

## Research Article

# Deep Learning-Driven Gaussian Modeling and Improved Motion Detection Algorithm of the Three-Frame Difference Method

Dingchao Zheng, Yangzhi Zhang, and Zhijian Xiao 

Digital Engineering Academy, Zhejiang Dongfang Polytechnic, Wenzhou 325200, China

Correspondence should be addressed to Zhijian Xiao; xiaozhijiandigital@126.com

Received 17 March 2021; Revised 24 April 2021; Accepted 19 May 2021; Published 8 June 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Dingchao Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To enhance the effect of motion detection, a Gaussian modeling algorithm is proposed to fix holes and breaks caused by the conventional frame difference method. The proposed algorithm uses an improved three-frame difference method. A three-frame image sequence with one frame interval is selected for pairwise difference calculation. The logical “OR” operation is used to achieve fast motion detection and to reduce voids and fractures. The Gaussian algorithm establishes an adaptive learning model to make the size and contour of the motion detection more accurate. The motion extracted by the improved three-frame difference method and Gaussian model is logically summed to obtain the final motion foreground picture. Moreover, a moving target detection method, based on the U-Net deep learning network, is proposed to reduce the dependency of deep learning on the number of training datasets. It helps the algorithm to train models on small datasets. Next, it calculates the ratio of the number of positive and negative samples in the dataset and uses the reciprocal of the ratio as the sample weight to deal with the imbalance of positive and negative samples. Finally, a threshold is set to predict the results for obtaining the moving object detection accuracy. Experimental results show that the algorithm can suppress the generation and rupture of holes and reduce the noise. Also, it can quickly and accurately detect movement to meet the design requirements.

## 1. Introduction

Motion detection refers to the process of extracting and identifying objects in image sequences or videos of which spatial position changes as foreground motion [1]. Due to the dynamic changes of the scene, it is not easy to reliably and effectively detect and segment motion. In the field of computer vision, it has always been a common research subject. Operations such as tracking, processing, recognition, and behavior discrimination can be performed through effective segmentation of moving objects. These are widely applied to live production, such as video surveillance, motion recognition, and multimedia applications, among others [2].

At present, the frame difference method [3], optical flow method [4], and background subtraction method [5] are standard methods for motion detection, in addition to the feature matching method, K-nearest neighbor method, and ViBe algorithm [6]. Two adjacent frames of the image

sequence are taken to perform pixel-based time difference and binarization to extract the motion area in the image for the frame difference method. The advantages of this method are as follows: its simple principle and a fast calculation speed. However, the noise resistance is poor, and holes and breaks are easy to be produced [7]. The ViBe algorithm stores a sample set for each pixel. In the sample collection, the sampling value is the pixel's previous pixel value and the pixel values of its neighbors. The first frame of the image can be taken to complete the creation of the background model's sample set. The sample set can then be compared to each new pixel value to see if it belongs to the context point. The calculation efficiency is high, but it is easy to cause breaks and other problems. In the optical flow method, the motion information of objects between adjacent frames is determined using changes in the time domain of pixels in the image series and the correlation between adjacent frames, based on the relationship between the previous frame and

the current frame. Although the entire area of motion can be detected, there is a large amount of calculation. With a complex structure, it can be easily affected by external factors.

In this study, we propose an optimized Gaussian model and improved three-frame difference method for motion detection. The main contributions are as follows.

- (1) We present an optimized Gaussian model and improved three-frame difference method to solve the shortcomings of the aforementioned algorithm, which is used for this task till date
- (2) We compare the optimized Gaussian model with the improved three-frame difference method. First, the traditional three-frame difference method is improved to obtain the foreground motion. Then, the Gaussian model is optimized. According to the motion obtained by the Gaussian model, logical operation and morphological processing are performed on the two motions to obtain the final foreground motion. This algorithm can effectively overcome holes and adapt to environmental changes with good robustness.

The rest of this study is organized as follows. In Section 2, we discuss the improved three-frame difference method followed by the optimized Gaussian modeling in Section 3. In Section 4, the improved U-Net model is discussed. In Section 5, the experimental results and analysis are provided. Finally, the study is concluded and future research directions are provided in Section 6.

## 2. Improved Three-Frame Difference Method

In the traditional three-frame difference method, the first two frames and the last two frames of the adjacent three frames of images are taken for the difference operation. The two frames of difference images are then subjected to the “AND” operation to obtain the result [8]. Due to the continuity of the video sequence, the change between adjacent frames will be minimal if there is no motion in the area. If there is a motion, there will be apparent changes between consecutive frames [9]. Although motions can be quickly detected by this method, problems such as holes and double shadows are likely to appear.

To solve the aforementioned problems, the traditional three-frame difference method is improved in this study. The difference operation is conducted on the three-frame images of the video sequence interval. The “OR” operation instead of the “AND” operation is conducted for different images. The specific steps are as follows:

- (1) Three-frame images with one frame interval are collected and converted into grayscale images. Then, filtering processing to remove noise is performed to obtain the corresponding three-frame images, denoted as  $f_{k-2}(x, y)$ ,  $f_k(x, y)$ ,  $f_{k+2}(x, y)$ .
- (2) The  $k$  frame of the image is taken as the current frame, and the difference operation is performed with the  $k+2$  and  $k-2$  frames, respectively, to obtain

the absolute value of difference  $D_1(x, y)$  and  $D_2(x, y)$  as shown in the following equations.

$$D_1(x, y) = |f_{k-2}(x, y) - f_k(x, y)|, \quad (1)$$

$$D_2(x, y) = |f_{k+2}(x, y) - f_k(x, y)|. \quad (2)$$

- (3) The obtained difference of the above equations is binarized, and an OR operation is performed as shown in the following equation.

$$D(x, y) = D_1(x, y) \oplus D_2(x, y). \quad (3)$$

- (4) Finally, appropriate morphological processing is performed on the results

According to the above principles, the code is written to realize the motion detection of a video image by the traditional three-frame difference and the improved three-frame difference. The detection effect is shown in Figure 1.

In Figures 1(a) and 1(b), images at different periods in the same video are shown. The left side image in each figure is the original video image. The middle is the motion detected by the traditional three-frame difference method. The motion observed by the enhanced three-frame difference method is seen on the right. The detection results show that improving the three-frame difference approach will help solve the holes’ problem to some degree. However, the motion is deformed to a certain extent, so it needs to be improved or combined with other algorithms for processing.

## 3. Optimized Gaussian Modeling

The value or feature of each pixel in the image is distributed within a specific range around a certain central point in a short time, and this distribution is regular. According to statistical laws, if the amount of data is large enough, it will be normally distributed, known as Gaussian distribution. According to this feature, if the pixel’s value is far from the center point, the pixel belongs to the foreground; otherwise, it belongs to the background [10]. In motion detection, the Gaussian modeling is taken for each background pixel by mixing  $K$  Gaussian distributions (3–5 is generally taken for  $K$ ). Clearer motions can be extracted by Gaussian modeling, and the background can be updated and adaptive. The improved three-frame difference method can be combined to achieve complementary advantages and disadvantages and enhance the detection of motions [11].

**3.1. Background Model Building.** The grayscale histogram of an image reflects the frequency of a specific grayscale value in the image. Suppose the motion area in the image is larger than the background area. There is a certain difference in grayscale between the background and motion area. In that case, the grayscale histogram of the image presents a double peak-valley shape. One of the peaks corresponds to the motion, and the other peak corresponds to the central grayscale of the background. For complex images, they are generally multipeak. By treating the multipeak characteristics of the histogram as the superposition of multiple

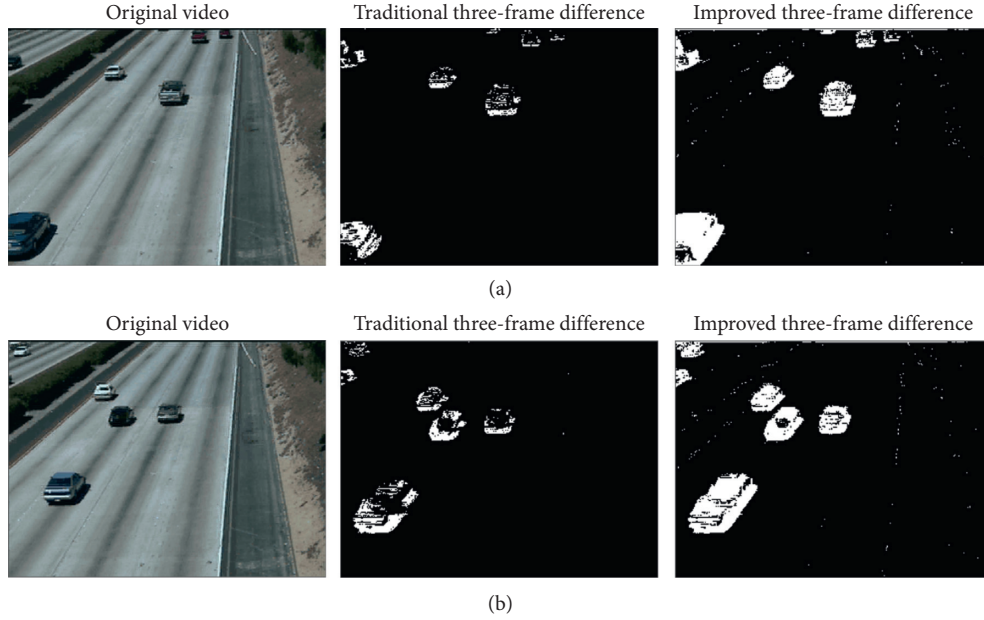


FIGURE 1: Motion detection by the three-frame difference method.

Gaussian distributions, the image segmentation problem can be solved [12].

Each pixel of the background image is modeled by a Gaussian mixture model composed of  $K$  Gaussian distributions as shown in the following equation.

$$P(x_t) = \sum_{i=1}^k \omega_{i,t} \eta(x_t, \mu_{i,t}, \tau_{i,t}), \quad (4)$$

where  $k$  is the total number of distribution patterns, and  $\omega_{i,t}$  represents the estimated value of the weight coefficient of the  $i^{\text{th}}$  Gaussian distribution in the Gaussian mixture model at time  $t$ ,  $\eta(x_t, \mu_{i,t}, \tau_{i,t})$  is the  $i^{\text{th}}$  Gaussian distribution at time  $t$ , and  $\eta$  represents the probability density function of the Gaussian distribution.  $x_t$  represents the value of  $i^{\text{th}}$  pixel at time  $t$ .  $\mu_{i,t}$ ,  $\tau_{i,t}$  represent the mean vector and covariance matrix of the  $i^{\text{th}}$  Gaussian distribution in the Gaussian mixture model at time  $t$ .

The patterns are arranged in descending order of the value of  $\omega/\delta^2$ . The larger the value, the more likely the  $k^{\text{th}}$  distribution is the background model. Then, the first  $B$  Gaussian distributions were selected as the background models. As shown in the following equation,  $T$  represents the proportion of the background.

$$B = \arg \left( \min \left( \sum_{k=1}^b \omega_k > T \right) \right), \quad (5)$$

**3.2. Foreground Motion Detection.** After the background model is established, each pixel value  $x_t$  of the new input image frame is matched with  $B$  Gaussian distributions in the background model. If the pixel value matches any Gaussian distribution, the pixel is the background point; otherwise, it is the foreground, thus completing the motion detection.

Each new pixel value  $x_t$  is compared with the current  $K$  models. As generally shown in equation (6), the mean deviation coefficient is generally 2.5.

$$|x_t - \mu_{i,t-1}| \leq 2.5\sigma_{i,t-1}. \quad (6)$$

**3.3. Background Update.** The context model must be changed during the motion detection process. The context model is learning to enable the detection to adapt to changes in the environment, and  $a$  is the learning rate. Suppose a frame of image does not match any Gaussian distribution during the detection process. In that case, the pattern with the most negligible weight is replaced. The weight of the pattern is represented by equation (7), and its mean and standard deviation remain unchanged.

$$\omega_{k,t} = (1 - \alpha) * \omega_{k,t-1}. \quad (7)$$

Suppose, it is detected that at least one Gaussian distribution in the background model matches the pixel. In that case, the parameters of the Gaussian model are updated using the following equations.

$$\omega_{k,t} = (1 - \alpha) * \omega_{k,t-1} + \alpha, \quad (8)$$

$$\rho = \alpha * \eta(x_t | \mu_k, \sigma_k), \quad (9)$$

$$\mu_t = (1 - \rho) * \mu_{t-1} + \rho * x_t, \quad (10)$$

$$\sigma^2 = (1 - \rho) * \sigma_{t-1}^2 + \rho * (x_t - \mu_t)^T (x_t - \mu_t). \quad (11)$$

The effect of motion detection by Gaussian modeling is shown in Figure 2. The left side is the original video image, and the right side is the motion detection image. From figure, it can be seen that the Gaussian model can better



FIGURE 2: Motion detection by Gaussian modeling.

overcome the problem of holes. Simultaneously, the appearance of the motion is not elongated, and the outline is evident [13].

*3.4. Combining the Improved Three-Frame Difference Method.* The frame difference approach has a simple operation and quick speed. However, due to its low noise resistance, holes are easy to appear. For the enhanced three-frame difference process, three nonadjacent frames are taken, and the OR operation is used instead of the original, which may not only eliminate gaps but also trigger issues such as elongated movements and blurred outlines. Therefore, the Gaussian model is combined to establish a learning background model for motion detection. The outline of the motion detected by the Gaussian model method is relatively straightforward, and the appearance is not elongated.

In the detection process of motions, first, the motions are extracted by the improved three-frame difference method and Gaussian modeling. Then, the logical AND operation is conducted for the motion images obtained by these two algorithms. Next, connectivity detection and morphological operations are taken to the process, and finally, the completed motion is achieved. The flow chart of the algorithm is shown in Figure 3.

#### 4. Improved U-Net Model

The top-down image detection algorithm based on deep learning [14–18] extracts global image information and high-level semantic information by self-learning image features. In contrast, the bottom-up image detection algorithm based on low-level visual features extracts low-level image color and texture. Image features can extract rich image local and spatial information. Therefore, fuse the semantic information and global information of the top-down image detection algorithm with the bottom-up saliency detection algorithm's spatial information. The local information can help the high-level network obtain the local image information to locate the salient target and improve the image detection effect.

Therefore, this study adds the saliency map extracted by the bottom-up motion detection algorithm in the decoder stage of U-Net and realizes the accurate prediction of the saliency image by fusing high-level global information and

low-level local information. Specifically, the algorithm first reduces the saliency map extracted by the bottom-up image detection algorithm, generates a priori map of the same size as the feature map of the U-Net network decoder stage, and then combines the a priori map with upsampling. Feature fusion is performed on the feature map of the stage. Furthermore, the channel attention module is added after sampling at each layer to extract more important channel information. Finally, the saliency map output of each layer is upsampled to construct a weighted loss function to continuously optimize the learning process of the neural network [16, 19–22]. The specific structure of the model is shown in Figure 4.

#### 5. Experimental Results and Analysis

To verify the algorithm's effectiveness in the study, a highway surveillance video for testing was selected for the experiment. The hardware environment of the experimental platform is an ordinary computer, with an I3 processor, main frequency 2.1 GHz, and memory 4G. The software environment is Visual Studio 2019 and OpenCV3 computer vision open source library. The programming language is C#. Gaussian modeling, improved three-frame difference method, and combination of Gaussian modeling and improved three-frame difference method were taken to the test. The effect diagram of the test is shown in Figure 5. For the improved U-Net model, this article uses python language, and the deep learning framework uses Keras.

In Figure 5, the original video sequence is in the first column. The result of the improved three-frame difference method is in the second column. The third column is the result obtained by Gaussian modeling. The last column is the motion obtained by combining the two methods. Figure 5(a)–5(c) are the detection effect diagrams of the same video at different periods.

In Figure 6, we compare the accuracy of our work against the conventional Gaussian model and improved three-frame different model. We compare the accuracy of these schemes for different time frames. This figure shows that our proposed work has much better accuracy as compared to the existing schemes.

Figure 7 shows the comparison experiment results of the average time complexity under different time frames. Our scheme outperforms the existing models, and it has much better results for varying number of frames.

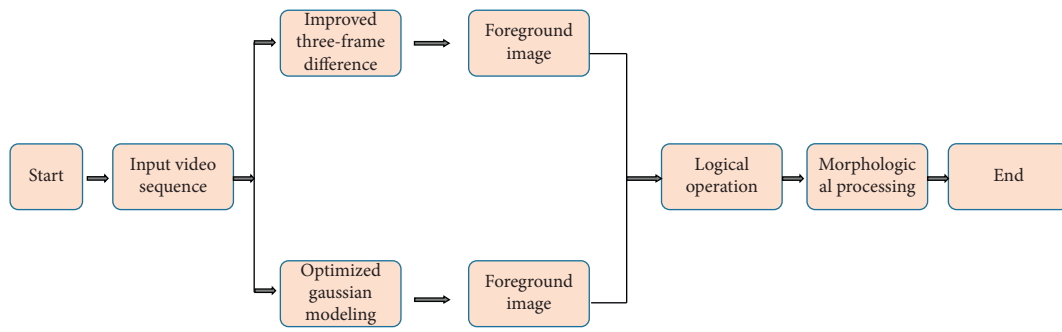


FIGURE 3: Flow chart of the algorithm.

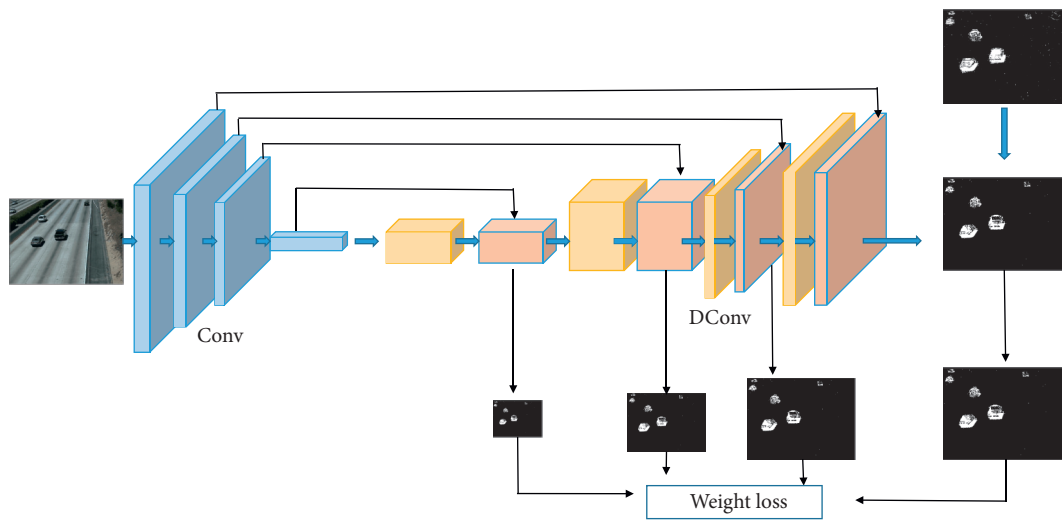


FIGURE 4: Improved U-Net architecture diagram.

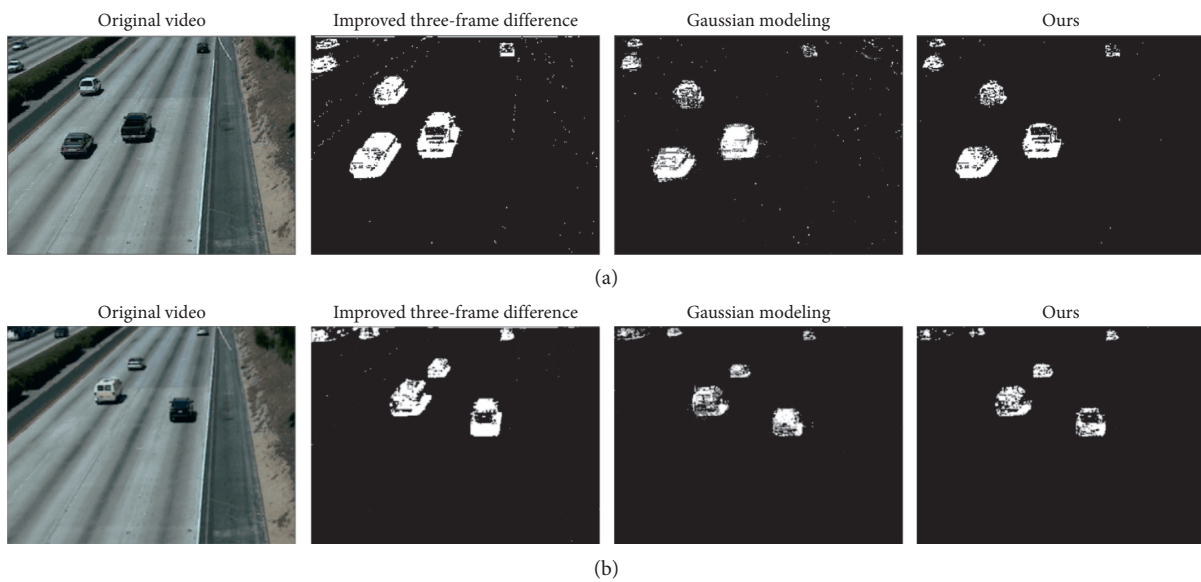


FIGURE 5: Continued.

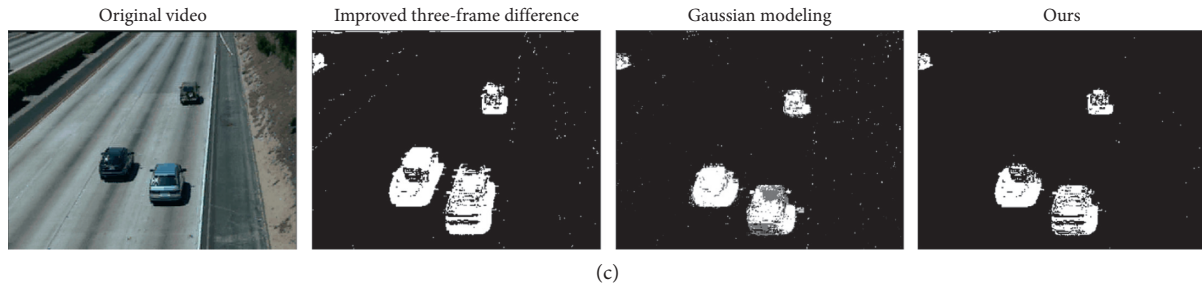


FIGURE 5: Effect diagram for three kinds of motion detection.

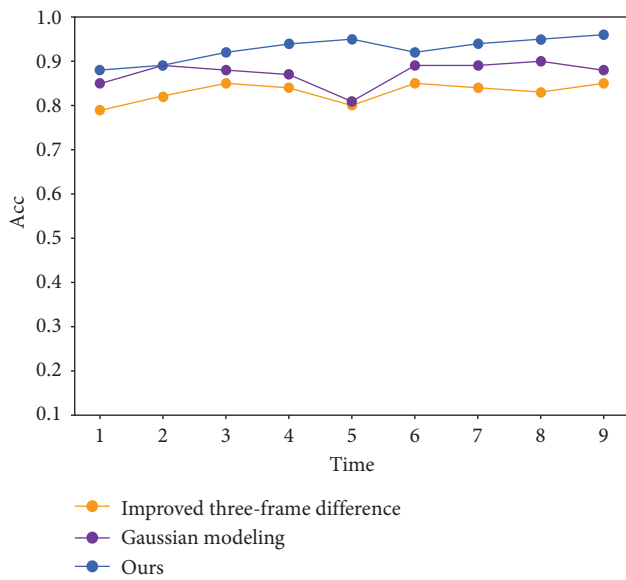


FIGURE 6: Accuracy comparison experiment results in different time frames.

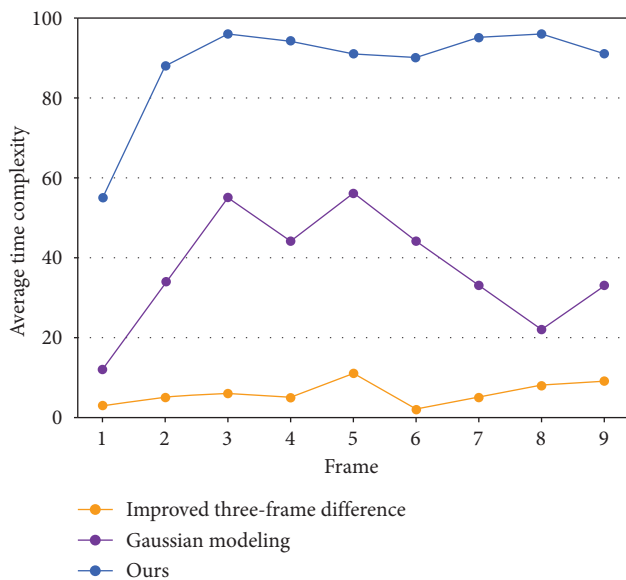


FIGURE 7: Comparison of experimental results of average time complexity in different time frames.

## 6. Conclusions

A motion detection algorithm based on Gaussian modeling and an improved three-frame difference method is proposed in the study. Adjacent frames of images are taken in the traditional three-frame difference method. Images of non-adjacent frames are taken in the improved algorithm, and “OR” operation is performed upon them. We use Gaussian modeling to adapt to environmental changes. After the two algorithms are tested, improved, and combined, the deformation, holes, noise, and other motion problems caused by the frame difference method can be effectively improved. It helps in raising the detection efficiency of motion detection. The antiinterference is enhanced, which can meet the requirements of motion detection, with certain practicability and value.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] H. Hu, *Video Motion Detection in Complex Traffic Scenes*, Anhui University, Hefei, China, 2020.
- [2] Y. Zhang, *Research on Video-Based Motion Detection Algorithm*, Shanghai Normal University, Shanghai, China, 2020.
- [3] L. Xiao, Y. Yang, and Y. Xu, “Motion detection by four-frame difference and improved Gaussian mixture model,” *Science Technology and Engineering*, vol. 20, no. 15, pp. 6141–6150, 2020.
- [4] Y. Jin, Y. Yang, and Y. Du, “Motion tracking algorithm based on optical flow feature points in power monitoring scenes,” *Electric Power Science and Engineering*, vol. 36, no. 5, pp. 40–47, 2020.
- [5] S. Sui, “Overview of pixel-based background subtraction technology[J],” *Science and Technology Innovation*, vol. 9, pp. 63–66, 2019.
- [6] Y. Mou and L. Jin, “Improved ViBe algorithm for motion detection[J],” *Information Technology and Informatization*, vol. 6, pp. 75–77, 2020.
- [7] H. Gao and X. Duan, “Motion detection at sea based on frame difference method and Gaussian mixture,” *Computer and Digital Engineering*, vol. 47, no. 5, pp. 1140–1144, 2019.

- [8] Q. Sun, X. Zhao, and X. Zhang, "Research on the detection method of small motion trajectory based on improved three-frame difference method," *Equipment Manufacturing Technology*, vol. 3, pp. 135–137, 2019.
- [9] W. Jiang, *Research on Conspicuous Video Motion Detection Algorithm Based on Biological Vision Mechanism*, University of Electronic Science and Technology of China, Chengdu, China, 2020.
- [10] R. Cao and H. Xiaoli, "Improved vibe algorithm based on Gaussian mixture model and threshold adaptation," *Journal of Taiyuan University of Technology*, vol. 51, no. 4, pp. 516–521, 2020.
- [11] C. Zhang, Y. Zhang, Y. Chen, Y. Wang, and D. Zhao, "Motion detection and recognition based on five-frame difference and template matching," *Laser and Infrared*, vol. 48, no. 12, pp. 1554–1560, 2018.
- [12] Y. Yang, B. Hui, Z. Chang, and Y. Ma, "Motion detection algorithm at sea based on Gaussian mixture model," *Application Research of Computers*, vol. 37, no. 1, pp. 310–313, 2020.
- [13] Z. Li and X. Tan, "Motion detection algorithm based on Gaussian mixture model and five-frame difference," *Computer and Digital Engineering*, vol. 46, no. 2, pp. 284–288, 2018.
- [14] L. Zhang, W. Li, X. Ning, X. Dong, and W. Liu, "A finger vein recognition method based on histogram of oriented lines and (2D)2FPCA," *Journal of Computer-Aided Design & Computer Graphics*, vol. 30, no. 2, pp. 254–261, 2018.
- [15] X. Zhang, Y. Yang, Z. Li, X. Ning, Y. Qin, and W. Cai, "An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field," *Entropy*, vol. 23, no. 4, p. 435, 2021.
- [16] S. Li, X. Ning, L. Yu et al., "Multi-angle head pose classification when wearing the mask for face recognition under the COVID-19 coronavirus epidemic," in *Proceedings of the 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, pp. 1–5, Shenzhen, China, May 2020.
- [17] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, 2020.
- [18] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, "A biomimetic covering learning method based on principle of homology continuity," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 8–15, 2021, <https://www.aspress.org/index.php/tpris/article/view/9>.
- [19] W. Cai, B. Liu, Z. Wei, M. Li, and J. Kan, "TARDB-Net: triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 11291–11312, 2021.
- [20] J. Chen, C. Du, Y. Zhang, P. Han, and W. Wei, "A clustering-based coverage path planning method for autonomous heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 1–11, 2021.
- [21] Z. L. Yang, S. Y. Zhang, Y. T. Hu, Z. W. Hu, and Y. F. Huang, "VAE-Stega: linguistic steganography based on variational auto-encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880–895, 2020.
- [22] X. Ning, X. Wang, S. Xu et al., "A review of research on co-training, concurrency and computation: practice and experience," , 2021.