unknown columns. It is necessary to introduce a new variable to replace the augmented matrix, and the variable splitting method is used to separate the objective function into: $Z = [A_{sub}, A]$.

Augmented Lagrangian functions are built for the above optimization problem to be solved:

$$L(Z, E, Y) = ||Z||_* + \lambda ||E||_1 + \langle Y, D - Z - E \rangle + \mu ||D - Z - E||_F^2 / 2$$

where Y is a Lagrangian multiplier; $\mu ||D - Z - E||_F^2 / 2$ is a penalty function term, and μ is a penalty function factor. Repeat the iteration of the following update formula until convergence:

$$\begin{aligned} Z_{k+1} &= \arg \min_Z L(Z, E_{k+1}, Y_k, \mu_k) = D_{1/\mu_k}(D - Z_{k+1} + Y_k/\mu_k), \\ E_{k+1} &= \arg \min_E L(Z_{k+1}, E, Y_k, \mu_k) = S_{1/\mu_k}(D - E_{k+1} + Y_k/\mu_k). \end{aligned}$$

Among them, $S_\tau(x)$ is a soft threshold operator, $S_\tau(x) = \text{sgn}(x) \times \max(|x| - \tau, 0)$; $D_\tau(x)$ is a singular value threshold operator, $D_\tau(x) = US_\tau(\Sigma)V_*$ ($X = U\Sigma V_*$ indicates singular value decomposition for X). The convergence condition is $||D - Z - E||_F \leq \sigma ||D||_F$, where $\sigma = 10^{-7}$ [24].

4.3 Background Model Updates

In order to obtain a new background model, it is used to process the next batch of frames, after each calculation, it is necessary to use the currently obtained A to update A_b . This article uses the incremental SVD method to update A_b [48].

$$\begin{aligned} A_b &\approx U_p D_p V_p^T, [U^{new}, D^{new}] = \text{svd}([\omega_b A_b, \omega_a A]) \\ &\approx \text{svd}[\omega_b U_p D_p V_p^T, \omega_a A] = \text{svd}[\omega_b U_p D_p, \omega_a A] = \text{svd}[\omega_b A_{sub}, \omega_a A]. \end{aligned}$$

A new background model A_b^{new} is obtained:

$$A_{sub}^{new} = U_p^{new} D_p^{new} \tag{9}$$

wherein, U_p^{new} is composed of the first P columns of U^{new} ; D_p^{new} is a diagonal matrix which are composed of P largest singular values. ω_a and ω_b are the weights of the control updating speed ($\omega_a, \omega_b \in [0, 1]$). The larger ω_a , the faster the updating of the background model; the larger ω_b , the slower the updating of the background model, so the multimode scenes with complex backgrounds can choose $\omega_a > \omega_b$. At the same time, the matrix V does not need to be calculated in the above formula operation, which effectively reduces the time complexity when updating the model.

5 EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

The experimental simulation hardware environment is: Intel Core i3 dual-core processor, clocked at 3.4 GHz, and main memory is 4 GB. The software environment is: Window10 operating system, Matlab R2018a.

5.1 Experiments on CDnet and Wallflower Datasets

The test data selects the standard video libraries CDnet and Wallflower for motion target detection. CDnet 2014 includes 11 series of basic scenes, dynamic backgrounds, camera shakes, etc. Wallflower includes 7 scenes such as motion background and lighting gradient. At the same time, the data set also gives the groundtruth of each frame segmentation foreground, which facilitates the comparative analysis of the algorithm. Two scenes, Running and Highway, were selected. The following algorithm was compared with the GMM algorithm, the PCP algorithm, and the DECOLOR algorithm. Figure 2 shows the motion foreground recognition effect of each algorithm.

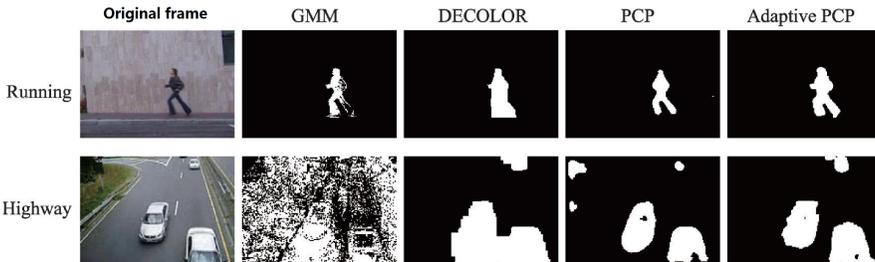


Figure 2. Experiment results of 4 algorithms in different scenes

The running scene of the first row in Figure 2 is a single-moving target. The motion scene in a simple background gives the experimental results of the 17th frame.

The GMM algorithm uses multiple Gaussian models to fit the changes in pixel values. The identified motion foreground contours are more detailed, but there are holes in the right and left arms of the motion object. The DECOLOR algorithm uses the MRF to remove noise and small background motions, but it makes the motion foreground smooth and loses contour details. Both the PCP and the method of this paper can completely identify the motion prospects. The second line Highway is a complex scene with multiple moving objects and multi-modal backgrounds coexisting with light changes. The branches are shaken in the upper left corner, that are the sports background. In the figure, the video frame 76 is taken in the truncated light. Obviously, the speed of updating the background model of GMM cannot adapt to the sudden change of light, leading to a large area of misjudgment. The DECOLOR algorithm is robust to sudden changes in light, but due to the loss of contour details, it leads to adhesion of the two motion foregrounds. Although the PCP algorithm can accurately identify all the moving targets, it mistakenly identifies the branches that are sloshing in the upper-left corner as motion foreground, because the PCP algorithm assumes that the background is completely stationary. The method in this paper is able to cope with sudden light changes and multi-mode background disturbances, and it has the best recognition effect in this scene. Because the inexact Lagrangian multiplier method is used for optimization, for the running scenes with relatively simple background and foreground, the algorithm can achieve convergence with 33 iterations, and the complex scene with multiple moving targets has a complex background. It takes 71 iterations to achieve convergence. That is, the simpler the scene, the faster the algorithm converges.

In quantitative analysis of the algorithm, the indicators are recall (Re), accuracy (precision, Pr) and comprehensive performance (F-measure, F1). Among them, TP (true positive) indicates true positive, that is, the number of pixels are correctly detected as the front sight; FN (false negative) indicates false negative, that is, the number of pixels are detected as background points by mistake; FP (false positive) indicates false positive, that is, the number of pixels are detected as the previous sighting mistake.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Table 1 gives a quantitative comparison of the performance parameters of each algorithm in the two scenarios. It can be seen that the introduction of an adaptive process in the PCP algorithm significantly improves the ability of the algorithm to handle complex backgrounds.

Algorithm	Scene	Re	Pr	F1
GMM	Running	0.69	0.82	0.87
	Highway	0.62	0.42	0.5
DECOLOR	Running	0.82	0.76	0.79
	Highway	0.83	0.71	0.77
PCP	Running	0.89	0.88	0.88
	Highway	0.87	0.79	0.83
Adaptive PCP	Running	0.9	0.87	0.88
	Highway	0.9	0.88	0.89

Table 1. Comparison of average performance

Note: GMM is Gaussian mixed model; PCP is principal component pursuit; The entire process of DECOLOR is detecting contiguous outliers in the low-rank representation.

Figure 3 compares the time efficiency of PCP and Adaptive PCP. Experimental results show that by loading part of the video frame and singular value decomposition of the background model, the delay of the algorithm is effectively reduced. At the same time, since the PCP algorithm needs to load all the frames at once. Compared to the Running ($180 \times 144 \times 42$ frames) scene, Highway scenes with a low resolution but a large total number of frames ($160 \times 112 \times 119$ frames) require more average processing time. While Adaptive PCP loads the video frames to be calculated in batches, the average processing time of a single frame is inversely proportional to the resolution of the video.

5.2 Experiments on the VOT2014 Dataset

In order to intuitively compare the experimental results of the algorithm, the two running and highway modes with no lower resolution and fewer frames are selected. If we choose the PETS2006 scene of the dataset ($320 \times 240 \times 1200$ frames), the memory overflow will occur in the traditional PCP algorithm, and the algorithm can still obtain the detection result. Figure 4 shows the foreground detection result of the 740th frame of the scenario.

The algorithm has been implemented. We have tested the algorithm with both simulated and actual sequences of images of vehicles in different landscapes. The system can detect and track targets in real-time. We achieved the frame rate of 4 frames/second for detection and 15 frames/second for tracking of 100×90 pixels target in 352×288 pixels video frames. We tested the proposed algorithm with wide variety of image sequence. Figure 5 shows some results for detection of various objects in arbitrary background. As it is shown the algorithm has successfully detected the targets. In Figure 6 the tracking results for a vehicle are shown. In this example, the tracked vehicle turns over the road and its shape and size change. As it shown in the pictures both the size and the shape of the vehicle varies but our tracking algorithm can successfully track it. Results showed the accuracy of

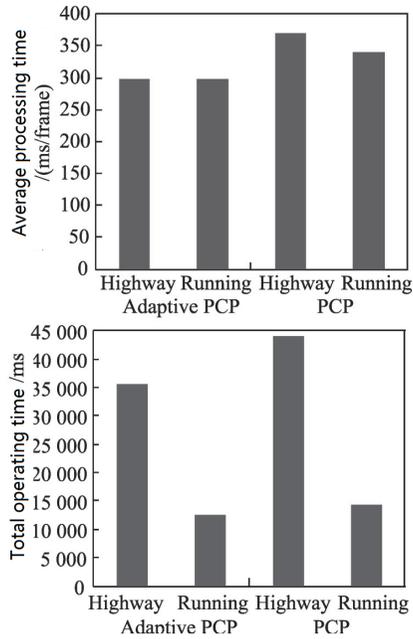


Figure 3. Comparison of running time between PCP and Adaptive PCP

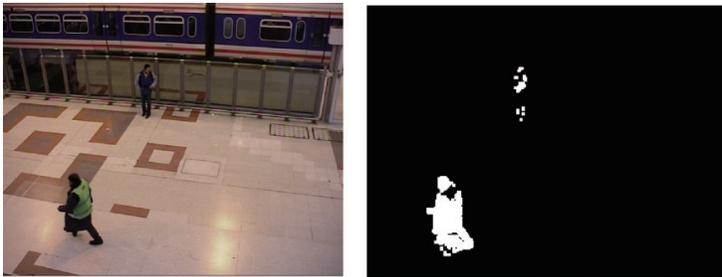


Figure 4. Detection result of 740th frame of this paper method

the method in detecting and tracking of moving objects. Comparison of results generated by the proposed method with those of other methods showed that more reliable results could be obtained using the proposed method in real-time.

Further, the experimental data is taken from the VOT2014 dataset (<https://www.votchallenge.net/vot2014/dataset.html>) [49]. In order to ensure the effectiveness and stability of the algorithm, the moving target is selected here, where the dataset name is jogging. Detect and track the moving object with the largest moving area in the first 22 frames of video of each data set. This video is a video of the movement of two joggers. The feature is that the background is relatively single,



Figure 5. Various targets are detected in the detection algorithm

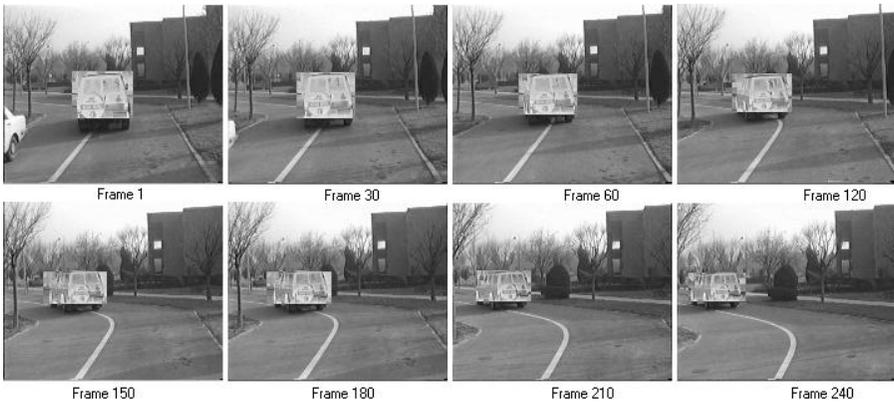
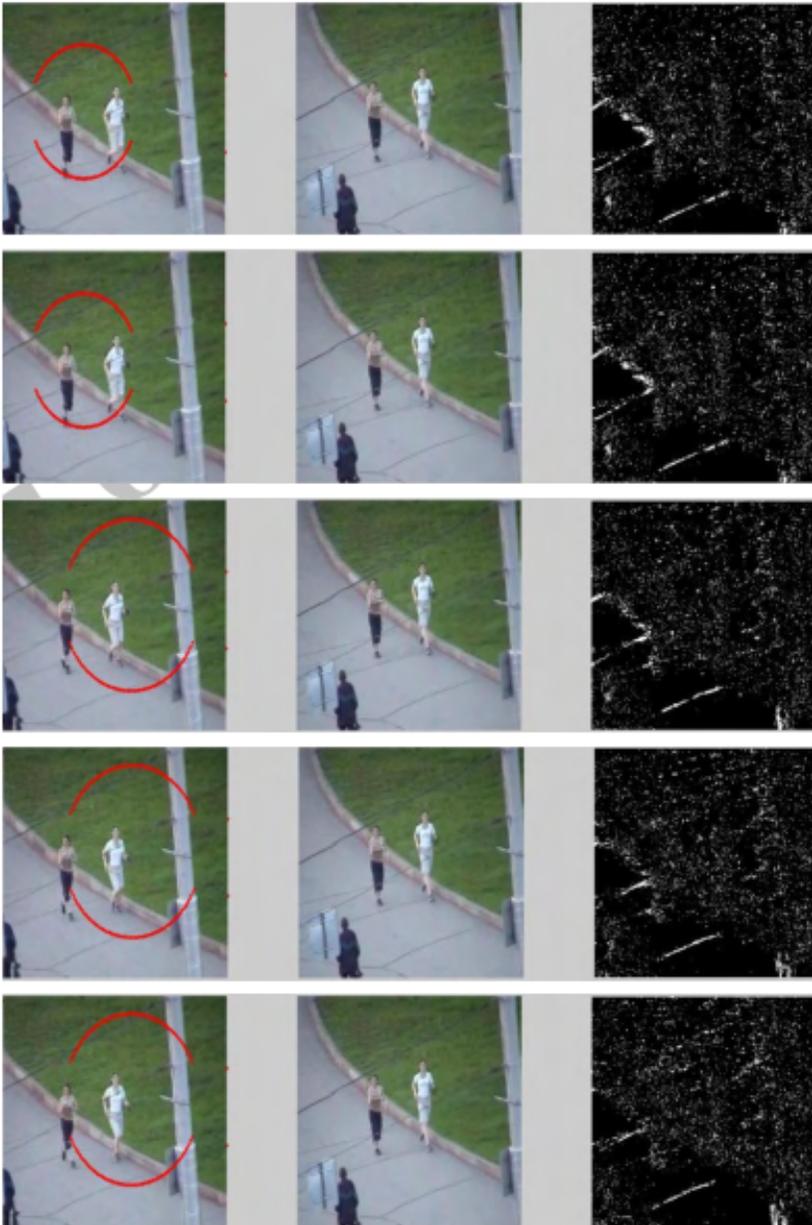


Figure 6. Tracking a vehicle while it turns over and changes its size and shape

but the two targets appear too close to each other. At this time, they are treated as a moving target for tracking and detection. The first five frames are shown in Figure 7. The experimental results show that the background of the entire video is relatively stable. When using the algorithm in this article to track two joggers, the target is almost marked in the ellipse.

6 CONCLUSIONS

In this paper, an adaptive low-rank sparse decomposition algorithm is proposed or it is known as adaptive principal component pursuit (A-PCP). First, a part of video frames are loaded, and the initial background A_b is obtained by low-rank sparse decomposition, and singular value decomposition is performed to obtain a low-dimensional subspace A_{sub} , A_{sub} connects with a new number of frames of the calculated video background vector, an augmented low-rank matrix is constructed. Then, a low-rank sparse decomposition of the objective function containing \hat{A} results in a new low-rank background matrix A . Finally, the low-dimensional background



a) Current frame image b) Background extracted by the algorithm in this paper c) Sports targets of interest

Figure 7. Jogging target tracking effect map

is updated with the current A model A_{sub} , the above process is repeated until all frames are processed.

This adaptive low-rank sparse decomposition algorithm is applied to the detection of moving targets in intelligent video surveillance. In the algorithm, the background model A_b is established by low-rank sparse decomposition, and it is reduced to the low-dimensional space A_{sub} . The background of the new frame is obtained by performing low-rank sparse decomposition on the augmented matrix containing A_{sub} , and then the current background vector is used to update the background model. In this paper, an adaptive process is introduced in the traditional PCP algorithm, experiments show that the robustness of the algorithm is improved to complex backgrounds, and it can achieve better detection results. At the same time, compared to the PCP algorithm, all frame vectors are loaded at once, this adaptive update mechanism can improve the robustness of the algorithm. The mechanism of partially loading video frames can also reduce the delay of the algorithm and avoid memory overflow at the same time, thus the overall performance of moving target recognition is improved.

Acknowledgements

The project is supported by the Scientific Research Fund of Hunan Provincial Education Department (No. 16C0804, Research on Personalized Push Algorithm of Unstructured Education Resources Based on Mobile Micro-Lectures), China.

REFERENCES

- [1] GARCIA-GARCIA, B.—BOUWMANS, T.—ROSALES SILVA, A. J.: Background Subtraction in Real Applications: Challenges, Current Models and Future Directions. *Computer Science Review*, Vol. 35, 2020, Art.No. 100204, doi: 10.1016/j.cosrev.2019.100204.
- [2] BABAEI, M.—DINH, D. T.—RIGOLL, G.: A Deep Convolutional Neural Network for Video Sequence Background Subtraction. *Pattern Recognition*, Vol. 76, 2018, pp. 635–649, doi: 10.1016/j.patcog.2017.09.040.
- [3] LIM, L. A.—KELES, H. Y.: Foreground Segmentation Using Convolutional Neural Network for Multiscale Feature Encoding. *Pattern Recognition Letters*, Vol. 112, 2018, pp. 256–262, doi: 10.1016/j.patrec.2018.08.002.
- [4] CHOO, S.—SEO, W.—JEONG, D.—CHO, N. I.: Multi-Scale Recurrent Encoder-Decoder Network for Dense Temporal Classification. 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 2018, pp. 103–108, doi: 10.1109/icpr.2018.8545597.
- [5] CHOO, S.—SEO, W.—JEONG, D.—CHO, N. I.: Learning Background Subtraction by Video Synthesis and Multi-Scale Recurrent Networks. In: Jawahar, C., Li, H., Mori, G., Schindler, K. (Eds.): *Computer Vision – ACCV 2018*. Springer, Cham,

- Lecture Notes in Computer Science, Vol. 11366, 2018, pp. 357–372, doi: 10.1007/978-3-030-20876-9_23.
- [6] GRACEWELL, J.—JOHN, M.: Dynamic Background Modeling Using Deep Learning Autoencoder Network. *Multimedia Tools and Applications*, Vol. 79, 2020, pp. 4639–4659, doi: 10.1007/s11042-019-7411-0.
 - [7] YANG, Y.—ZHANG, T.—HU, J.—XU, D.—XIE, G.: End to End Background Subtraction via a Multi-Scale Spatio-Temporal Model. *IEEE Access*, Vol. 7, 2019, pp. 97949–97958, doi: 10.1109/access.2019.2930319.
 - [8] QIU, M.—LI, X.: A Fully Convolutional Encoder-Decoder Spatial-Temporal Network for Real-Time Background Subtraction. *IEEE Access*, Vol. 7, 2019, pp. 85949–85958, doi: 10.1109/access.2019.2925913.
 - [9] MINEMATSU, T.—SHIMADA, A.—UCHIYAMA, H.—TANIGUCHI, R.: Analytics of Deep Neural Network-Based Background Subtraction. *MDPI Journal of Imaging*, Vol. 4, 2018, No. 6, Art. No. 78, 19 pp., doi: 10.3390/jimaging4060078.
 - [10] GARCÍA-GONZÁLEZ, J.—ORTIZ-DE-LAZCANO-LOBATO, J. M.—LUQUE-BAENA, R. M.—LÓPEZ-RUBIO, E.: Background Modeling by Shifted Tilings of Stacked Denoising Autoencoders. In: Ferrández Vicente, J., Álvarez-Sánchez, J., de la Paz López, F., Toledo Moreo, J., Adeli, H. (Eds.): *From Bioinspired Systems and Biomedical Applications to Machine Learning (IWINAC 2019)*. Springer, Cham, *Lecture Notes in Computer Science*, Vol. 11487, 2019, pp. 307–316, doi: 10.1007/978-3-030-19651-6_30.
 - [11] GARCÍA-GONZÁLEZ, J.—ORTIZ-DE-LAZCANO-LOBATO, J. M.—LUQUE-BAENA, R. M.—MOLINA-CABELLO, M. A.—LÓPEZ-RUBIO, E.: Foreground Detection by Probabilistic Modeling of the Features Discovered by Stacked Denoising Autoencoders in Noisy Video Sequences. *Pattern Recognition Letters*, Vol. 125, 2019, pp. 481–487, doi: 10.1016/j.patrec.2019.06.006.
 - [12] STAUFFER, C.—GRIMSON, W. E. L.: Adaptive Background Mixture Models for Real-Time Tracking. *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, USA, Vol. 2, 1999, pp. 246–252, doi: 10.1109/cvpr.1999.784637.
 - [13] ELGAMMAL, A. M.—HARWOOD, D.—DAVIS, L. S.: Nonparametric Model for Background Subtraction. In: Vernon, D. (Ed.): *Computer Vision – ECCV 2000*. Springer, Berlin, Heidelberg, *Lecture Notes in Computer Science*, Vol. 1843, 2010, pp. 751–767, doi: 10.1007/3-540-45053-x_48.
 - [14] KIM, K.—CHALIDABHONGSE, T.—HARWOOD, D.—DAVIS, L.: Real-Time Foreground-Background Segmentation Using Codebook Model. *Real-Time Imaging*, Vol. 11, 2005, No. 3, pp. 172–185, doi: 10.1016/j.rti.2004.12.004.
 - [15] MADDALENA, L.—PETROSINO, A.: A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. *IEEE Transactions on Image Processing*, Vol. 17, 2008, No. 7, pp. 1168–1177, doi: 10.1109/tip.2008.924285.
 - [16] MADDALENA, L.—PETROSINO, A.: The SOBS Algorithm: What Are the Limits? *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Providence, USA, 2012, pp. 21–26, doi: 10.1109/CVPRW.2012.6238922.

- [17] BARNICH, O.—VAN DROOGENBROECK, M.: ViBe: A Universal Background Subtraction Algorithm for Video Sequences. *IEEE Transactions on Image Processing*, Vol. 20, 2011, No. 6, pp. 1709–1724, doi: 10.1109/tip.2010.2101613.
- [18] HOFMANN, M.—TIEFENBACHER, P.—RIGOLL, G.: Background Segmentation with Feedback: The Pixel-Based Adaptive Segmenter. *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision Pattern Recognition Workshops, 2012*, pp. 38–43, doi: 10.1109/cvprw.2012.6238925.
- [19] DONOHO, D. L.: Compressed Sensing. *IEEE Transactions on Information Theory*, Vol. 52, 2006, No. 4, pp. 1289–1306, doi: 10.1109/tit.2006.871582.
- [20] CANDÈS, E. J.—ROMBERG, J. K.—TAO, T.: Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Communications on Pure and Applied Mathematics*, Vol. 59, 2006, No. 8, pp. 1207–1223, doi: 10.1002/cpa.20124.
- [21] VASWANI, N.—ZAHZAH, E.: Robust PCA and Robust Subspace Tracking. Preprint, Arxiv 2017.
- [22] BOUWMANS, T. et al.: Robust PCA via Principal Component Pursuit: A Review for a Comparative Evaluation in Video Surveillance. *Special Issue on Background Models Challenge, Computer Vision and Image Understanding (CVIU)*, Vol. 122, 2014, No. 2, pp. 22–34, doi: 10.1016/j.cviu.2013.11.009.
- [23] CANDÈS, E. J.—LI, X. D.—MA, Y.—WRIGHT, J.: Robust Principal Component Analysis? *Journal of the ACM (JACM)*, Vol. 58, 2011, No. 3, Art.No. 11, 37 pp., doi: 10.1145/1970392.1970395.
- [24] ZHOU, X. W.—YANG, C.—YU, W. C.: Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, 2013, No. 3, pp. 597–610, doi: 10.1109/tpami.2012.132.
- [25] GAO, Z.—CHEONG, L. F.—WANG, Y. X.: Block-Sparse RPCA for Salient Motion Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, 2014, No. 10, pp. 1975–1987, doi: 10.1109/tpami.2014.2314663.
- [26] LIU, X.—ZHAO G. Y.—YAO J. W.—QI, C.: Background Subtraction Based on Low-Rank and Structured Sparse Decomposition. *IEEE Transactions on Image Processing*, Vol. 24, 2015, No. 8, pp. 2502–2514, doi: 10.1109/tip.2015.2419084.
- [27] BOUWMANS, T.—SOBRAL, A.—JAVED, S.—JUNG, S. K.—ZAHZAH, E.: Decomposition into Low-Rank Plus Additive Matrices for Background/Foreground Separation: A Review for a Comparative Evaluation with a Large-Scale Dataset. *Computer Science Review*, Vol. 23, 2017, pp. 1–71, doi: 10.1016/j.cosrev.2016.11.001.
- [28] JOLLIFFE, I. T.: Graphical Representation of Data Using Principal Components. Chapter 5. In: Jolliffe, I. T.: *Principal Component Analysis*. Springer, New York, Springer Series in Statistics, 2010, pp. 41–64, doi: 10.1007/978-1-4757-1904-8_5.
- [29] WRIGHT, J.—PENG, Y.—MA, Y.—GANESH, A.—RAO, S.: Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (Eds.): *Advances in Neural Information Processing Systems 22 (NIPS 2009)*. 2009, 9 pp.

- [30] BECKER, S.—CANDES, E.—GRANT, M.: TFOCS: Flexible First-Order Methods for Rank Minimization. Low-Rank Matrix Optimization Symposium, SIAM Conference on Optimization, 2011.
- [31] TANG, G.—NEHORAI, A.: Robust Principal Component Analysis Based on Low-Rank and Block-Sparse Matrix Decomposition. 2011 45th Annual Conference on Information Sciences and Systems (CISS 2011), Baltimore, MD, USA, 2011, doi: 10.1109/ciss.2011.5766144.
- [32] WOHLBERG, B.—CHARTRAND, R.—THEILER, J.: Local Principal Component Pursuit for Nonlinear Datasets. International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), Kyoto, Japan, 2012, pp. 3925–3928, doi: 10.1109/icassp.2012.6288776.
- [33] LIN, Z.—CHEN, M.—WU, L.—MA, Y.: The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. arXiv preprint arXiv:1009.5055, 2010.
- [34] YUAN, X.—YANG, J.: Sparse and Low-Rank Matrix Decomposition via Alternating Direction Methods. Pacific Journal of Optimization, Vol. 9, 2009, No. 1, pp. 1–11.
- [35] RODRÍGUEZ, P.—WOHLBERG, B.: Fast Principal Component Pursuit Via Alternating Minimization. IEEE International Conference on Image Processing (ICIP 2013), Melbourne, Australia, 2013, pp. 69–73, doi: 10.1109/icip.2013.6738015.
- [36] GUYON, C.—BOUWMANS, T.—ZAHZAH, E.: Moving Object Detection via Robust Low Rank Matrix Decomposition with IRLS Scheme. In: Bebis, G. et al. (Eds.): Advances in Visual Computing (ISVC 2012). Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 7431, 2012, pp. 665–674, doi: 10.1007/978-3-642-33179-4_63.
- [37] QIU, C.—VASWANI, N.: Real-Time Robust Principal Components' Pursuit. 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2010, pp. 591–598, doi: 10.1109/allerton.2010.5706961.
- [38] NARAYANAMURTHY, P.—VASWANI, N.: MEDRoP: Memory-Efficient Dynamic Robust PCA. Preprint, December 2017, arXiv:1712.06061v1.
- [39] RODRIGUEZ, P.—WOHLBERG, B.: Incremental Principal Component Pursuit for Video Background Modeling. Journal of Mathematical Imaging and Vision, Vol. 55, 2016, No. 1, pp. 1–18, doi: 10.1007/s10851-015-0610-z.
- [40] JAVED, S.—MAHMOOD, A.—BOUWMANS, T.—JUNG, S. K.: Background-Foreground Modeling Based on Spatiotemporal Sparse Subspace Clustering. IEEE Transactions on Image Processing, Vol. 26, 2017, No. 12, pp. 5840–5854, doi: 10.1109/tip.2017.2746268.
- [41] SOBRAL, A.—BOUWMANS, T.—ZAHZAH, E.: Double-Constrained RPCA Based on Saliency Maps for Foreground Detection in Automated Maritime Surveillance. ISBC 2015 Workshop conjunction with AVSS 2015, Karlsruhe, Germany, 2015, pp. 1–6, doi: 10.1109/avss.2015.7301753.
- [42] JAVED, S.—BOUWMANS, T.—JUNG, S. K.: Combining ARF and OR-PCA Background Subtraction of Noisy Videos. In: Murino, V., Puppo, E. (Eds.): Image Analysis and Applications – ICIAP 2015. Springer, Cham, Lecture Notes in Computer Science, Vol. 9280, 2015, pp. 340–351, doi: 10.1007/978-3-319-23234-8_32.

- [43] ZHOU, P.—FENG, J.: Outlier-Robust Tensor PCA. 2017 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 2017, pp. 3938–3946, doi: 10.1109/cvpr.2017.419.
- [44] HU, W.—YANG, Y.—ZHANG, W.—XIE, Y.: Moving Object Detection Using Tensor Based Low-Rank and Saliently Fused-Sparse Decomposition. IEEE Transactions on Image Processing, Vol. 26, 2017, No. 2, pp. 724–737, doi: 10.1109/tip.2016.2627803.
- [45] SOBRAL, A.—JAVED, S.—JUNG, S. K.—BOUWMANS, T.—ZAHZAH, E.: Online Stochastic Tensor Decomposition for Background Subtraction in Multispectral Video Sequences. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 2015, pp. 946–953, doi: 10.1109/iccvw.2015.125.
- [46] JAVED, S.—BOUWMANS, T.—JUNG, S. K.: Stochastic Decomposition into Low Rank and Sparse Tensor for Robust Background Subtraction. 6th International Conference on Imaging for Crime Prevention and Detection (ICDP 2015), 2015, doi: 10.1049/ic.2015.0105.
- [47] YANG, F.—JIANG, H.—SHEN Z. W.—DENG, W.—METAXAS, D.: Adaptive Low Rank and Sparse Decomposition of Video Using Compressive Sensing. Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 2013, pp. 1016–1020, doi: 10.1109/icip.2013.6738210.
- [48] BRAND, M.: Fast Low-Rank Modifications of the Thin Singular Value Decomposition. Linear Algebra and Its Applications, Vol. 415, 2006, No. 1, pp. 20–30, doi: 10.1016/j.laa.2005.07.021.
- [49] LI, W.—LI, X.: Multiple Object Tracking Based on Modified Algorithm of GMMCP Tracker. IEEE International Conference on Signal and Image Processing (ICSIP), 2017, pp. 11–15, doi: 10.1109/siprocess.2016.7888214.



Jiang CHONG received her Bachelor's degree in computer science and technology from Hunan University in 2002. Then she obtained her Master's degree and her Ph.D. in computer application technology from Central South University in Changsha, China. Now she is Researcher at the School of Computer Science and Engineering, Hunan Women's University, China. Her research interests include machine learning, digital image and video processing directions.