



Unsupervised Change Detection in Wide-Field Video Images Under Low Illumination

Shi, B., Jia, Z., Yang, J., & Kasabov, N. (2022). Unsupervised Change Detection in Wide-Field Video Images Under Low Illumination. *IEEE Transactions on Circuits and Systems for Video Technology*, 1-14. Advance online publication. <https://doi.org/10.1109/TCSVT.2022.3216457>

[Link to publication record in Ulster University Research Portal](#)

Published in:
IEEE Transactions on Circuits and Systems for Video Technology

Publication Status:
Published online: 21/10/2022

DOI:
[10.1109/TCSVT.2022.3216457](https://doi.org/10.1109/TCSVT.2022.3216457)

Document Version
Author Accepted version

General rights
Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

Unsupervised Change Detection in Wide-Field Video Images Under Low Illumination

Baoqiang Shi, Zhenhong Jia*, Jie Yang, and Nikola Kasabov, *Fellow, IEEE*

Abstract—In low-illumination environments such as at night, due to factors such as the large monitoring field of an eagle eye, short sensor exposure time, and high-density random noise, the video images collected by image sensors generally have poor visual quality and low signal-to-noise ratio, which makes it difficult for surveillance systems to detect weak changes. To solve this problem, we propose a method for image change detection (CD) in surveillance video based on optimized k-medoids clustering and adaptive fusion of difference images (DIs). First, for the input multitemporal video surveillance images, two DIs are obtained by log-ratio and extremum pixel ratio operators. Then, the two DIs are adaptively fused by combining the local energy of DIs and the Laplacian pyramid. Simultaneously, the fused DI is compressed by the normalization function, and the final DI is obtained via the improved adaptive median filter. Finally, the changed image is obtained by using the optimized k-medoids clustering algorithm. The experimental results show that the proposed method can accurately and effectively detect weak changes in the eagle eye surveillance picture in a low-illumination environment. Compared with those of other methods, the accuracy and robustness of the proposed method are higher, and the running time of the algorithm is shorter. Moreover, it will not generate a false alarm due to the influence of noise in unchanged scenes.

Index Terms—Wide field of view, low illumination change detection, surveillance video, k-medoids, difference image.

I. INTRODUCTION

THE eagle eye, a video monitoring device with a wide field of view, plays a very important role in personnel monitoring, control, and evidence collection. However, at night or in a low-light environment with a light brightness value below 0.8 lux, the video images captured by an eagle eye vision sensor are reduced in clarity, resulting in poor image quality. Police officers in a monitoring room may have to watch multiple surveillance monitors, and it is difficult for them to detect slow-moving weak targets in surveillance video

in time, which can lead to severe consequences if criminals are present. Therefore, it is particularly important to study the weak CD technology of video images under a wide field of view and low illumination. Similar to anomaly detection in surveillance video [1], its main purpose is to detect weak changes between different video image frames collected by eagle eye vision sensors in the same scene. However, for the CD of surveillance video images under low-illumination conditions, the main challenges we face are to solve the real-time problem of the detection algorithm, effectively suppress random noise, and improve the robustness of the detection algorithm. With the above analysis, it is important to study CDs for video images with a wide field of view and low illumination.

Although most CD methods for remote sensing images have successfully suppressed the influence of noise, they have little significance in the CD of video images under low illumination because the noise of remote sensing image is mainly speckle noise, and the image noise under low illumination is mainly complex random noise. At this stage, research on low illumination images is mainly focused on enhancement [2] and denoising [3], while research on the CD of low illumination images is relatively limited. However, many studies have been performed on the detection of pathological changes in medical images [4], video image anomaly detection [5], and video salient object detection [6], and they all must consider the impact of noise. To better suppress the influence of speckle noise, some scholars have proposed new methods, such as the combination of unsupervised k-means++ and CNNs [7], the location of change regions by the dual-region-of-interest network model [8], the modification of objective functions by introducing prior knowledge [9] and the salient object detection strategy [10]. Although infrared cameras [11] have been able to meet the needs of current monitoring systems under low illumination conditions, they cannot be used on a large scale in daily life due to their high cost. Therefore, we mainly focus on ordinary eagle eye devices to study video image CDs under low illumination conditions. However, in a low illumination environment, the imaging process is polluted by various mixed noises, such as speckle noise and dark current noise, whose removal requires special means. For example, Zhu et al. [12] used a morphological structure filter to effectively remove random noise from video images under low illumination conditions, thereby improving the accuracy of CD.

Similar to the CD of remote sensing images, the generation of DIs remains the most critical step in the CD of low-illumination images, which directly affects the final detection

This work was supported by the National Science Foundation of China (No. U1803261 and 62261053).

Baoqiang Shi is with College of Information Science and Engineering, Xinjiang University, Key Laboratory of Signal Detection and Processing, Xinjiang Uygur Autonomous Region, Xinjiang University, Urumqi 830046, China; e-mail: (m18298820751@163.com).

Zhenhong Jia is with College of Information Science and Engineering, Xinjiang University, Key Laboratory of Signal Detection and Processing, Xinjiang Uygur Autonomous Region, Xinjiang University, Urumqi 830046, China; Corresponding author (e-mail: jzh@xju.edu.cn).

Jie Yang is with Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200400, China; (e-mail: jieyang@sjtu.edu.cn).

Nikola Kasabov is with Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland 1020, New Zealand. And George Moore Chair in Data Analytics, Ulster University, Maggy campus, BT48 7JL, the UK. (e-mail: nkasabov@aut.ac.nz).

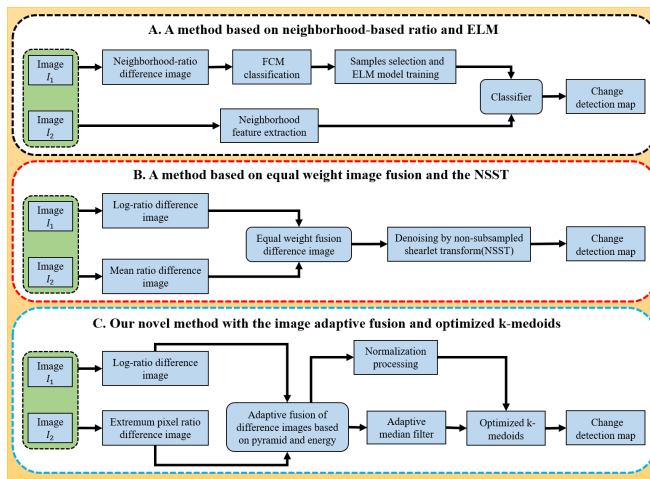


Fig. 1. Differences between our novel method and the other methods.

result. At present, the methods of generating DI mainly include log-ratio [13] and mean ratio [14]. However, the single DI generation method not only is susceptible to noise but also loses details. Reference [15] proposed a CD framework based on a neighborhood-based ratio and ELM (Fig. 1-A); although that method is simple in structure and easy to operate, the accuracy of the detection results is poor. To further suppress the noise between images, some scholars choose to fuse the DI by setting weights [16]-[17]. Reference [18] proposed a CD framework (Fig. 1-B) based on the fusion of DI with equal weight. Although it can make up for the lack of missing detail information (Fig. 1-A), the weight is fixed in the fusion process so that the quality of the DI is not very high. However, these methods cannot completely remove the influence of noise in the process of CD, so scholars have proposed some new methods for the generation of DI, such as stacked autoencoders (SAEs) [19] and deep pyramid feature learning networks (DPFL-Nets) [20]. Moreover, multiscale decomposition of DI is also a common method, such as principal component analysis (PCA) [21], nonsubsampled shearlet transform (NSST) [18], discrete 3D-wavelet transform (3-DWT) [22], and scale-invariant feature transform (SIFT) [23]. Although these methods can better overcome the influence of noise, the disadvantage is that the algorithm runs for a long time. In addition, saliency detection methods have been applied to image detection. Reference [10] proposed a computationally efficient and consistently accurate spatiotemporal salient object detection method to identify the most noticeable object in a video sequence, and reference [24] suppresses speckle noise and highlights the difference between two images by a saliency-guided method. For video image detection, the introduction of a saliency detection strategy [25]-[27] can not only effectively improve the detection performance but also alleviate the computational burden further. In this paper, the two DIs are adaptively fused by combining the energy feature and Laplace pyramid to improve the quality of DI.

In image CD, the final detection result is usually obtained by clustering or threshold segmentation. Common clustering algorithms include k-means [28] and FCM [29]. k-means

clustering is widely used in CD due to its advantages including a simple structure and low time complexity. Nevertheless, it is sensitive to the initial cluster center and outliers, and the initial cluster center is randomly selected. When the cluster center is not ideal, incorrect classification will occur. Although FCM clustering has the advantages of a simple principle and flexible clustering result, it is easily affected by noise and outliers and falls into local optima [30]. Threshold segmentation mainly extracts the difference between the target and the background in gray values by setting the threshold to achieve the separation of the target and the background, but the optimal threshold is difficult to obtain, as shown in [31] and [32]. Therefore, we intend to use k-medoids clustering [33] to obtain the final detection result, which is insensitive to noise and outliers.

Compared with ordinary video images, low illuminance video images under a wide field of view have the characteristics of large monitoring viewing angles, small detection objects, and high-density random noise, which bring greater challenges to the CD of low illuminance video images. If the image filtering is not handled properly, the structure and edge of the detected target will be seriously damaged, compromising the accuracy of the detected result. To address the high noise density under a low illumination environment, we propose a novel CD method based on image adaptive fusion and optimized k-medoids (Fig. 1-C), which effectively combines the energy features of DI and Laplacian pyramids to improve the quality of DI. Our contributions are mainly as follows:

- To improve the fusion accuracy of DI, we propose an adaptive fusion method based on DI energy and a Laplacian pyramid for the first time, which effectively improves the quality of DI.
- Inspired by the log-ratio and the neighborhood-ratio operators, we propose a novel method to generate the DI by the extremum pixel ratio, which can compensate for the lack of information lost by the log-ratio operator due to the compression of the CD.
- A new k-medoids clustering algorithm is proposed to optimize and improve the calculation of the similarity distance of sample points, which can effectively shorten the running time of the algorithm while ensuring the clustering effect.

The remainder of this paper is organized as follows. In Section II, we introduce the basic ideas of CD theory and discuss related works. The proposed video image CD method is presented in Section III. The evaluation results are provided in Section IV. Section V concludes the paper.

II. RELATED WORK

Generally, CD algorithms can be divided into three categories according to whether the label information is used: supervised [34]-[36], semisupervised [37]-[39], and unsupervised [40]-[43]. Supervised CD methods rely on the supervised classification to obtain the changed and unchanged regions. This process requires the ground truth to provide the labeled samples for classifier learning. However, the labeled samples are not available and difficult to obtain in many practical

applications. The semisupervised CD methods also require a small number of labeled samples for training the network [44]. In contrast, the unsupervised classifier does not require any training samples, so the unsupervised classification methods are more widely used in the CD. Therefore, we mainly study the unsupervised CD of video images under low illumination conditions. The traditional unsupervised CD methods can be divided into three processes: image preprocessing, DI generation, and DI analysis. In preprocessing, geometric correction is usually accomplished by image-to-image registration to ensure that the corresponding pixels in the multitemporal images refer to the same geographic location. In the second process, the two registered images are compared to generate the DI, which aims to increase the contrast between changed and unchanged areas. In the last process, the DI can be divided into the changed class and unchanged class to obtain the binary change map. In this paper, multitemporal surveillance images are collected by a fixed eagle eye camera in a short time, and the geometric correction has little effect on the CD. Consequently, we assume that the input multitemporal surveillance images have been preprocessed, and only the second and third steps must be studied to improve the accuracy of CD.

The noise types of remote sensing images and low-illumination video images are different. Speckle noise is the main type of interference for remote sensing images. However, in low illumination conditions, the video images captured by an eagle eye monitor device usually contain high-density random noise due to the short exposure time of the image sensor. To achieve higher detection accuracy, image denoising is an indispensable step in the process of CD. Reference [45] first utilized the discrete wavelet transform strategy to fuse the DIs and then reduced the random noise in the DI by the nonsubsampling contourlet transform. Reference [46] adopted a combination of frequency-domain analysis and multiple random images to suppress speckle noise in SAR images. Reference [47] proposed a CD method based on convolutional wavelet neural networks (CWNNs) that effectively overcame the speckle noise in SAR images. To solve the problem of high-density random noise and the low signal-to-noise ratio of video images under low illumination conditions, we decide to adopt the strategy of combining an adaptive fusion of DI, normalization of DI, and an improved adaptive median filter for denoising.

For image CD, the quality of the DI determines the CD performance. Common methods for generating DI mainly include the subtraction operator [17] and log-ratio operator [12], but it is difficult to obtain more details of the changed regions by using a single method to generate DI. Therefore, some scholars are exploring how to obtain high-quality DI. Zheng et al. [48] proposed a SAR image CD method based on fusion DI and k-means clustering, which fuses the DIs obtained by the subtraction operator and log-ratio operator. Compared with the DI generated by the subtraction operator or log ratio operator, the DI generated by the fusion method retains more detailed information and is less affected by noise. Zhou et al. [49] proposed a SAR image CD method based on fusion DI and fuzzy C-means (FCM) clustering, which also fuses the DIs obtained by the log-ratio and mean ratio operators. Its

purpose is to overcome the influence of noise to improve the performance of subsequent clustering.

The above analysis shows that common CD methods are not suitable for video image CDs in a low-illumination environment. Moreover, false alarms easily occur due to the difference in noise in scenes without obvious change. To accurately and effectively detect the changed areas of surveillance video images under low illumination environments and avoid false alarms in unchanged scenes, in our method, we mainly explore low illumination video image denoising, generation DI, and clustering algorithm. The differences are as follows:

- We adopt an adaptive median filter that can dynamically change the size of the filtering window and consider both denoising and protection of detail information, and its denoising effect is better than other methods.
- We propose an adaptive fusion method of DI, and the DI obtained by this method has richer detailed information.
- The k-medoids clustering algorithm is optimized to shorten the runtime of the algorithm.
- The proposed method has strong applicability under different low illumination scenes, and its robustness is higher than that of other methods.

III. METHODOLOGY

In this paper, we propose an unsupervised CD method based on improved k-medoids clustering and adaptive fusion of energy features of DI. I_1 and I_2 are two multitemporal surveillance images obtained at different times in the same scene, and they are called $I_1 = \{I_1(i, j) | 1 \leq i \leq W, 1 \leq j \leq H\}$ and $I_2 = \{I_2(i, j) | 1 \leq i \leq W, 1 \leq j \leq H\}$, respectively, where W and H are the width and height of the image, respectively. Here, W and H are set to 500 and 380, respectively. The flowchart in Fig. 2 shows the four main steps of the algorithm: 1) Two DIs are generated by using the log-ratio and extremum pixel ratio. 2) The energy of the two DIs is calculated, and then the two DIs are adaptively fused according to their respective energy and Laplace pyramid. 3) The normalization function and improved adaptive median filter are applied for denoising. 4) Optimized k-medoids clustering is executed to obtain the CD results.

A. Generation of Difference Images

Among the methods of DI generation, the log-ratio operator has the advantage of converting multiplicative coherent noise into additive noise, and the background information of DIs obtained by the log-ratio operation is relatively flat. Consequently, we also use the log-ratio method to generate the DI. Inspired by the mean ratio [14] and the neighborhood ratio [15], we propose a novel method for generating DI by the extremum pixel ratio, as shown in Eq. 2.

$$DI_1(i, j) = \left| \log \frac{I_1(i, j) + 1}{I_2(i, j) + 1} \right| \quad (1)$$

$$DI_2(i, j) = 1 - \frac{\min(I_1(i, j), I_2(i, j))}{\max(I_1(i, j), I_2(i, j))} \quad (2)$$

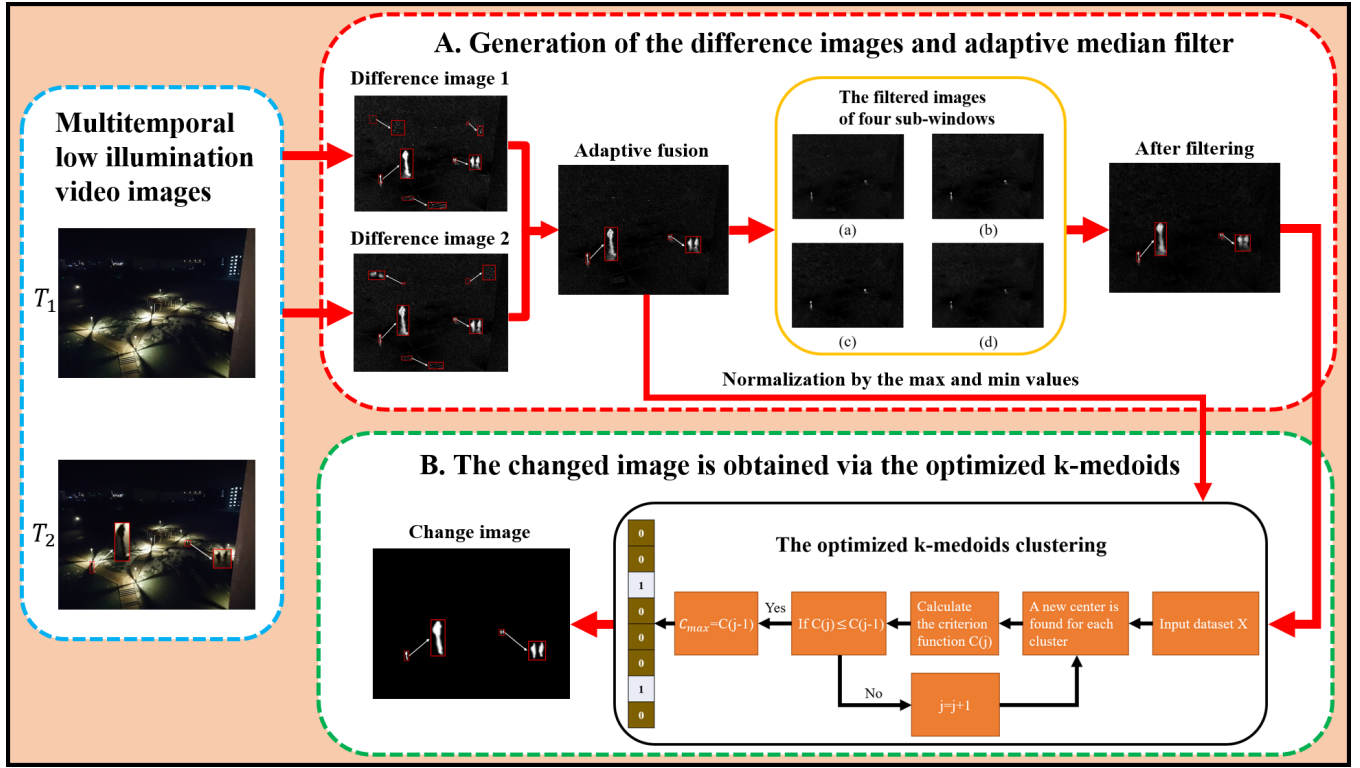


Fig. 2. Flowchart of the proposed method. First, in part A, the two DIs are generated by the log-ratio and mean ratio operators, and the DIs are adaptively fused by combining the energy feature of the DI and the Laplace pyramid. Then, the fused DI is processed by an adaptive median filter, where (a) - (d) are the filtered images of the filter subwindows in the 0° , 45° , 90° , and 135° directions, respectively. Finally, the changing image is obtained by the optimized k-medoids in part B. Here, the DI generation, DI fusion, and adaptive median filter in part A are presented in Section III-A, III-B, and III-C, respectively, and the k-medoids in part B are in Section III-D.

where $I_1(i, j)$ and $I_2(i, j)$ represent the pixel values of two multitemporal video images at points (i, j) , $I_1(i, j)$ and $I_2(i, j)$ are replaced by $I_1(i, j) + 1$ and $I_2(i, j) + 1$ to prevent the case where $I_1(i, j)$ and $I_2(i, j)$ are 0.

The quality of the DI determines the accuracy of the CD. Although the log-ratio operator has strong antinoise performance and the background information of the generated DI is relatively flat, it compresses the variation range of the DI and cannot reflect the actual change to the maximum extent, which may lead to the loss of the changed area. However, the extremum pixel ratio does not have such a problem. Compared with the log-ratio operator, although the extremum pixel ratio is more sensitive to noise, it can retain more detailed information in the changing region. Therefore, the fusion of the DIs obtained by the log-ratio operator and the extremum pixel ratio operator can effectively reduce noise interference. Inspired by equal-weight fusion method adopted by Zhu et al. [12], we combine the energy of the DI and the Laplacian pyramid to extract more detailed information at multiple scales for achieving a better fusion effect. The fused DI can be obtained by Eq. 4 and Eq. 8.

B. Difference Image Fusion Based on Local Energy and Laplacian Pyramid

The DI is a typical gradient image, except for the large gray value of pixels in the difference area, the gray values are close to 0. Thus, we first calculate the local energy of the DI and

then combine the Laplacian pyramid to adaptively fuse the DI. The energy calculation method is shown in Eq. 3, which takes the square sum of all pixel values centered on point (i, j) in the $m \times n$ window, where m and n are the length and width of the local window, respectively.

$$E(i, j) = \sum_m \sum_n DI(i + m, j + n)^2 \quad (3)$$

Since the Laplacian pyramid can extract more detailed information at multiple scales. To achieve a better fusion effect, we combine the DI energy features and the Laplacian pyramid to extract more detailed information at multiple scales. The specific steps of the algorithm are as follows: 1) First, $DI_1(i, j)$ and $DI_2(i, j)$ are decomposed into two N -layer Laplacian image pyramids, and $DI_1^k(i, j)$ and $DI_2^k(i, j)$ are set as the k -levels of the DI pyramid, where $k = 0, 1, \dots, N - 1$; 2) Two brightness values of $DI_1^{N-1}(i, j)$ and $DI_2^{N-1}(i, j)$ are calculated, and then the image with the largest brightness value is selected as the top-level pyramid fusion image $F_{N-1}(i, j)$; 3) $DI_1^k(i, j)$ and $DI_2^k(i, j)$ are fused according to the fusion rules, where $k = 0, 1, \dots, N - 2$. The specific process is shown in Algorithm 1.

In the fusion process, it is necessary to set a specified threshold t , and its value is usually between 0 and 1. The experimental results in Fig. 6 show that the fusion effect is best when the value of t is 0.6, so we take $t = 0.6$. When $M < t$, the image with higher energy is selected at this point,

Algorithm 1 Adaptive fusion of DI.

Input: DI_1 and DI_2 , threshold t .

Output: The k -level fusion image (F_k).

- 1: Calculate the energy of DI_1 and DI_2 by using Eq. 3;
- 2: The matching degree (M) of the k -layer Laplace pyramid is obtained by Eq. 5;
- 3: **if** $M < t$ **then**
- 4: F_k is obtained by Eq. 4;
- 5: **else**
- 6: Calculate V_{min}^k, V_{max}^k with Eqs. 6, and 7;
- 7: F_k is obtained from Eq. 8;
- 8: **end if**

the rest are discarded, and F_k denotes the image with k -layer horizontal fusion, which can be obtained by Eq. 4.

$$F_k = \begin{cases} DI_1^k, & E_{DI_1}^k \geq E_{DI_2}^k \\ DI_2^k, & E_{DI_1}^k < E_{DI_2}^k \end{cases} \quad (4)$$

$$M = \frac{(\sum_m \sum_n D_1(i+m, j+n) \cdot D_2(i+m, j+n))^2}{E_{DI_1}(i, j) \cdot E_{DI_2}(i, j)} \quad (5)$$

where “.” indicates multiplication, D_1 , and D_2 represent two DIs obtained by log ratio and pixel ratio, E_{DI_1} and E_{DI_2} represent the energy of the two DIs, M represents the local region matching degree of the k th layer of the Laplacian pyramid, which is similar to the covariance of two local windows. The calculation method is shown in Eq. 5. E_{DI}^k represents the local regional energy of the k th layer of the Laplacian pyramid. The calculation method is shown in Eq. 3. When $M \geq t$, F_k can be obtained by Eq. 8.

$$V_{min}^k = 0.5 \times \left(1 - \frac{1-M}{1-t}\right) \quad (6)$$

$$V_{max}^k = 1 - V_{min}^k \quad (7)$$

$$F_k = \begin{cases} V_{max}^k \cdot DI_1^k + V_{min}^k \cdot DI_2^k, & E_{DI_1}^k \geq E_{DI_2}^k \\ V_{min}^k \cdot DI_1^k + V_{max}^k \cdot DI_2^k, & E_{DI_1}^k < E_{DI_2}^k \end{cases} \quad (8)$$

where “.” indicates multiplication, V_{min}^k and V_{max}^k represent the fusion weight of the maximum energy and minimum energy of the k th layer of the pyramid, respectively. DI_1^k and DI_2^k represent the two DIs of the k th layer of the pyramid, respectively. Finally, the fused DI is obtained by inverse transformation of pyramid image.

C. Improved Adaptive Median Filter and the Normalization

In a low illumination environment, our purpose is to obtain the changing area of multitemporal video images captured by eagle eye surveillance equipment. Especially at night, the video images collected by the visual sensor are affected by high-density random noise. If CD is carried out on these images directly, the accuracy of the detection results will be reduced to a certain extent. Therefore, we adopt the improved adaptive median filter for denoising, which can effectively

remove the noise while retaining the edge and detail information of the image. The filtering algorithm is divided into two processes: noise detection and noise removal.

In the filter window area, the width and height of the window are set as w and h , respectively; the pixel value at pixel point $I(i, j)$ is $y(i, j)$; and the maximum and minimum values of the pixel value in the whole area are y_{max} and y_{min} , respectively. When $y(i, j) = y_{max}$ or $y(i, j) = y_{min}$, pixel point $I(i, j)$ is identified as a candidate noise point, where $f(i, j)$ represents the candidate noise points, D and $M(i, j)$ is used to denote the set of pixel values and the mean of the set of pixel values after removing extreme points in the filter window, respectively. The brief process of image filtering is shown in Algorithm 2.

Algorithm 2 Improved adaptive median filter.

Input: Noisy image $I(i, j)$, width (w) and height (h) of filter window.

Output: Filtered image $G(i, j)$.

- 1: **if** $f(i, j) == 1$ **then**
- 2: The pixel point $I(i, j)$ is considered as a candidate noise point;
- 3: Calculate $M(i, j)$, T with Eqs. 10, and 12;
- 4: **if** $|y(i, j) - M(i, j)| > T$ **then**
- 5: $F(i, j) = f(i, j)$, the candidate noise point $f(i, j)$ is the actual noise point;
- 6: Calculate d , r with Eqs. 13, and 14;
- 7: Calculate $M_k(i, j)$, c_k with Eqs. 15, and 16;
- 8: The filtered image $G(i, j)$ is obtained by Eq. 17;
- 9: **else**
- 10: $f(i, j)$ is not a noise point;
- 11: **end if**
- 12: **end if**

$$f(i, j) = \begin{cases} 1, & y(i, j) = y_{max} \text{ or } y(i, j) = y_{min} \\ 0, & \text{other} \end{cases} \quad (9)$$

$$M(i, j) = \frac{1}{k} \sum_{y(i+w, j+h) \in D} y(i+w, j+h) \quad (10)$$

The pixel value of the candidate noise point is compared with the mean; if the absolute value of the difference is greater than the threshold, the candidate noise point is the actual noise point. The calculation method is shown in Eq. 11.

$$F(i, j) = \begin{cases} f(i, j), & |y(i, j) - M(i, j)| > T \\ 0, & \text{other} \end{cases} \quad (11)$$

$$T = \sqrt{\frac{1}{k} \sum_{y(i+w, j+h) \in D} (y(i+w, j+h) - M(i, j))^2} \quad (12)$$

where T is the threshold and k is the number of pixels in the filter window.

To better suppress the noise and protect the details of the images, we introduce the local noise density (the probability

of noise points) to determine the size of the filter window, which is expressed as Eq. 13.

$$d = \frac{N}{w \times h} \quad (13)$$

where N is the number of noise points in the filter window and w and h are the width and height of the filter window, respectively. When the local noise density is small, the filter window is correspondingly small to ensure the image denoising effect and the ability to protect the details. Conversely, when the local noise density becomes larger, a larger filter window is needed. Consequently, the size of the filter window is adaptively determined by the density of noise points, and the expression is shown in Eq. 14:

$$r = \begin{cases} 3, & d \leq 0.2 \\ 5, & 0.2 < d \leq 0.5 \\ 7, & 0.5 < d \leq 1 \end{cases} \quad (14)$$

Assuming the pixel point $I(i, j)$ is polluted by noise, we select four specific directions in the filter window for filtering: 0° , 45° , 90° , and 135° ; and four filter subwindows can be obtained: $W_1(i, j)$, $W_2(i, j)$, $W_3(i, j)$, and $W_4(i, j)$. Then, the medians $M_1(i, j)$, $M_2(i, j)$, $M_3(i, j)$, and $M_4(i, j)$ of the pixel values in the four filter subwindows are obtained by Eq. 15.

$$M_k(i, j) = \text{med}[W_k(i, j)], k = 1, 2, 3, 4 \quad (15)$$

$$c_k = \frac{M_k(i, j)}{\sum_{k=1}^4 M_k(i, j)} \quad (16)$$

where $\text{med}[\cdot]$ denotes the median filter; c_k is the weighting coefficient, and its value depends on the filtering results in four directions. Finally, the noise points in the four subwindows are filtered out via the median filter. The calculation method is shown in Eq. 17.

$$G(i, j) = \sum_{k=1}^4 c_k M_k(i, j) \quad (17)$$

Although most noise in the image is effectively suppressed by filtering, there is still a small amount of noise that will lead to false alarms in the surveillance system. Inspired by reference [12], we compress the fused DI to between [0,1] by a normalization function, which is mainly to suppress the random noise caused by sensors and improve the performance of subsequent clustering. As shown in Fig. 3, some pixels polluted by noise have large differences in gray values. If the k-medoids are directly used for clustering, these unchanged pixels with large differences in pixel values will be incorrectly classified into changed classes. However, after processing by the normalization function, the correct classification can be obtained by k-medoids. The fused DI is normalized by Eq. 18.

$$\text{Nor}(i, j) = \frac{I(i, j) - I_{\min}(i, j)}{I_{\max}(i, j) - I_{\min}(i, j)} \quad (18)$$

where $I_{\max}(i, j)$ and $I_{\min}(i, j)$ are the maximum and minimum pixel values at point (i, j) of two multitemporal video images, respectively.

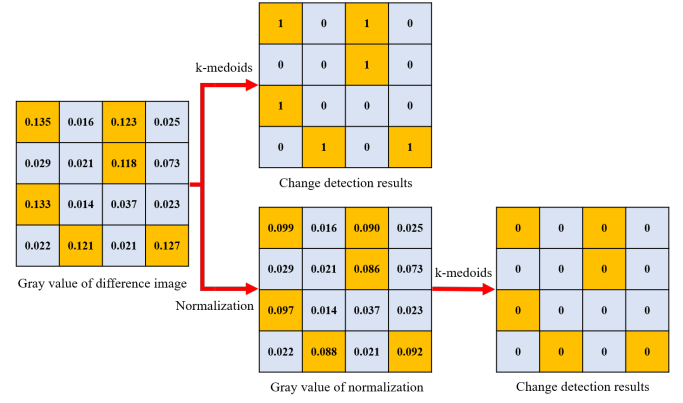


Fig. 3. Normalized and unnormalized detection results under noise conditions. The yellow part has a large difference in gray value due to noise interference, but there is no obvious change. If k-medoids clustering is carried out directly, there will be misclassification, but the correct results can be obtained after normalization.

D. Optimization of the K-medoids Clustering Algorithm

When extracting change information from DI, the classical algorithms mainly include the threshold method [32] and the clustering method [28]. The threshold method must establish a complex image statistical model for classification, while the clustering algorithm does not need it, so we adopt the clustering algorithm for classification. The traditional k-medoids clustering [50] algorithm has strong antinoise performance and a good clustering effect, but it runs for a long time. Accordingly, we optimize and improve the clustering algorithm to reduce the time cost of the algorithm while ensuring the clustering effect.

In the process of searching for the optimal solution, the traditional k-medoids clustering algorithm must repeatedly update the central point. This operation has great blindness and uncertainty, which increases the time cost of the algorithm. Therefore, we introduce the data density [51] and a new criterion function to optimize the traditional k-medoids clustering. Suppose the dataset $X = \{x_1, x_2, \dots, x_n\}$, and each sample x_i has m-dimensional features: $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, $i = 1, 2, \dots, n$. Here, the Euclidean distance between samples, data density, and criterion function is defined by Eqs. 19, 20, and 21, respectively.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (19)$$

$$D(x_i) = \sum_{j=1}^n \frac{d(x_i, x_j)}{\sum_{l=1}^n d(x_l, x_j)} \quad (20)$$

$$C(j) = \frac{\sum_{j=1}^k n_j \|x_c^j - x_c\|^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} \|x_i^j - x_c^j\|^2} \cdot \log(k) \quad (21)$$

where n_j is the number of samples in the j th cluster, k is the number of clusters, x_c^j is the centroid of the j th cluster, x_c is the centroid of all samples, $\|x_c^j - x_c\|$ is the distance between x_c^j and x_c , $\|x_i^j - x_c^j\|$ is the distance between the i th sample x_i^j of the j th cluster and the centroid x_c^j of the j th cluster,

and $\log(k)$ is the adjustment factor ($k \geq 2$), which is mainly used to prevent local extrema due to an excessively large or small value of k . Its value mainly depends on the ratio of the intercluster distance to the sample centroid distance. Thus, the larger the value of $C(j)$ is, the better the clustering effect. The optimization process of k-medoids clustering is shown in Algorithm 3.

Algorithm 3 Optimization process of k-medoids clustering.

Input: Dataset X and sample distance matrix.

Output: The k clusters of the dataset X and C_{max} .

- 1: Initialization: $C(1) = 0$, $j = 2$, $S = \emptyset$;
- 2: A new center is found for each cluster;
- 3: Calculate $C(j)$ with Eq. 21;
- 4: **if** $C(j) > C(j-1)$ **then**
- 5: $j = j + 1$;
- 6: Go to step 2 to continue the iteration;
- 7: **else**
- 8: When $C(j) \leq C(j-1)$, the iteration is stopped;
- 9: $C_{max} = C(j-1)$;
- 10: **end if**

After the optimization of the algorithm, the calculation of the distance between samples is moved out of all loop bodies, the calculation of only the distance must be called, and the time complexity of k-medoids clustering becomes $O(k^2dnt)$. Compared with the time complexity of the traditional k-medoids clustering ($O(k(n-k)^2dt)$), the time complexity of the improved algorithm is greatly reduced, which shortens the running time of the algorithm.

IV. EXPERIMENTAL ANALYSIS

To verify the superiority of the proposed method, image data in different scenes are selected for experiments. The proposed method is compared with eight other image CD algorithms: NR-ELM [15], DWT [49], CWNN [47], NSST [18], FDA-RMG [46], NPSG [53], Zhu [12], and CDI-K [48], which are compared experimentally, and subjective analysis and objective index analysis are conducted on the experimental results. The experimental environment includes an Intel Core i5-1135G7@2.40 GHz processor with 16 GB of memory, and the software used is MATLAB2018b (64-bit). Objective evaluation indexes of detection results include the overall error (OE), percentage correct classification (PCC), Kappa coefficient, F1-score (F1), and run time (T).

This study focuses on the weak CD of multitemporal surveillance video images taken by an eagle eye camera under low illumination conditions. Since the detected target is too small, the distribution of changed and unchanged pixels is unbalanced. Consequently, using only PCC is not sufficient to evaluate the detection accuracy of the algorithm. Thus, we introduce the Kappa coefficient and F1 to comprehensively evaluate the algorithm. The Kappa coefficient is an indicator used to measure the consistency of detection results, while F1 represents the weighted average of detection accuracy and recall. The calculation method is shown in Eqs. 22 and 23.

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (22)$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \quad (23)$$

where P_o and P_e represent the accuracy and accidental consistency of detection, respectively. TP represents the number of samples that can be correctly detected, and FN and FP represent false negatives and false positives, respectively.

A. Performance Analysis of Difference Images Generated by Different Methods

The log-ratio operator has strong antinoise performance, and the background information of the generated DI is relatively flat, but it compresses the variation range of the DI, which may lead to the loss of the changed area. In contrast, the pixel ratio can make up for the disadvantage of the log ratio. Although the pixel ratio is more sensitive to noise, it can retain richer details. Hence, the fusion of the DIs obtained by the log-ratio operator and the pixel ratio operator can effectively remove the noise and better preserve the changing area. We select three datasets in different scenes for experiments. Fig. 4 shows the detection results.

TABLE I
AVERAGE EVALUATION INDEXES OF
EXPERIMENTAL DATA

Methods	$\overline{OE} \uparrow$	$\overline{PCC} \uparrow$	$\overline{Kappa} \uparrow$	$\overline{F1} \uparrow$	$\overline{T(s)} \downarrow$
Log ratio	11.9	0.9993	0.8374	0.8374	0.233
Pixel ratio	12.3	0.9993	0.8348	0.8349	0.239
Log + Pixel	10.3	0.9994	0.8393	0.8393	0.248
Proposed	8.8	0.9995	0.8430	0.8431	0.311

It is obvious from Fig. 4 that the CD results obtained by using the log-ratio operator alone are poor. Especially in experimental data 1 and 3, the edges of the detected object become incomplete, mainly because the log-ratio operator compresses the DI, resulting in a serious loss of detailed information in the generated DI, and the detection result in experimental data 2 is also affected by a small number of noise points. Compared with the log-ratio operator, although the CD results obtained by the pixel ratio operator can better retain the detailed information of the changing area, there are a small number of false change points in the detection results because the pixel ratio operator is more sensitive to noise. Therefore, the detection result obtained by using the log-ratio or pixel ratio alone to generate DI is not ideal. We adopt the common equal-weighted fusion method to fuse the DIs generated by the log ratio and pixel ratio. It can be seen from the detection results in Fig. 4 that the equal-weighted fusion method can improve the quality of the DI to a certain extent and compensate for the lack of missing detail information. However, this method cannot effectively suppress noise, and the detection results in experimental data 2 and 3 are still affected by the noise. Consequently, the detection result obtained by fusing the DI with equal weight is not optimal. However, in this study, we utilize the adaptive fusion method based on the energy feature of the DI and the Laplacian pyramid to generate a DI with higher quality, and

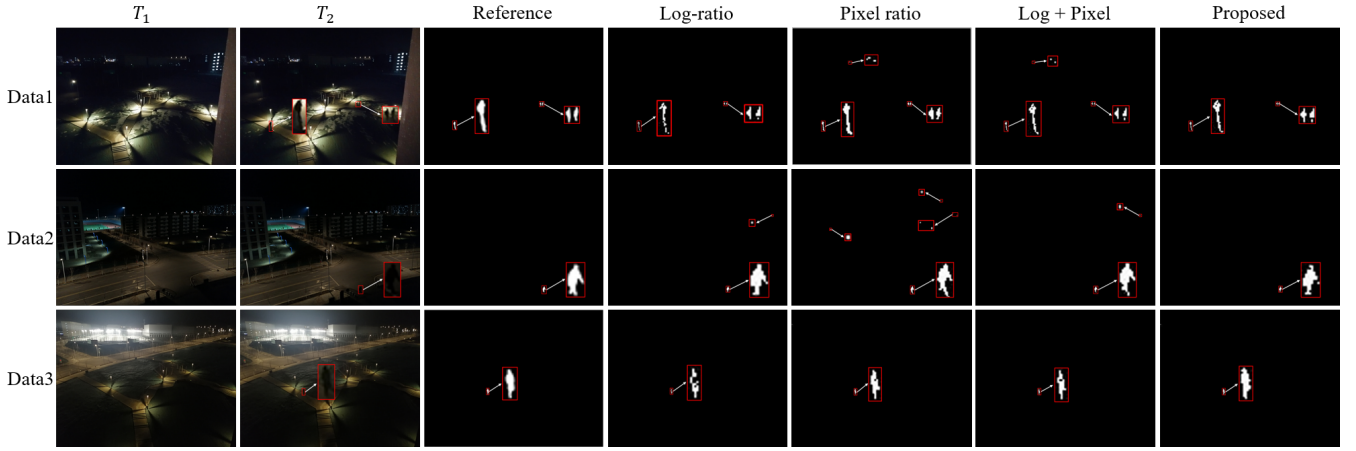


Fig. 4. The CD results obtained by using different DI generation methods. The CD result obtained by log-ratio or pixel operators is not ideal. It is not only susceptible to noise interference, but also loses some important details. The DIs generated by log-ratio and pixel ratio operators are fused with equal weight, which can improve the CD performance to a certain extent, but it is still affected by a little noise. However, the CD result obtained by our method is closest to the reference.

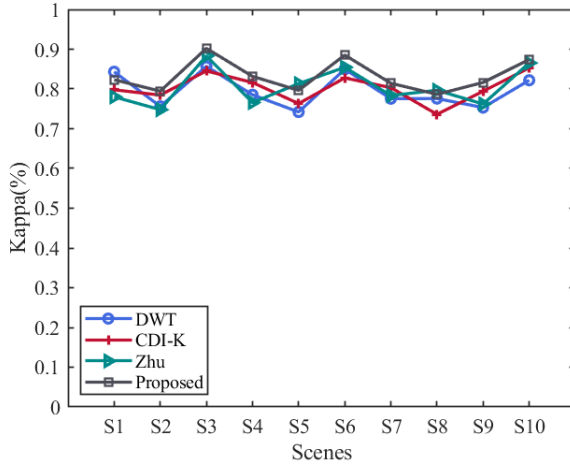


Fig. 5. The changing trend of Kappa under different datasets. The experimental results show that the overall performance of our method is better than WDT, CDI-K, and Zhu.

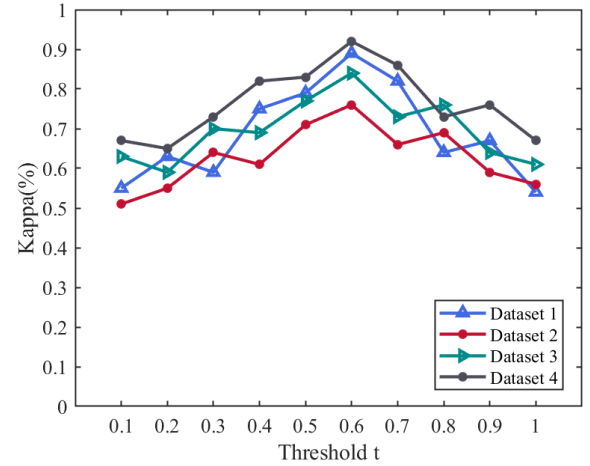


Fig. 6. Analysis on threshold t . The experimental results are obtained on four datasets for different thresholds t . We finally choose $t = 0.6$ as the threshold.

the DI generated by this method not only retains more detailed information but also effectively suppresses the influence of noise on the detection result. Table I shows the average evaluation indexes of the detection results of experimental data 1, 2, and 3. The evaluation indexes in Table I show that the DIs obtained by the log-ratio, pixel ratio, and equal-weighted fusion can shorten the running time of the algorithm, but the accuracy of the detection results is not high. However, in our method, the other evaluation indexes except time are the best, which further proves the effectiveness of the method adopted in this study.

To further verify the effectiveness of our method, we separately select three comparison algorithms that are similar to our method framework for experiments, and the experimental results are shown in Fig. 5. From the detection results under 10 different scenes, we can see that the overall robustness of our algorithm is the highest, which proves the superiority

of our method. In the process of DI fusion, it is necessary to set a threshold in advance, and choosing the appropriate threshold parameter is crucial to the CD results. When the threshold parameter is too small, the detailed information of the changed area may be lost, and when the threshold is too large, the fusion results of the DI will be distorted. According to the experimental results in Fig. 6, we finally choose $t = 0.6$ as the threshold.

B. Accuracy Analysis of the Change Detection Algorithm

To verify the accuracy of the proposed algorithm, three groups of video images in different scenes are selected for relevant CD experiments, and from the subjective analysis, and objective evaluation, it is verified that the robustness of the algorithm in this paper is better than the comparison algorithm in terms of Kappa coefficient, PCC and T. Experimental data 4 are shown in Fig. 7 (a) and (b), where (a) and (b) show the video images of the previous phase and the later phase,

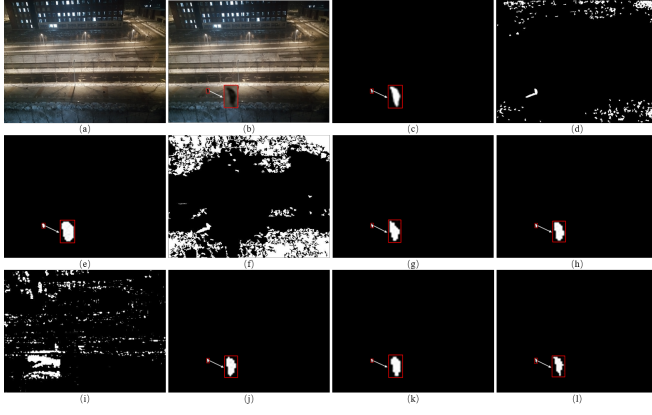


Fig. 7. Change detection results of experimental data 4, (a) Original image, (b) Changed image, (c) Reference image, (d) NR-ELM [15], (e) DWT [49], (f) CWNN [47], (g) FDA-RMG [46], (h) NSST [18], (i) NPSG [53], (j) Zhu [12], (k) CDI-K [48], (l) Proposed method.

TABLE II
OBJECTIVE EVALUATION INDEXES OF
EXPERIMENTAL DATA 4

Algorithms	OE \uparrow	PCC \uparrow	Kappa \uparrow	F1 \uparrow	T(s) \uparrow
NR-ELM [15]	7067	0.9628	0.0132	0.0137	15.21
DWT [49]	37	0.9981	0.7174	0.7176	1.082
CWNN [47]	47457	0.7502	0.0015	0.0021	39.80
FDA-RMG [46]	13	0.9993	0.8631	0.8632	2.249
NSST [18]	12	0.9994	0.8749	0.8750	47.26
NPSG [53]	11346	0.9403	0.0081	0.0086	33.87
Zhu [12]	13	0.9993	0.8659	0.8660	0.478
CDI-K [48]	15	0.9992	0.8235	0.8236	1.891
Proposed	8	0.9996	0.9148	0.9149	0.299

respectively, with a difference of approximately 5 seconds, and (c) shows the reference image of the CD. Here, we set the size of the image to 500×380 .

Fig. 7 and Table II show that although some comparison algorithms have also achieved good detection results, the results obtained by our proposed algorithm are better, and the detection results are closer to the reference image. This is mainly because we use the energy feature adaptive fusion method of DIs to generate high-quality DI and use a multiwindow and multiscale filter to effectively suppress the influence of noise. As shown in Fig. 7, since NR-ELM, CWNN and NPSG do not take special measures to suppress the high-density random noise under low illumination, a large number of pseudo-change points appear in the unchanged region of (d), (f), and (i), resulting in a higher value of OE in Table II than other methods. As seen from (e), (g), (h), (j), and (k) in Fig. 7, DWT, FDA-RMG, NSST, Zhu, and CDI-K can effectively overcome the influence of noise, but these methods enlarge the range of the effective change region to a certain extent and reduce the accuracy of algorithm detection. According to the evaluation indexes in Table II, our method is superior to other algorithms in terms of Kappa coefficient, F1, and PCC, and the running time of our algorithm is shorter.

Experimental data 5 are shown in Fig. 8 (a) and (b), which show the multitemporal video images differing by approximately 8 seconds, and (c) shows the reference for CD.

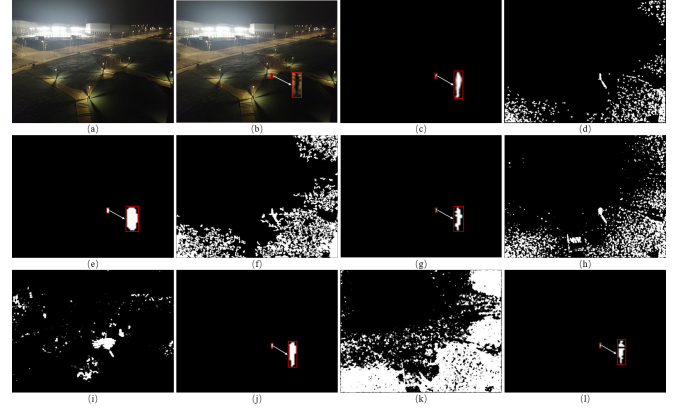


Fig. 8. Change detection results of experimental data 5, (a) Original image, (b) Changed image, (c) Reference image, (d) NR-ELM [15], (e) DWT [49], (f) CWNN [47], (g) FDA-RMG [46], (h) NSST [18], (i) NPSG [53], (j) Zhu [12], (k) CDI-K [48], (l) Proposed method.

TABLE III
OBJECTIVE EVALUATION INDEXES OF
EXPERIMENTAL DATA 5

Algorithms	OE \downarrow	PCC \downarrow	Kappa \uparrow	F1 \uparrow	T(s) \uparrow
NR-ELM [15]	12039	0.9366	0.0051	0.0055	15.29
DWT [49]	61	0.9968	0.5039	0.5041	1.454
CWNN [47]	32420	0.8294	0.0017	0.0020	27.23
FDA-RMG [46]	9	0.9995	0.8578	0.8579	2.156
NSST [18]	23752	0.8750	0.0024	0.0028	46.87
NPSG [53]	5860	0.9692	0.0108	0.0111	34.05
Zhu [12]	15	0.9992	0.8050	0.8052	0.335
CDI-K [48]	70120	0.6309	0.0005	0.0011	2.137
Proposed	10	0.9994	0.8581	0.8582	0.312

Compared with the previous group, the background of this group of images is more complex, the visibility is lower, and the illumination distribution is more uneven. Similarly, we still set the size of the image to 500×380 .

Similar to the detection results in Fig. 7, since NR-ELM, CWNN and NPSG do not adopt the effective denoising means, a large number of pseudo-change regions appear in (d), (f) and (i). NSST and CDI-K both adopt a mean filter to remove noise, but the denoising effect is not good, and false alarms are generated in the unchanged regions. Therefore, the OE values of these methods are higher than those of the other methods in Table III. In contrast, (e), (g), (j), and (l) is not affected by the noise, and they can detect the changed area better. However, DWT enlarges the range of the detection area to a certain extent, resulting in a reduction in the robustness of the algorithm. It can be seen from the evaluation indicators in Table III that although the OE and PCC of FDA-RMG are superior to our method, the Kappa and F1 of our method are higher and the running time of our method is shorter, which further shows that our method can meet the real-time requirements while completing the detection task.

Experimental data 6 are shown in Fig. 9 (a) and (b), which show the multitemporal video images differing by approximately 6 seconds, and (c) shows the reference for CD. Compared with that of the previous groups, although the image illumination of this group is more uniform, the visibility is

TABLE IV
OBJECTIVE EVALUATION INDEXES OF
EXPERIMENTAL DATA 6

Algorithms	OE \uparrow	PCC \uparrow	Kappa \uparrow	F1 \uparrow	T(s) \uparrow
NR-ELM [15]	6322	0.9667	0.0224	0.0232	16.16
DWT [49]	112	0.9941	0.5723	0.5725	1.815
CWNN [47]	2392	0.9874	0.0583	0.0590	25.60
FDA-RMG [46]	10	0.9995	0.9333	0.9333	2.074
NSST [18]	17	0.9991	0.8930	0.8931	48.01
NPSG [53]	4810	0.9747	0.0295	0.0302	34.35
Zhu [12]	14	0.9993	0.9027	0.9028	0.488
CDI-K [48]	53	0.9972	0.7133	0.7135	1.942
Proposed	8	0.9996	0.9480	0.9480	0.337

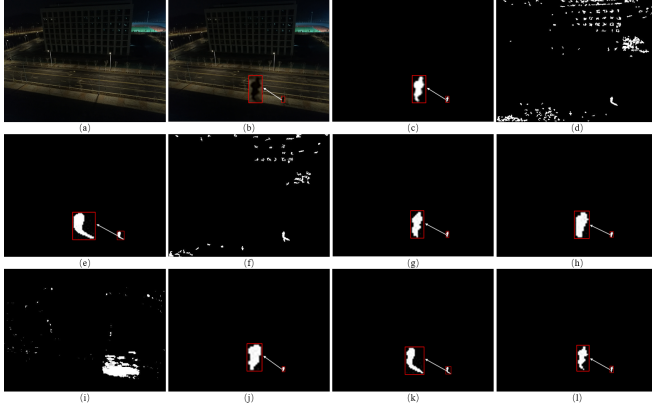


Fig. 9. Change detection results of experimental data 6, (a) Original image, (b) Changed image, (c) Reference image, (d) NR-ELM [15], (e) DWT [49], (f) CWNN [47], (g) FDA-RMG [46], (h) NSST [18], (i) NPSG [53], (j) Zhu [12], (k) CDI-K [48], (l) Proposed method.

lower. Again, we set the size of the image to 500×380 .

The detection results in Fig. 9 show that NR-ELM, CWNN, and NPSG have errors in the classification of changed pixels and unchanged pixels due to their sensitivity to noise, and a large number of pseudo-change areas appear in (d), (f), and (i). Thus, NR-ELM, CWNN and NPSG in Table IV have higher values of OE. Although (e), (h), (j), and (k) is not affected by noise, the detected target edge is larger than the real edge, which reduces the detection performance of the algorithm. Both FDA-RMG and our method have achieved good results, but the detection result obtained by our method is closer to the reference. This is mainly because our algorithm utilizes an adaptive fusion method based on the energy feature of the DI and the Laplacian pyramid to generate high-quality DI, and the final detection accuracy is higher than that of the other comparison algorithms. The evaluation indicators in Table IV also show that the Kappa coefficient, PCC, and F1 of our algorithm are higher than those of the other algorithms, and our algorithm has the shortest running time.

C. Artificial Change Detection Analysis

To further verify the robustness of the proposed algorithm, we select experimental data 7 for artificial image CD. Both (a) and (b) in Fig. 10 add uniformly distributed random noise with a mean value of 0 and a standard deviation of 80; (a) shows the time-phased image of the monitoring equipment

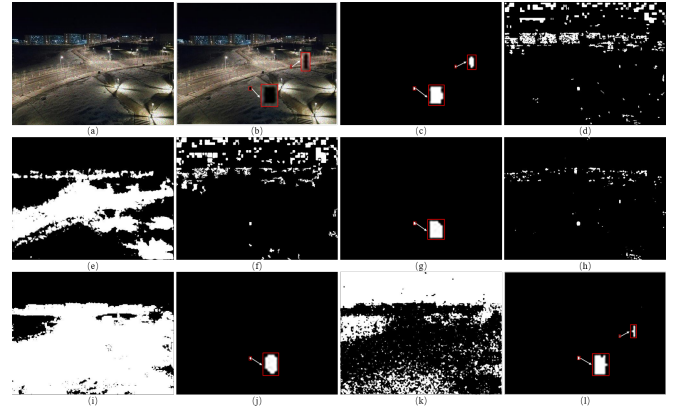


Fig. 10. Change detection results of experimental data 7, (a) Original image, (b) Changed image, (c) Reference image, (d) NR-ELM [15], (e) DWT [49], (f) CWNN [47], (g) FDA-RMG [46], (h) NSST [18], (i) NPSG [53], (j) Zhu [12], (k) CDI-K [48], (l) Proposed method.

at a certain time, and (b) shows the time-phased image after adding the corresponding change area in (a). The size of the two multitemporal video images is 500×380 , and (c) shows a reference image of the changing area.

From the detection results in Fig. 10 and the evaluation indicators in Table V, DWT, NPSG, and CDI-K are the most sensitive to noise, resulting in a large number of pseudo-change areas in the detection results. Although NR-ELM, CWNN, and NSST can detect the approximate change area, they are still affected by noise and cause a large number of false alarms. Accordingly, the detection results obtained by these methods in Table V all have higher values of OE. In contrast, FDA-RMG and Zhu effectively suppress noise and better detect the changed area, mainly because FDA-RMG and Zhu adopt frequency-domain saliency detection and multiscale morphological filters to suppress noise, respectively. However, at the same time, they regard another small change area as noise and filter it out, thus reducing the detection accuracy of the algorithm. Compared with the comparison algorithms, our proposed method achieves the best detection results under noise interference conditions, and the evaluation indicators in Table V also show that our algorithm is better than the compared algorithms in terms of Kappa, F1, and T.

We also artificially add Poisson noise to detect the changed region of relevant video images for the above research. The experimental results show that our algorithm can still effectively suppress Poisson noise and accurately detect the corresponding change area, which further verifies the strong robustness of the proposed algorithm in a noisy environment.

D. Robustness Analysis of Change Detection Methods

In most cases of low illumination, the multitemporal video images collected by the eagle eye surveillance device are scenes with no obvious changes. Due to the small difference in the pixel value of the corresponding position of the two multitemporal video images, the noise will become abnormally obvious in this case, resulting in false alarms. The proposed algorithm not only can effectively detect the changed area but also has good robustness to the unchanged area.

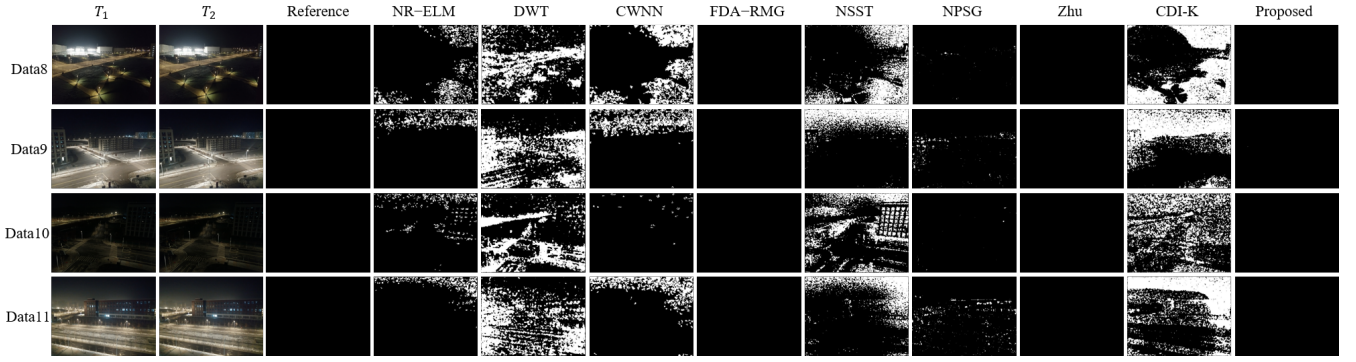


Fig. 11. Detection results in unchanged scenes. It can be seen from the detection results that DWT, CWNN, NSST and CDI-K are very sensitive to noise, resulting in a large number of white areas of false detection. Although NR-ELM and NPSG have less false detection areas, they are still affected by noise. However, FDA-RMG and Zhu effectively suppress noise via saliency detection and multi-scale morphological filter, respectively, while our method also achieves good detection results.

TABLE V
OBJECTIVE EVALUATION INDEXES OF
EXPERIMENTAL DATA 7

Algorithms	OE \uparrow	PCC \uparrow	Kappa \uparrow	F1 \uparrow	T(s) \uparrow
NR-ELM [15]	14457	0.9239	0.0075	0.0081	14.70
DWT [49]	66613	0.6809	0.0013	0.0019	1.922
CWNN [47]	16056	0.9155	0.0048	0.0052	17.53
FDA-RMG [46]	20	0.9989	0.8076	0.8077	2.267
NSST [18]	2339	0.9877	0.0474	0.0480	47.23
NPSG [53]	123981	0.3475	0.0003	0.0009	31.40
Zhu [12]	29	0.9985	0.6881	0.6882	0.418
CDI-K [48]	78905	0.5844	0.0002	0.0009	2.174
Proposed	17	0.9991	0.8495	0.8496	0.306

TABLE VI
AVERAGE EVALUATION INDEXES OF DETECTION
RESULTS IN UNCHANGED SCENE

Algorithms	$\overline{OE} \downarrow$	$\overline{PCC} \downarrow$	$\overline{T(s)} \uparrow$
NR-ELM [15]	11807	0.9378	15.19
DWT [49]	67561	0.6444	4.397
CWNN [47]	24933	0.8683	22.52
FDA-RMG [46]	0.3	0.9998	11.02
NSST [18]	42420	0.7757	46.46
NPSG [53]	1686	0.9911	31.25
Zhu [12]	4	0.9991	0.399
CDI-K [48]	71836	0.6219	1.784
Proposed	0.5	0.9997	0.312

We selected four sets of experimental datasets of different scenes for relevant experiments. Fig. 11 shows that the detection results of NR-ELM, DWT, CWNN, NSST, and CDI-K are not ideal for unchanged scenes, mainly because the noise becomes abnormally obvious due to the lack of large difference change areas as a contrast, resulting in the pixels polluted by noise being wrongly divided into change pixels. Though NPSG does not have large pseudo-change areas, it is still affected by a small amount of noise. However, FDA-RMG, Zhu, and our method effectively suppress the interference of noise under the unchanged condition. From the average indexes in Table VI, although the OE and PCC values of FDA-RMG are better than those of our method, our method runs in the shortest time. Table VI is the average evaluation

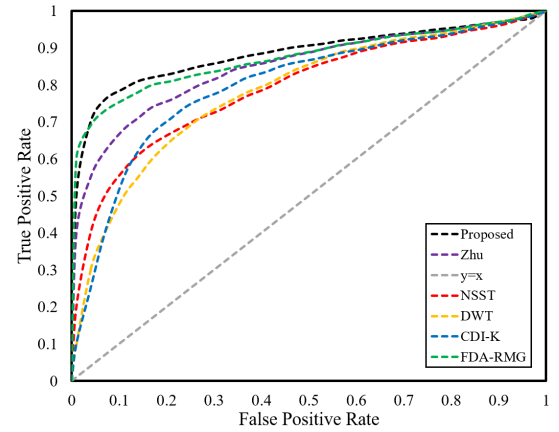


Fig. 12. ROC curves of different comparison algorithms.

TABLE VII
AVERAGE EVALUATION INDICATORS OF
EXPERIMENTAL DATA

Algorithms	$\overline{OE} \downarrow$	$\overline{PCC} \downarrow$	$\overline{Kappa} \uparrow$	$\overline{F1} \uparrow$	$\overline{T(s)} \uparrow$
NR-ELM [15]	10313	0.9457	0.0103	0.0108	15.11
DWT [49]	39.7	0.9973	0.7719	0.7726	1.814
CWNN [47]	25739	0.8645	0.0159	0.0164	28.75
FDA-RMG [46]	14.1	0.9993	0.8043	0.8043	2.113
NSST [18]	14979	0.9212	0.3359	0.3362	46.87
NPSG [53]	10918	0.9425	0.0124	0.0129	32.40
Zhu [12]	21.6	0.9989	0.8042	0.8043	0.422
CDI-K [48]	55	0.9964	0.7125	0.7131	2.098
Proposed	14.9	0.9992	0.8546	0.8547	0.302

index of the experimental results. Since there are no changed pixels in the reference image under the unchanged scenes, the Kappa coefficient is meaningless, and since TP is 0, F1 is also 0.

E. Analysis of Average Experimental Results

To comprehensively evaluate the stability and robustness of the proposed algorithm, we selected 75 sets of multitemporal video images in different scenes to conduct experiments on our method and comparison methods. Each set of experimental

TABLE VIII
THE RUNNING TIME OF EACH PART
OF THE PROPOSED METHOD

Process	Running time(s)	Process	Running time(s)
Log-ratio	0.0031	Filter	0.0127
Mean ratio	0.0014	K-medoids	0.2276
Image fusion	0.0727	The total	0.3175

data is tested 10 times, and finally we obtain the average CD results shown in Table VII. To account for the imbalance between changed and unchanged pixels, we plot the receiver operating characteristic (ROC) curve as shown in Fig. 12, from which we can see that FDA-RMG, Zhu, and our proposed method all have good detection performance, mainly because FDA-RMG and Zhu adopt saliency detection and multiscale morphological filter strategies, respectively, which improve the quality of DI. Compared with CDI-K, our method makes full use of the energy feature of DI to improve detection accuracy.

Simultaneously, from the average evaluation index in Table VII, the overall detection performance of NR-ELM, CWNN, and NPSG is inferior due to the influence of noise, so they cannot complete the CD task. The unstable detection results of NSST lead to higher OE values and lower Kappa and F1 values. However, the overall detection performance of DWT, FDA-RMG, Zhu, and CDI-K is outstanding, and the changed areas can be well detected in different low illumination scenes. Though the OE and PCC of FDA-RMG are superior to those of our algorithm, our algorithm performs better in terms of Kappa and F1, and the overall runtime of the algorithm is the shortest.

F. Time Complexity Analysis

The flowchart of the proposed method is shown in Fig. 2, which mainly includes four steps: using the log-ratio and pixel ratio operators to generate two DIs; calculating the energy of the two DIs, and combining the Laplacian pyramid to perform an adaptive fusion of two DIs; the normalization function and improved adaptive median filter are used for compressing DI and denoising, respectively; the results of CD are obtained via the optimized k-medoids. First, we analyze the time complexity of each step and then add the results of each step to obtain the final time complexity, where n is the total number of pixels of the video image, w is the window size of the improved adaptive median filter, k is the cluster center, t is the iteration number of k-medoids clustering, and d is the calculation distance complexity of k-medoids.

In the process of generating the DIs, the time complexity of using the log ratio is $O(n)$, and the time complexity of the pixel ratio is $O(n)$, then the time complexity of the fusion DI is $O(2n)$. The time complexity of the normalization function is $O(n)$, the time complexity of the improved adaptive median filter is $O(w^2n)$, and the time complexity of optimized k-medoids clustering is $O(k^2dnt)$. Consequently, the time complexity of the whole algorithm is $O(3n + w^2n + k^2dnt)$. From the perspective of high-order terms, the time complexity of the whole algorithm can be approximated as $O(n + k^2dnt)$.

To comprehensively analyze the running time of the algorithm, we analyze the running time of each part of the algorithm. The experimental results in Table VIII show that k-medoids clustering takes the most time, while it takes less time to generate DI and adaptive median filter.

V. CONCLUSION

To accurately and effectively detect the weak changes of video images under low illumination conditions, we propose a method for image CD in surveillance video based on optimized k-medoids clustering and adaptive fusion of DI. First, two DIs are generated by the log-ratio operator and pixel ratio operator. Then, the two DIs are adaptively fused by using the local energy features of the DIs and the Laplacian pyramid. In addition, the method of maximum and minimum normalization is used to compress the fused DI, and the fusion DI is filtered by the improved adaptive median filter. Finally, the CD results are obtained by the optimized k-medoids clustering algorithm. The experimental results show that the proposed method can accurately and effectively detect the weak changes of video images in low illumination environment. Compared with those of the existing comparison algorithms, the accuracy and robustness of the proposed algorithm are higher, and the running time is shorter. At present, the CD of image structure consistency [52], [53] is a new direction, Therefore, we will explore and study the image structure consistency in the next work, and use its advantage of being insensitive to noise, illumination and other interference factors to detect low-illumination video image CD.

REFERENCES

- [1] S. Zhang et al., "Influence-aware Attention Networks for Anomaly Detection in Surveillance Videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5427-5437, Aug. 2022.
- [2] Y. Xu and B. Sun, "A Novel Variational Model for Detail-Preserving Low-Illumination Image Enhancement," *Signal Processing*, vol. 195, p. 108468, Jun. 2022.
- [3] H. Lv and H. Li, "Denoising method of low illumination underwater motion image based on improved canny," *Microprocessors and Microsystems*, vol. 82, p. 103862, Apr. 2021.
- [4] K. M. Adal, P. G. van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, "An Automated System for the Detection and Classification of Retinal Changes Due to Red Lesions in Longitudinal Fundus Images," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1382-1390, Jun. 2018.
- [5] J. T. Zhou, L. Zhang, Z. Fang, J. Du, X. Peng, and Y. Xiao, "Attention-Driven Loss for Anomaly Detection in Video Surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4639-4647, Dec. 2020.
- [6] M. Xu, B. Liu, P. Fu, J. Li, Y. H. Hu, and S. Feng, "Video Salient Object Detection via Robust Seeds Extraction and Multi-graphs Manifold Propagation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 2191-2206, Jul. 2020.
- [7] H. Dong, W. Ma, L. Jiao, F. Liu, and L. Li, "A Multiscale Self-Attention Deep Clustering for Change Detection in SAR Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1-16, 2022.
- [8] P. Han, C. Ma, Q. Li, P. Leng, S. Bu, and K. Li, "Aerial image change detection using dual regions of interest networks," *Neurocomputing*, vol. 349, pp. 190-201, Jul. 2019.
- [9] D. Tian and M. Gong, "A novel edge-weight based fuzzy clustering method for change detection in SAR images," *Information Sciences*, vol. 467, pp. 415-430, Oct. 2018.
- [10] F. Guo, W. Wang, Z. Shen, J. Shen, L. Shao, and D. Tao, "Motion-Aware Rapid Video Saliency Detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4887-4898, Dec. 2020.

- [11] Q. S. Hamad, M. S. Croock, and S. Al Qaraawi, "Efficient infrared sensor and camera based monitoring system," in *2013 International Conference on Electrical Communication, Computer, Power, and Control Engineering (ICECCPCE)*, Mosul, Iraq, Dsec. 2013, pp. 20–25.
- [12] Y. Zhu, Z. Jia, J. Yang, and N. K. Kasabov, "Change Detection in Multitemporal Monitoring Images Under Low Illumination," *IEEE Access*, vol. 8, pp. 126700–126712, 2020.
- [13] J. Wang, X. Yang, X. Yang, L. Jia, and S. Fang, "Unsupervised change detection between SAR images based on hypergraphs," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 61–72, Jun. 2020.
- [14] Z. Li, Z. Jia, luyang liu, J. Yang, and N. Kasabov, "A method to improve the accuracy of SAR image change detection by using an image enhancement method," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 137–151, May 2020.
- [15] F. Gao, J. Dong, B. Li, Q. Xu, and C. Xie, "Change detection from synthetic aperture radar images based on neighborhood-based ratio and extreme learning machine," *J. Appl. Remote Sens.*, vol. 10, no. 4, p. 046019, Dec. 2016.
- [16] S. Kalaiselvi and V. Gomathi, " α -cut induced Fuzzy Deep Neural Network for change detection of SAR images," *Applied Soft Computing*, vol. 95, p. 106510, Oct. 2020.
- [17] L. Liu, Z. Jia, J. Yang, and N. K. Kasabov, "SAR Image Change Detection Based on Mathematical Morphology and the K-Means Clustering Algorithm," *IEEE Access*, vol. 7, pp. 43970–43978, 2019.
- [18] Z. Wenyan, J. Zhenhong, Y. Yu, J. Yang, and N. Kasabov, "SAR image change detection based on equal weight image fusion and adaptive threshold in the NSST domain," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 785–794, Jan. 2018.
- [19] H. Chen, L. Jiao, M. Liang, F. Liu, S. Yang, and B. Hou, "Fast unsupervised deep fusion network for change detection of multitemporal SAR images," *Neurocomputing*, vol. 332, pp. 56–70, Mar. 2019.
- [20] M. Yang, L. Jiao, F. Liu, B. Hou, S. Yang, and M. Jian, "DPFL-Nets: Deep Pyramid Feature Learning Networks for Multiscale Change Detection," *IEEE Trans. Neural Netw. Learning Syst.*, pp. 1–15, 2021.
- [21] W. Zhang, J. Li, F. Zhang, J. Sun, and K. Zhang, "Unsupervised Change Detection of Multispectral Images Based on PCA and Low-Rank Prior," *IEEE Geosci. Remote Sensing Lett.*, vol. 19, pp. 1–5, 2022.
- [22] X. Wang, C. Xing, Y. Feng, R. Song, and Z. Mu, "A Novel Hyperspectral Image Change Detection Framework Based on 3D-Wavelet Domain Active Convolutional Neural Network," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium, Jul. 2021, pp. 4332–4335.
- [23] Y. Wang, L. Du, and H. Dai, "Unsupervised SAR Image Change Detection Based on SIFT Keypoints and Region Information," *IEEE Geosci. Remote Sensing Lett.*, vol. 13, no. 7, pp. 931–935, Jul. 2016.
- [24] Y. Zheng, L. Jiao, H. Liu, X. Zhang, B. Hou, and S. Wang, "Unsupervised saliency-guided SAR image change detection," *Pattern Recognition*, vol. 61, pp. 309–326, Jan. 2017.
- [25] C. Chen, J. Song, C. Peng, G. Wang, and Y. Fang, "A Novel Video Salient Object Detection Method via Semisupervised Motion Quality Perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2732–2745, May 2022.
- [26] C. Chen, G. Wang, C. Peng, X. Zhang, and H. Qin, "Improved Robust Video Saliency Detection Based on Long-Term Spatial-Temporal Information," *IEEE Trans. on Image Process.*, vol. 29, pp. 1090–1100, 2020.
- [27] C. Chen, S. Li, H. Qin, and A. Hao, "Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency clues and low-rank analysis," *Pattern Recognition*, vol. 52, pp. 410–432, Apr. 2016.
- [28] L. Ma, Z. Jia, Y. Yu, J. Yang, and N. K. Kasabov, "Multi-Spectral Image Change Detection Based on Band Selection and Single-Band Iterative Weighting," *IEEE Access*, vol. 7, pp. 27948–27956, 2019.
- [29] L. T. Ngo, D. S. Mai, and W. Pedrycz, "Semi-supervising Interval Type-2 Fuzzy C-Means clustering with spatial information for multi-spectral satellite image classification and change detection," *Computers & Geosciences*, vol. 83, pp. 1–16, Oct. 2015.
- [30] M. A. Khalilia, J. Bezdek, M. Popescu, and J. M. Keller, "Improvements to the relational fuzzy c-means clustering algorithm," *Pattern Recognition*, vol. 47, no. 12, pp. 3920–3930, Dec. 2014.
- [31] L. Ke, Y. Lin, Z. Zeng, L. Zhang, and L. Meng, "Adaptive Change Detection With Significance Test," *IEEE Access*, vol. 6, pp. 27442–27450, 2018.
- [32] Z. Wang, C. Peng, Y. Zhang, N. Wang, and L. Luo, "Fully convolutional siamese networks based change detection for optical aerial images with focal contrastive loss," *Neurocomputing*, vol. 457, pp. 155–167, Oct. 2021.
- [33] Z. Chen, G. Li, J. He, Z. Yang, and J. Wang, "A new parallel adaptive structural reliability analysis method based on importance sampling and K-medoids clustering," *Reliability Engineering & System Safety*, vol. 218, p. 108124, Feb. 2022.
- [34] S. Saha, B. Banerjee, and X. X. Zhu, "Trusting Small Training Dataset for Supervised Change Detection," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul. 2021, pp. 2031–2034.
- [35] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–16, 2022.
- [36] Z. Zheng, Y. Liu, S. Tian, J. Wang, A. Ma, and Y. Zhong, "Weakly Supervised Semantic Change Detection via Label Refinement Framework," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul. 2021, pp. 2066–2069.
- [37] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A Semisupervised Convolutional Neural Network for Change Detection in High Resolution Remote-Sensing Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.
- [38] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised Change Detection Using Graph Convolutional Network," *IEEE Geosci. Remote Sensing Lett.*, vol. 18, no. 4, pp. 607–611, Apr. 2021.
- [39] J. Shi, T. Wu, A. K. Qin, Y. Lei, and G. Jeon, "Semisupervised Adaptive Ladder Network for Remote Sensing Image Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–20, 2022.
- [40] G. Liu, Y. Yuan, Y. Zhang, Y. Dong, and X. Li, "Style Transformation-Based Spatial-Spectral Feature Learning for Unsupervised Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [41] J. Liu, M. Gong, A. K. Qin, and K. C. Tan, "Bipartite Differential Neural Network for Unsupervised Image Change Detection," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 31, no. 3, pp. 876–890, Mar. 2020.
- [42] C. Ren, X. Wang, J. Gao, X. Zhou, and H. Chen, "Unsupervised Change Detection in Satellite Images With Generative Adversarial Network," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 12, pp. 10047–10061, Dec. 2021.
- [43] C. Wu, H. Chen, B. Du, and L. Zhang, "Unsupervised Change Detection in Multitemporal VHR Images Based on Deep Kernel PCA Convolutional Mapping Network," *IEEE Trans. Cybern.*, pp. 1–15, 2021.
- [44] L. Liu, D. Hong, L. Ni, and L. Gao, "Multilayer Cascade Screening Strategy for Semi-Supervised Change Detection in Hyperspectral Images," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 15, pp. 1926–1940, 2022.
- [45] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised Change Detection in SAR Image Based on Gauss-Log Ratio Image Fusion and Compressed Projection," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.
- [46] F. Gao, X. Wang, J. Dong, and S. Wang, "Synthetic aperture radar image change detection based on frequency-domain analysis and random multigraphs," *J. Appl. Rem. Sens.*, vol. 12, no. 1, p. 016010, Jan. 2018.
- [47] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea Ice Change Detection in SAR Images Based on Convolutional-Wavelet Neural Networks," *IEEE Geosci. Remote Sensing Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019.
- [48] Y. Zheng, X. Zhang, B. Hou and G. Liu, "Using Combined Difference Image and k - Means Clustering for SAR Image Change Detection," *IEEE Geosci. Remote Sensing Lett.*, vol. 11, no. 3, pp. 691–695, Mar. 2014.
- [49] M. Gong, Z. Zhou and J. Ma, "Change Detection in Synthetic Aperture Radar Images based on Image Fusion and Fuzzy Clustering," *IEEE Trans. on Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.
- [50] P. D'Urso, M. Mucciardi, E. Otranto, and V. Vitale, "Community mobility in the European regions during COVID-19 pandemic: A partitioning around medoids with noise cluster based on space-time autoregressive models," *Spatial Statistics*, p. 100531, Jul. 2021.
- [51] H. Jiang, Y. Wu, K. Lyu, and H. Wang, "Ocean Data Anomaly Detection Algorithm Based on Improved k-medoids," in *2019 Eleventh International Conference on Advanced Computational Intelligence (ICACI)*, Guilin, China, Jun. 2019, pp. 196–201.
- [52] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure Consistency-Based Graph for Unsupervised Change Detection With Homogeneous and Heterogeneous Remote Sensing Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–21, 2022.
- [53] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognition*, vol. 109, p. 107598, Jan. 2021.



Baoqiang Shi received the B.S. degree from the School of Information and Engineering, Longdong University, Qingyang, China, in 2019. He obtained the M.S. degree from the School of Information Science and Engineering, Xinjiang University, Urumqi, China, in 2023. His research direction is the change detection of the wide-field video images under low illumination.



Zhenhong Jia received the B.S. degree from Beijing Normal University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1987 and 1995, respectively. He is currently a Professor with the Autonomous University Key Laboratory of Signal and Information Processing Laboratory, Xinjiang University, China. His research interests include digital image processing, optical information detection, and machine learning.



Jie Yang received the B.S. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1982 and 1985, respectively, and the Ph.D. degree from the Department of Computer Science, Hamburg University, Hamburg, Germany, in 1994. He is currently a Professor with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. His major research interests are object detection and recognition, data fusion and data mining, and medical image processing.



Nikola Kasabov (M'93–SM'98–F'10) received MSc. degree in computing and electrical engineering and his Ph.D. degree in mathematical sciences from the Technical University of Sofia, Bulgaria, in 1971 and 1975, respectively. He is the Founding Director of KEDRI and Professor of knowledge engineering at Auckland University of Technology. His major research interests include computational intelligence, neural networks, bioinformatics, neuroinformatics, with more than 690 publications.