updated B_i, and then perform the sub-block background judgment until all M×N sub-blocks in I_b generate backgrounds.

Normally, the background modeling process will enter the i++ loop many times, that is, the symmetric frame difference processor will be called many times before the background image can be successfully generated, as shown in figure 5.

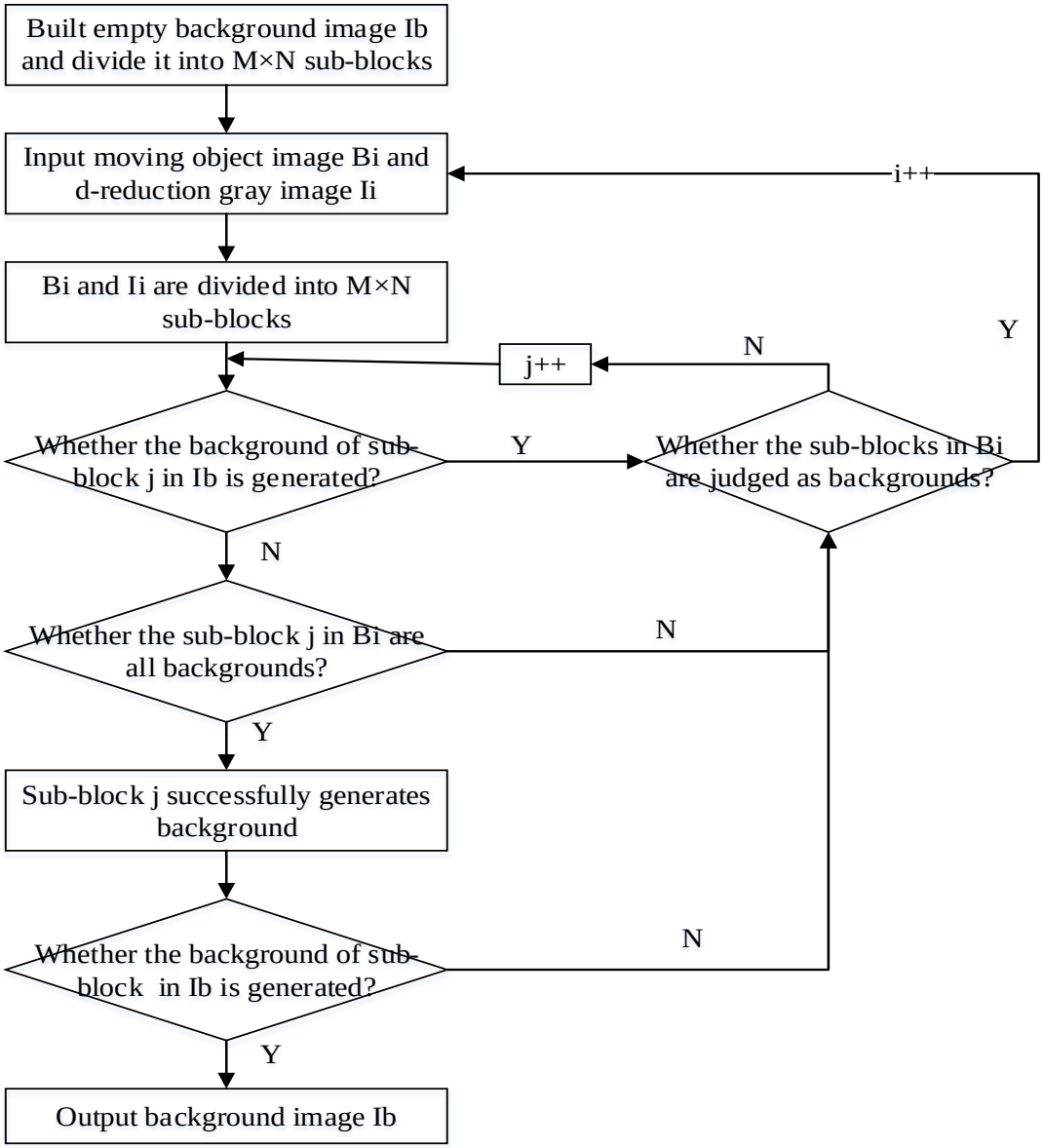


Figure 3. Background Modeling flowchart

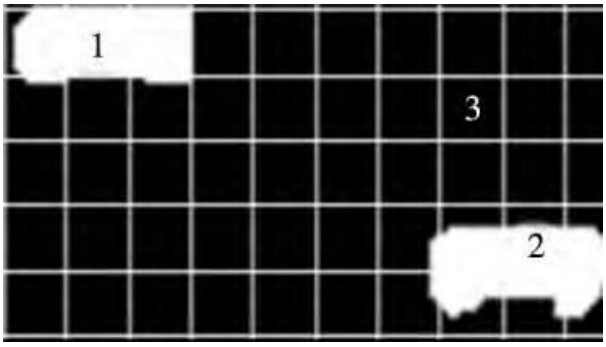
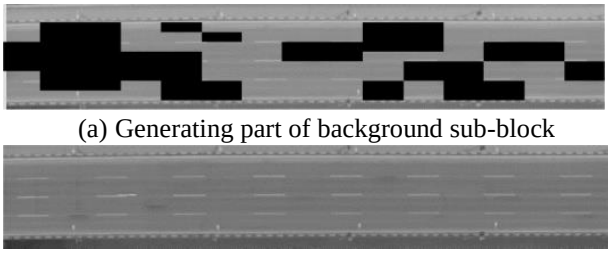


Figure 4. sub-block division of B_i



(a) Generating part of background sub-block

(b) Successfully generated background image

Figure 5. Example of background modeling process

3. Low rank sparse decomposition model

The low rank sparse decomposition is to decompose the observation matrix $D \in R^{m\times n}$ of video is decomposed into the sum between low rank matrix $L \in R^{m\times n}$ and sparse matrix $S \in R^{m\times n}$. The mathematical model can be expressed as:

$$\min_{L,S} R_{rank}(L) + \lambda \| S \|_0, s.t.\ \ D = L + S \qquad (4)$$

Where $R_{rank}(\cdot)$ represents the rank function of matrix L, that is, the rank size of matrix L. $\| \cdot \|_0$ represents the quasi zero norm of matrix S, that is, the number of non-zero elements in matrix S. λ>0 is a penalty factor. Because the rank function and quasi-zero norm in equation (4) are non-convex and non-smooth, it is an NP hard problem.

The convex relaxation of equation (4) is transformed into a convex optimization problem, and its mathematical model can be expressed as:

$$\min_{L,S} \| L \| + \lambda \| S \|_1, s.t.\ \ D = L + S \qquad (5)$$

Where $\| \cdot \|_*$ represents the kernel norm of matrix L, that is, the sum of singular value of matrix L. $\| \cdot \|_1$ represents the L1 norm of matrix S, that is, the sum of the absolute values of all elements in matrix S.

3.1. Low rank modeling

In RPCA algorithm, the low rank matrix L is constrained by kernel norm, which is not the best approximation of rank function. The weighted schatten-p norm in reference [24] is a more accurate rank function approximation and can suppress the noise generated during measurement. Therefore, the weighted Schatten-p norm is used in this study to carry out the low-rank constraint, and its expression is:

$$\| L \|_{w,S_p} = (\sum_{i=1}^{\min\{m,n\}} \omega_i \sigma_i^p)^{1/p} \qquad (6)$$

Where 0<p≤1 is the weighting parameter. σ_i represents the i-th singular value of matrix L. $w = [\omega_1, \cdots, w_i, \cdots, w_{\min(m,n)}]$ represents the weight matrix of corresponding singular value σ_i, it is a non-negative vector. $\omega_i \geq 0$ satisfies non decrement, i.e. $0 \leq \omega_1 \leq \cdots \leq \omega_i \leq \cdots \leq \omega_{\min(m,n)}$. Then, the weighted schatten-p norm is given to the power of P, and the result is obtained.

$$\| L \|_{w,S_p}^p = \sum_{i=1}^{\min\{m,n\}} \omega_i \sigma_i^p \qquad (7)$$

Where, the weight calculation formula $\omega_i = C\sqrt{mn} / [\sigma_i(L) + \varepsilon]$. C is a constant greater than zero. ε represents a minimal constant, usually $\varepsilon = 10^{-16}$.

3.2. Sparse modeling

In RPCA algorithm, the sparse matrix S is constrained by L1 norm, which processes each element in the matrix independently, but the movement of foreground targets is continuous in space. Using this structural prior information, reference [20] proposed a structured sparse norm based on overlapping groups.

As shown in figure 6, assuming the sparse foreground on the 8×8pixel image has two different distributions, as shown in figure 6 (a) and figure 6 (b). In the figure, white indicates that the pixel gray value is large, and black indicates that the pixel gray value is small. Since L1 norm represents the sum of the absolute values of all elements, L1 norm will have similar values in both cases. Structured sparse norm is a t×t pixel window slides on the image line

by line and column by column to obtain the l_∞ norm corresponding to each window. This paper uses 3×3pixel window, six pixels overlap between adjacent windows. On the image with 8×8 pixel, 36 groups sub-space g of 3×3pixel can be obtained in advance, that is, G={g1,g2,...,g36}. Then it takes the l_∞ norm of each subspace, that is, the maximum value of each subspace. Therefore, two significantly different values will be generated. For figure6 (a) and figure6 (b), since there are more groups of pixels with larger values in figure 6 (d), the value of figure 6 (c) will be much smaller than that of figure 6 (d). Under the requirement of minimum sparsity of foreground, figure6 (a) will be more likely to be considered as foreground.

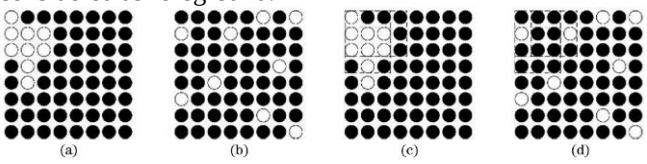


Figure 6. Structured figures based on 3×3 pixel overlapping group. (a) (b) Two different foreground distributions; (c) (d) overlapping group of two different distributions.

In the foreground matrix, each frame image is regarded as a column of elements, then the structured sparse norm constraint for the whole foreground matrix S can be expressed as:

$$\Omega(S)=\sum_{j=1}^{n}\sum_{g\in G}\| s_g^{(j)} \|_\infty \tag{8}$$

Where $s_g^{(j)}$ represents the element covered by the window in j-th frame. $s^{(j)}\in R^m$ represents the j-th column element of the foreground matrix S, that is, the j-th frame of the video. g represents the subspace of the elements covered by the window. G represents the set of g, and $\| s_g^{(j)} \|$ represents the l_∞ norm of $s_g^{(j)}$.

3.3. Model establishment and solution

Based on the above two norm constraints, the low rank sparse decomposition model is obtained by replacing the low rank part and sparse part in equation (5), which can be expressed as:

$$\min_{L,S}\| L \|_{w,S_p}^p+\lambda\Omega(S), s.t.\ D=L+S \tag{9}$$

The above optimization problem is solved by ALM algorithm. The specific solution steps are as follows.

1) The augmented Lagrange function is constructed.

$$\begin{aligned} L(L,S,Y;\mu)=&\| L \|_{w,S_p}^p+\lambda\Omega(S)\\ &+<Y,D-L-S>+\frac{\mu}{2}\| D-L-S \|_F^2 \end{aligned} \tag{10}$$

Where μ is a positive penalty factor. Y is the Lagrange multiplier vector. $\|\cdot\|_F$ represents the F norm of the matrix.

2) Solve the following three sub problems.

$$L=\arg\min_L L(L,S,Y;\mu) \tag{11}$$

$$S=\arg\min_S L(L,S,Y;\mu) \tag{12}$$

$$Y=Y+\mu(D-L-S) \tag{13}$$

3) Fix S and Y and solve L.

$$\begin{aligned} L_{k+1}=&\arg\min_L\frac{1}{\mu_k}\| L \|_{w,S_p}^p\\ &+0.5\| L-(D-S_k+\frac{Y_k}{\mu_k}) \|_F^2 \end{aligned} \tag{14}$$

The generalized soft threshold algorithm (GST) is used to solve equation (14).

4) Fix L and Y and solve S.

$$\begin{aligned} S_{k+1}=&\arg\min_S\frac{\lambda}{\mu_k}\Omega(S)\\ &+0.5\| D-L_{k+1}+\frac{Y_k}{\mu_k}-S \|_F^2 \end{aligned} \tag{15}$$

5) Fix L and S and solve Y.

$$Y_{k+1}=Y_k+\mu_k(D-L_{k+1}-S_{k+1}) \tag{16}$$

6) Update μ_k.

$$\mu_{k+1}=\min(\rho\mu_k,10^7\mu_k) \tag{17}$$

4. Experimental results and analysis

4.1. Experimental environment

In order to verify the effectiveness of the proposed algorithm, nine videos with complex background characteristics in three mainstream databases: I2R [25], CDnet2014 [26] and Wallflower [27] are selected for comparative experiments. I2R: Airport, Bootstrap, Curtain, SwitchLight and WaterSurface; CDnet2014: Highway, Office; Wallflower: CamouFlage, WavingTrees. The characteristics of the selected dataset are shown in Table 1.

Table 1. Information of 9 video datasets for comparison

Dataset	Video resolution	Scene feature
Airport	176×144	Dynamic background, changing time information on the frame
Bootstrap	160×120	Static background, some moving people in the room
Curtain	160×128	Dynamic background, the curtains moving slowly and light change gradually
SwitchLight	160×128	Dynamic background, light mutation
WaterSurface	160×128	Dynamic background, watersurface fluctuation
Highway	360×240	Dynamic background, the high speed vehicle and the moving shadow
Office	360×240	Staticback ground, a people moving locally with strong light
CamouFlage	160×120	Dynamic background, background with a changing screen
WavingTrees	160×120	Dynamic background, the swinging branch

100 frames are selected equidistantly for each video data set, and compared with the mainstream GoDec, PCP, LSD, WNNM and WSNM algorithms. The running environment of all experiments in this paper is Intel (R) core (TM) i7-8550U, 1.80GHz 8GB memory and matlab2017a [28-30]. All the experimental results are processed by median filter.

According to the experience of existing algorithms, the termination condition used in this experiment is $\| D-L_{k+1}-S_{k+1} \|_F^2 / \| D \|_F^2 \le 10^{-7}$, and the value of some parameters are: ε=10^{-16}, ρ=1.05, $\lambda = 1/\sqrt{m}$, $\mu_1 = 12.6/\sigma_1(D)$, $\sigma_1(D)$ represents the maximum singular value of matrix D. For the value of parameter p, according to the quantitative experiment in literature [18], when p=0.7, the recovery effect is the best and the sensitivity to rank and noise is the smallest. Therefore, p is taken as 0.7 in this paper. Please refer to references [13,14,15,17,20] for the specific experimental parameters of the comparison algorithm. The value of C will be discussed in 4.2.

4.2. Evaluation index

The comprehensive measurement index F-measure is used to evaluate the separation effect, and its expression is:

$$F = \frac{2rp}{r+p} \tag{18}$$

Where $p = \frac{TP}{TP+FP}$ denotes precision, that is, the ratio of foreground pixels correctly recovered by the algorithm to all foreground pixels recovered by the algorithm. $r = \frac{TP}{TP+FN}$ denotes recall, that is, the ratio of foreground pixels correctly recovered by the algorithm to real foreground pixels. TP indicates that the foreground pixels are judged as foreground pixels, that is, the number of foreground pixels judged correctly by the algorithm. FP indicates that the background pixel is judged as the foreground pixel, that is, the number of foreground pixels judged wrong by the algorithm. FN indicates that the foreground pixels are judged as background pixels, that is, the number of foreground pixels not judged by the algorithm [31]. According to equation (14), the higher the value of F-measure, the better the separation effect.

Taking the airport dataset as an example, other parameters are fixed, the value of C is between [0.1,1], and the analysis is carried out in steps of 0.1. As can be seen from figure 7, when C= 0.1, the F-measure value is the highest, so C in this paper is 0.1.

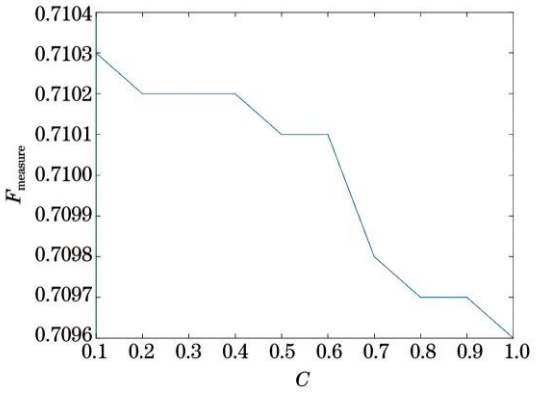


Figure 7. F-measure values under the different C values on Airport dataset.

4.3. Comparison experiments

The number of real foreground frames provided by different databases also varies. I2R provides 20 frames, CDnet2014 provides all the real foreground frames, and Wallflower only provides the most representative one frame. The experimental results obtained are shown in figure 8, and F values under quantitative experiment are given in Table 2.

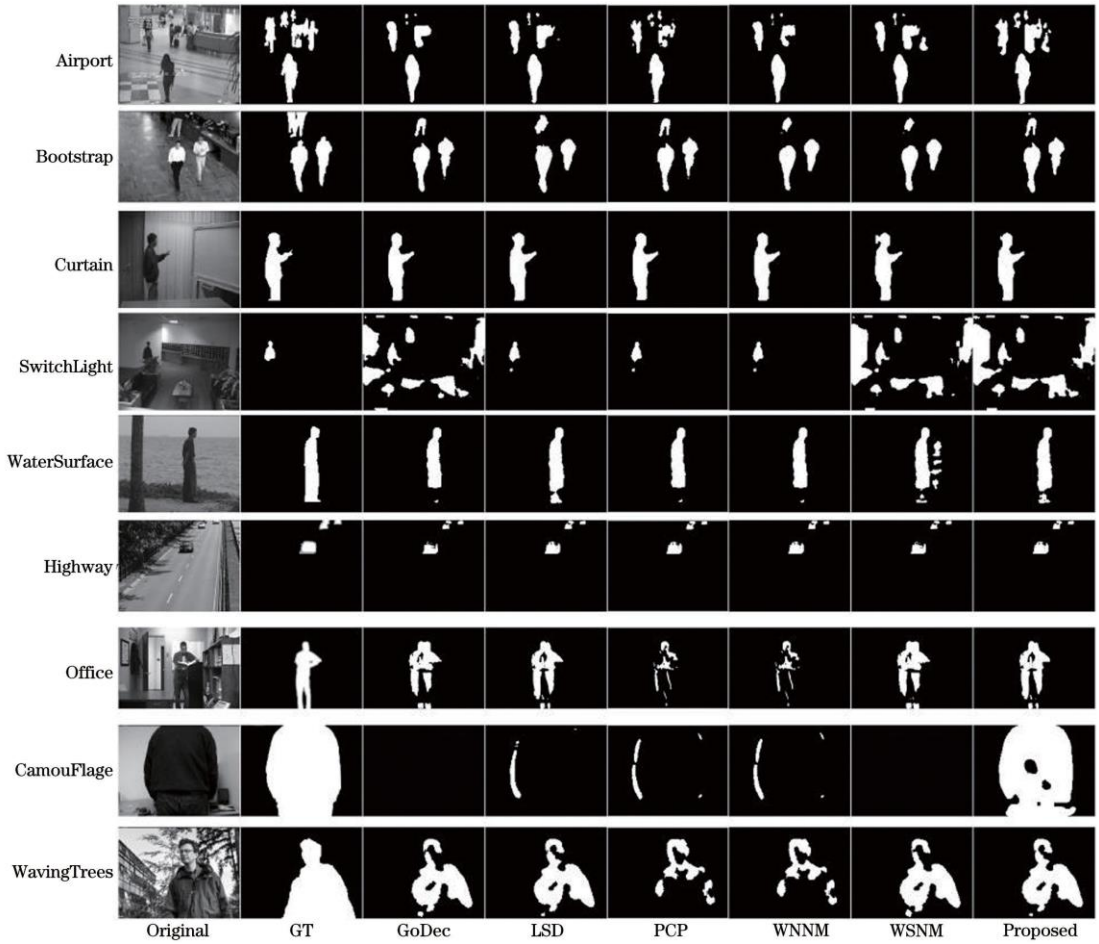


Figure 8. Results of six algorithms on nine sequences

Table 2. F values with different methods

video	F-measurement					
	GoDec	LSD	PCP	WNNM	WSNM	Proposed
airport	0.6712	**0.7384**	0.6712	0.6712	0.6277	0.7214
bootstrap	0.5221	0.5237	0.5199	0.5199	0.4977	**0.5246**
curtain	0.8418	0.8707	0.8420	0.8418	0.8381	**0.8725**
switchlight	0.1379	0.7995	0.8077	**0.8100**	0.1401	0.1433
watersurface	0.8850	**0.9281**	0.8788	0.8772	0.8523	0.9207
highway	0.8163	0.8973	0.8998	**0.9005**	0.8254	0.8980
office	0.5024	0.4758	0.3888	0.3903	0.5054	0.4759
camouflage	01221	0.1767	0.2264	0.2259	0.0478	**0.9598**
wavingtrees	0.7168	**0.7324**	0.4240	0.4257	0.7192	0.7319
average	0.5672	0.6726	0.6176	0.6180	0.5574	**0.6942**

Figure 8 shows the foreground effects of GoDec, LSD, PCP, WNNM, WSNM and the proposed in this paper in nine scenarios (airport, bootstrap, curtain, switchlight, watersurface, highway, office, camouflage and wavingtrees). It can be seen from the bootstrap data set that the proposed algorithm is not much different from other methods for front background separation under static background. It can be seen from the switchlight dataset that the method in this paper cannot be applied to the case of sudden illumination change, but it has excellent separation effect under other complex and changeable dynamic backgrounds. Especially on the camouflage dataset, its background is a changing computer screen [32-34]. When other algorithms fail, the proposed algorithm can still correctly separate the foreground and background, It shows the excellent performance. From the measurement index F values in Table 2, it can be seen that most F values with dynamic background in the nine data sets belong to the highest value or the second highest value, and the average F value of the proposed in this paper is the highest in all methods [35-40]. Therefore, in a comprehensive view, the performance of new method is better than other comparison algorithms on the whole.

5. Conclusion

Based on the above two constraints, the low rank and sparse parts are modeled respectively, and a low rank sparse decomposition model for video foreground and background separation is proposed. The low rank part adopts the symmetrical frame difference method, which can more accurately approximate the rank function, suppress the noise generated during measurement. It is more suitable for background modeling. The sparse part adopts the structured sparse norm based on overlapping groups, and uses the foreground structure information to judge the foreground target more accurately, which is more conducive to foreground modeling. Experimental results show that the proposed algorithm is not robust to sudden illumination scenes, but it can obtain good separation effect in static background and better separation effect in complex and diverse dynamic background. In the next work, not only the sudden change of illumination should be considered to make it more robust, but also the movement of camera should be considered.

Acknowledgements.

The author would like to thank the reviewers and the editor for their valuable comments

References

[1] Shoulin Yin, Ye Zhang, Shahid Karim. Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model[J]. IEEE Access. volume 6, pp: 26069 - 26080, 2018.

[2] Yin Shoulin, Liu Jie, Teng Lin. A new krill herd algorithm based on SVM method for road feature extraction[J]. Journal of Information Hiding and Multimedia Signal Processing, v 9, n 4, p 997-1005, July 2018.

[3] Yin K, Li J, Li L, et al. Adaptive Feature Update Object-Tracking Algorithm in Complex Situations[J]. Acta Optica Sinica, 2019, 39(11):1115002.

[4] Q. Wang, Siyuan, et al. Approach for recognizing and tracking beacon in inter-satellite optical communication based on optical flow method.[J]. Optics express, 2018, 26(21):28080-28090.

[5] Shoulin Yin, Ye Zhang. Singular value decomposition-based anisotropic diffusion for fusion of infrared and visible images[J]. International Journal of Image and Data Fusion, 10(2), pp: 146-163, 2019.

[6] Peng B, Cai X, Zhang Y, et al. Automatic vehicle detection from UAV videos based on symmetrical frame difference and background block modeling[J]. Journal of Southeast University(Natural Science Edition), 47(4):685-690, 2017.

[7] Elkabetz A, Yitzhaky Y. Background modeling for moving object detection in long-distance imaging through turbulent medium[J]. Applied Optics, 2014, 53(6):1132.

[8] Zhu X, Zhang C, Xue J P, et al. Background subtraction via time continuity and texture consistency constraints[J]. Journal of the Optical Society of America A, 2019, 36(9):1495.

[9] Boyer C, Bigot J, Weiss P. Compressed sensing with structured sparsity and structured acquisition[J]. Applied and Computational Harmonic Analysis, 2019, 46(2):312-350.

[10] D Sjoerd, Christian J H, Holger R. One-bit compressed sensing with partial Gaussian circulant matrices[J]. Information and Inference: A Journal of the IMA, 2019(3):3.

[11] S. Daei, F. Haddadi, A. Amini and M. Lotz, "On the Error in Phase Transition Computations for Compressed Sensing," in IEEE Transactions on Information Theory, vol. 65, no. 10, pp. 6620-6632, Oct. 2019, doi: 10.1109/TIT.2019.2920640.

[12] Liu S S, Ying G G, Liu S, et al. Analysis of 21 progestagens in various matrices by ultra-high-performance liquid chromatography tandem mass spectrometry (UHPLC-MS/MS) with diverse sample pretreatment[J]. Analytical & Bioanalytical Chemistry, 2014, 406(28):7299-7311.

[13] Zhou, T., Tao, D.: GoDec: Randomized low-rank & sparse matrix decomposition in noisy case. In: International Conference on Machine Learning, pp. 33–40 (2011)

[14] Gu S, Lei Z, Zuo W, et al. Weighted Nuclear Norm Minimization with Application to Image Denoising[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2014.

[15] D. Zhang, Y. Hu, J. Ye, X. Li and X. He, "Matrix completion by Truncated Nuclear Norm Regularization," 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2192-2199, doi: 10.1109/CVPR.2012.6247927.

[16] F. Nie, H. Wang, X. Cai, H. Huang and C. Ding, "Robust Matrix Completion via Joint Schatten p-Norm and lp-Norm Minimization," 2012 IEEE 12th International Conference on Data Mining, 2012, pp. 566-574, doi: 10.1109/ICDM.2012.160.

[17] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang and L. Zhang, "Weighted Schatten p -Norm Minimization for Image Denoising and Background Subtraction," in IEEE Transactions on Image Processing, vol. 25, no. 10, pp. 4842-4857, Oct. 2016, doi: 10.1109/TIP.2016.2599290.

[18] Teng Lin, Hang Li and Shoulin Yin. Modified Pyramid Dual Tree Direction Filter-based Image De-noising via Curvature Scale and Non-local mean multi-Grade remnant multi-Grade Remnant Filter [J]. International Journal of Communication Systems. v 31, n 16, November 10,pp. e.3486.1-e.3486.12, 2018.

[19] C. Guyon, T. Bouwmans and E. Zahzah, "Foreground detection based on low-rank and block-sparse matrix decomposition," 2012 19th IEEE International Conference on Image Processing, 2012, pp. 1225-1228, doi: 10.1109/ICIP.2012.6467087.

[20] Liu X, Zhao G, Yao J, et al. Background Subtraction Based on Low-Rank and Structured Sparse Decomposition.[C]// IEEE International Conference on Electronics. IEEE, 2013.

[21] Shi F, Qiu F, X Li, et al. Detecting and Tracking Moving Airplanes from Space Based on Normalized Frame Difference Labeling and Improved Similarity Measures[J]. Remote Sensing, 2020, 12(21):3589.

[22] Wang S, F Yu, Zhou C, et al. Straw Burning Detection Method Based on Improved Frame Difference Method and Deep Learning[C]// 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC). IEEE, 2020.

[23] Nan Zhao, Xiao-Wei Wang, and Shou-Lin Yin. Research of Fire Smoke Detection Algorithm Based on Video [J]. International Journal of Electronics and Information Engineering. Vol. 13, No. 1, pp. 1-9, 2021.

[24] Zha Z, Zhang X, Wu Y, et al. Non-Convex Weighted Schatten p-Norm Minimization based ADMM Framework for Image Restoration. 2017. arXiv:1704.07056

[25] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu and Qi Tian, "Statistical modeling of complex backgrounds for foreground object detection," in IEEE Transactions on Image Processing, vol. 13, no. 11, pp. 1459-1472, Nov. 2004, doi: 10.1109/TIP.2004.836169.

[26] Wang Y, Jodoin P M, Porikli F, et al. CDnet 2014:an expanded change detection benchmark dataset [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press,2014: 393-400.

[27] Yufeng Wei, Mingli Jing, Lan Li, et al. Video Foreground-Background Separation via Weighted Schatten-p Normand Structured Sparsity Decomposition[J]. Laser & Optoelectronics Progress, 58(8), pp.0815008, 2021.

[28] Shoulin Yin, Hang Li, Asif Ali Laghari, et al. A Bagging Strategy-Based Kernel Extreme Learning Machine for Complex Network Intrusion Detection[J]. EAI Endorsed Transactions on Scalable Information Systems. 21(33), e8, 2021. http://dx.doi.org/10.4108/eai.6-10-2021.171247

[29] Qingwu Shi, Shoulin Yin, Kun Wang, et al. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems (2021). https://doi.org/10.1007/s12530-021-09392-3

[30] S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.

[31] Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation [J]. Multimedia Tools and Applications. Vol. 79, pp. 31049-31068, 2020.

[32] Luong H V, Joukovsky B, Eldar Y C, et al. A Deep-Unfolded Reference-Based RPCA Network For Video Foreground-Background Separation[C]// 2020 28th European Signal Processing Conference (EUSIPCO). 2021.

[33] Yang Z, Fan L, Yang Y, et al. Generalized nuclear norm and Laplacian scale mixture based low-rank and sparse decomposition for video foreground-background separation[J]. Signal processing, 2020, 172(Jul.):107527.1-107527.10.

[34] Yin, S., Li, H. & Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images [J]. Sensing and Imaging, vol. 21, 2020. https://doi.org/10.1007/s11220-020-00314-2

[35] Laghari, A.A., Wu, K., Laghari, R.A. et al. A Review and State of Art of Internet of Things (IoT). Arch Computat Methods Eng (2021). https://doi.org/10.1007/s11831-021-09622-6

[36] Laghari A A, Laghari M A. Quality of experience assessment of calling services in social network[J]. ICT Express, 2021(2).

[37] Laghari A A, Laghari K, Memon K A, et al. Quality of Experience (QoE) Assessment of Games on workstations and Mobile[J]. Entertainment Computing, 2020, 34:100362.

[38] A. A. Laghari, H. He, A. Khan, N. Kumar and R. Kharel, "Quality of Experience Framework for Cloud Computing (QoC)," in IEEE Access, vol. 6, pp. 64876-64890, 2018, doi: 10.1109/ACCESS.2018.2865967.

[39] Laghari, A.A., He, H., Shafiq, M. et al. Application of Quality of Experience in Networked Services: Review, Trend & Perspectives. Syst Pract Action Res 32, 501–519 (2019). https://doi.org/10.1007/s11213-018-9471-x

[40] Laghari, A.A., Jumani, A.K. & Laghari, R.A. Review and State of Art of Fog Computing. Arch Computat Methods Eng 28, 3631–3643, 2021.