



Video Compressed Sensing Reconstruction Based on Multi-Dimensional Reference Frame Multi Hypothesis Rediction

Hua Li^{1,*}, Yuchen Yue² and Jianhua Luo³

¹Army Armored Force Academy, Weapon and Control Department, Beijing, 100072, China

²Academy of Military Sciences, Beijing, 100091, China

³Army Armored Force Academy, Drill Training Center, Beijing, 100072, China

*Corresponding Author: Hua Li. Email: iuanne@163.com

Received: 24 January 2022; Accepted: 06 March 2023

Abstract: In this paper, a video compressed sensing reconstruction algorithm based on multidimensional reference frames is proposed using the sparse characteristics of video signals in different sparse representation domains. First, the overall structure of the proposed video compressed sensing algorithm is introduced in this paper. The paper adopts a multi-reference frame bidirectional prediction hypothesis optimization algorithm. Then, the paper proposes a reconstruction method for CS frames at the re-decoding end. In addition to using key frames of each GOP reconstructed in the time domain as reference frames for reconstructing CS frames, half-pixel reference frames and scaled reference frames in the pixel domain are also used as CS frames. Reference frames of CS frames are used to obtain higher quality assumptions. The method of obtaining reference frames in the pixel domain is also discussed in detail in this paper. Finally, the reconstruction algorithm proposed in this paper is compared with video compression algorithms in the literature that have better reconstruction results. Experiments show that the algorithm has better performance than the best multi-reference frame video compression sensing algorithm and can effectively improve the quality of slow motion video reconstruction.

Keywords: Video compressed sensing; multi-dimensional reference frame; reconstruction algorithm

1 Introduction

The classical traditional compression frameworks (e.g., ISO-MPEG and ITU-H.26x) [1–3] usually sample the signal first and then compress it according to Nyquist's sampling law with a large number of motion estimation, compensation, and other techniques with high computational complexity. It is only suitable for video signal processing occasions with strong hardware support at the coding end (e.g., broadcasting, video-on-demand systems, etc.). In recent years, many video systems also need terminal equipment to achieve real-time encoding, transmission and other functions (such as wireless video surveillance and mobile video phones and other distributed video systems). Due to the client's wireless,



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.

portable and other factors, the terminal hardware devices have weak processing power and limited power consumption, and the traditional video coding framework is no longer suitable for this area.

The compressed sensing (CS) [4,5] theory proposed by Donoho and Tao breaks through the limitations of the Nyquist sampling method. It can simultaneously sample and compress sparse signals, which effectively solves the problem of resource wastage in the signal acquisition process in the traditional signal compression framework. At present, compressed sensing theory has been intensively studied in the fields of speech coding, nuclear magnetic resonance, optical imaging and radar detection, image and video compression coding, etc. It has good prospects for development in wireless communication systems in the upcoming fifth generation mobile communication era and intelligent video surveillance and traffic control in the future artificial intelligence era.

Since video signals with different motion characteristics have different sparsity in different transform domains, using a single video sparse representation will make the video sparse representation unsatisfactory, thus affecting the quality of video compression-aware reconstruction. Taking slow-motion sequences as an example, the video compression-aware reconstruction algorithm with the best comprehensive performance in existing studies is 2sMHR proposed in the literature [6]. This algorithm not only has good reconstruction quality for the reconstruction of slow-motion sequences, but also maintains a very low time complexity due to the fixed number of hypothesis blocks, which directly solves the linear weighting problem in multi-hypothesis reconstruction. However, objects in slow-motion video sequences tend to produce half-pixel displacements, and moving objects also produce random and non-periodic scaling change phenomena. Since 2sMHR searches only similar blocks in the reference frame in the time dimension, the target blocks cannot get better similar blocks when half-pixel displacement or size scaling deformation occurs during motion. In addition, 2sMHR uses the method of fixing the number of hypothetical blocks, so the object surface in the slow-motion video sequence is likely to introduce low-quality hypothetical blocks. Therefore, in this paper, based on the multi-reference frame video compressed sensing reconstruction algorithm, we additionally introduce half-pixel dimensional reference frames and scaled dimensional reference frames [7] to provide better hypothesis blocks and propose a multidimensional reference frame multi-hypothesis prediction reconstruction algorithm.

2 Multi-Dimensional Reference Frame Multi Hypothesis Prediction Reconstruction Algorithm

2.1 Overall Framework of the Algorithm

The overall framework of multi hypothesis prediction and reconstruction algorithm based on dimensional reference frame is shown in Fig. 1.

At the encoding end, the input sequence is divided into multiple GOPs. The first frame of the GOP is set to be a Key frame, and the rest are CS frames, where the Key frame has a higher sampling rate and the CS frame has a lower sampling rate.

At the decoding, the Key frames are reconstructed using the intra-frame GSR in reference [7]. These CS frames are chunked in the measurement domain, and then reconstructed using a multi-dimensional reference frame multi-hypothesis prediction reconstruction algorithm.

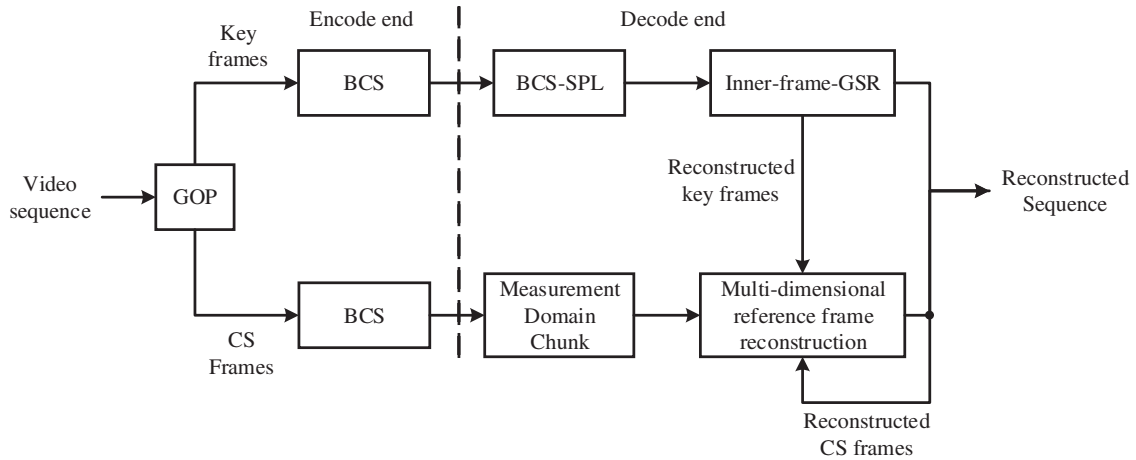


Figure 1: The framework of multi reference frame bidirectional prediction hypothesis optimization algorithm

2.2 Multi-Dimensional Reference Frame Generation Method

The semi-pixel reference frame is obtained by semi-pixel interpolation of the time domain reference frame. In order to ensure that the interpolated reference frame does not bring additional distortion phenomena such as edge sawtooth, contour blur and excessive smoothing, affect the quality of the semi-pixel reference frame, and maintain high interpolation efficiency. The sixth-order interpolation filter proposed in reference [8] is used to realize the interpolation of the time domain reference frame. Expand the interpolation reference frame size to 4 times the original size, then extract and separate the semi-pixel points at the same position. Finally recombine to obtain three semi-pixel reference frames with the same size as the original video frame to provide more power estimation assumptions. This process is shown in Fig. 2.

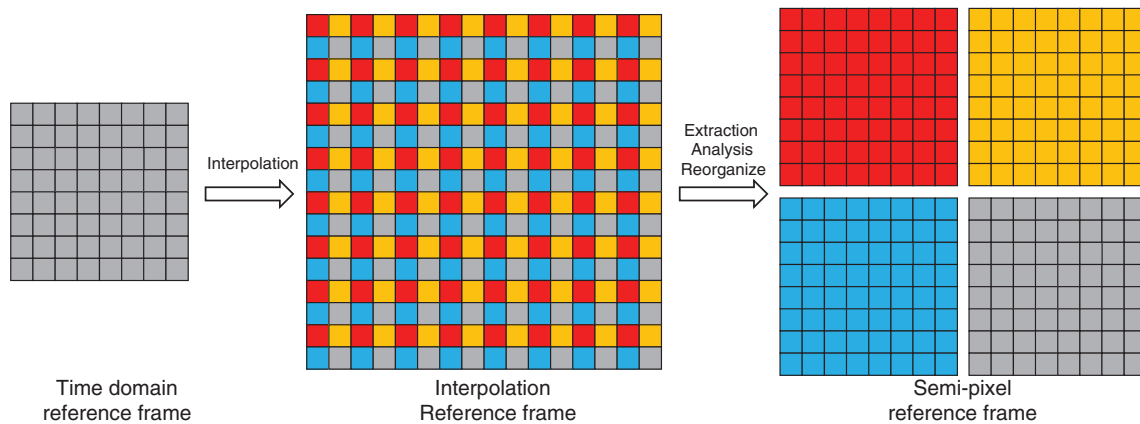


Figure 2: Schematic diagram of semi-pixel reference frame generation

The scaling dimension reference frame is generated after the time domain reference frame is reduced or enlarged by cubic convolution interpolation [9], and the scaling rate is set to 15%. There are 2 scaling dimension reference frames are generated for each reference frame. Using the 0 element to supplement the reduced reference frame edge, eliminate the enlarged reference frame edge, and keep

the scaled reference frame size consistent with the original reference frame. This process is shown in Fig. 3.

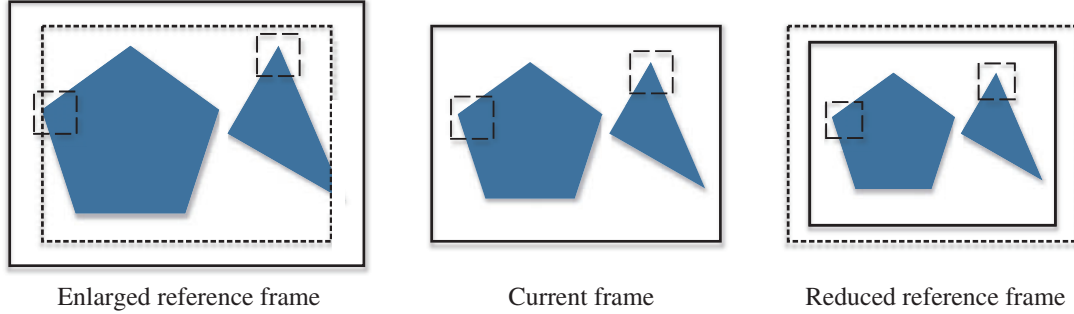


Figure 3: Schematic diagram of zoom reference frame generation

Note that RF, 1, RF, 2 and RF, 3 are three semi-pixel reference frames, ZF, 1 and ZF, and 1 are two scaling reference frames. Taking $\text{GOP} = 8$ as an example, the video frames in GOP are numbered from 1 to 8 in chronological order. Frame 1 is the key frame, the other frames are CS frames, and frame 9 is the key frame of the next GOP. The reference frame selection scheme of the slow-motion block is shown in Table 1.

Table 1: Example of slow-motion block reference frame selection scheme

Serial number	2	3	4	5	6	7	8
Time domain reference frame	1, 9	1, 2, 9	1, 3, 9	1, 4, 6, 9	1, 7, 9	1, 8, 9	1, 9
Semi-pixel reference frame	$\mathbf{R}_{1,1} \sim \mathbf{R}_{1,3}, \mathbf{R}_{9,1} \sim \mathbf{R}_{9,3}$						
	/	$\mathbf{R}_{2,1} \sim \mathbf{R}_{2,3}$	$\mathbf{R}_{3,1} \sim \mathbf{R}_{3,3}$	$\mathbf{R}_{4,1} \sim \mathbf{R}_{4,3}, \mathbf{R}_{8,1} \sim \mathbf{R}_{8,3}$	$\mathbf{R}_{6,1} \sim \mathbf{R}_{6,3}$	$\mathbf{R}_{7,1} \sim \mathbf{R}_{7,3}$	/
Scaling dimension reference	$\mathbf{Z}_{1,1} \sim \mathbf{Z}_{1,2}, \mathbf{Z}_{9,1} \sim \mathbf{Z}_{9,2}$						
	/	$\mathbf{Z}_{2,1} \sim \mathbf{Z}_{2,2}$	$\mathbf{Z}_{3,1} \sim \mathbf{Z}_{3,2}$	$\mathbf{Z}_{4,1} \sim \mathbf{Z}_{4,2}, \mathbf{Z}_{7,1} \sim \mathbf{Z}_{7,2}$	$\mathbf{Z}_{6,1} \sim \mathbf{Z}_{6,2}$	$\mathbf{Z}_{8,1} \sim \mathbf{Z}_{8,2}$	/
Reconstruction order	1	3	5	7	6	4	2

Although the three-dimensional reference frames can more comprehensively provide matching blocks for the CS frame to be reconstructed, the efficiency of searching matching blocks by full search is too low. In the meanwhile, because of the algorithm in this section is for slow moving video blocks, usually the matching blocks of such video blocks will be located at almost the same position in the reference frame or near the same position. Therefore, the four step fast search algorithm is adopted to realize the large-scale fast search of the optimal matching block of the video block to be reconstructed

in multiple reference frames. The matching criteria are as follows:

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [\mathbf{x}(i,j) - \hat{\mathbf{x}}(i,j)]^2 \times 100\% \quad (1)$$

The similarity between slow motion blocks and corresponding matching blocks is high, so only a few matching blocks are needed to complete the reconstruction. In addition, the reconstruction quality is also affected by the sampling rate. With the improvement of CS frame sampling rate, the number of optimal matching blocks increases. To sum up, in order to ensure the reconstruction efficiency, the fixed number of matching blocks in document [6] can be adopted. Through a large number of experimental verification, and the reconstruction time and reconstruction quality can be compromised to achieve better results. It is concluded that when the sampling rates are 0.1, 0.2 and 0.3 respectively, the corresponding number of matching blocks are 244, 300 and 362 respectively.

3 Comparison of Reconstruction Performance with Multi Hypothesis Prediction Algorithm

By comparing the rate distortion performance with three excellent video compressed sensing reconstruction algorithms MH-wElasticnet [10], 2sMHR [6] and MH-LE [11] Based on multi hypothesis prediction framework, the performance superiority of the proposed algorithm is verified. The subjects were hall, mother-daughter, TEMPETE and foreman. Comparing the experimental results of the algorithm, referring to the experimental data given in the corresponding literature, the experimental parameters of the proposed algorithm are consistent with the above literature. In order to ensure that the experiment is closest to the actual situation, $GOP = 16$ is set, and the rate distortion performance experiments of the four algorithms are carried out using the first 96 frames of the above experimental sequence. Both the proposed algorithm and MH-LE use intra GSR to reconstruct the key frame, 2sMHR and MH-wElasticnet use MH-BCS-SPL to reconstruct the key frame, repeat the key frame reconstruction experiment for 50 times, and obtain the average PSNR of the reconstructed key frame, as shown in Table 2.

Table 2: Average PSNR/DB of reconstructed key frame

Video sequence	MH-BCS-SPL	Inner-frame-GSR
Hall	43.12	47.57
Mother-daughter	47.11	51.19
Tempete	43.22	47.26
Foreman	43.57	47.32

As can be seen from Table 2, the average PSNR of key frame reconstructed by Intra-frame-GSR is about 4 dB higher than that of MH-BCS-SPL. With the improvement of reconstructed key frame is used as the reference frame for CS frame reconstruction, it can improve the reconstruction quality of CS frame. When the CS frame sampling rates are 0.1, 0.2 and 0.3 respectively, the average PSNR results of the reconstructed CS frame of the proposed algorithm and the other three comparison algorithms are shown in Fig. 4, and the specific results are shown in Table 3.

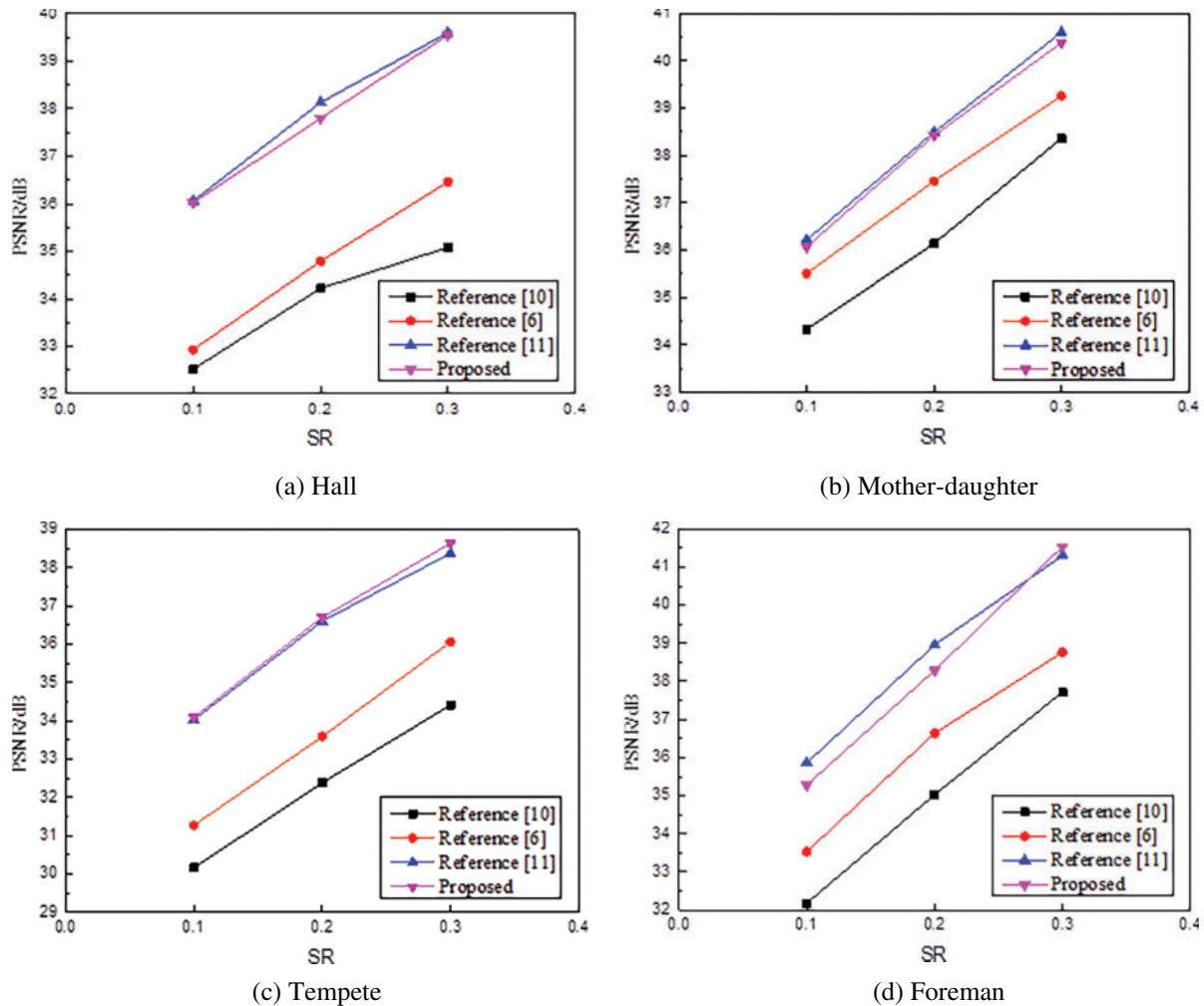


Figure 4: The reconstructed CS frame average PSNR results of the proposed algorithm and the other three comparison algorithms

Table 3: Proposed algorithm and three multi hypothesis prediction algorithms to reconstruct the average PSNR/DB of CS frame

Video sequence	Reconstruction algorithm	SR		
		0.1	0.2	0.3
Hall	Reference [10]	32.52	34.22	35.08
	Reference [6]	32.93	34.79	36.46
	Reference [11]	36.06	38.13	39.60
	Proposed	36.02	37.80	39.55
Mother-daughter	Reference [10]	34.33	36.14	38.36
	Reference [6]	35.51	37.46	39.26

(Continued)

Table 3: Continued

Video sequence	Reconstruction algorithm	SR		
		0.1	0.2	0.3
Tempete	Reference [11]	36.22	38.49	40.60
	Proposed	36.07	38.42	40.38
	Reference [10]	30.16	32.38	34.41
	Reference [6]	31.27	33.59	36.06
	Reference [11]	34.03	36.60	38.37
	Proposed	34.09	36.70	38.63
Foreman	Reference [10]	32.18	35.03	37.72
	Reference [6]	33.53	36.64	38.76
	Reference [11]	35.86	38.96	41.30
	Proposed	35.28	38.29	41.53

It can be seen from Fig. 4 and Table 3, because of the sequence hall, mother-daughter, tempete and foreman with slow average motion speed, the average PSNR of the proposed algorithm and the algorithm in reference [10] are greatly improved compared with the algorithm in reference [11] and the algorithm in reference [6]. Due to both algorithms use Intra-frame-GSR to reconstruct the key frame, the reconstruction quality of key frame is greatly improved. When it is used as reference frame, it can provide better matching block to assist CS frame reconstruction. The performance of the proposed algorithm is close to that of the algorithm in reference [8], and the algorithm in reference [11] can accurately distinguish the stable block and the changing block in the video frame, and enhance and reconstruct the video block in the changing region, while the proposed algorithm can accurately judge the video blocks with different degrees of motion in the video frame and adopt appropriate reconstruction methods respectively. Therefore, the CS frame reconstruction quality of the two algorithms is high. The reconstruction quality of the proposed algorithm for the TEMPETE sequence at the sampling rates of 0.1, 0.2 and 0.3 is higher than that of the algorithm in the reference [11]. The reason is that the motion of the object in the tempete sequence is caused by the slow distance of the lens, so there is a lot of scaling information in the video sequence. The proposed algorithm using multiple scaling dimension reference frames when reconstructing the slow-motion block, and provide high-quality matching blocks for the video blocks with scaling changes in the video sequence. So the reconstruction quality is improved.

4 Conclusion

Based on the sparse characteristics of video signals in different sparse representation domains, this paper proposes a video compressed sensing reconstruction algorithm based on multi-dimensional reference frames. In the process of multi reference frame video compressed sensing reconstruction, semi-pixel reference frames and scaled dimensional reference frames are introduced. So as to obtain more high-quality assumptions and effectively improve the quality of slow-motion video reconstruction. Experiments show that the proposed algorithm has better performance than the algorithm using only time-domain reference frame.

Acknowledgement: We thank graduate student Cao Zhendi from Army Armored Force Academy for his help with the writing of the paper.

Funding Statement: The authors received no specific funding for this study.

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

References

- [1] D. L. Gall, "MPEG: A video compression standard for multimedia applications," *Communications of the ACM*, vol. 34, no. 4, pp. 46–58, 1991.
- [2] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301–317, 2001.
- [3] G. J. Sullivan and T. Wiegand, "Video compression-from concepts to the H. 264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18–31, 2005.
- [4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [5] E. J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [6] W. F. Ou, "Research on multi-hypothesis prediction-based reconstruction algorithms for compressed video sensing," Ph.D. Dissertation, South China University of Technology, China, 2016.
- [7] C. L. Yang and X. W. Zheng, "Dual-sparsity reconstruction algorithm based on multi-dimension reference frames in compressed video sensing," *Journal of South China University of Technology (Natural Science Edition)*, vol. 46, no. 8, pp. 1–10, 2008.
- [8] T. Wiegand, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H. 264| ISO/IEC 14496-10 AVC)," *JVT-G050*, 2003.
- [9] Z. B. Zheng, "Research on block classification based inter-frame group sparse reconstruction algorithm for compressed video sensing," Ph.D. Dissertation, South China University of Technology, China, 2020.
- [10] J. Chen, Y. Chen, D. Qin and Y. H. Kuo, "An elastic net-based hybrid hypothesis method for compressed video sensing," *Multimedia Tools & Applications*, vol. 74, no. 6, pp. 2085–2108, 2015.
- [11] R. D. Tang, C. L. Yang and Y. Y. Xuan, "Local enhancement recovery algorithm based on multi-hypothesis prediction," in *Compressed Video Sensing*, Beijing, China: Acta Automatica Sinica, 2021. [Online]. Available: <https://doi.org/10.16383/j.aas.c190408>