WILEY | Hindawi

Research Article

Video Error-Resilience Encoding and Decoding Based on Wyner-Ziv Framework for Underwater Transmission

Borui Li ⓘ **, Yang Zhang** ⓘ **, and Qingjuan Feng**

School of Automation, Beijing Information Science and Technology University, Beijing 100192, China

Correspondence should be addressed to Yang Zhang; zhangyang@bistu.edu.cn

Limited by the characteristics of underwater acoustic channels, the video transmission applications targeting deep-sea detection and operation tasks are facing severe challenges such as network failure and high delay, resulting in loss of video details, color distortion, blurring, and even bit errors, which seriously affect decoding quality of the video transmission and reception. In order to solve the problems of deep-sea long-distance wireless communication, this paper proposes an improved Wyner-Ziv coding scheme (UnderWater-WZ) for video transmission through acoustic channels. The implementation process includes controlling error range by using MJPEG coding, combining motion compensation time interpolation with calibration information to generate high-quality side information. And intraframe quantization matrix is designed to weaken the change of video scene. The experimental results show that under the highest packet loss rate of 20%, this scheme can achieve 2.6~3.5 dB improvements in terms of video reconstruction compared to the previous methods, which is close to the error-free level.

1. Introduction

In recent years, with the development of deep-sea exploration technology and the increasing demand for marine resources, various countries have begun to build their own deep-sea space stations. Using the visual perception in deep-sea environments, a variety of video applications can be transmitted through underwater acoustic channels, including seabed exploration, disaster prevention, mine reconnaissance, and environmental monitoring [1–4]. To fulfill the transmission needs for underwater video data, coding and transmission techniques have been proposed over underwater acoustic networks [5–9]. However, due to the characteristics of acoustic signals, an underwater acoustic communication system typically has three defects: narrow bandwidth, long-distance propagation delay, and error-prone transmission. These limitations provide no guarantees of real-time video transmission and end-to-end reproduction quality. An increased desire for higher quality and reliability is also arising in underwater multimedia applications. Most underwater video coding algorithms for underwater acoustic communication are aimed at improving the com-

pression efficiency. These algorithms can be divided into two categories: improved algorithms based on standard encoders and self-designed coding algorithms. In [10], the authors compare the coding performance of different standard encoders on underwater acoustic networks. For underwater video streaming, the standard encoder is improved to cooperate with acoustic links. On the basis of H.264, Avrashi et al. [11] realize efficient compression of 36 kbps code rate by adaptively reducing the frame rate and use motion vector interpolation to reconstruct the missing frames at the decoding end to restore the normal frame rate. Aiming at the transmission of underwater forward-scan sonar video, Mirizzi et al. [12] propose a coding algorithm based on region segmentation, which adopts different coding strategies for image foreground and background, respectively, and embeds the algorithm into an HEVC encoder. This method exploits the characteristics of 2-D sonar imagery to remove redundancy more effectively over standard encoders for low-bandwidth underwater acoustic transmission, but it is not designed for optical video/images with high details. Hoag et al. [13] propose the underwater video compression techniques start using the discrete wavelet transform as coding

basis to fully eliminate the spatial correlation between pixels. Negahdaripour and Khamene [14] combine wavelet transform with interframe motion compensation, and the average compression ratio of the algorithm reaches 150 : 1. In order to further eliminate the time redundancy between video frames, Nagrale et al. [15] use a hybrid global and local motion compensation strategy to achieve efficient underwater video compression. Based on the wavelet coding tree proposed by Li and Wang [16], Song et al. [17] adopt an improved wavelet transform algorithm with stronger signal representation ability while improving coding efficiency and decoding quality. Dong et al. [18] propose a fast algorithm for VVC from two aspects of mode selection and prediction terminating to reduce coding complexity. For the mode selection, adaptive mode pruning (AMP) is proposed to remove nonpromising modes. What is more, mode-dependent termination (MDT) is proposed to select an appropriate model through the optimal mode and terminate unnecessary intrapredictions of remaining depth levels. For the new QTMT segmentation structure and two-way coded intraprediction mode, Yang et al. [19] propose a low-complexity CTU partition decision method and a fast intra-mode decision method to reduce the computational burden of VVC intracoding.

In the current research, most of the methods are not aimed at the characteristics of deep-sea video transmission and cannot be applied to videos containing complex motion information under the ocean. The most important thing is that the current video error-resilience research cannot control the bit error rate and improve the video quality at the same time. In this paper, we propose video error-resilience scheme based on Wyner-Ziv framework (UnderWater-WZ) for underwater transmission.

UnderWater-WZ coding system is aimed at improving the efficiency of encoding and decoding and improving video quality. This scheme is based on Wyner-Ziv video coding system. By optimizing WZ coding scheme and using MJPEG coding algorithm to reduce data correlation, the bit error rate performance can be enhanced. Through the combination of motion-compensated temporal interpolation (MCTI) and check information, higher-quality side information (SI) is generated, and with the help of DCT transform intraquantization matrix and feedback channel, the video quality is significantly improved. To sum up, our video error-resilience research mainly is aimed at reducing the bit error rate and improving the reconstruction quality. This paper provides the following contributions:

(1) In this paper, by combining MJPEG coding with the Wyner-Ziv system, a high-efficiency video error-resilience method with low correlation is formed. When data loss occurs, the updated coding system will not cause error diffusion of reconstructed images and control the range of data packet loss, thereby improving the robustness of the system

(2) This paper proposes a macro-block-based side information generation method, which combines MCTI and verify information transmission to generate

higher-quality side information. Firstly, for the macroblocks in the reference frame, intraframe coding is performed through the verified information transmitted by the encoder and then using MCTI to generate side information. Finally, higher-quality side information is generated through motion-compensated quality enhancement techniques. The method can encode nonlinear moving objects in ocean video and improve the quality of reconstructed video

(3) In this paper, we design an intraframe quantization matrix based on DCT. The scheme adopts a new spatial frequency calculation method. In this way, we can weigh the different frequency coefficients. Efficient video coding reduces people's sensitivity to scene changes. The rest of this paper is organized as follows. Section 2 reviews the relevant works. Section 3 introduces the implementation details of the Wyner-Ziv video coding system and the proposed video error-resilience scheme, respectively. The experimental results are presented in Section 4, and the conclusions from this investigation are finally given in Section 5

2. Related Work

The need for emerging applications such as wireless low-power video surveillance has spurred the development of video encoding. The current research direction of WZ video coding mainly focuses on the following aspects: one is to improve the coding efficiency and improve the RD performance of the system; the other is to provide error recovery capability and improve video quality.

2.1. Improving the Coding Efficiency and Improving the RD Performance of the System. Multiple description coding [20–26] and redundant image [27–29] techniques are also effective against channel transmission errors, but both come at the cost of increased bit consumption. The video error-resilience technology in the video compression algorithm needs to consume a part of the code stream so that the sender has to generate more data transmission to ensure the same reconstruction quality in the decoder. Ramon et al. [30] propose a new curved-based intraframe prediction method. It enhances the intraframe prediction process by modeling curved texture features. The proposal incurs a small bitstream overhead for transmitting the displacement information, which is offset by encoding efficiency gains. Singhadia et al. [31] establish an empirical relationship of PSNR and bitrate with quantization parameter, through optimization algorithm to find the best quantization parameter to achieve the highest PSNR. However, this method cannot perform efficient encoding when transmission errors occur. Marzuki et al. [32] propose a tile-level rate control algorithm for high-efficiency video coding on tile-parallelization case. The proposed rate control algorithm embeds a tile-level rate control by adaptively assigning the appropriate bits for each tile. However, for video sequences

of nonlinear moving objects, the rate control cannot be effectively performed, and thus, the video quality cannot be improved.

The above research is aimed at terrestrial video transmission. However, the bandwidth of underwater transmission is extremely limited. When high compression rate transmission is performed, it also means high correlation. A 1-bit data error during the transmission process will cause the data transmission of the entire frame to fail. Therefore, for the key frames, this paper combines the WZ system with the MJPEG coding technology to form a high-efficiency video error-resistance method with low correlation. In the case of ensuring the low correlation of data, each frame is independently compressed using the JPEG algorithm, which can control the data packet loss in a stable range and adapt to the special nature of video framing and marine environment transmission. Most studies improve RD performance by upgrading the way of generating side information; this paper combines MCTI and parity information transmission based on macroblocks to generate higher-quality side information. The Euler distance can be better mapped to a smaller Hamming map by using the gray mapping. In this way, it can improve the correlation between initial information and side information and handle underwater video sequences with dramatic scene changes efficiently.

2.2. Providing Error Recovery Capability and Improving Video Quality. The current research about improving video quality focuses on superresolution image reconstruction based on deep learning [33–39]. This method is realized on the basis of a large amount of data, but the underwater video transmission has the characteristics of narrow bandwidth and high delay. In order to avoid external interference during transmission, resulting in transmission errors, it is not suitable for processing a large number of video sequences. For small sample data processing, Luo et al. [40] propose a novel recursively adaptive perceptual nonlocal means preprocessing algorithm based on just noticeable distortion model. It can avoid discontinuous artifacts to occur in reconstructed video by reducing perceptual redundancy. The algorithm can effectively improve the perceptual quality of reconstructed video frames. However, when data loss occurs during underwater transmission, this method cannot improve the video quality according to the existing data. Liu et al. [41] propose a hierarchical motion estimation and compensation network for video compression. Video frames are marked as intraframes and interframes. Independent compression is employed for intraframes and hierarchical prediction by adjacent frames using a bidirectional motion prediction network for interframes. However, it will result in highly sparse and compressible residues, thereby reducing intra- and interframe correlation.

Aiming at the defects in the above research, this paper proposes a motion compensation quality enhancement technology with the rate adaptation. This method utilizes the temporal correlation between frames and improves the quality of intraframe coding by weighted averaging of reference frames and intraframe blocks.

3. Description of the Proposed Coding Solution

The UnderWater-WZ video coding system is mainly improved based on the Wyner-Ziv system framework. The side information generation method in the scheme is improved according to the motion compensation time interpolation adopted in the WZ system. In the framework proposed in this paper, we still use the bit plane-based method for video encoding and decoding while also retaining the feedback channel in the framework. In addition, we aim at the defect that the WZ system cannot effectively deal with nonlinear moving targets, improving the key frame coding method, the side information generation method, and the discrete cosine transform process.

3.1. Wyner-Ziv Video Coding System. In recent years, distributed video coding algorithm has been a hot research direction of video coding. Compared with other traditional video coding, its coding method has the characteristics of simple coding and high bit error robustness. It can well meet the needs of new video services such as underwater robot operation, UAV aerial photography, and wireless monitoring. The most typical distributed video coding system is Wyner-Ziv video coding system [42–44] proposed by Stanford University and prism system proposed by the University of California, Berkeley [45]. However, most researches are based on the WZ video coding system framework of Stanford University. Among them, the more typical scheme is the DISCOVER scheme [46] proposed by the research group funded by the European Union IST FP6 in 2007, which is one of the best schemes in the video coding system.

The Wyner-Ziv video coding system based on transform domain integrates a variety of excellent algorithms and technologies. Its framework is shown in Figure 1.

The Wyner-Ziv video coding system adopts the frame structure of intracoding and intercoding. According to whether there is discrete cosine transform (DCT) in the system, it can be divided into WZ video coding system based on pixel domain and WZ video coding system based on transform domain. Because DCT has spatial correlation, WZ video coding system based on transform domain has better rate-distortion (RD) performance, and its system framework is shown in Figure 1.

For the encoder, according to the order of parity, the input video sequence is divided into Wyner-Ziv frame and key frames. Among them, the key frames adopt the traditional intracoding. Next, macro-block-based discrete cosine transform is performed on each WZ frame. According to the position of the DCT coefficients in the macroblock, we combine all coefficients. And then, it forms a DCT coefficient band X_k, where k represents the coefficient. Secondly, 2^{Mk} level uniform quantization is carried out for each DCT coefficient band X_k, extracting the bit plane of all quantized values in the quantized coefficient band q_k and sending the bit plane to the LDPC encoder for coding. Finally, the system stores the verification information generated by each bit plane in the buffer, and the decoder requests to send the verification information through the feedback channel to assist the decoding process.
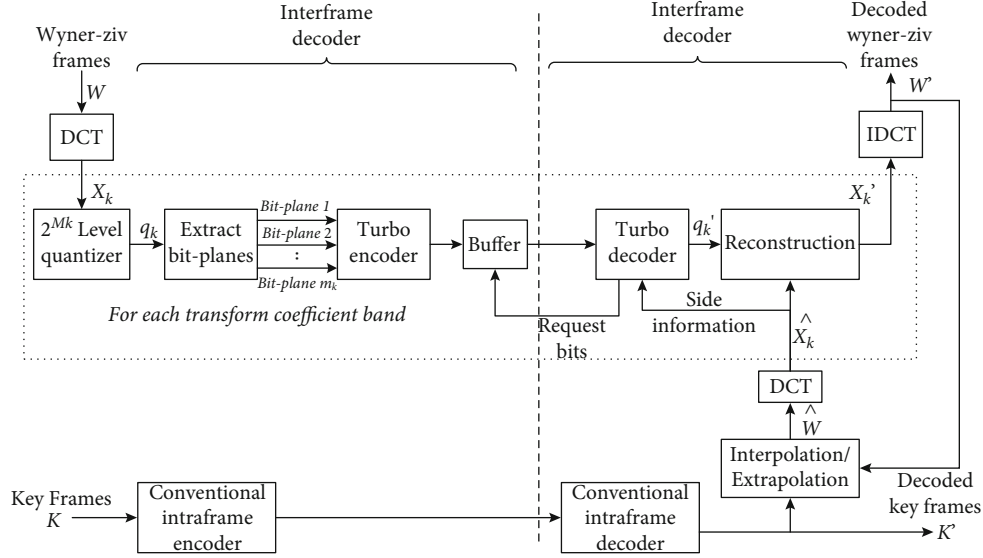
FIGURE 1: Transform-domain Wyner-Ziv video codec with intraframe encoding and interframe decoding.

For the decoder, first, the system utilizes key frames to generate side information SI frames through motion-compensated frame interpolation techniques or extrapolation techniques. Secondly, the Laplacian distribution model is constructed by the residual statistical properties between the DCT coefficients corresponding to the WZ frame and the side information frames. The system requests the encoder to send check information through the feedback channel and then uses the side information coefficient band for decoding. Next, combine the bit planes of all decoded coefficient bands to form a decoded quantized coefficient bandkq'. Through the SI coefficients, the DCT coefficient band is reconstructed to form a reconstructed coefficient band Xk'. And inverse discrete cosine transform (IDCT) is then performed on Xk' to obtain the decoded WZ frame. Finally, the system reassembles the decoded frames to obtain the decoded video sequence.

3.2. Side Information (SI) Generation Algorithm Based on MCTI.

The quality of SI largely determines the RD performance of Wyner-Ziv system. The higher the quality of SI, the higher the correlation with the initial WZ frame. To achieve the target decoding quality, the encoder needs to transmit fewer check bits to the decoder, so the overall RD performance of the system is higher.

At present, the typical side information generation method is motion-compensated time domain interpolation (MCTI) in the Wyner-Ziv coding framework. MCTI uses the motion estimation method, the motion trajectory of the interpolated frame is found based on the time domain and spatial domain information between the pre- and post-decoded reference frames, and then, the SI is generated through motion compensation. Finally, the weighted vector median filter is used to smooth the motion vector, so as to enhance the continuity of the motion vector in space and time domain. Other methods also include updating and iter-

ation of side information and sending check information from the encoder.

3.2.1. Algorithm Framework for Generating SI Based on MCTI.
MCTI used in the Wyner-Ziv coding system is based on the scheme proposed in [47]. The framework for generating side information using MCTI is shown in Figure 2.

The scheme mainly includes frame interpolation structure definition module, forward motion estimation module, bidirectional motion estimation module, spatial motion smoothing module, and bidirectional motion compensation module.

3.2.2. Frame Interpolation Algorithm.
MCTI uses the maximum interframe distance of decoded reference frames X_b' and X_f' to obtain side information by interpolating half the time interval.

3.2.3. Forward Motion Estimation.
MCTI takes X_b' as the reference frame and obtains the motion vector of each macroblock of X_f' by estimating the correlation between the key frames. The main parameters include search window size, search range, and search step. Its structure is shown in Figure 3, where mv_b and mv_f are the motion vectors.

3.2.4. Bidirectional Motion Estimation.
Due to forward motion estimation, side information frames generate overlapping and uncovered regions. Bidirectional motion estimation further improves the motion vector, and its process is shown in Figure 4. The interpolation frame is divided into nonoverlapping frames; all motion vectors passing through this frame are used as candidate motion vectors. And then, the center of the interpolation block is selected as the motion vector of the subblock of the interpolation frame. To improve the quality of the interpolated frame and increase the linearity of the motion vector, the subblock of the
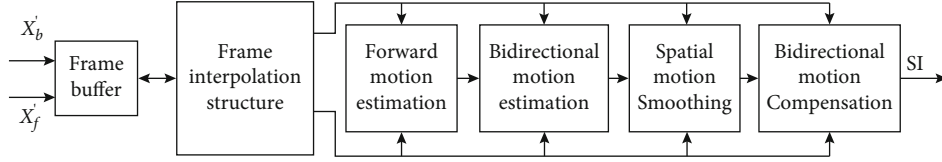
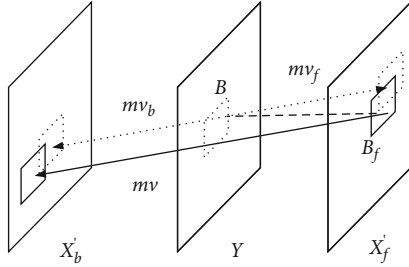FIGURE 2: Motion-compensated time interpolation framework.

FIGURE 3: Forward motion estimation. The block effect caused by the quantization process needs to be removed by low-pass filtering first, so as to improve the reliability of the vector.
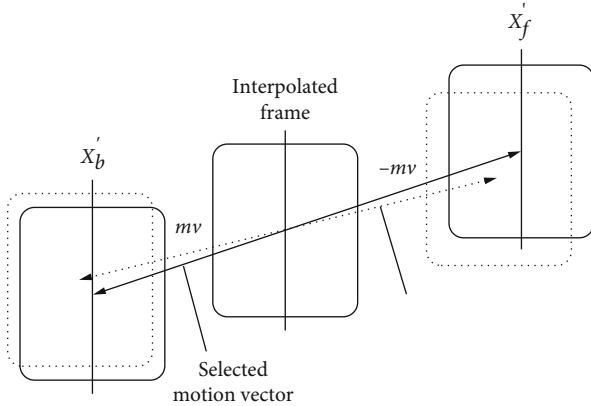

FIGURE 4: Bidirectional motion estimation module. The method is to further improve the motion vector obtained by forward motion estimation.

interpolated frame can be used as the midpoint of the motion vector.

The vector of forward motion estimation is mv. According to the target motion and the temporal distance from the interpolation frame to X_b' and X_f' which are equal, the bidirectional motion vector of the corresponding interpolation block in the interpolation frame is $(mv/2, -mv/2)$.

3.2.5. Spatial Motion Smoothing.
The motion vectors obtained by bidirectional motion estimation have low spatial correlation, which leads to errors in the interpolated frames, so the error can be reduced by the spatial smoothing algorithm. A weighted motion vector median filter (WMVMF) is used to reduce the influence of error vectors on the quality of the interpolated frame. WMVMF maintains the spatial correlation of the motion field by finding candidate motion vectors in the neighboring blocks, and its structure is shown

in Figure 5. The motion vectors of 8 macroblocks surrounding the filter block are selected for calculation.

mv_c represents the motion vector of the block to be filtered, and mv_1 to mv_8 represent the motion vectors of eight macroblocks adjacent to the block to be filtered.

3.2.6. Bidirectional Motion Compensation.
MCTI fills the interpolated frame by bidirectional motion compensation. Assuming that the temporal distance between the interpolated frame and the preceding and following key frames is the same, when bidirectional motion compensation is performed, each reference frame has the same weight. Let the size of the motion vector of the interpolation block be mv $= (mv_x, mv_y)$. Motion compensation is performed as follows:

$$Y(x,y) = \frac{X_f'(x + mv_x, y + mv_y) + X_b'(x - mv_x, y - mv_y)}{2},$$

(1)

where $Y(x, y)$ represents the motion compensation obtained at (x, y) and X_b' and X_f' represent the decoded before and after reference frame.

3.3. Video Error-Resilience Scheme of Video Transmission under Underwater Acoustic Channel.
The UnderWater-WZ system framework under the underwater acoustic channel proposed in this paper is shown in Figure 6. It is based on the frequency domain Wyner-Ziv framework of Stamford Bridge University. The difference is that the UnderWater-WZ framework designed in this paper includes a gray mapping module, an LDPCA codec module, a side information quality enhancement module, and an optimal design of the intraframe quantization matrix for DCT. The encoding and decoding process is as follows:

(1) Encoder side: first, the encoder divides the video sequence into frames in the order of parity. For key frames, format conversion is performed to convert video frames in BMP format into data in YUV format required for encoding and transmission. Then, the MJPEG algorithm is used for intraframe coding. For the WZ frames, the same format conversion is performed first, followed by discrete cosine transform and DCT intraframe quantization matrix processing. For the two adjacent reference frames (X_b and X_f), first we perform 4×4 DCT on each macroblock in the WZ frames and group colocated coefficients together in each macroblock to form coefficient band b_k, $k = 0, \cdots, 15$. 2^{Mk} level

mv_1	mv_2	mv_3
mv_4	mv_c	mv_5
mv_6	mv_7	mv_8

FIGURE 5: Adjacent motion vectors in WMVMF. This process can reduce the impact of errors in motion vectors.

quantization is performed on each coefficient band. Afterwards, the quantization interval of the DCT coefficients is mapped into a binary codeword by using gray mapping, and the bit plane is extracted. The extracted bit planes are sequentially sent to the LDPCA encoder for encoding in the order from MSB to LSB, and the verification information generated by the encoding is first stored in the buffer. The encoder transmits part of the verification information to the packet loss channel through the feedback channel according to the request of the decoder. This channel environment is simulated according to the marine and finally enters the decoder for decoding

(2) Decoder side: since the key frames adopt the MJPEG algorithm encoding mode and the bilinear interpolation algorithm is a typical algorithm in spatial error concealment, this algorithm is used to mask the error of the key frame with errors. For decoded reference frames X_b' and X_f', MCTI is used to generate an estimate of the initial WZ. Side information quality can be significantly improved by utilizing MCTI-generated estimates, intradecoding blocks, and motion-compensated quality enhancement modules. And the correlation noise between WZ frames and side information is modeled by the correlation noise model based on the Laplacian distribution, and the side information is converted into soft input information. The parameters of the Laplacian distribution are estimated online at the coefficient layer. After that, the improved side information is sent to the LDPCA decoder, and the gray code is demapped, the code stream correction and interpolation restoration are processed, and finally, the refactor module is entered. When all the DCT coefficient bands are reconstructed, the decoded WZ frame is obtained by IDCT and then merged with the decoded key frames, obtaining the video sequence processed by the video error-resilience scheme designed in this paper

3.4. Optimization of WZ Frame Encoding and Decoding Method.
In the Wyner-Ziv system, the key frames are usually compressed using the H.264/AVC coding standard. Its advantage is that H264/AVC algorithm removes interframe redundancy and performs intra- and interframe compres-

sion. Therefore, the algorithm can provide better compression ratio.

In general video compression and coding algorithms, the compression action depends on the correlation of data. The higher the compression rate, the stronger the correlation between video data. However, the wireless underwater acoustic communication method used for video transmission in marine environment has three defects: narrow bandwidth, high delay, and error-prone transmission. This leads to strong data correlation not suitable for wireless underwater acoustic communication, and it causes a small part of packet loss to result in a large part. For example, in the H.264/AVC compression coding method in underwater acoustic channel, the transmission error of 1-bit data will lead to the loss of 1 second of video sequence data.

The video error-resilience scheme of UnderWater-WZ is mainly aimed at controlling the bit error rate and image restoration. Therefore, for controlling the bit error rate, we combine the motion joint photographic experts group (MJPEG) compression coding algorithm with the Wyner-Ziv system, and we design a video error-resilience system based on low correlation coding. MJPEG was originally developed for multimedia PC applications. Now, it is used by video capture devices such as digital cameras, IP cameras, webcams, and nonlinear video editing systems. MJPEG processes a continuous sequence of video frames as continuous still image frames, and each frame is compressed by JPEG algorithm. Without considering the correlation between frames, there is a lack of interprediction, and frames are compressed independently, so it is suitable for processing video sequence framing.

Although the compression efficiency of MJPEG is limited, it can control data packet loss in a stable range and adapt to the special nature of the marine environment. Due to the harsh marine environment, H.264/AVC relies more on hardware devices in transmission, but MJPEG does not have high hardware requirements.

3.5. Improved Generation Method Based on Macroblock Side Information.
MCTI is only applicable to the simple linear motion of the target. And in the marine environment, there are often violent collisions and obvious light changes, so the original scheme cannot be applied to the marine environment.

Most WZ video coding frameworks use MCTI to generate side information. In order to adapt to the video sequence of complex scenes under the marine environment, UnderWater-WZ system improves the SI generation method in WZ system. We combine MCTI with additional information of coding segment transmission, which can cope with video transmission in complex environments, so as to produce higher-quality side information.

Combined with DISCOVER coding technology, the scheme framework proposed in this section includes gray mapping, adaptive LDPCA codec, and side information quality enhancement module.

3.5.1. Gray Mapping.
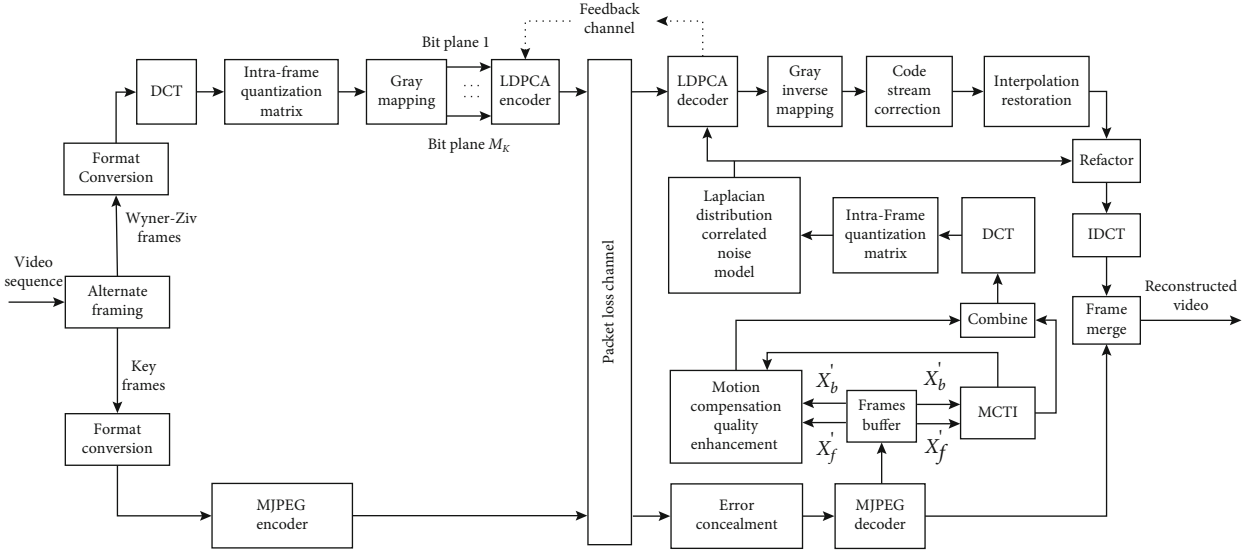Gray mapping is widely used in digital communication, which can significantly improve the

FIGURE 6: The video error-resilience scheme (UnderWater-WZ) based on the Wyner-Ziv video coding system under the underwater acoustic channel designed in this paper.

TABLE 1: Comparison of natural binary codes and gray mapping codewords for coefficient values around zero.

Coefficient value	Natural binary code	Gray code
-2	11111110	10000001
-1	11111111	10000000
0	00000000	00000000
1	00000001	00000001
2	00000010	00000011

correlation between initial information and side information and can better map smaller Euler distances to smaller Hamming distances, making encoded data less susceptible to complex environments.

The natural binary code is usually used in the WZ codec to describe the quantization symbols. The Hamming distance between the binary codes is obviously higher than that of the gray mapping. The corresponding relationship between the two coefficients near zero is shown in Table 1.

The quantized DCT coefficient q is converted into a binary form using natural binary as follows:

$$q_{\mathrm{NB}} = \left[q^0 q^1 q^2 \cdots q^{M_k-1} \right], \qquad (2)$$

where M_K represents the number of bits required for the DCT coefficient band b_k and q^0 and q^{Mk-1} represent the most significant and least significant bits, respectively.

Then, the above binary code is converted into a gray code by the following equation as follows:

$$\begin{aligned} q_G^0 &= q^0, \\ q_G^i &= q^{i-1} \oplus q^i, \\ q_G &= \left[q_G^0 \quad q_G^1 \quad q_G^2 \cdots \quad q_G^{M_k-1} \right], \end{aligned} \qquad (3)$$

where $i \in [1, M_k - 1]$, and the XOR calculation has low-complexity characteristics, so the added complexity can be ignored.

At the decoding end, after the M_k bit planes in the DCT coefficient band are decoded, the natural binary code shown in equation (4) is obtained by inverse gray mapping as follows:

$$\begin{aligned} q'0 &= q_G'0, \\ q'i &= q'i-1 \oplus q_G'i, \\ q_{\mathrm{NB}}' &= \left[q'0 q'1 q'2 \cdots q'M_k-1 \right], \end{aligned} \qquad (4)$$

where $q_G'0$ indicates the decoded most significant bit.

The smaller the Hamming distance between the codewords, the smaller the bit error rate (BER) of the bit stream of the correlation between the initial WZ information and the corresponding side information. Replacing the natural binary code with gray mapping can significantly reduce BER and improve the DSC compression efficiency.

3.5.2. Rate-Adaptive LDPCA Codec. Since LDPC codes have poor source compression performance, a rate-adaptive LDPC codec composed of LDPC code concomitant encoders and accumulators is proposed, and its system block diagram is shown in Figure 7.

First, add the syndrome code stream $(S, S_2 \cdots S_n \cdots S_{N-1}, S_N)$ output by LDPC encoding to modulo 2 to obtain $(a_1, a_2 \cdots a_n \cdots a_{N-1}, a_N)$. And stored in the buffer area, according to the request of the decoder, the check bits in the encoder are segmented and transmitted to the decoder. During decoding, the decoder first performs modulo 2 differential on the received parity bits to obtain a concomitant code stream and then performs LDPC iterative encoding.

3.5.3. Quality Enhancement Module for Side Information. When there is severe shaking and obvious illumination changes in the deep-sea environment, the quality of side

information will deteriorate accordingly. The MCTI scheme will fail. In this paper, a quality enhancement module based on motion compensation is proposed to be applied in the UnderWater-WZ system. The basic idea is to improve the quality of the intraframe coding fast I_i, thereby improving the quality of SI. The framework is shown in Figure 8, which is mainly composed of three parts, namely, T_i block estimation, U_i block estimation, and generation of side information Y_i.

(1) T_i Block Estimation. T_i blocks are estimated by exploiting the temporal correlation between frames of a video sequence. For intracoded block I_i, the symmetrical motion vector pair obtained by bidirectional motion estimation is $\{(vx, vy), (-vx, -vy)\}$. According to the classical linear motion model, the initial frame X_i is linked with the macroblocks in the same position in the preceding and following reference frames X_b and X_f as follows:

$$X_i(x, y) = X_b(x + v_x, y + v_y) + Z_b(x, y), \tag{5}$$

$$X_i(x, y) = X_f(x - v_x, y - v_y) + Z_f(x, y), \tag{6}$$

where $Z_b()$ and Z_f both obey a normal distribution with a mean value of 0 and the decoded reference frames Xb' and X_f' before and after the quantization noise Q processing as follows:

$$X_b'(x, y) = X_b(x, y) + Q_b(x, y), \tag{7}$$

$$X_f'(x, y) = X_f(x, y) + Q_f(x, y), \tag{8}$$

where $Q_b(x, y)$ and $Q_f(x, y)$ represent the quantization distortion at (x, y).

For a uniform quantizer with a quantization step size of Δ, the quantization distortion is approximately uniformly distributed in the interval $[-\Delta/2, \Delta/2]$, and the variance is $\Delta 2/12$. Under the action of the quantization distortion, the quality of the macroblock is estimated by the variance of the quantization distortion, as follows:

$$\sigma_{X_b'}^2 = E\left[\left(X_b'(x, y) - X_b(x, y)\right)^2\right] = E\left[Q_b^2(x, y)\right] = \frac{\Delta^2}{12}, \tag{9}$$

$$\sigma_{X_f'}^2 = E\left[\left(X_f'(x, y) - X_f(x, y)\right)^2\right] = E\left[Q_f^2(x, y)\right] = \frac{\Delta^2}{12}. \tag{10}$$

According to the vector obtained by bidirectional motion estimation and the decoded reference frames, the motion-compensated version X_b^{MC} and X_f^{MC} of block I_i is

obtained, as shown in formulas (13) and (14).

$$\begin{aligned} X_b^{MC}(x, y) &= X_b'(x + v_x, y + v_y) + \Delta X_b(x, y) \\ &= X_b(x + v_x, y + v_y) + Q_b(x + v_x, y + v_y) \\ &\quad + \Delta X_b(x, y), \end{aligned} \tag{11}$$

$$\begin{aligned} X_f^{MC}(x, y) &= X_f'(x - v_x, y - v_y) + \Delta X_f(x, y) \\ &= X_f(x - v_x, y - v_y) + Q_f(x - v_x, y - v_y) \\ &\quad + \Delta X_f(x, y), \end{aligned} \tag{12}$$

where ΔX_b and ΔX_f correspond to the residuals of the motion estimation process, respectively.

By combining the motion compensation of the reference frames with the corresponding intraframe decoding block, the corresponding T_i block can be obtained, as follows:

$$T_i(x, y) = \frac{1}{3}\left(X_b^{MC}(x, y) + X_f^{MC}(x, y) + I_i(x, y)\right). \tag{13}$$

The mean square error between T_i and I_i blocks is smaller than that between I_i and X_i blocks. Therefore, when the mean error is zero, the T_i block should satisfy the following:

$$\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} \sigma_{T_i}^2 \leq \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} \frac{\varphi_I^2}{12}, \tag{14}$$

where N is the macroblock size and φ_I is the quantization step size of the intraframe macroblock I_i. Then, the error variance of T_i block is as follows:

$$\sigma_{T_i}^2 = E\left[\left(T_i(x, y) - X_i(x, y)\right)^2\right]. \tag{15}$$

When the I_i block satisfies the condition of (16), that is, the sequence correlation is low or the intraframe decoding quality is low, motion compensation is performed, and the obtained macroblock is used for T_i estimation. The rest will do U_i block estimation.

(2) U_i Block Estimation. For macroblocks that do not satisfy (14), the corresponding MCTI estimation block S_i will be combined with the weighted average to obtain U_i, as shown in (16).

$$U_i(x, y) = \omega I_i(x, y) + (1 - \omega)S_i(x, y), \tag{16}$$

where ω represents the weight of the intraframe decoding block I_i, and the higher the quality of the intraframe decoding block, the greater the weight of the I_i block in the estimation of the corresponding U_i block. The quality of I_i block is evaluated as follows:

$$\sigma_{I_i}^2 = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} \frac{\varphi_I^2}{12}. \tag{17}$$
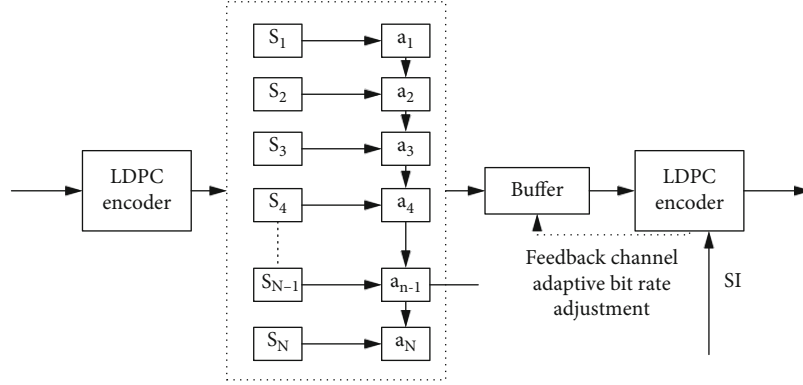
Figure 7: Codec based on LDPCA code. The system consists of LDPC code concomitant encoder and accumulator.
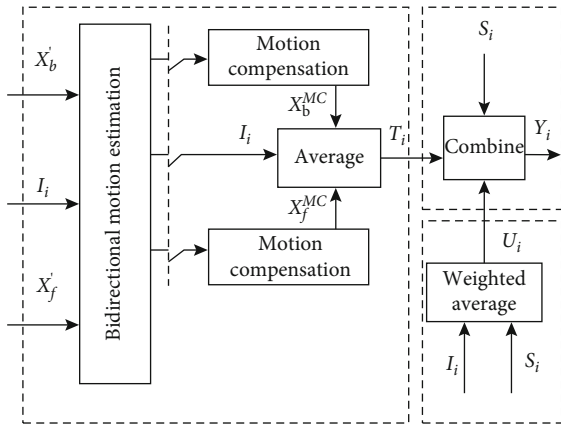

Figure 8: Motion compensation quality enhancement framework. This module consists of T_i block estimation, U_i block estimation, and side information Y_i generation.

The variance of the matching error Z_i between the front and rear reference frames can be calculated as follows:

$$\sigma_{z_i}^2 = E\left[\left(X_b'(x + u_x, y + u_y) - X_f'(x - u_x, y - u_y)\right)^2\right]. \quad (18)$$

Finally, the weight ω is calculated as follows:

$$\sigma_{Z_i}^2 = E\left[\left(X_b'(x + u_x, y + u_y) - X_f'(x - u_x, y - u_y)\right)^2\right]. \quad (19)$$

Among them, quality of estimated $S_i \sigma_{S_i}^2 \approx \sigma_{Z_i}^2$.

(3) The Side Information Y_i. For the generation of side information, only the corresponding macroblock in the reference frame needs to be used as the data block of the side information frames. The generation of SI in this scheme can reduce the number of parity bits that need to be transmitted to the decoding end, thereby improving the RD performance. Secondly, the data amount of the LDPCA encoder can also be reduced, which is beneficial to improve the decoding performance.

The motion compensation quality enhancement is to select a suitable candidate block from the two reference frames before and after and combine the low-quality intra-frame block I_i to obtain a higher-quality macroblock, thereby generating high-quality side information.

3.6. DCT Intraframe Quantization Matrix Design. Because there are many interference factors such as creatures, light, and water flow in marine videos, the scenes of the videos change frequently and violently. In order to reduce the visual sensitivity to scene changes, so as to obtain better visual quality and compression effect, Bossen et al. [48] propose a spatial frequency calculation method, and we design an intraframe quantization matrix based on DCT according to the weighted quantization matrix proposed by Shang et al. [49]. Using $H(u, v)$ to weigh different change frequency coefficients, people will weaken the changes to the scene. $H(u, v)$ represents the sensitivity corresponding to the (u, v) position coefficient, and its calculation is as follows:

$$H(u, v) = \begin{cases} a(b + cf(u, v)) \exp\left(-(df(u, v))^{1.1}\right), & f(u, v) > f_{\max}, \\ 1, & f(u, v) \leq f_{\max}, \end{cases} \quad (20)$$

where a, b, c, and d are all constants and $f(u, v)$ represents the spatial frequency corresponding to the (u, v) position, and its expression is as follows:

$$f(u, v) = \frac{\pi}{180 \sin^{-1}\left(1/\sqrt{1 + \text{dis}^2}\right)} \times \sqrt{f(u)^2 + f(v)^2}. \quad (21)$$

After different frequency coefficients are weighted and adjusted, it has the same sensitivity to the scene, so different coefficients should be quantized with the same length after weighting, as shown below:

$$\begin{aligned} \text{Level}(u, v) &= \text{Round}\left(\frac{\text{Coeff}(u, v) \times H(u, v)}{Q_{\text{step}}}\right) \\ &= \text{Round}\left(\frac{\text{Coeff}(u, v)}{Q_{\text{step}}/H(u, v)}\right), \end{aligned} \quad (22)$$

FIGURE 9: Different categories of video frames from the Brackish Dataset. The video categories include Crab, Fish-big, Jellyfish, Fish-school, and Fish-small-shrimp.

TABLE 2: The corresponding value of QP and q.

Video	QP = 22	QP = 24	QP = 26	QP = 28
Crab	$q = 7$	$q = 9$	$q = 12$	$q = 15$
Fish-big	$q = 7$	$q = 9$	$q = 12$	$q = 15$
Jellyfish	$q = 7$	$q = 9$	$q = 12$	$q = 15$
Fish-school	$q = 7$	$q = 9$	$q = 12$	$q = 15$
Fish-small-shrimp	$q = 7$	$q = 9$	$q = 12$	$q = 15$

TABLE 3: Video quality comparison under DCT intraframe quantization matrix.

Video	QP	Bit rate (kbps)		PSNR (dB)	
		Without matrix	With matrix	Without matrix	With matrix
Crab	22	62208	61057	38.28	39.35
	24	34190	33542	32.54	33.34
	26	16321	15973	25.31	26.47
	28	7953	7806	21.46	22.15
Fish-big	22	61732	59823	37.23	38.21
	24	30132	27959	29.39	30.31
	26	12783	10833	21.98	23.01
	28	6292	4802	18.72	19.43
Jellyfish	22	59831	57932	35.43	36.41
	24	28934	26382	28.23	28.82
	26	12072	11283	24.12	24.89
	28	4827	3987	16.87	17.21
Fish-school	22	60123	58263	37.14	38.24
	24	27092	25873	31.95	32.92
	26	11039	9802	24.21	25.02
	28	6029	4192	20.24	21.94
Fish-small-shrimp	22	63219	61220	38.02	38.92
	24	38378	37218	31.90	32.52
	26	15384	13982	25.42	25.89
	28	8739	6273	20.93	21.67

where $\text{Coeff}(u, v)$ represents the transform coefficient value, the uniform quantization step size Q_{step} is used for different weighted coefficients, and the quantized coefficient is Level (u, v).

This process is equivalent to quantizing the corresponding coefficients $\text{Coeff}(u, v)$ with different quantization step sizes $Q_{\text{step}}/H(u, v)$, so the DCT coefficient intraframe quantization matrix QM_{DCT} can be obtained as shown below:

$$QM_{\text{DCT}}(u, v) = \frac{Q_{\text{step}}}{H(u, v)}. \tag{23}$$

4. Experiments

In order to simulate the real underwater acoustic channel environment, it is necessary to establish a simulator to realize the real environment simulation of underwater video transmission, so as to verify the UnderWater-WZ. The framework is mainly composed of three main modules, namely, underwater acoustic channel simulator, packet loss channel simulator, and video encoder/decoder. Under this framework, the video packet loss situation in the deep-sea can be simulated, and the specific packet loss rate can be set, taking into account the interference of network and bandwidth on video transmission.

4.1. Simulation of Marine Video Transmission Environment

(1) The video transmission process uses Wi-Fi to transmit video frames. The input end is responsible for encoding and sending video frames, and the output end is responsible for decoding and saving video frames

(2) In order to simulate the packet loss channel, the experiment uses the UDP protocol, which uses its nontransportability and does not retransmit after packet loss. The input end preprocesses the frames before sending each video frame, dividing the frames into multiple parts. When data is transmitted over Wi-Fi, some of these chunks are randomly discarded. After the output end receives the data with missing chunks, then the frames enter the video
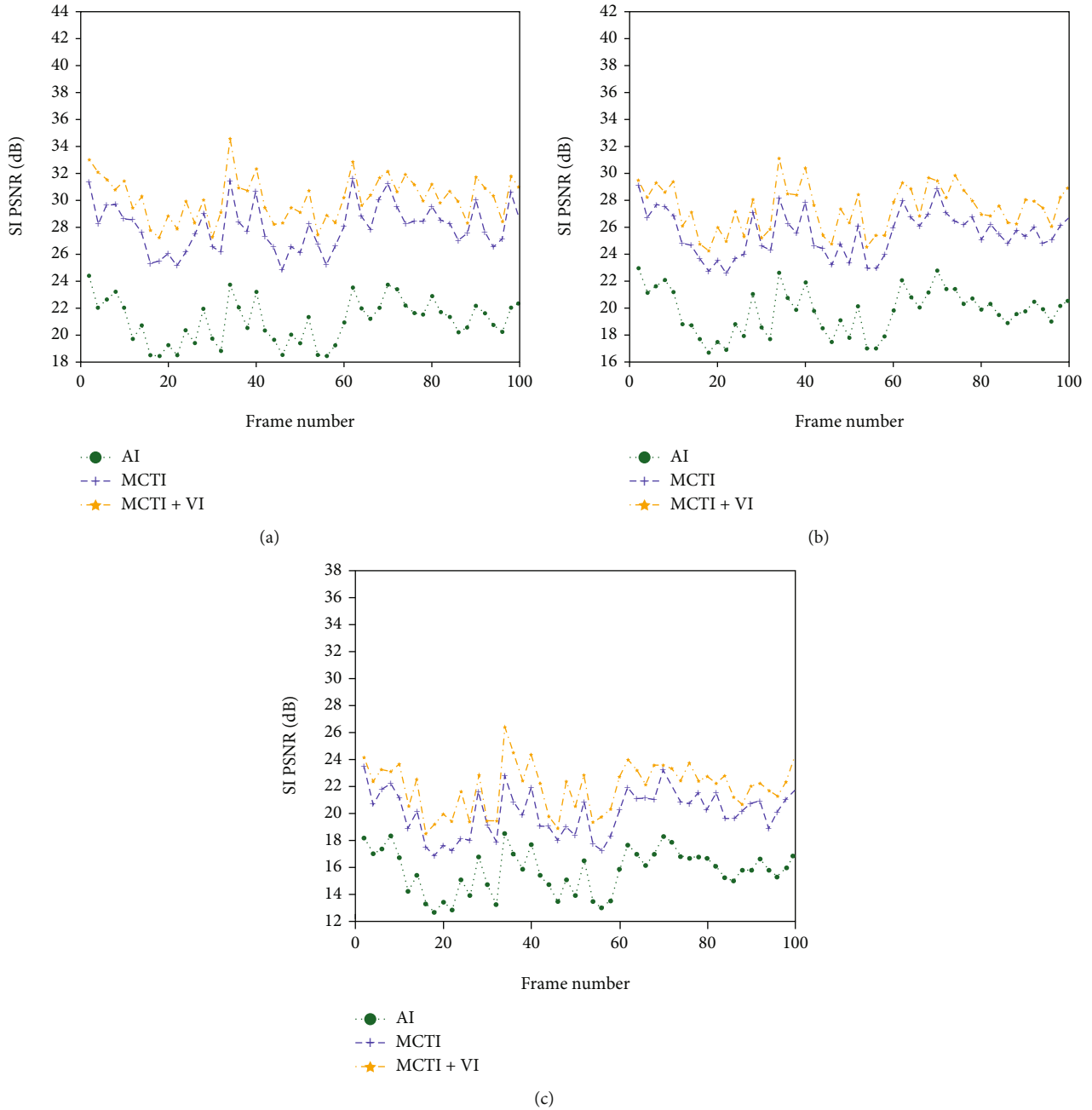
(a)

(b)

(c)

Figure 10: Quality of side information generated in the first 100 frames under the resolution ratio of 1280 × 720. (a) 5% PLRs. (b) 10% PLRs. (c) 15% PLRs.

error-resilience system designed in this paper to restore the frames to form a new video sequence

(3) To evaluate video latency and video quality, we set up one video encoder node to be underwater and another node to be above water. In order to simulate the underwater video transmission environment, the experiment adopts the bandwidth of 200 K, the protocol which is 802.11 N, and the rate of MCS7

4.2. Experimental Parameter Settings. We evaluate the proposed UnderWater-WZ error-resistant coding system on

the Brackish Dataset [50]. The simulation experiment uses five video sequences, Crab, Fish-big, Jellyfish, Fish-school, and Fish-small-shrimp, with different motion characteristics in the dataset. Some video frames in the dataset are shown in Figure 9.

In the experiment, the first 100 frames under different bit error rate environments were tested. The video size is CIF (352 × 288), video format Y : U : V is 4 : 0 : 0, only the luminance component is tested, and the video frame rate is 15 bps.

In UnderWater-WZ, traditional MJPEG intraframe coding is used for key frames, and LDPCA coding is used for
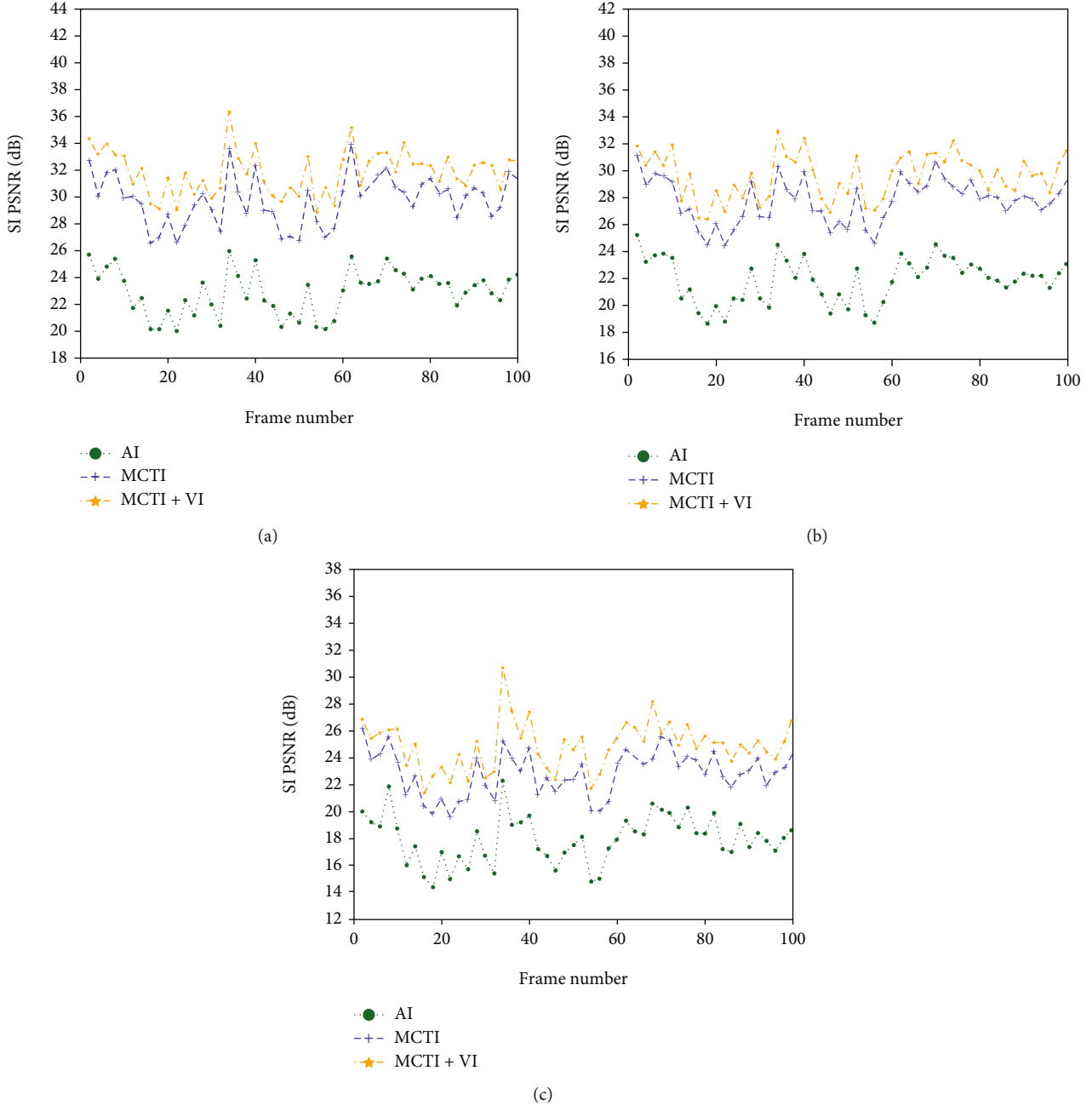
(a)

(b)

(c)

FIGURE 11: Quality of side information generated in the first 100 frames under the resolution ratio of 1600×900. (a) 5% PLRs. (b) 10% PLRs. (c) 15% PLRs.

WZ frames. In the coding scheme designed in this paper, the block size is 4×4, and the quantization matrix is 2. The adopted quantization step size is represented by q, wherein different quantization parameters (QP) correspond to different quantization step sizes. Through a large number of simulation experiments, the corresponding value of QP and q is shown in Table 2 so that a good rate-distortion effect can be obtained.

In the next section, the experimental results are compared in detail and comprehensively, so as to verify the performance and video quality of the scheme proposed in this paper in wireless channel transmission.

4.3. Evaluation Settings. To evaluate the reconstructed video quality when receiving data, we use a video quality evaluation metric: average peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

MSE is the mean squared error between all pixels of the decoded video and the original video, calculated as follows:

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \| I(i,j) - K(i,j) \|^2, \qquad (24)$$

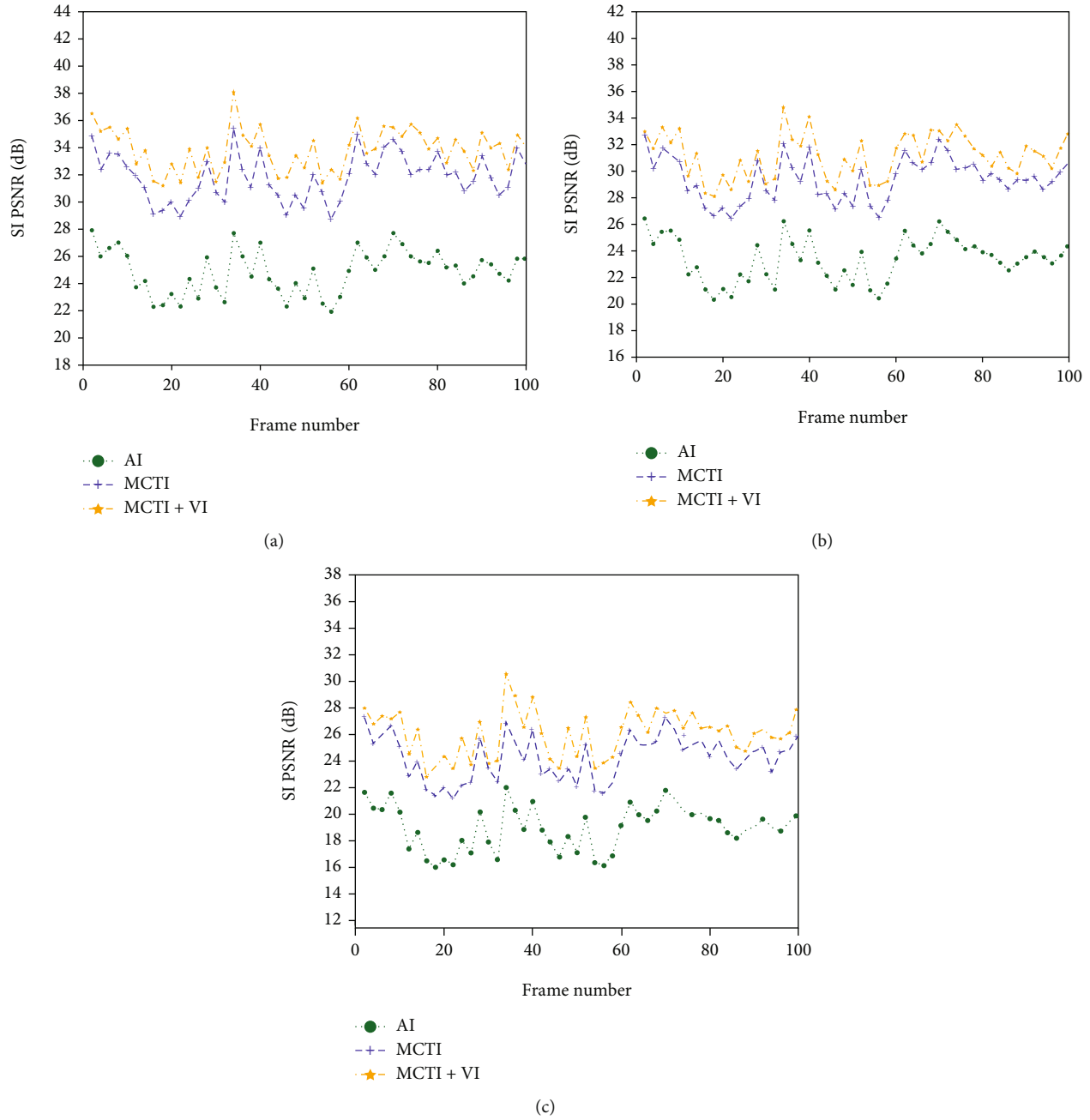where $I(i,j)$ and $K(i,j)$ represent the pixel values of the

(a)

(b)

(c)

FIGURE 12: Quality of side information generated in the first 100 frames under the resolution ratio of 1920×1080. (a) 5% PLRs. (b) 10% PLRs. (c) 15% PLRs.

TABLE 4: Running time of three video coding methods.

Video	Time		
	HEVC	VVC	UnderWater-WZ
Crab	6.43	26.72	18.25
Fish-big	7.05	29.54	20.15
Jellyfish	5.82	24.54	16.83
Fish-school	5.67	23.68	16.27
Fish-small-shrimp	5.35	22.34	15.81

original image and the decoded image at the coordinate position (i, j), respectively, and $M \times N$ represents the image size.

In our case, we also calculated the average video peak signal-to-noise ratio between the reconstructed received video and the original video. As shown in [51], for channel error or data loss conditions, the average peak signal-to-noise ratio is calculated as follows:
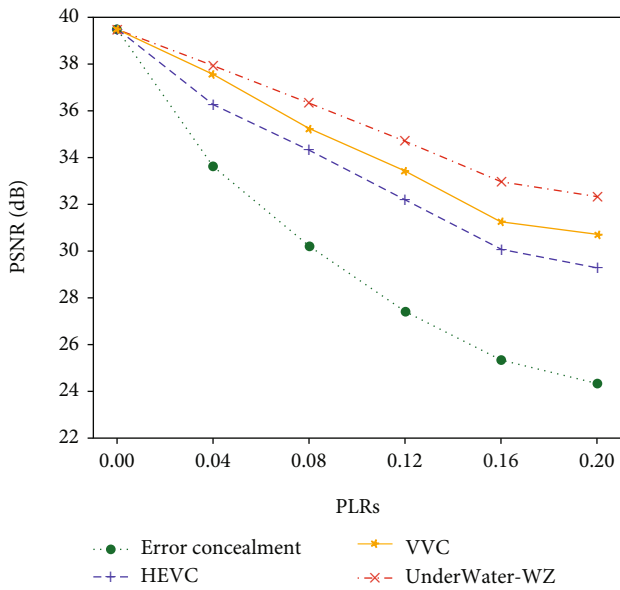
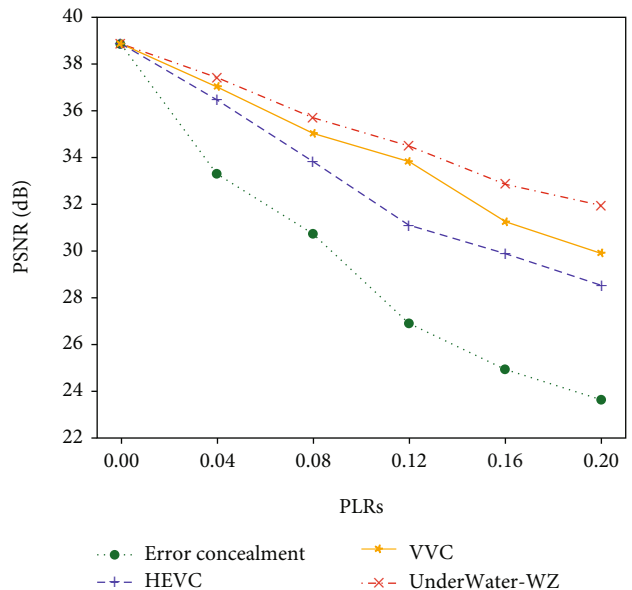$$PSNR_{avg} = 10 \log_{10} \left(\frac{(2^n - 1)^2}{\mu_{MSE}} \right), \tag{25}$$
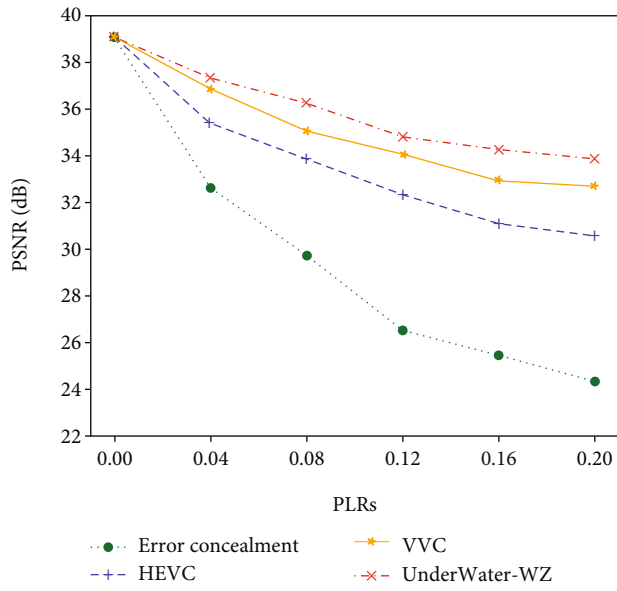
(a)
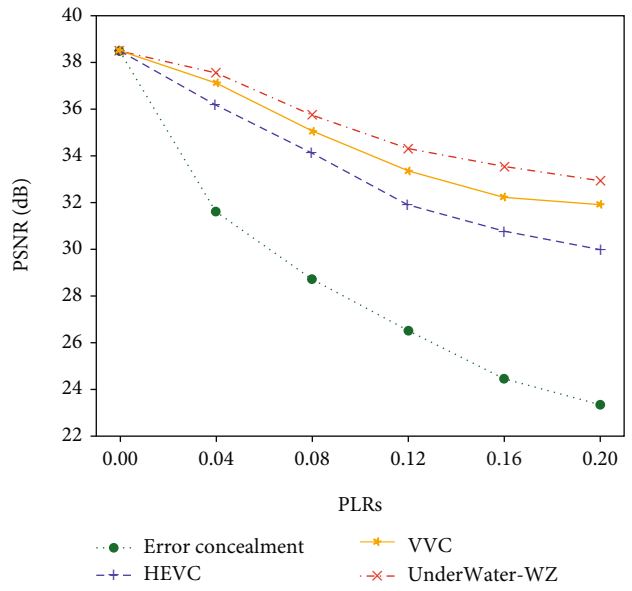
(b)

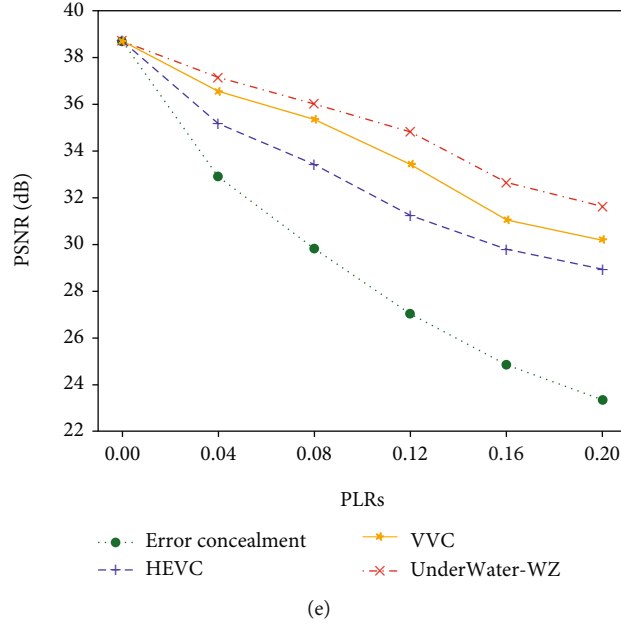(c)

(d)

FIGURE 13: Continued.

(e)

FIGURE 13: Video reconstruction quality under different PLRs. (a) Crab. (b) Fish-big. (c) Jellyfish. (d) Fish-school. (e) Fish-small-shrimp (QP = 26).

where n is the number of bits used to encode the brightness of the pixel, and in the standard case of the image, the components of the pixel are encoded in 8 bits. μ_{MSE} is the average of the frame-by-frame MSE values.

SSIM usually evaluates the similarity of two images through the following three aspects: luminance, contrast, and structure. Given two images x and y, the structural similarity of the two images can be obtained as follows:

$$\text{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}, \quad (26)$$

where μ_x is the mean of x, μ_y is the mean of y, $\sigma_x{}^2$ is the variance of x, $\sigma_y{}^2$ is the variance of y, and σ_{xy} is the covariance of x and y. And $c_1 = (k_1 L)^2$, and $c_2 = (k_2 L)^2$, where L is the dynamic range of pixel values, $k_1 = 0.01$, and $k_2 = 0.03$.

4.4. DCT Intraframe Quantization Matrix. The DCT-based intraframe quantization matrix designed in this paper reduces visual sensitivity and enhances video quality by processing different frequency coefficients. The experiment is verified by Crab under different quantization step size QP, and the relationship between QP, bit rate, and PSNR is obtained, and the experimental comparison is carried out. The experimental results are shown in Table 3.

Summarizing the experimental results, compared with the nonquantization matrix, the quantization matrix designed in this paper further reduces the average bit rate of the test video while improving the PSNR quality of the image. Under different quantization step sizes, the Crab video sequence bit rate is reduced by about 1.93%, 1.89%, 2.13%, and 1.99%, respectively. The video quality PSNR

gains were 1.07 dB, 0.8 dB, 1.16 dB, and 0.69 dB. Experiments show that, without changing the structure of the coding model, by designing an intraframe quantization matrix for DCT, the average video bit rate can be further reduced while improving the video quality.

4.5. Side Information Quality Comparison. The video error-resilience scheme designed in this paper includes the improvement of the side information generation method; that is, the MCTI is combined with the check information transmitted by the encoder to generate high-quality side information. The side information generation methods of the video coding system based on the Wyner-Ziv framework are all through MCTI.
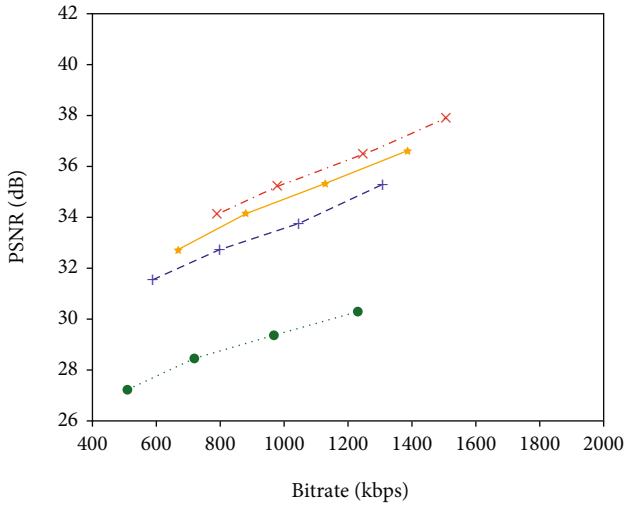
This experiment is verified by the first 100 frames of Crab video sequence under three different resolutions (1280×720, 1600×900, and 1920×1080). Under the premise of 5%, 10% and 15% packet loss rate, compare with the Average interpolation (AI), Motion compensated time interpolation(MCTI) and MCTI+ Verification information (VI) method. The obtained side information PSNR (SI PSNR) quality curve is shown the figure below.

For the resolution ratio of 1280×720, the quality of side information generated in the first 100 frames under different packet loss rates is shown in Figure 10.

For the resolution ratio of 1600×900, the quality of side information generated in the first 100 frames under different packet loss rates is shown in Figure 11.
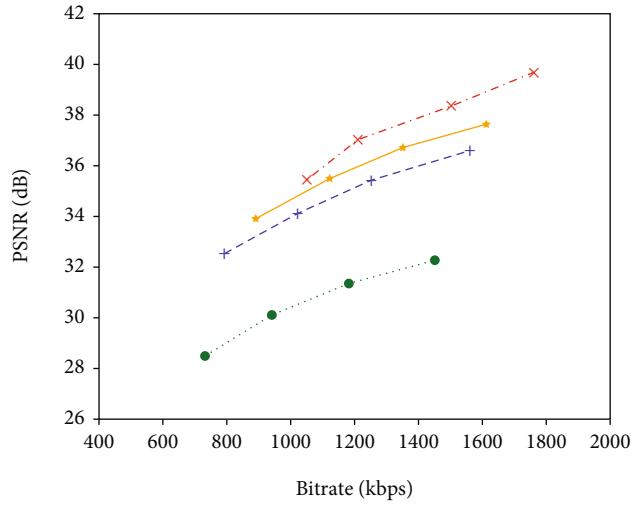
For the resolution ratio of 1920×1080, the quality of side information generated in the first 100 frames under different packet loss rates is shown in Figure 12.

It can be concluded from the experimental results that the quality of side information obtained by MCTI and MCTI+VI methods is obviously better than that of AI, so
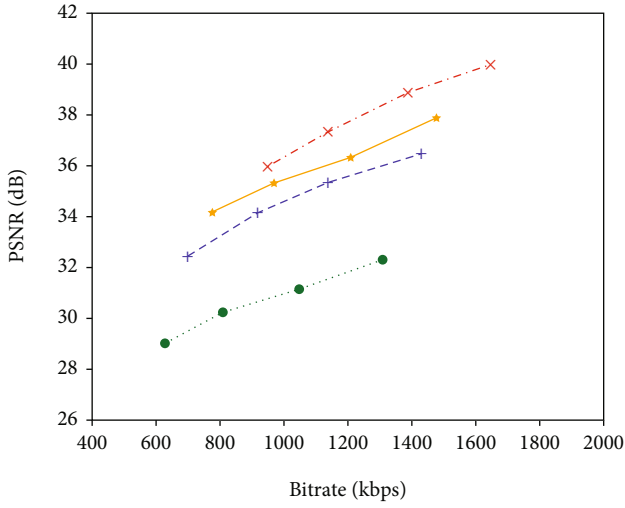
(a)

(b)

(c)

(d)

Figure 14: Continued.

(e)

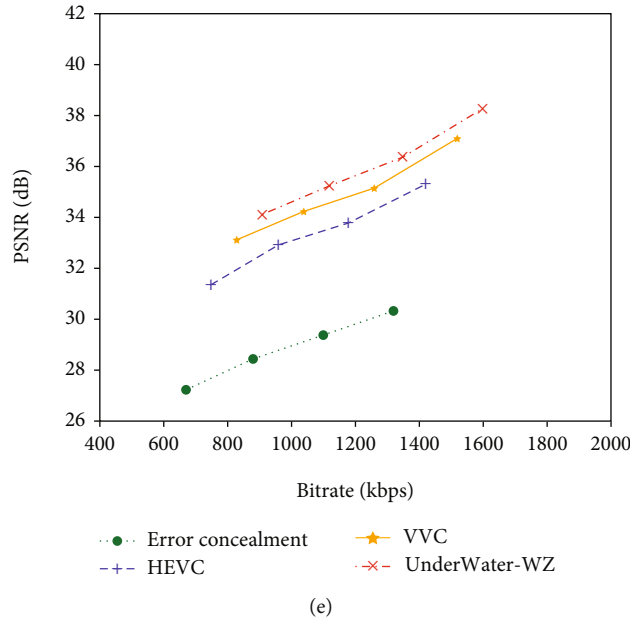FIGURE 14: Rate-distortion performance of reconstructed video. (a) Crab. (b) Fish-big. (c) Jellyfish. (d) Fish-school. (e) Fish-small-shrimp (PLRs = 0.05).

TABLE 5: SSIM comparison of different video sequences processed by different coding systems.

| Video | QP | SSIM | | |
		HEVC	VCC	UnderWater-WZ
Crab	22	0.927	0.957	0.965
	24	0.913	0.931	0.939
	26	0.894	0.918	0.923
	28	0.883	0.904	0.912
Fish-big	22	0.918	0.952	0.967
	24	0.902	0.941	0.948
	26	0.889	0.922	0.936
	28	0.872	0.908	0.916
Jellyfish	22	0.916	0.943	0.951
	24	0.891	0.926	0.932
	26	0.882	0.907	0.914
	28	0.868	0.892	0.904
Fish-school	22	0.925	0.945	0.952
	24	0.908	0.928	0.931
	26	0.892	0.912	0.914
	28	0.879	0.893	0.897
Fish-small-shrimp	22	0.923	0.962	0.965
	24	0.912	0.948	0.949
	26	0.900	0.931	0.937
	28	0.891	0.919	0.921

MCTI must be used to obtain high-quality side information. In addition, with the improvement of resolution ratio, the quality of side information is gradually improved. Therefore, improving the resolution ratio of video sequences is also one of the ways to improve the quality of side information. The scheme in this paper combines MCTI with check information to generate better side information. Compared with AI, the MCTI+VI method can achieve a maximum SI PSNR gain of about 10 dB in the first 100 frames of video sequences. Compared with MCTI, SI PSNR can be improved by about 4 dB at most. In addition, the lower the packet loss rate, the more obvious the effect of the MCTI+VI method. In addition, the resolution ratio is changed from 1280×720 to 1920×1080 in the experiment; the SI PSNR can be increased by about 3.8 dB.

The way of generating side information determines the RD performance of a video coding system. The more similar the side information generated by the encoding end and the decoding end, the higher the RD performance of the reconstructed video, and the better the effect of video error resilience.

4.6. Coding Complexity. In the experiment, we evaluate the coding complexity of the UnderWater-WZ encoder from the running time of the deep-sea video and compare the processing time of HEVC and VVC coding. In this section, the experiment verifies five types of test sequences. For test conditions, the resolution ratio is 1920×1080, the QP parameter is 22, and the experimental results of coding complexity are shown in Table 4.

From the experimental results, we can see that the running time of HEVC coding is nearly 5 times that of VVC coding. It can be seen that compared with HEVC coding, the coding complexity of VVC is quite high, and the main reason of high complexity is that recursive multitype tree partitioning is used in coding. However, for UnderWater-WZ coding, although the running time cannot reach the level of HEVC, the running time is significantly reduced compared with VVC coding.

Through the above experimental results, it shows that the low correlation coding method used in UnderWater-WZ coding effectively reduces the complexity compared with VVC while ensuring that the coding efficiency does not decline significantly.

4.7. Video Quality. In order to verify the effectiveness of the video error-resilience scheme based on underwater acoustic channel designed in this paper, set QP = 26, *q* is 12, and verify that under different PLRs, the five video sequences, Crab, Fish-big, Jellyfish, Fish-school, perform video error-resilience transmission through the simulated marine video communication environment, and then, we obtain reconstructed video with the error-resilience processing. Finally, the experimental results under the error concealment algorithm and HEVC, VVC, and UnderWater-WZ coding systems are compared. The experimental results are shown in Figure 13.

It can be seen from the above experiments that with the increase of the packet loss rate, the PSNR after reconstruction by the error concealment algorithm has the most obvious downward trend, indicating that the video quality after this method is the worst. Based on the HEVC and VVC coding system and the UnderWater-WZ coding system designed in this paper, it is obvious that the quality of the reconstructed video image does not decrease significantly. In contrast, in the simulated deep-sea environment, the video quality after the UnderWater-WZ video encoding system is the best. Even when the packet loss rate is as high as 20%, the scheme proposed in this paper can improve the PSNR of the reconstructed image in the case of bit errors to varying degrees. PSNR can be improved by about 2.6~3.5 dB at most.

Figure 14 shows the rate-distortion performance curves of the reconstructed videos after the error concealment algorithm and HEVC, VVC, and UnderWater-WZ coding processing for three video sequences when the packet loss rate is set to 5%. Here, the error concealment processing adopts the typical bilinear interpolation algorithm in the spatial domain error concealment algorithm. Compared with the error concealment algorithm, the other three schemes can significantly improve the video quality after reconstructing the video. In particular, as the bit rate increases, the video quality improves most obviously when the quantization step size increases. In the deep-sea environment, compared with HEVC, the UnderWater-WZ scheme can improve the PSNR of the video sequence by about 2.8~3.4 dB at most. Compared with the VVC, the PSNR of the video sequence can also be improved by about 0.8~2.2 dB at most.

In order to further verify the improvement of video quality, we evaluate the SSIM indicators processed by the HEVC, VVC, and UnderWater-WZ systems for three video sequences with different step lengths. The data are shown in Table 5. It can be seen from the data in the table that as the step size increases, the SSIM is also decreasing. Compared with the HEVC system, the video sequences processed by the UnderWater-WZ system have higher SSIM. For the Fish-school and the Fish-small-shrimp sequences, since the target of the video is too small, the bit rate obtained is also

small, resulting in insufficient quality improvement. In the follow-up work, the work will also focus on small target videos for further research.

The experimental results show the video error-resilience scheme (UnderWater-WZ) designed in this paper under the underwater acoustic channel. Through the coding scheme that reduces the data correlation and the improved macroblock-based side information generation method, with the assistance of the feedback channel, the bit error rate of video transmission in the deep-sea environment is successfully reduced, and the video quality is significantly improved. Compared with other video error-resilience schemes, the reconstructed video quality obtained by UnderWater-WZ is better, indicating that the video transmission under the underwater acoustic channel is more suitable for processing by the UnderWater-WZ system.

5. Conclusions

In order to solve the problems caused by wireless video transmission in the marine environment, in this paper, a new video error-resilience scheme (UnderWater-WZ) is formed by improving the Wyner-Ziv coding system for high-quality multiview video transmission in marine environment. The implementation process includes controlling error range by using MJPEG coding, combining motion compensation time interpolation with calibration information to generate high-quality side information. And intra-quantization matrix is designed to weaken the change of video scene.

The experimental evaluation results show that in the simulated underwater video transmission environment, after processing the proposed scheme, higher-quality side information can be obtained. The reconstructed video quality obtained by the scheme is significantly better than the Wyner-Ziv system. The PSNR of the video can gain up to 2.6~3. 5 dB. In future work. we will discuss a scheme for multiview video transmission with low latency. We will divide each panoramic video frame into multiple tiles and send the tiles based on the display probability of each tile. In the follow-up work, the intraframe quantization matrix designed in this paper can be extended to the interframe quantization matrix to further improve the coding efficiency of HEVC. In this way, video error-resilience processing in low-latency transmission of underwater panoramic video is realized.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Aracri, F. Giorgio-Serchi, G. Suaria et al., "Soft robots for ocean exploration and offshore operations: a perspective," *Soft Robotics*, vol. 8, no. 6, pp. 625–639, 2021.

[2] Y. Liu, Z. Jing, Q. Liu et al., "All-silica fiber-optic temperature-depth-salinity sensor based on cascaded EFPIs and FBG for deep sea exploration," *Optics Express*, vol. 29, no. 15, pp. 23953–23966, 2021.

[3] J. Kim, "Cooperative localisation for deep-sea exploration using multiple unmanned underwater vehicles," *IET Radar, Sonar & Navigation*, vol. 14, no. 8, pp. 1244–1248, 2020.

[4] B. T. Phillips, S. Licht, K. S. Haiat et al., "DEEPi: a miniaturized, robust, and economical camera and computer system for deep-sea exploration," *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 153, article 103136, 2019.

[5] H. Luo, X. Wang, Z. Xu, C. Liu, and J.-S. Pan, "A software-defined multi-modal wireless sensor network for ocean monitoring," *International Journal of Distributed Sensor Networks*, vol. 18, Article ID 15501477211068389, 2022.

[6] M. Murad, I. A. Tasadduq, and P. Otero, "Coded-GFDM for reliable communication in underwater acoustic channels," *Sensors*, vol. 22, no. 7, p. 2639, 2022.

[7] S. A. El-Meadawy, H. M. Shalaby, N. A. Ismail et al., "Performance analysis of 3D video transmission over deep-learning-based multi-coded N-ary orbital angular momentum FSO system," *IEEE Access*, vol. 9, pp. 110116–110136, 2021.

[8] R. Ma, R. Wang, G. Liu, W. Meng, and X. Liu, "UAV-aided cooperative data collection scheme for ocean monitoring networks," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13222–13236, 2021.

[9] J.-L. Duan, B. Lin, L. X. Cai, Y.-X. Liu, and Y. Wu, "Node deployment of marine monitoring networks: a multiobjective optimization scheme," *Sensors*, vol. 20, no. 16, article 4480, 2020.

[10] F. Campagnaro, R. Francescon, D. Tronchin, and M. Zorzi, "On the feasibility of video streaming through underwater acoustic links," in *2018 Fourth Underwater Communications and Networking Conference*pp. 1–5, Lerici, Lerici, Italy, 2018.

[11] G. Avrashi, S. Museri, Y. Bucris, A. Sinai, and A. Amar, "Underwater video streaming using adaptive frame decimation," in *Oceans 2014 - Taipei*, pp. 1–5, Taipei, Taiwan, 2014.

[12] N. Mirizzi, S. Negahdaripour, and C. Guaragnella, "Underwater forward-scan sonar video coding by background modeling and synthesis for real-time transmission," in *OCEANS 2018 MTS/IEEE Charleston*, pp. 1–7, Charleston, SC, USA, 2018.

[13] D. F. Hoag, V. K. Ingle, and R. J. Gaudette, "Low-bit-rate coding of underwater video using wavelet-based compression algorithms," *IEEE Journal of Oceanic Engineering*, vol. 22, no. 2, pp. 393–400, 1997.

[14] S. Negahdaripour and A. Khamene, "Motion-based compression of underwater video imagery for the operations of unmanned submersible vehicles," *Computer Vision and Image Understanding*, vol. 79, no. 1, pp. 162–183, 2000.

[15] V. N. Nagrale, P. S. Kulkarni, and N. K. Nagrale, "Fast algorithm for underwater video compression," *International Journal of Electronics Engineering*, vol. 2, no. 2, pp. 253–257, 2010.

[16] Q.-Z. Li and W.-J. Wang, "Low-bit-rate coding of underwater color image using improved wavelet difference reduction," *Journal of Visual Communication and Image Representation*, vol. 21, no. 7, pp. 762–769, 2010.

[17] C.-M. Song, B. Fu, X.-H. Wang, and M.-Z. Fu, "A wavelet video coding algorithm with balanced significance probability tree based on energy weighting," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 30877–30893, 2019.

[18] X. Dong, L. Shen, M. Yu, and H. Yang, "Fast intra mode decision algorithm for versatile video coding," *IEEE Transactions on Multimedia*, vol. 24, pp. 400–414, 2022.

[19] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1668–1682, 2020.

[20] X. Li, L. Meng, Y. Tan, J. Zhang, W. Wan, and H. Zhang, "Deep semantic segmentation-based multiple description coding," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 10323–10337, 2021.

[21] F. Boumehrez and R. Brai, "Quality of experience enhancement of high efficiency video coding video streaming in wireless packet networks using multiple description coding," *Journal of Electronic Imaging*, vol. 27, no. 1, p. 1, 2018.

[22] J. Zong, L. Meng, Y. Tan, J. Zhang, Y. Ren, and H. Zhang, "Adaptive reconstruction based multiple description coding with randomly offset quantizations," *Multimedia Tools and Applications*, vol. 77, no. 20, pp. 26293–26313, 2018.

[23] D. Bhardwaj, K. Kant, and D. S. Chauhan, "QoS-aware routing protocol using adaptive retransmission of distorted descriptions in MDC for MANETs," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 28, no. 1, p. 55, 2018.

[24] L. Meng, U. Jie Liang, Y. Z. Samarawickrama, H. Bai, and A. Kaup, "Multiple description coding with randomly and uniformly offset quantizers," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 582–595, 2014.

[25] N. Kamnoonwatana, D. Agrafiotis, and C. N. Canagarajah, "Flexible adaptive multiple description coding for video transmission," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 1, pp. 1–11, 2012.

[26] M. A. Labiod, M. Gharbi, F.-X. Coudoux, P. Corlay, and N. Doghmane, "Cross-layer scheme for low latency multiple description video streaming over Vehicular Ad-hoc NETworks (VANETs)," *AEU-International Journal of Electronics and Communications*, vol. 104, pp. 23–34, 2019.

[27] W. Xu and Y. Chen, "End-to-end rate-distortion optimisation of multiple description coding utilising multiple redundant pictures," *Electronics Letters*, vol. 50, no. 21, pp. 1520–1522, 2014.

[28] D. Zhang, H. Li, and C. W. Chen, "Robust transmission of scalable video coding bitstream over heterogeneous networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 2, pp. 300–313, 2015.

[29] J. Jia, H.-C. Choi, J.-G. Kim, H.-K. Kim, and Y. Chang, "Improved redundant picture coding using polyphase downsampling for H.264," *ETRI Journal*, vol. 29, no. 1, pp. 18–26, 2007.

[30] R. Fernandes, G. Sanchez, R. Cataldo, L. Agostini, and C. Marcon, "Using curved angular intra-frame prediction to

improve video coding efficiency," *Journal of Visual Communication and Image Representation*, vol. 80, article 103291, 2021.

[31] A. Singhadia, P. S. Pati, C. Singhal, and I. Chakrabarti, "Efficient HEVC encoding to meet bitrate and PSNR requirements using parametric modeling," *Circuits, Systems, and Signal Processing*, vol. 41, no. 8, pp. 4479–4511, 2022.

[32] I. Marzuki, Y.-J. Ahn, and D. Sim, "Tile-level rate control for tile-parallelization HEVC encoders," *Journal of Real-Time Image Processing*, vol. 16, no. 6, pp. 2107–2125, 2019.

[33] M. Zhang, Q. Wu, J. Guo, Y. Li, and X. Gao, "Heat transfer-inspired network for image super-resolution reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2022.

[34] X. Zhang, "Superresolution reconstruction of remote sensing image based on middle-level supervised convolutional neural network," *Journal of Sensors*, vol. 2022, Article ID 2603939, 14 pages, 2022.

[35] X. Li and Z. Chen, "Single image super-resolution reconstruction based on fusion of internal and external features," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 1589–1605, 2022.

[36] B. Wang, Y. Zou, L. Zhang et al., "Low-light-level image super-resolution reconstruction based on a multi-scale features extraction network," *Photonics*, vol. 8, no. 8, p. 321, 2021.

[37] R. Jia and X. Wang, "Research on super-resolution reconstruction algorithm of image based on generative adversarial network," *Journal of Physics: Conference Series*, vol. 1944, no. 1, article 012014, 2021.

[38] R. Ramya and M. Senthilmurugan, "An effective super resolution image reconstruction using spatial entropy based normalized deep convolutional neural network," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 18737–18753, 2022.

[39] H. Ahn and C. Yim, "Convolutional neural networks using skip connections with layer groups for super-resolution image reconstruction based on deep learning," *Applied Sciences*, vol. 10, no. 6, p. 1959, 2020.

[40] X. Luo, J. Huang, G. Xiang, H. Fan, Y. Hou, and W. Yan, "Recursively adaptive perceptual non-local means for video coding," in *Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*, pp. 455–462, Singapore, 2021.

[41] Y. Liu, P. Du, and Y. Li, "Hierarchical motion-compensated deep network for video compression," in *Big Data III: Learning, Analytics, and Applications*, pp. 124–131, United States, 2021.

[42] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[43] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002*, vol. 1, pp. 240–244, Pacific Grove, CA, USA, 2002.

[44] A. Aaron, S. D. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *SPIE Proceedings*, pp. 520–528, San Jose, CA USA, 2004.

[45] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: a video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[46] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in *Picture Coding Symposium (PCS'07)*, pp. 513–516, Lisbon Portugal, 2007.

[47] J. Ascenso, C. Brites, and F. Pereira, "Content adaptive Wyner-ZIV video coding driven by motion activity," in *2006 International Conference on Image Processing*, pp. 605–608, Atlanta, GA, USA, 2006.

[48] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1685–1696, 2012.

[49] X. Shang, G. Wang, X. Zhao, Y. Zuo, J. Liang, and I. V. Bajic, "Weighting quantization matrices for HEVC/H.265-coded RGB videos," *IEEE Access*, vol. 7, pp. 36019–36032, 2019.

[50] M. Pedersen, J. Bruslund Haurum, R. Gade, and T. B. Moeslund, "Detection of marine animals in a new underwater dataset with varying visibility," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 18–26, Long Beach, CA, 2019.

[51] A. T. Nasrabadi, M. A. Shirsavar, A. Ebrahimi, and M. Ghanbari, "Investigating the PSNR calculation methods for video sequences with source and channel distortions," in *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1–4, Beijing, China, 2014.