Foreground Segmentation via Fusion using for Low Illumination Scene

Yang Li (liyang_19901222@163.com)
Jiangsu Vocational College of Information Technology

Research Article

Keywords: Foreground Segmentation, Features Extraction, Super-Pixel, Neural Network

Posted Date: April 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1539863/v1

Foreground Segmentation via Fusion using for Low Illumination Scene

Li Yang[1*]

[1*]School of IoT Engineering (School of Information Security), Jiangsu Vocational College of Information Technology, Qianou Road, Wuxi, 214153, Jiangsu, China.

Corresponding author(s). E-mail(s): liyang_19901222@163.com;

Abstract

Foreground segmentation (FS) plays a fundamental and important role in computer vision. In recent years, scholars have put forward many effective methods. However, these methods are limited in low illumination scenes. In order to improve the performance of FS in low illumination scene, a simple but effective method is proposed via feature extraction and RGB-D camera. Firstly, the coined Iterated Robust CUR (IRCUR) is used to get candidate foreground for depth image sequence. At the same time, the RGB image sequence is segmented using the simple linear iterative clustering (SLIC). Then, feature extraction is performed on the candidate matrix region corresponding to the super-pixel block. Then, the neural network is trained by using the acquired super-pixel features. Experiments show that the average F-measure value of this method is 35% higher than that of other methods only based on RGB images.

Keywords: Foreground Segmentation, Features Extraction, Super-Pixel, Neural Network

1 Introduction

Foreground segmentation (FS) technology has attracted extensive attention in video processing and lays an important foundation for subsequent target detection and recognition. In recent years, many excellent algorithms have been proposed for different scenes [16, 17, 19, 22, 27]. They have achieved good results in different application scenes, such as slow motion, shaking branches, fluctuating lake surface, etc.

The existing methods can be divided into three categories: pixel based method, deep learning based method and subspace learning based method. Some famous pixel based methods are Vibe [2], Gaussian Mixture Model (GMM) [24], pixel-based adaptive segmenter (PBAS) [12] which use spatial domain or neighborhood pixels to establish a model to realize FS. These methods are simple and easy to understand. However, when the background is complex, the detection performance will be affected because it only depends on the pixel value. The deep learning based methods [16, 17, 19, 22, 27] extract features on the basis of pixels and time complexity is relatively high and their performance is better than pixel level based methods. However, the deep learning based methods need a lot of annotation data to train model and their generalization ability needs to be improved. The methods based on subspace learning mainly includes robust principal component analysis (RPCA) [5] and its extended models [8, 10, 28] which can accurately extract the foreground in a simple scene. If the background is complex, the foreground matrix contains a lot

of background information. In addition, their time complexity is high. The performance of the above methods in low illumination scenes will be limited.

In order to improve the performance of FS in low illumination scenes, a simple but effective method is proposed based on feature extraction and RBG-D camera. The method is divided into three steps. Firstly, the coined Iterated Robust CUR (IRCUR) is used to extract candidate foreground matrix. Meanwhile, the RGB image sequence is segmented using the simple linear iterative clustering (SLIC). Secondly, the one-dimensional image and two-dimensional features are extracted from the candidate foreground matrix. Finally, the extracted features are used to train the neural network classifier.

The contributions of this paper are as follows:

- ◇ A dataset for low illuminance FS is made using RGB-D camara, which contains 1730 annotated RGB-D images and three challenging scenarios, such as: illumination change, intermittent movement and similar color.
- ◇ A method is proposed to effectively extract the features of super pixels from the candidate foreground matrix. It includes one-dimensional and two-dimensional feature extraction of outliers in super-pixel blocks corresponding to matrix decomposition.
- ◇ Based on the proposed feature extraction for super-pixel , a FS algorithm based on feature selection is presented. Although this algorithm belongs to supervised algorithm, it has strong generalization ability based on feature classification of outliers.

2 Related Works

2.1 Pixel Based Method

Stauffer et al. [24] propose a gaussian mixture model which is established for each pixel, and the model is updated by linear approximation. Though GMM has been widely used, and the problem of converging to the worse solution is usually encountered if the main mode is stretched to over control the weak distribution. An appropriate splitting operation and the corresponding criterion for the selection of candidate modes is applied to solve this problem. A new idea is adopted for target detection, and the random principle is introduced into target detection. The basic idea is to randomly sample each pixel within the radius of R as the background model of the pixel, and the default sampling point is 20. Then, for the new pixel, compare its value with the background model [2]. In order to improve the detection ability of the model to complex models such as such as illumination variations, rippling water, camera jitter, color features and texture features are fused using Choquet fuzzy integral [9].

2.2 Deep Learning Based Method

In recent years, the method based on deep learning [3, 19, 23, 26, 29] has achieved satisfactory results in FS. Braham [3] is the first researcher to apply convolutional neural network to FS. The structure of ConvNet model [3] is similar to lenet-5. The steps of ConvNet are as follows: background extraction, data set generation, model training and background subtraction. Wang et al. [29] adopts an improved CNN model with different patch size. The size of the patch is 31×31. A new model is proposed using multi-scale full revolutionary network (MFCC) architecture, rather than patch wise algorithm. In order to capture the motion characteristics of foreground target in time domain, Sakkos [23] presents a model to track the temporal changes in the video sequence by applying 3D convolution to the latest frame of the video. To deal with various difficult scenes, such as lighting change, background or camera motion, camouflage effect, shadow, etc. To solve these problems, a two robust encoder decoder neural networks is proposed, which generate multi-scale feature coding in different ways, and only a small number of training samples can be used for end-to-end training [18]. Some of the best performance architectures, such as FgSegNet [18], are usually too suitable for training video based on the detection of single frame spatial appearance cues. A new compact multi-threaded automatic encoder depth structure is presented for robust moving target detection. It has better generality than FgSegNet [18], and the required weight parameters are nearly 30 times less than FgSegNet [18]. Motion and change cues are estimated using a multimodal background subtraction module combined with flux tensor motion estimation [21].

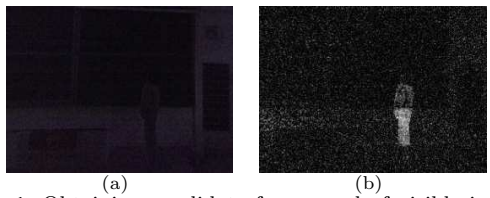
2.3 Subspace Learning Based Method

Recently, the method based on subspace learning [6] has attracted scholars' attention in the field of FS because of its excellent effect. Cans et al. [6] Propose Robust Principal Component Analysis (RPCA) which is a subspace learning method and is widely used in FS. When the scene is complex, RPCA cannot well realize moving target detection. The concept of Common Vector Approach with Gram-Schmidt orthogonalization is applied to sperate foreground [13]. In order to highlight the foreground information, the original video is divided into three parts: background information, noise information and foreground information. Then the optical flow method performs significance detection to detect the candidate foreground information, and the candidate information is added to the RPCA model [20]. To handle dynamic background and slow motion, segmentation and saliency are used to constrain the RPCA model [15]. The performance of these methods deteriorates in the case of dynamic background scenes, camera jitter, camouflage moving objects and / or lighting changes. This is because of a basic assumption that the elements in the sparse component are independent of each other, so the spatio-temporal structure of the moving object is lost. In order to solve this problem, a spatiotemporal sparse RPCA moving target detection algorithm is proposed, which regularizes the sparse components in the form of graph Laplacian [14]. RPCA is widely used in background separation, but its time complexity is high. To solve this problem, coined Iterated Robust CUR (IRCUR) is applied to improve the computational efficiency by employing CUR decomposition when updating the low rank component [4].

3 The Proposed Algorithm

Facing the FS task in the weak illumination scene, it is impossible to use simple pixel feature. In the weak illumination scene, the detection effect will be limited by relying only on the visible image. In order to improve the detection effect, more features must be integrated into the task.

In recent years, Microsoft's Azure Kinect DK camera has been successfully applied in target detection and motion recognition. This device can produce well calibrated RGB and depth frames, usually capturing 30 frames per second, which is suitable for motion detection algorithms.



(a) (b)

Fig. 1 Obtaining candidate foreground of visible image. (a) An input visible image frame. (b) Obtained candidate foreground using IRCUR [4].

3.1 Candidate Foreground Matrix

The original RPCA model is a good helper for extracting background, but it has high time complexity. To this end, IRCUR [4] is proposed by employing CUR decomposition. In this paper, IRCUR [4] is applied to obtain candidate foreground matrix from depth image sequence.

For a given video composed of multiple image frames, each frame can be stacked as a column of the matrix D at first, then D is decomposed into a low-rank part A and a sparse component E.

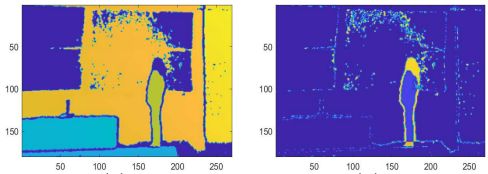
$$\min_{A,E} \|A\|_* + \lambda \|E\|_1, \text{ s.t. } D = A + E, \qquad (1)$$

where $\|\cdot\|_*$ is the nuclear norm of a matrix, $\|\cdot\|_1$ presents the ℓ_1 norm of a vector, and $\lambda > 0$ is a parameter.

Formula 1 can be solved by IRCUR [4]. Through the test, the decomposition speed of IRCUR is ten times that of RPCA.

As can be seen from the figure 1, the foreground candidate matrix contains a lot of noise due to the weak illumination. The upper body of pedestrians is lost because the color of pedestrian coat is similar to that of background. Thus IRCUR is unable to process complex scenes. For complex scenes, scholars have proposed extended RPCA models. Although the performance of these models has been improved, the time complexity is very high. In order to solve this problem, the depth information of the scene is obtained to improve the detection performance.

From the figure 2, it can be seen that the depth candidate matrix also contains a lot of noise



(a) (b)

Fig. 2 Obtaining candidate foreground of depth image. (a) An input depth image frame. (b) Obtained candidate foreground using IRCUR [4].

information, but it can detect pedestrian area relatively completely. Next, we will extract and fuse the features of the candidate foreground matrix of RGB and depth images.

image segmentation, and usually do not destroy the boundary information of objects in the image.

3.2 Feature Extraction

From figure 1 (b) and figure 2 (b), it is obvious that only using a single threshold can not complete the task of distinguishing between foreground and background due to the weak illumination. Therefore, more information needs to be fused to obtain more robust features.

3.2.1 Super Pixel Segmentation

By observing the candidate matrix of the depth image, it can be found that if a region belongs to the foreground, the difference of each value in the region is very small, and most of the value in the region are generally larger as shown in figure 2 (b). If a region belongs to the background, the difference of each value in the region is large, and most of the value in the region are smaller. In order to obtain the feature of the region, the simple linear iterative clustering (SLIC) [30] is adopted. A super pixel is a small area composed of a series of pixels with adjacent positions and similar characteristics, such as color, brightness and texture. Most of these small areas retain effective information for further. Figure 3 is the result obtained by SLIC.

3.2.2 One-Dimensional Feature

From figure5, it is can be found that the values in the background super pixels are distributed around 0 and the value in the foreground super pixel is relatively large. Therefore, histogram,

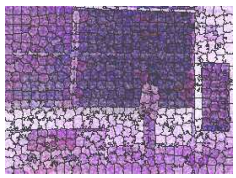
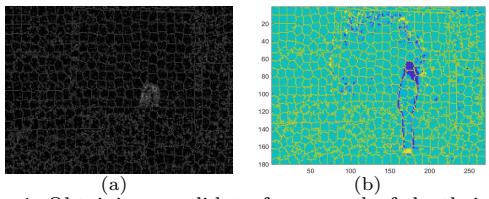
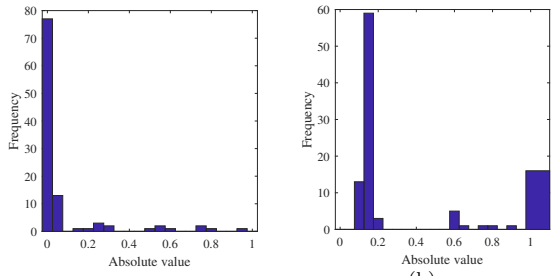


Fig. 3 Obtaining segmentation using SLIC [30]



(a) (b)

Fig. 4 Obtaining candidate foreground of depth image. (a) An input depth image frame. (b) Obtained candidate foreground using IRCUR [4].



(a) (b)

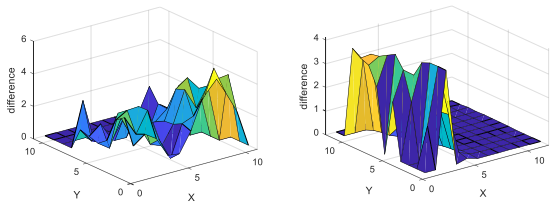
Fig. 5 Super pixel histogram. (a) Histogram for background super pixel. (b) Histogram for foreground super pixel.

mean and variance are used to calculate one-dimensional features [25]. The group distance of the histogram is 0.1, there are 10 groups in total, and the range is 0-1. The formulas for computing mean and variance are given in formula 2 and formula 3.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{2}$$

where $\bar{x}$ represents the means, n represents the number of a super pixel block and x_i denotes the ith value of a super pixel block.

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} \tag{3}$$



(a) (b)

Fig. 6 Flatness. (a) Flatness for background super pixel. (b) Flatness for foreground super pixel.

where S^2 the variance of the value in a super pixel block.

3.2.3 Two-Dimensional Feature

Although one-dimensional feature is effective, it ignores two-dimensional information. We propose a two-dimensional measurement index called *flatness*. The so-called flatness is the difference between the pixel and the surrounding pixels. The flatness of foreground super pixel is relatively low because their values differ less from those of their neighbors as shown in figure 6. The flatness of background super pixel is relatively high because their values differ more from those of their neighbors. Figure 6 (a) is rougher than figure 6 (b). The steps for solving flatness is given in algorithm 1.

Algorithm 1 Solving Flatness

Require: candidate foreground matrix of depth image D_c.

Ensure: the flatness G of the candidate foreground matrix.

1: Convolute the image using the formula by (4) and get the horizontal gradient H.
2: Convolute the image using the formula by (5) and get the vertical gradient V.
3: Calculate the flatness of each pixel by (6) and get the gradient of the candidate foreground matrix G. **return** G

$$H = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{4}$$

$$V = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{5}$$

$$G = \sqrt{H.^2 + V.^2} \tag{6}$$

3.2.4 Super Pixel Thinning

Because the visible image is calibrated with the depth image, there will be pixel offset. In order to reduce the error of feature statistics caused by offset, image corrosion is used to refine super pixels as shown in formula 7.

$$A - B = \{x \mid B_x \subseteq A\} \tag{7}$$

where A is the image to be corroded, and B is the convolution kernel of corrosion as shown in formula 8.

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{8}$$

3.3 Background Update

Existing RPCA models [6, 15] are prone to voids when performing slow-moving foreground segmentation. To this end, the background updating strategy of the proposed algorithm is as follows: firstly, ten frames without foreground are selected as the background of the model. If a new frame needs to be detected, this frame and ten frames of the background are put into the IRCUR [4] model for decomposition. After detection, if the model contains foreground, the background template will not be updated. Otherwise, replace the oldest frame in the background with that detected frame.

3.4 Procedures of Proposed Algorithm

The steps for model training are given in algorithm 2.

Algorithm 2 Model Training

Require: visible images $\{I_k\}_{k=1}^N$, depth images $\{D_k\}_{k=1}^N$, ground truth $\{G_k\}_{k=1}^N$ of visible images $\{I_k\}_{k=1}^N$.
Ensure: the trained model.
1: initialization: $region = 20$, $rank = 1$.
2: get candidate foreground $\{DF_k\}_{k=1}^N$ of depth images $\{D_k\}_{k=1}^N$ using IRCUR [4].
3: get segmentation results $\{C_k\}_{k=1}^N$ of visible images $\{D_k\}_{k=1}^N$ using SLIC [30].
4: get flatness $\{F_k\}_{k=1}^N$ of candidate foreground $\{DF_k\}_{k=1}^N$ using subsubsection 3.2.3.
5: **while** not converged **do**
6: take out the coordinate information of a super-pixel from $\{C_k\}_{k=1}^N$.
7: refine coordinates using 3.2.4.
8: take out the corresponding pixel according to the refined coordinate information.
9: Judge whether the super-pixel has a foreground according to ground truth $\{G_k\}_{k=1}^N$.
10: **if** true **then**
11: compute mean, variance and histogram of candidate foreground $\{DF_k\}_{k=1}^N$ and flatness $\{F_k\}_{k=1}^N$ respectively.
12: add the features obtained in the previous step to the foreground feature F
13: **else**
14: compute mean, variance and histogram of candidate foreground $\{DF_k\}_{k=1}^N$ and flatness $\{F_k\}_{k=1}^N$ respectively.
15: add the features obtained in the previous step to the background feature B
16: **end if**
17: **end while**
18: train BP neural network [7] using F and B.
return the trained model.

4 Results and Evaluation

4.1 Data-set Information

This paper collects three image sequences through Azure Kinect DK sensor. The information of the three image sequences is shown in the table 1. The three image sequences are *Corridor*, *Sofa* and *Classroom* respectively. The *corridor* sequence is mainly a group of people walking along the corridor. There is a problem of illumination change in this image sequence. The *sofa* sequence includes some people sitting on the sofa, and the depth information changes very weakly because of intermittent movement. The *classroom* sequence mainly includes some people walking through the podium. The TOF camera is sensitive to the blackboard, causing a lot of noise. In addition, the color of some pedestrians' clothes is similar to that of the blackboard.

Table 1 Dataset information.

dataset name	Corridor	Sofa	Classroom
Resolution	180×270	180×270	180×270
Number	597	382	751

4.2 Parameter Setting

The *region* and *rate* of SLIC [30] are set to 12 and 0.1 respectively. The BP neural network contains three hidden layers, each with 10 neurons. The size of input layer is $2 \times (1 + 1 + 10) = 24$. The size of input layer is 2.

4.3 Evaluation Index

Recall (Re), Precision (Pre), Accuracy (Acc) and F-measure (F) [15] are used for objective evaluation and their calculation methods are shown as follow.

$$Re = \frac{TP}{TP + FN} \tag{9}$$

$$Pre = \frac{TP}{TP + FP} \tag{10}$$

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{11}$$

$$F = 2\frac{Re * Pre}{Re + Pre} \tag{12}$$

where, False Negative (TN) represents that the pixel is determined to be a negative sample, but it is actually a positive sample. False Positive (FP) represents that the pixel is determined to be a positive sample, but it is actually a negative sample.

True Negative (TN) represents that the pixel is determined to be a negative sample, which is also a negative sample in fact. True Positive (TP) represents that the pixel is determined to be a positive sample, which is actually a proof sample.

4.4 Results and Discussion

From table 2 we can see that the proposed algorithm achieves the highest values for Re, Pre, Acc and F. It means that my fusion method of visible image and depth image is effective. It is 0.6 higher than the common method in F-measure such as: MOG, VIBE, IA-IA3DFHRI.

Figure 7, figure 8 and figure 9 give some visualization results. From figure 7 it can be seen because the light intensity is low, VIBE and IA-IA3DFHRI does not detect the area of travelers. MOG detects partial area. From figure 8 it can be seen VIBE, IA-IA3DFHRI, MOG and TRPCA detect all the changes of illumination, which shows that they have poor robustness to illumination. Figure 9 presents VIBE, IA-IA3DFHRI and MOG lose most areas of the human body, which means can't handle intermittent motion. Because the ten frames in the background template do not contain background information, if there is a moving target in the foreground, it will be completely detected, so there will be no hole when dealing with intermittent motion. The proposed method can deal with the problem of similar color between foreground and background, illumination change because the depth image is not sensitive illumination. Table 3 gives the time consumption for processing per frame. The time consumption for the proposed method is twice as much as that of MOG and IA-IA3DFHRI. To sum up, the proposed method achieves the best effect.

Table 3 Time consumption for Per Frame (seconds).

Methods	Pro	MOG	VIBE	IA-IA3DFHRI	TRPCA
Time	0.86	0.48	1.20	0.39	3.45

5 Conclusion

In order to deal with the FS in low illumination scene, we integrate the depth information into the algorithm. Through the feature extraction of super pixel, the proposed method can effectively deal with the problems of intermittent motion, illumination change and so on. Although the proposed algorithm has been greatly improved compared with the algorithm of only using visible image, the method in this paper loses the features of time domain. More features of time domain will be used to improve the performance of the model in the feature.

Acknowledgments. This work was supported in part by the Jiangsu Provincial Colleges of Natural Science General Program under Grant 21KJB520006, in part by Research Project of Jiangsu Vocational College of Information Technology under Grant 10072020028(001), in part by Higher Vocational Education Teaching Fusion Production Integration Platform Construction Projects of Jiangsu Province under Grant No.2019(26), in part by The General Project fund of Natural Science Research in Universities of Jiangsu Province (18KJD510011), in part by Excellent Teaching Teams of "Qinglan Project" in Universities of Jiangsu Province (SuJiaoShi[2020] *).

References

[1] Change detection based on tensor rpca for longitudinal retinal fundus images. *Neurocomputing*, 387:1–12, 2020.

[2] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing*, 20(6): 1709–1724, 2010.

[3] Marc Braham and Marc Van Droogenbroeck. Deep background subtraction with scene-specific convolutional neural networks. In *2016 international conference on systems, signals and image processing (IWSSIP)*, pages 1–4. IEEE, 2016.

[4] HanQin Cai, Keaton Hamm, Longxiu Huang, Jiaqi Li, and Tao Wang. Rapid robust principal component analysis: Cur accelerated inexact low rank estimation. *IEEE Signal Processing Letters*, 28:116–120, 2020.

Table 2 Evaluation results on fourteen videos from the CDCNET 2014 database.

Video	Measure	Pro	MOG	VIBE	IA-IA3DFHRI	TRPCA
Classroom	Re	0.834	0.018	0.923	0.454	0.280
	Pre	0.734	0.020	0.011	0.179	0.966
	Acc	0.985	0.946	0.982	0.972	0.985
	F	0.781	0.019	0.021	0.257	0.434
Sofa	Re	0.723	0.567	0.647	0.462	0.291
	Pre	0.820	0.156	0.142	0.168	0.910
	Acc	0.946	0.912	0.899	0.906	0.956
	F	0.768	0.245	0.233	0.247	0.442
Corridor	Re	0.950	0.165	0.078	0.205	0.218
	Pre	0.768	0.787	0.430	0.765	0.969
	Acc	0.970	0.816	0.753	0.862	0.945
	F	0.850	0.273	0.131	0.324	0.356

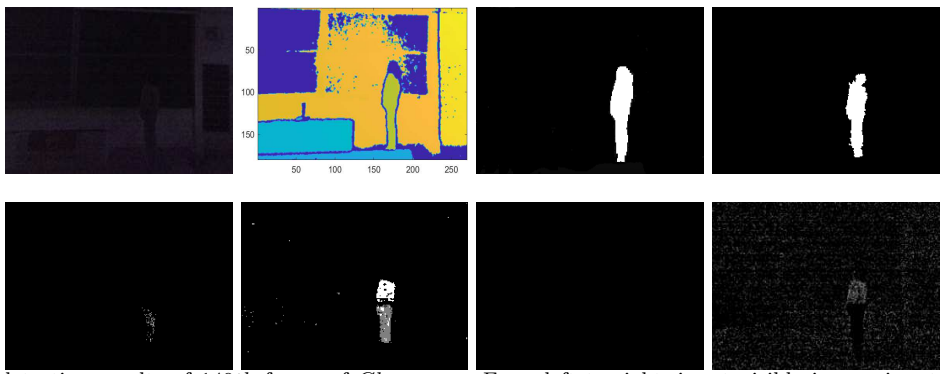


Fig. 7 The detection results of 140*th* frame of *Classroom*. From left to right: input visible image, input depth image, ground truth, the proposed method, IA-IA3DFHRI [11], MOG [24], VIBE [2], TRPCA [1]. For each video, we pick one frame to present.

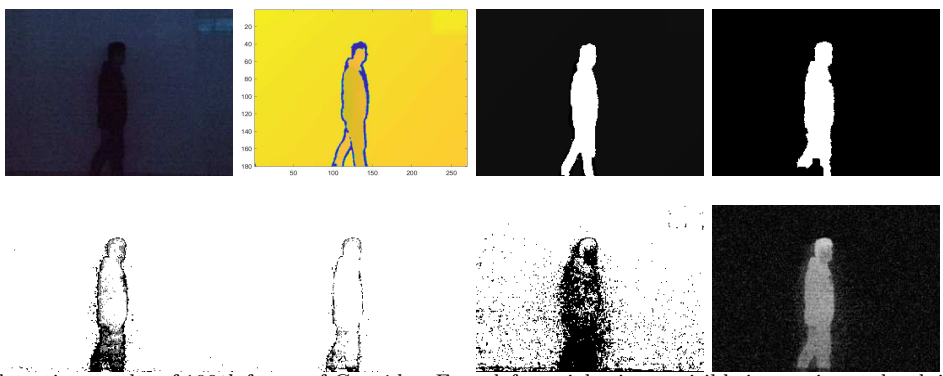


Fig. 8 The detection results of 190*th* frame of *Corridor*. From left to right: input visible image, input depth image, ground truth, the proposed method, IA-IA3DFHRI [11], MOG [24], VIBE [2], TRPCA [1]. For each video, we pick one frame to present.

[5] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

[6] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

Fig. 9 The detection results of 175*th* frame of *Sofa*. From left to right: input visible image, input depth image, ground truth, the proposed method, IA-IA3DFHRI [11], MOG [24], VIBE [2], TRPCA [1]. For each video, we pick one frame to present.

[7] Limin Chen, Vishal Jagota, and Ajit Kumar. Research on optimization of scientific research performance management based on bp neural network. *International Journal of System Assurance Engineering and Management*, pages 1–10, 2021.

[8] S. E. Ebadi and E. Izquierdo. Foreground segmentation with tree-structured sparse rpca. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2273–2280, 2018. doi: 10.1109/TPAMI.2017.2745573.

[9] Davar Giveki. Robust moving object detection based on fusing atanassov's intuitionistic 3d fuzzy histon roughness index and texture features. *International Journal of Approximate Reasoning*, 135:1–20, 2021.

[10] Yifan Guo, Guisheng Liao, Jun Li, and Xixi Chen. A novel moving target detection method based on rpca for sar systems. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6677–6690, 2020.

[11] Wei He, Wujing Li, Guoyun Zhang, Bing Tu, Yong Kwan Kim, Jianhui Wu, and Qi Qi. Detection of moving objects using adaptive multi-feature histograms. *Journal of Visual Communication and Image Representation*, 80:103278, 2021.

[12] Martin Hofmann, Philipp Tiefenbacher, and Gerhard Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In *2012 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 38–43. IEEE, 2012.

[13] Şahin Işık, Kemal Özkan, and Ömer Nezih Gerek. Cvabs: moving object segmentation with common vector approach for videos. *IET Computer Vision*, 13(8): 719–729, 2019.

[14] Sajid Javed, Arif Mahmood, Somaya Al-Maadeed, Thierry Bouwmans, and Soon Ki Jung. Moving object detection in complex scene using spatiotemporal structured-sparse rpca. *IEEE Transactions on Image Processing*, 28(2):1007–1022, 2019.

[15] Yang Li, Guangcan Liu, Qingshan Liu, Yubao Sun, and Shengyong Chen. Moving object detection via segmentation and saliency constrained rpca. *Neurocomputing*, 323:352–362, 2019.

[16] Ye-Fei Li, Lei Liu, Jia-Xiao Song, Zhuang Zhang, and Xu Chen. Combination of local binary pattern operator with sample consensus model for moving objects detection. *Infrared Physics & Technology*, 92: 44–52, 2018.

[17] Jian Liao, Guanjun Guo, Yan Yan, and Hanzi Wang. Multiscale cascaded scene-specific convolutional neural networks for background subtraction. In *Pacific Rim Conference on Multimedia*, pages 524–533. Springer, 2018.

[18] Long Ang Lim and Hacer Yalim Keles. Foreground segmentation using convolutional

neural networks for multiscale feature encoding. *Pattern Recognition Letters*, 112:256–262, 2018.

[19] Long Ang Lim and Hacer Yalim Keles. Learning multi-scale features for foreground segmentation. *Pattern Analysis and Applications*, 23(3):1369–1380, 2020.

[20] Omar Oreifej, Xin Li, and Mubarak Shah. Simultaneous video stabilization and moving object detection in turbulence. *IEEE transactions on pattern analysis and machine intelligence*, 35(2):450–462, 2012.

[21] Gani Rahmon, Filiz Bunyak, Guna Seetharaman, and Kannappan Palaniappan. Motion u-net: Multi-cue encoder-decoder network for motion segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 8125–8132. IEEE, 2021.

[22] Dimitrios Sakkos, Heng Liu, Jungong Han, and Ling Shao. End-to-end video background subtraction with 3d convolutional neural networks. *Multimedia Tools and Applications*, 77 (17):23023–23041, 2018.

[23] Dimitrios Sakkos, Heng Liu, Jungong Han, and Ling Shao. End-to-end video background subtraction with 3d convolutional neural networks. *Multimedia Tools and Applications*, 77 (17):23023–23041, 2018.

[24] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*, volume 2, pages 246–252. IEEE, 1999.

[25] Piotr Sulewski. Equal-bin-width histogram versus equal-bin-count histogram. *Journal of Applied Statistics*, 48(12):2092–2111, 2021.

[26] Maryam Sultana, Arif Mahmood, Sajid Javed, and Soon Ki Jung. Unsupervised deep context prediction for background estimation and foreground segmentation. *Machine Vision and Applications*, 30 (3):375–395, 2019.

[27] Ozan Tezcan, Prakash Ishwar, and Janusz Konrad. Bsuv-net: a fully-convolutional neural network for background subtraction of unseen videos. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2774–2783, 2020.

[28] Junpu Wang, Guili Xu, Chunlei Li, Zhengsheng Wang, and Fuju Yan. Surface defects detection using non-convex total variation regularized rpca with kernelization. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

[29] Yi Wang, Zhiming Luo, and Pierre-Marc Jodoin. Interactive deep learning method for segmenting moving objects. *Pattern Recognition Letters*, 96:66–75, 2017.

[30] Jiaxing Zhao, Ren Bo, Qibin Hou, Ming-Ming Cheng, and Paul Rosin. Flic: Fast linear iterative clustering with active search. *Computational Visual Media*, 4(4):333–348, 2018.