

## Key Frame Extraction of Surveillance Video Based on Frequency Domain Analysis

Yunzuo Zhang<sup>1,\*</sup>, Shasha Zhang<sup>1</sup>, Jiayu Zhang<sup>1</sup>, Kaina Guo<sup>1</sup> and Zhaoquan Cai<sup>2</sup>

<sup>1</sup>School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, 050043, China

<sup>2</sup>Department of Computer Science and Engineering, Huizhou University, Huizhou, 516007, China

\*Corresponding Author: Yunzuo Zhang. Email: zhangyunzuo888@sina.com

Received: 23 January 2021; Accepted: 25 March 2021

**Abstract:** Video key frame extraction, reputed as an essential step in video analysis and content-based video retrieval, and meanwhile, also serves as the basis and premise of generating video synopsis. Video key frame extraction means extracting the meaningful parts of the video by analyzing their content and structure to form a concise and semantically expressive summary. Up to now, people have achieved many research results in key frame extraction. Nevertheless, because the surveillance video has no specific structure, such as news, sports games, and other videos, it is not accurate enough to directly extract the key frame with the existing effective key frame extraction method. Hence, based on frequency domain analysis, this paper proposed a key frame extraction method for surveillance video, which obtains the frequency spectrum and phase spectrum by performing Fourier transform on the surveillance video frames. Using the frequency domain information of two adjacent frames can accurately reflect the global motion state changes and local motion state changes of the moving target. Experimental results show that the proposed method is correct and effective, and the extracted key frames can more accurately capture the changes in the global and local motion states of the target compared with the previous methods.

**Keywords:** Frequency domain analysis; frequency spectrum; Fourier transform; local motion states; key frame extraction

### 1 Introduction

In many fields, such as the national economic construction and public security information construction, surveillance video plays an increasingly important role. The infrastructure of the video surveillance system has begun to take shape and is still developing rapidly. With thousands of surveillance cameras monitoring and recording round the clock, the amount of video data has exploded, so finding the required information in many surveillance videos is undoubtedly a needle in a haystack [1–6]. For this reason, we need to express this information concisely and clearly, which involves the key frame extraction method.

Key frame extraction is an essential step in video analysis and content-based video retrieval. Simultaneously, it is also the basis and premise for generating video synopsis [7–9]. Video key frame



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

extraction is to extract the meaningful parts of the video by analyzing their content and structure, thereby forming a concise summary capable of expressing semantics to eliminate redundancy, shorten the video time, and improve the browsing and query efficiency. At present, the technology of extracting key frames in surveillance video has made a significant breakthrough [10–18]. Nevertheless, applying the existing effective video key frame extraction directly to surveillance video will result in inaccurate keyframe extraction. Unlike news, sports programs, or other videos with structured information, since surveillance video does not have a specific structure, many algorithms for extracting key frames are not suitable for surveillance video [19–21].

Therefore, this paper proposed a key frame extraction method with surveillance video based on frequency domain analysis [22–29]. The method can perform frequency domain analysis on the video frame to obtain its frequency spectrum and phase spectrum. Furthermore, the changes in the frequency domain information of adjacent video frames, which means the changes in the frequency spectrum and phase spectrum of the video frame can reflect the changes in the global and local motion status of moving targets in the video. Using the change of the frequency domain information of two adjacent frames can accurately reflect the change of the local motion state of the target in the video to achieve the purpose of accurately extracting key frames.

## 2 Frequency Domain Analysis

The frequency-domain processing of an image is to convert the image into a frequency domain and then modify and delete the frequency domain information to optimize the image quality. Two-dimensional frequency domain processing provides a new perspective for target recognition and detection of video images. These problems, which are backbreaking to solve in the airspace, may be relatively easier to solve by the transformation domain. The method can provide fresh ideas to deal with the problem [30,31]. The image frequency domain transforms methods mainly include: Fourier transform, discrete cosine transform, two-dimensional orthogonal transform, and wavelet transform. The following will introduce the two-dimensional Fourier transform involved in this paper and analyze its feasibility.

### 2.1 Two-Dimensional Discrete Fourier Transform

In general, the formula of the two-dimensional Fourier transform is:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy \quad (1)$$

The inverse formula is:

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{j2\pi(ux+vy)} du dv \quad (2)$$

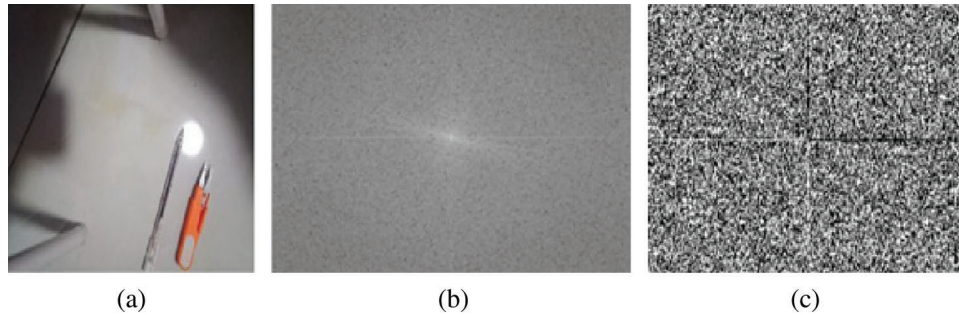
The above formula can verify that the Fourier transform has separation, periodicity, conjugate symmetry, linearity, rotation, and proportionality.

For an image of size  $m \times n$ , the equation of the discrete Fourier transform is as follow:

$$F(u, v) = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x, y) e^{-j2\pi(ux/m+vy/n)} \quad (3)$$

Among them, use variables  $u$  and  $v$  to define its frequency. The frequency-domain system is a coordinate system formed by  $F(u, v)$ , and the frequency variable is the sum. The space domain is a coordinate system defined by  $f(x, y)$ . The spectral components, along with the four corners and directions of the spectrogram,

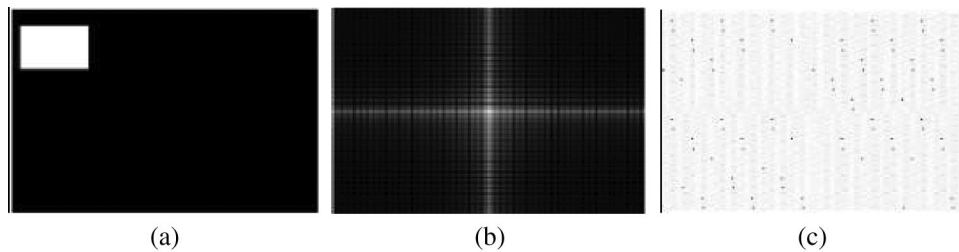
are all zero. The following figure is an example; figures (b) and (c) in Fig. 1 show the frequency spectrum and phase spectrum of the original image, respectively.



**Figure 1:** The Frequency Spectrum and Phase Spectrum of the Image (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum

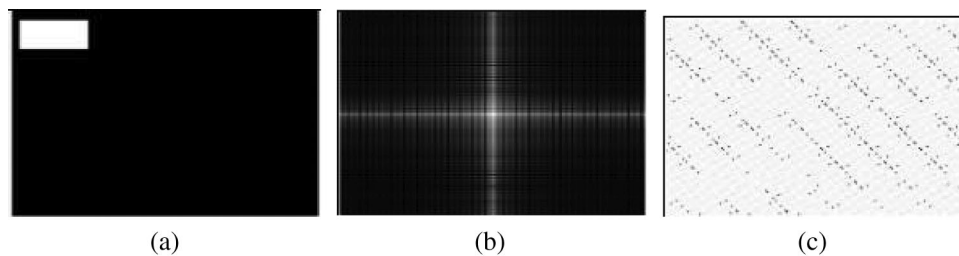
**2.2 Feasibility Analysis**

After the two-dimensional Fourier transform, the image will obtain its corresponding frequency spectrum and phase spectrum. Fig. 2 shows the frequency spectrum and phase spectrum of a rectangular.



**Figure 2:** The Frequency Spectrum and Phase Spectrum of the Rectangular (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum

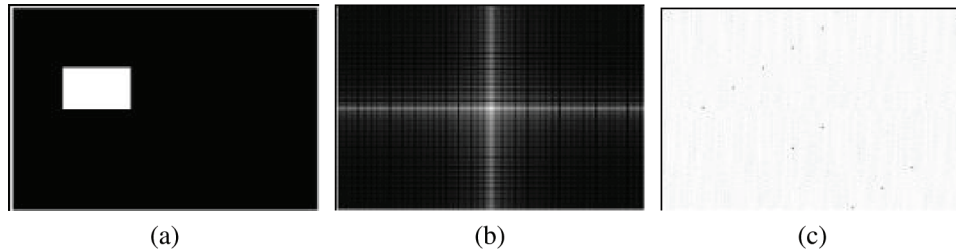
Fig. 3 shows the frequency spectrum and phase spectrum of the rectangular after changing the image shape.



**Figure 3:** The Frequency Spectrum and Phase Spectrum of the Rectangular (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum

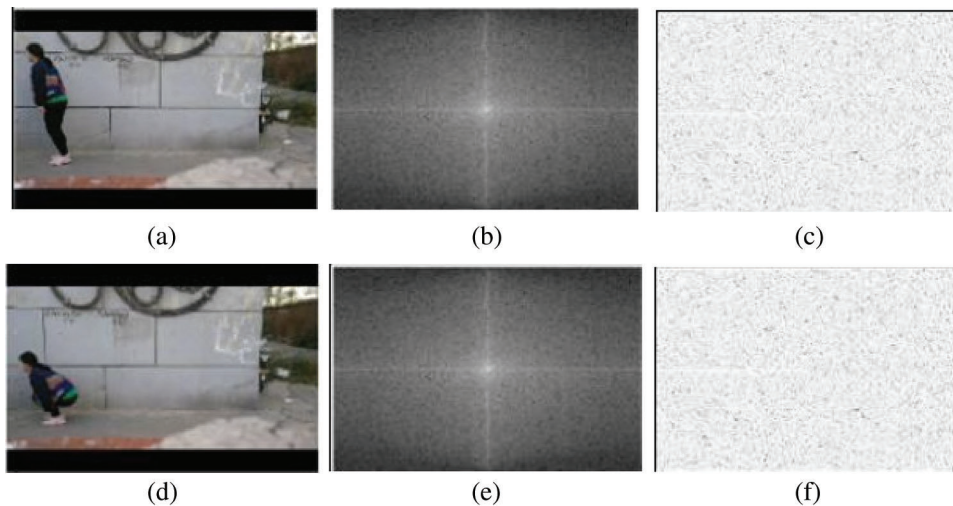
Comparing Figs. 2 and 3, we can find that when the rectangle changes, although the shape of the frequency spectrum and phase spectrum does not change significantly, the value of the graphic matrix has changed.

According to the Fourier transform property, when the rectangle in the image has rotated, its frequency spectrum and phase spectrum will change. When the rectangle in the image has translated, the frequency spectrum will not change, but the phase spectrum will, as shown in Fig. 4.



**Figure 4:** The Frequency Spectrum and Phase Spectrum of the Rectangular (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum

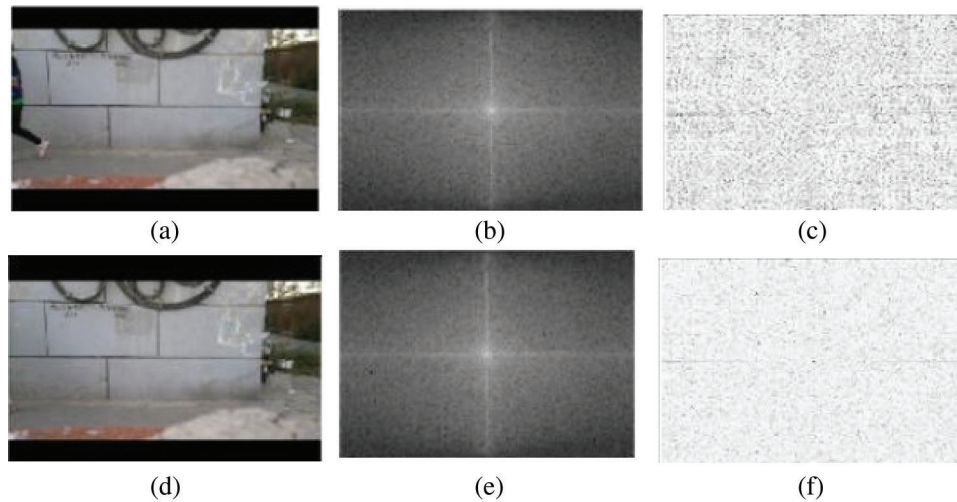
Take the moving target as an example; in the actual video, when the moving target goes from upright to squatting, the frequency spectrum and phase spectrum of the two video frames are shown in Fig. 5.



**Figure 5:** The Frequency Spectrum and Phase Spectrum of Two Video Frames (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum (d) Original Image (e) Frequency Spectrum (f) Phase Spectrum

Fig. 5 shows that after the target squats, the frequency spectrum and phase spectrum have changed. Compared with the two frequency spectra that the shape does not have many changes; phase spectra changes can be observed directly from the image.

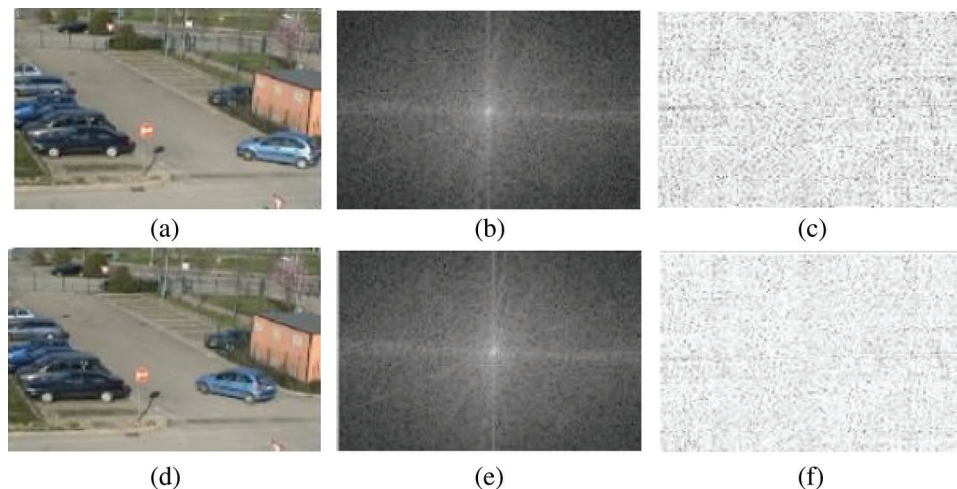
When the target appears or disappears, the frequency spectrum and phase spectrum will change. Take the target disappearing as an example; Fig. 6 shows the frequency spectrum and phase spectrum of the two video frames.



**Figure 6:** The Frequency Spectrum and Phase Spectrum of Two Video Frames (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum (d) Original Image (e) Frequency Spectrum (f) Phase Spectrum

Compare the frequency spectrum and phase spectrum of the two video frames in Fig. 6, the difference between the two video frames appeared. The shape of the frequency spectrum has not changed significantly, but the data have. Comparing the frequency spectrum and phase spectrum of the four video frames in Figs. 6 and 5, almost all target movements will cause changes in the frequency spectrum and phase spectrum. Unless the target keeps unchanged except translation, at this time, the frequency spectrum does not change, but the phase spectrum will also change.

During the target movement, the moving target is making a turn. Fig. 7 shows the frequency spectrum and phase spectrum of the two video frames.



**Figure 7:** The Frequency Spectrum and Phase Spectrum of Two Video Frames (a) Original Image (b) Frequency Spectrum (c) Phase Spectrum (d) Original Image (e) Frequency Spectrum (f) Phase Spectrum

Observing Fig. 7, we can find that the frequency spectrum and phase spectrum have changed. Compared with the frequency spectra, the phase spectra vary more obviously.

According to the above analysis, when the moving target in the video changes from static to motion or from motion to static, stretching, squatting, turning, the global motion states and the local motion states have changed, the frequency spectrum and phase spectrum also will change. Meanwhile, when there are multiple targets in the videos, changes in any one target will also cause changes in the frequency spectrum or phase spectrum. Therefore, it is feasible to use the frequency domain information of adjacent video frames to reflect the changes in the target motion state in the surveillance video. Based on this, the paper proposes a key frame extraction based on frequency domain analysis.

### 3 Algorithm Description

From the perspective of frequency domain analysis, this paper uses the frequency domain information change to reflect the change of the target motion state. By analyzing the principle of the two-dimensional discrete Fourier transform, we can find that it is reasonable to use the frequency domain information to measure the target motion change. Fig. 8 shows the basic framework of the proposed key frame extraction method.

From Fig. 8, we can find that the following steps of the input surveillance video sequence are as follow:

Step1: Image preprocessing. Perform grayscale processing on the surveillance video sequence.

Step2: Fourier transform. For surveillance video frame  $f(x, y)$ , perform a two-dimensional discrete Fourier transform by formula (4) to obtain  $F(u, v)$ :

$$F(u, v) = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x, y) e^{-j2\pi(ux/m+vy/n)} \quad (4)$$

Step3: Frequency spectrum and phase spectrum. According to  $F(u, v)$ , the Fourier frequency spectrum  $|F(u, v)|$  is as follow.

$$|F(u, v)| = [R^2(u, v) + I^2(u, v)]^{1/2} \quad (5)$$

In the formula,  $R(u, v)$  and  $I(u, v)$  denote the real and imaginary parts of  $F(u, v)$ , respectively. According to the  $F(u, v)$ , the phase angle  $\varphi(u, v)$  is as follow.

$$\varphi(u, v) = \arctan \left[ \frac{I(u, v)}{R(u, v)} \right] \quad (6)$$

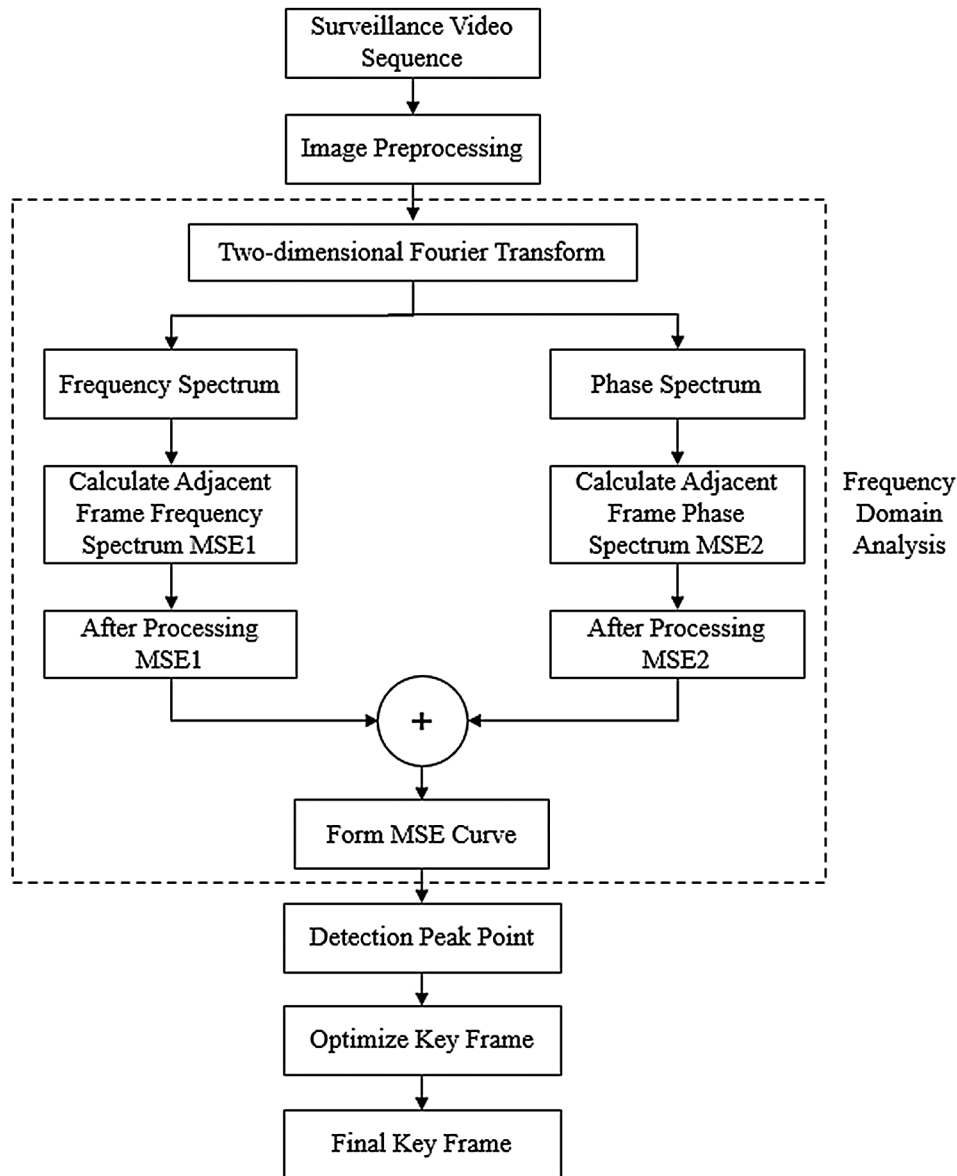
Step4: Calculate the mean square error MSE1 and MSE2, respectively. According to the obtained frequency spectrum and phase spectrum, we can calculate the mean square error MSE1 of the adjacent phase spectrum and the mean square error MSE2 of the adjacent frame frequency spectrum, respectively. Assuming the width and height of the video frame of the video sequence is  $W$  and  $H$ , respectively, the current frame spectrum is  $f(x, y)$ , and the previous frame spectrum is  $b(x, y)$ , then their mean square error MSE1 is defined as:

$$MSE1 = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |f(x, y) - b(x, y)|^2 \quad (7)$$

Use the formula (7) to calculate MSE1 and MSE2.

Step 5: This step is to weight MSE1 and MSE2. To make the mean square error change more evident than before, we expended MSE1 by five times. The frequency spectrum contains the grey information, and the phase spectrum contains the information of the edge and the overall structure of the original image.

Compared with the frequency spectrum, the phase spectrum includes more visual information, containing more critical information [32,33]. Because the phase spectrum is meaningful, we expended MSE2 by a factor of 10. Due to weighting, the curve will have a sudden shift.



**Figure 8:** The Basic Framework of Video Key Frame Extraction Method

Step6: Form the MSE curve. According to the above steps, the calculation formula of the mean square error MSE of the frame is:

$$MSE = 5MSE1 + 10MSE2 \tag{8}$$

Calculate the MSE according to the formula (8) to form the MSE curve.

Step 7: Detect the peak point of the curve. The formed MSE curve detects the curve's peak point and extracts the video frame corresponding to the curve peak and the first frame and the last frame as candidate key frames.

Step 8: Extract the final key frame. Firstly, to reduce the redundancy of extracting key frames, extract the video frames where the peak changes, that is, extract the video frames whose current frame peak is  $N$  times the previous key frame peak (starting from the second peak). We set the MSE of the first frame of the video to 0, so the second frame must change suddenly. Because the first and second frames will be redundant, start from the second peak. The video frame extracted comprises the video frame at the peak mutation and the first and last frames. Then the extracted key frames are optimized using the key frame optimization criterion based on the peak signal-to-noise ratio visual discrimination mechanism [33,34], and finally, the final key frame is determined.

## 4 Experimental Results and Analysis

To comprehensively evaluate the correctness and effectiveness of the frequency domain-based method, we have conducted many experiments on public video sets and self-collected videos. In the following section, 4.1 extracts key frames from two typical videos and analyzes the extraction result to verify the correctness of the proposed key frame algorithm. 4.2 confirms the effectiveness of the algorithm proposed in this section.

### 4.1 Algorithm Correctness

By observing the experimental results, we can find that using the frequency domain information can accurately capture the changes in the global and local motion states of the target. This section analyzes the experimental results by verifying the correctness of the key frame extraction algorithm based on frequency domain analysis.

The moving target in the test video begins with the target appearing, then crouching, and ends with the target standing up. The changes in the motion state of the target in the video include changes in the global motion state, such as the appearance of the target, acceleration, deceleration, changes in local motion states during walking, and changes in topical motion states during walking. Furthermore, other local changes in motion status are incorporated, such as squatting down while walking and putting the bag in hand on the ground. In this experiment, we set the experimental parameter  $N = 1$ . Fig. 9 shows the key frame extraction result except for the first and last frames.

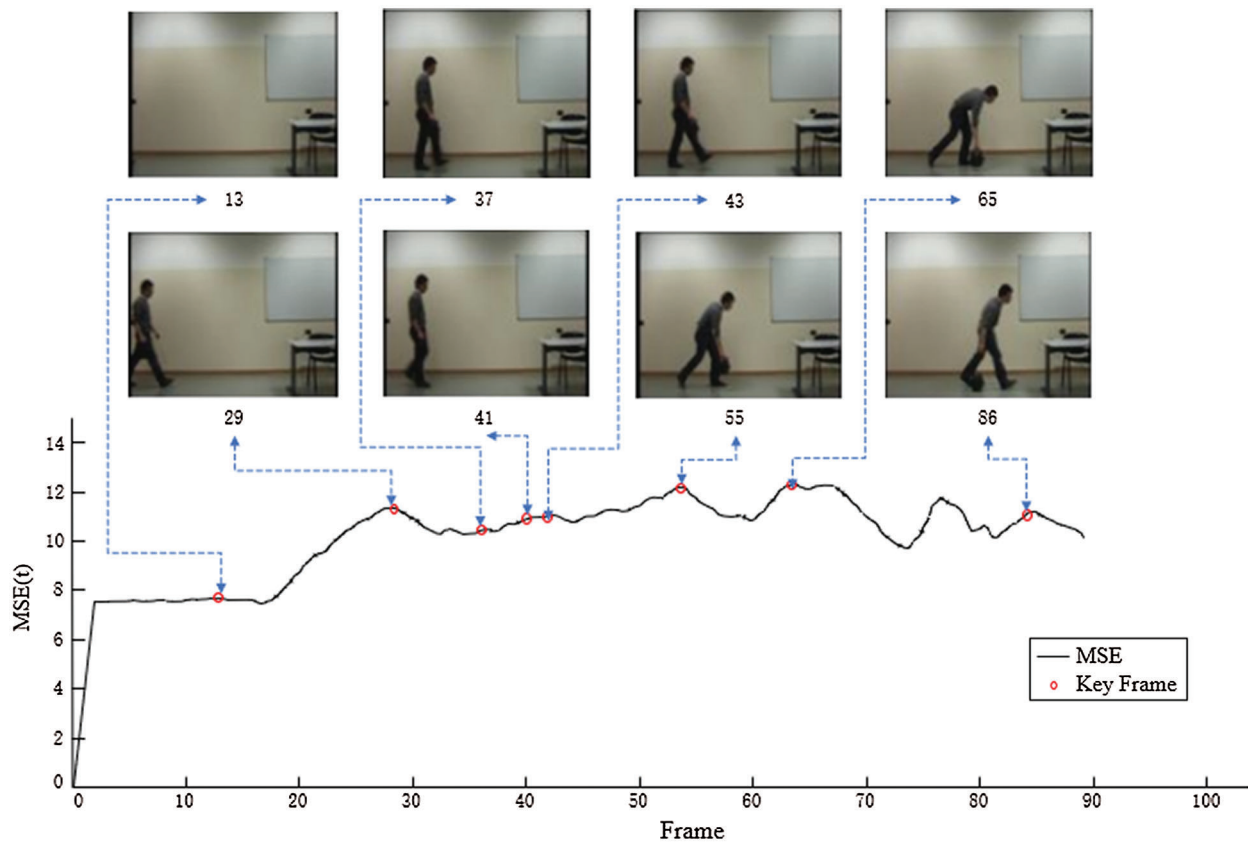
By observing the key frame extraction results, we can find that the proposed method extracts the 13<sup>th</sup> frame of the video, in which the target's feet begin to enter the surveillance scene. In the 29<sup>th</sup> frame, the movement of the target's leg changes more significantly. The 37<sup>th</sup>, 41<sup>st</sup>, and 43<sup>rd</sup> frames represent the change in the target's movement speed as it travels. The 55<sup>th</sup>, 65<sup>th</sup>, and 86<sup>th</sup> frames can describe the changes in the local state of the objective as it squats down and stands.

To further verify the correctness, this experiment selects other video clips for key frame extraction. Fig. 10 shows the fragment video sequence. In this video, between two targets who are fighting, one is lying on the ground, and the other starts to run away, and the third moving target starts to appear in the process of running away. In this experiment, we set the experimental parameter  $N = 1.11$ , and Fig. 10 shows the key frame extraction result of the test video except for the first and last frames.

Observing Fig. 10, we can find that this method will miss some video frames with more prominent peaks due to parameter  $N$  and the optimization criteria and will extract video frames with smaller crest values. Nevertheless, the final key frame determined can capture the change of the target motion state. The key frames extracted from the 3<sup>rd</sup> frame to the 75<sup>th</sup> frame describe the movement process of the two targets



exchanging positions before the battle, that is, the change of the global moving target. From frame 87 to frame 135, the four frames describe the changes in the target legs during the battle, which means the local motion state's variation. In frame 171, the white target knocked down the black target and fled. At this time, a third moving target appears. The four key frames extracted subsequently describe the global motion state changes of the second and third target's movement direction and speed.



**Figure 9:** The Key Frame Extraction Result of Test Video by the Proposed Method

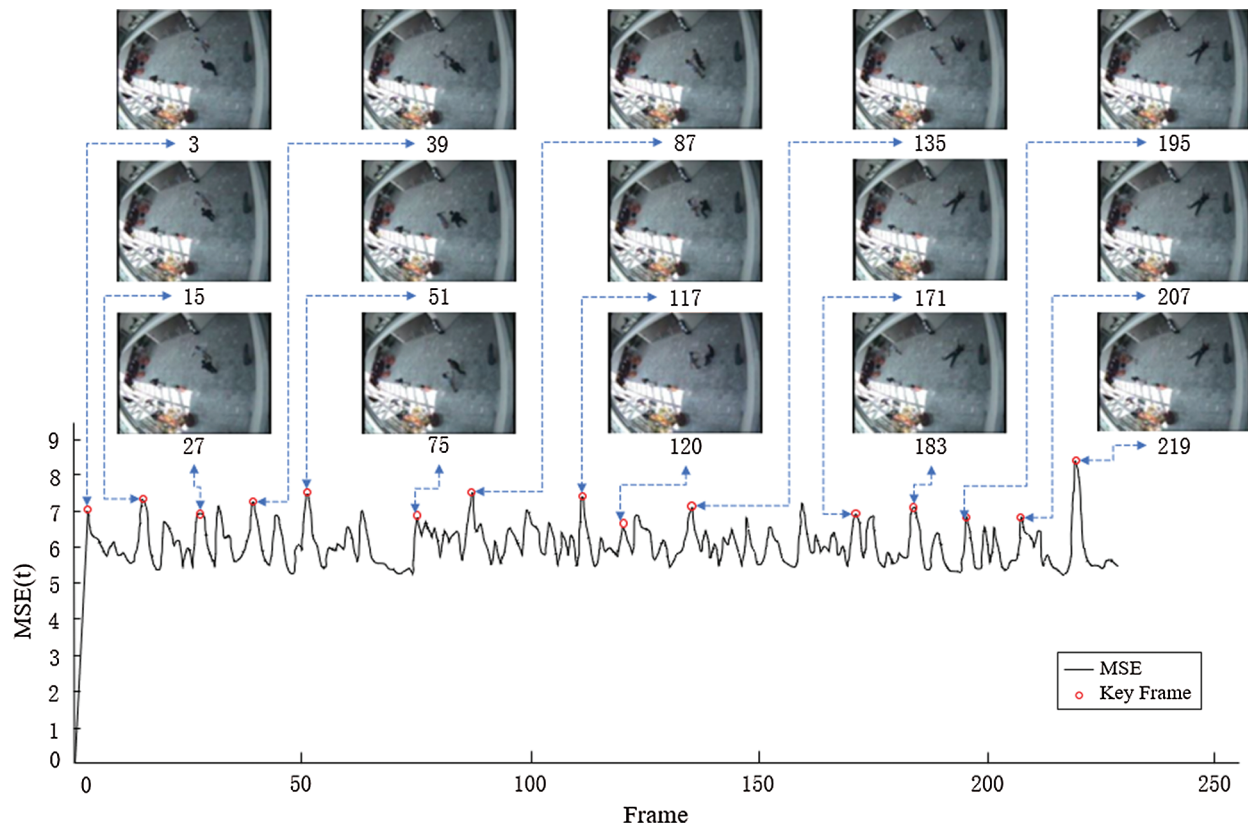
Based on the proposed method analysis, the key frame extraction can be summarized as follows: The changes in the global and local motion state of the target captured by the frequency domain information can more accurately extract the local motion state change of the target. What is more, the correctness of the key frame extraction method based on the frequency domain analysis proposed is verified.

#### 4.2 Algorithm Effectiveness

This paper compares the four key frame extraction methods: MA, MTSS, ME, and the center offset-based method (denoted as CO) to verify the effectiveness of the proposed method. We set the experimental parameter to  $N = 1$  and set the center offset-based method's experimental parameter to  $N = 5$ . We evaluate the extraction results of five key frame extraction methods from both subjective and objective aspects.

The following is a verification of the effectiveness of the proposed method from a subjective perspective. The number of key frames extracted from the same video by different methods is the same during the

experiment to ensure objectivity. There are six key frames for the test video except for the first and last frames of the video. Fig. 11 shows the key frame extraction result of the proposed method of test video.



**Figure 10:** The Key Frame Extraction Result of Test Video by the Proposed Method

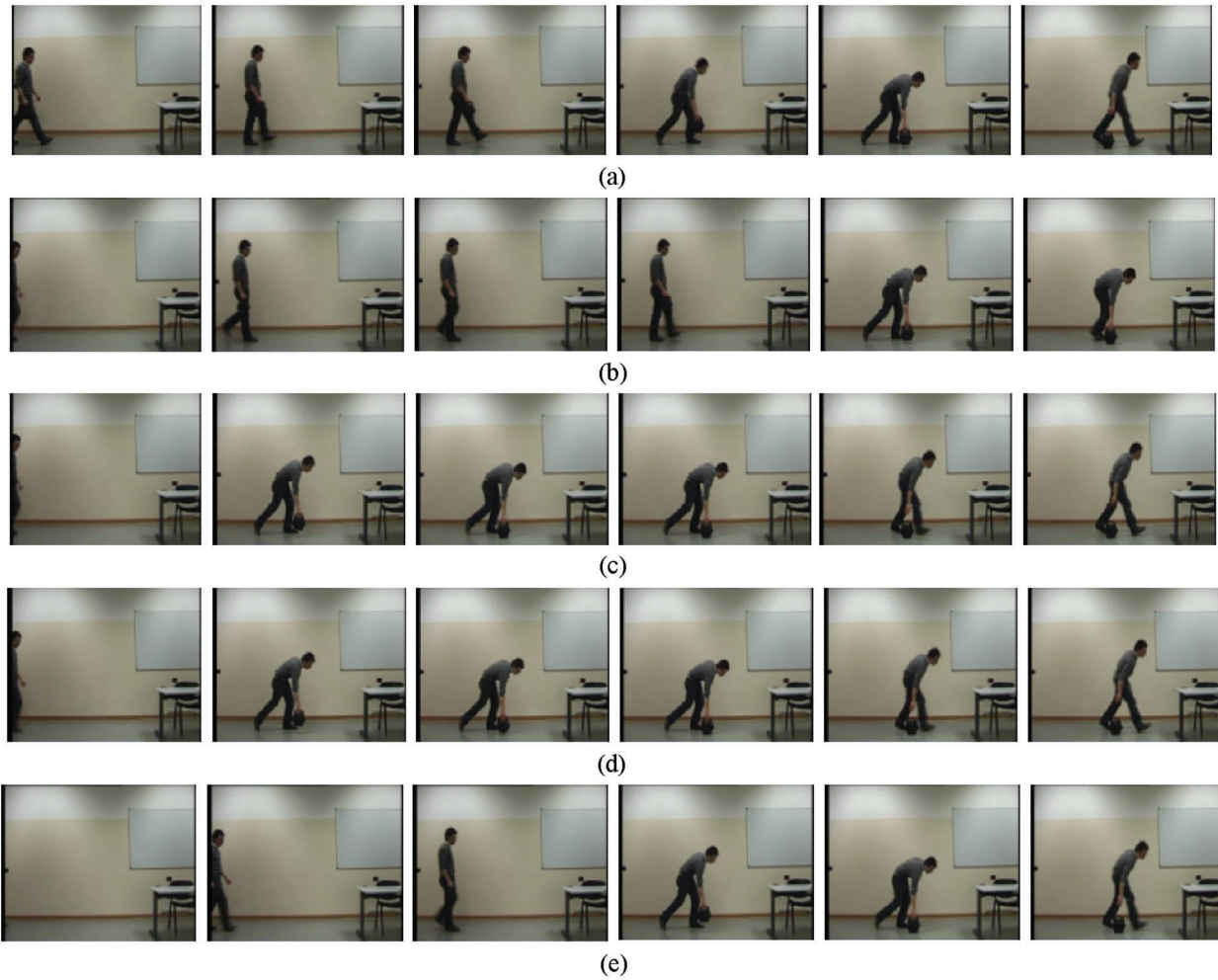
Observing Fig. 11, the five methods can extract the process of target entry and exit and put the bag. CO, MTSS, MA can all detect the appearance of the target in a relatively timely manner. However, compared with CO and other methods, the method proposed in this chapter is more sensitive to changes in the local movement of the target leg than other methods. Furthermore, it can clearly describe the movement state changes of the target in putting the bag.

Use SRD criteria to evaluate the effectiveness of the proposed method objectively. Fig. 12 shows the SRD of the proposed method and five key frame extraction methods of CO, MTSS, MA, and ME on test video.

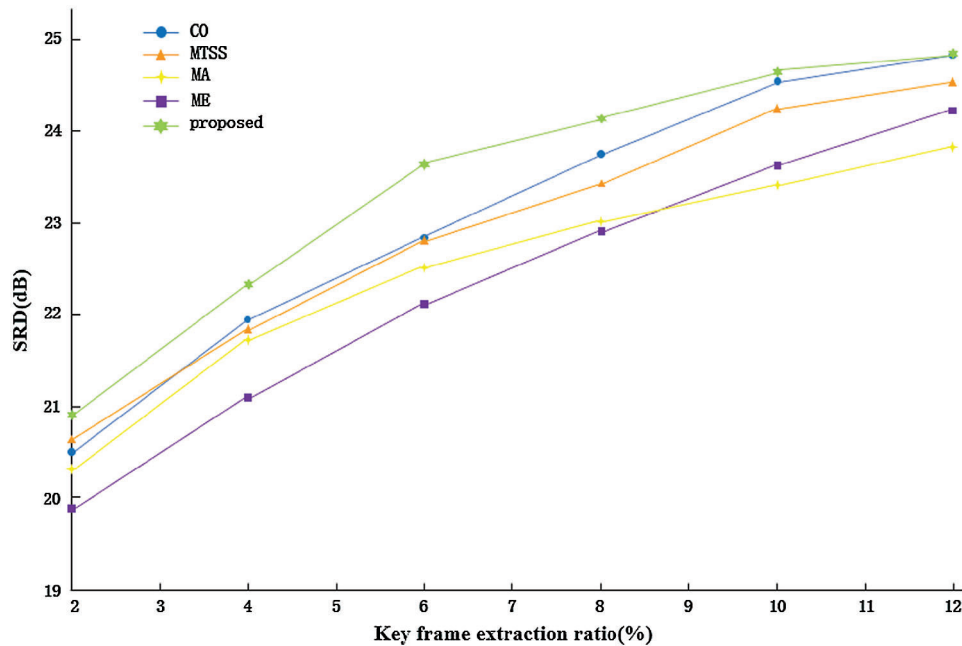
By observing the average SRD of different methods, we can find that when the key frame extraction ratio is 10%, the average SRD of the proposed method is about 0.5dB higher than that of CO. When the key frame extraction rate is 12%, the average SRD of the method proposed in this paper is that same as that of CO. In terms of overall trends, the method proposed in this paper has a higher advantage than the three methods of MA, MTSS, and ME. Therefore, as far as SRD is concerned, compared with other comparison methods, the method based on frequency domain analysis has more significant advantages, meanwhile, the ability to reconstruct video frames is more vital.

The correctness of the proposed key frame algorithm is verified by analyzing the results of extracting key frames from the test video. Meanwhile, the subjective and objective performance of the method

proposed in this paper and other motion-related key frame extraction methods are compared on the test video to verify the effectiveness of the method. The experimental results analysis shows that the method based on frequency domain analysis captures the changes in the local motion state of the target more accurately than other methods. This method realizes the precise extraction of key frames of the surveillance video.



**Figure 11:** Test Video Key Frame Extraction Result (a) Proposed Method (b) CO (c) MTSS (d) MA (e) ME



**Figure 12:** Average SRD of Different Method

## 5 Conclusion

The paper proposed a key frame extraction method based on frequency domain analysis to capture the changes in the local motion state of the target accurately. This method first performs a two-dimensional Fourier transform on the video frame to obtain the frequency spectrum and phase spectrum of the video frame. Then calculate the mean square error of the frequency spectrum and phase spectrum of the current frame and the previous frame, and process them separately. After processing, add the two to get the mean square error of the current frame, forming a mean square error curve. Then extract the candidate key frames and optimize them to obtain the final key frame. Experimental results show that the proposed method is correct and effective, and the extracted key frames can more accurately capture the changes in the target's local motion state.

**Funding Statement:** This work was supported by the National Nature Science Foundation of China (Grant Nos. 61702347, 61772225), Natural Science Foundation of Hebei Province (Grant Nos. F2017210161, F2018210148).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] S. E. Avila, A. B. Lopes, L. J. Antonio and A. A. Arnaldo, "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 56–68, 2011.
- [2] P. Jiang and X. Qin, "Key frame based on video summary using visual attention clues," *IEEE Trans. Multimedia*, vol. 17, no. 2, pp. 64–73, 2010.
- [3] C. R. Huang, P. C. Chung, D. K. Yang, H. C. Chen and G. J. Huang, "Maximum a posteriori probability estimation for online surveillance video synopsis," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 24, no. 8, pp. 1417–1429, 2014.

- [4] R. Zhong, R. Hu, Z. Wang and S. Wang, "Fast synopsis for moving objects using compressed video," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 834–838, 2014.
- [5] Y. Nie, C. Xiao, H. Sun and P. Li, "Compact video synopsis via global spatiotemporal optimization," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, no. 10, pp. 1664–1676, 2013.
- [6] W. Hu, N. Xie, X. Z. L. Li and S. Maybank, "A survey on visual content-based video indexing and retrieval," *IEEE Transactions on Systems*, vol. 41, no. 6, pp. 797–819, 2011.
- [7] N. Ejaz, I. Mehmood and S. W. Baik, "Efficient visual attention-based framework for extracting key frames from videos," *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 34–44, 2013.
- [8] L. J. Lai and Y. Yi, "Key frame extraction based on visual attention model," *Journal of Visual Communication and Image Representation*, vol. 23, no. 1, pp. 114–125, 2012.
- [9] S. Luo, S. J. Ma, J. Liang and L. M. Pan, "Method of key frame extraction based on sub-shot cluster," *IEEE Int. Conf. on Progress in Informatics and Computing*, vol. 31, no. 3, pp. 348–352, 2011.
- [10] H. Zeng and H. H. Yang, "A two phase video keyframe extraction method," *Computer and Modernization*, no. 6, pp. 33–35, 2011.
- [11] K. W. Sze, K. M. Lam and G. P. Qiu, "A new key frame representation for video segment retrieval," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 9, pp. 1148–1155, 2005.
- [12] W. Wolf, "Key frame selection by motion analysis," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Atlanta, GA, Vol. 2, pp. 1228–1231, 1996.
- [13] T. Liu, H. J. Zhang and F. Qi, "A novel video key frame extraction algorithm based on perceived motion energy model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 10, pp. 1006–1013, 2003.
- [14] T. Y. Liu, X. D. Zhang, J. Feng and K. T. Lo, "Shot reconstruction degree: A novel criterion for key frame selection," *Pattern Recognition Letters*, vol. 25, no. 1, pp. 1451–1457, 2004.
- [15] Y. Z. Ma, Y. L. Chang and H. Yuan, "Key-frame extraction based on motion acceleration," *Optical Engineering*, vol. 47, no. 9, pp. 1–3, 2008.
- [16] C. Li, Y. T. Wu, S. S. Yu and T. Chen, "Motion-focusing key frame extraction and video summarization for lane surveillance system," in *Proc. 16th IEEE ICIP*, Cairo, pp. 4329–4332, 2009.
- [17] T. Wang, Y. Wu and L. Chen, "An approach to video key-frame extraction based on rough set," in *2007 Int. Conf. on Multimedia and Ubiquitous Engineering*, Seoul, pp. 590–596, 2007.
- [18] J. M. Zhang and X. J. Jiang, "Key frame extraction based on particle swarm optimization," *Journal of Computer Applications*, vol. 31, no. 2, pp. 358–361, 2011.
- [19] S. C. Raikwar, C. Bhatnagar and A. S. Jalal, "A framework for key frame extraction from surveillance video," *Proc. 5th IEEE ICCCT*, Allahabad, pp. 297–300, 2014.
- [20] Z. Cernekova, I. Pitas and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 82–91, 2006.
- [21] Y. Zhang, R. Tao and Y. Wang, "Motion state adaptive video summarization via spatiotemporal analysis," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1340–1352, 2017.
- [22] X. L. Fan, J. J. Zhang, Z. H. Jia and X. S. Cai, "Estimation of water droplet movement direction based on frequency-domain analysis," *China Powder Technology*, vol. 15, no. 4, pp. 63–65, 2009.
- [23] Q. R. Zhang, G. C. Gu, H. B. Liu and H. M. Xiao, "Salient region detection using multi-scale analysis in the frequency domain," *Journal of Harbin Engineering University*, vol. 31, no. 3, pp. 361–365, 2010.
- [24] X. D. Yang, "Location algorithm for image matching based on frequency domain analysis," *Communication power technology*, vol. 29, no. 4, pp. 20–22+43, 2012.
- [25] R. Y. Chen, L. L. Pan, Y. Zhou and Q. H. Lei, "Image retrieval based on deep feature extraction and reduction with improved CNN and PCA," *Journal of Information Hiding and Privacy Protection*, vol. 2, no. 2, pp. 67–76, 2020.
- [26] J. Niu, Y. Jiang, Y. Fu, T. Zhang and N. Masini, "Image deblurring of video surveillance system in rainy environment," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 807–816, 2020.
- [27] A. Gumaei, M. Al-Rakhami and H. AlSalman, "DI-har: Deep learning-based human activity recognition framework for edge computing," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1033–1057, 2020.

- [28] W. Song, J. Yu, X. Zhao and A. Wang, "Research on action recognition and content analysis in videos based on DNN and MLN," *Computers, Materials & Continua*, vol. 61, no. 3, pp. 1189–1204, 2019.
- [29] C. Zhu, Y. K. Wang, D. B. Pu, M. Qi, H. Sun *et al.*, "Multi-modality video representation for action recognition," *Journal on Big Data*, vol. 2, no. 3, pp. 95–104, 2020.
- [30] F. Dou, "The research of traffic flow detection system based on image frequency spectrum analysis," M.S. dissertation. China University of Petroleum, China, 2017.
- [31] Z. S. Tang, "Image quality assessment based on deep learning, spatial and frequency domain analysis," M.S. dissertation. Xi'an University of Technology, China, 2019.
- [32] D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. Hoboken, NJ: Wiley, 1998.
- [33] Gonzalez, "Digital Image Processing," Beijing: Electronic Industry Press, 2005.
- [34] G. Zhang, H. Sun, Y. Zheng, G. Xia, L. Feng *et al.*, "Optimal discriminative projection for sparse representation-based classification via bilevel optimization," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1065–1077, 2020.