Article

Hierarchical QAM and Inter-Layer FEC for Multi-View Video Plus Depth Format in Two-Way Relay Channels

Dongho You [1] , **Sung-Hoon Kim** [2] **and Dong Ho Kim** [3,*]

1 Department of Information and Communication Engineering, Hannam University, Daejeon 34430, Republic of Korea; dongho.you@hnu.kr
2 Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea; steve-kim@etri.re.kr
3 Department of Smart ICT Convergence Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea
* Correspondence: dongho.kim@seoultech.ac.kr

Abstract: This paper presents an enhanced method for the transmission of 3D video in the Multi-view Video plus Depth (MVD) format over Two-Way Relay Channels (TWRC). Our approach addresses the unique challenges of MVD-based 3D video by combining Hierarchical Quadrature Amplitude Modulation (HQAM), a method that prioritizes data layers based on importance, and Inter-Layer Forward Error Correction (IL-FEC), which protects critical data from errors. These are specifically designed to handle the dual-layer data structure where color data and depth information require different levels of error protection, and it reduces transmission errors and enhances the quality of MVD-based 3D video over TWRC. In the TWRC scenario, the proposed scheme optimizes transmission by reducing the number of relayed bitstreams by half while maintaining high-quality requirements, as demonstrated by significant improvements in the Structural Similarity Index (SSIM) for virtually synthesized views. Furthermore, we identify and optimize the hierarchical modulation parameter (α), which controls the priority and protection levels of different data streams. Systematically varying α reveals its substantial impact on the quality of the reconstructed 3D video, as measured by SSIM. Our results demonstrate that the proposed combination of HQAM and IL-FEC not only maintains the target SSIM of 0.9 for the virtually synthesized view under various relay conditions but also reveals the optimal α value for balancing the error protection between the color and depth map data streams. Notably, while increasing α enhances the protection of critical data (such as color video streams), it may concurrently degrade the quality of less important streams (like depth maps), highlighting the importance of fine-tuning α to achieve the best overall video quality. These findings suggest that our method provides a flexible and effective solution for high-quality 3D video transmission in challenging communication environments, potentially advancing the development of future 3D video delivery systems.

Keywords: hierarchical QAM; inter-layer FEC; multi-view video plus depth; depth image-based rendering; two-way relay channel

1. Introduction

The rapid advancements in 3D video technologies, particularly in formats such as Multi-view Video plus Depth (MVD) [1,2], are transforming fields such as entertainment, virtual reality, telemedicine, and remote collaboration. Unlike 2D video, 3D video requires simultaneous transmission of multiple layers of data, including both color and depth information. This dual-layer structure necessitates specialized transmission techniques that can differentiate between more critical data (color) and less critical data (depth), ensuring robust quality under varying channel conditions. The proposed approach specifically addresses these unique requirements of 3D video transmission by leveraging Hierarchical

Quadrature Amplitude Modulation (HQAM) to prioritize data layers effectively and Inter-Layer Forward Error Correction (IL-FEC) to maintain data integrity in Two-Way Relay Channels (TWRC). With the increasing adoption of MVD content, the challenge of efficiently and reliably transmitting this complex, high-dimensional data becomes critical [3,4].

TWRC offers a promising solution for facilitating bidirectional communication between nodes exchanging MVD content by enhancing spectral efficiency and reducing latency, which is crucial for high-bandwidth applications like real-time 3D video communication [5–7]. However, the inherent high data rates and susceptibility to transmission errors in MVD content posed challenges that our method addressed by optimizing advanced modulation (i.e., HQAM) and error correction techniques (i.e., IL-FEC) to maintain quality and reliability.

HQAM presents a flexible approach to balancing data rate and robustness by organizing data hierarchically and prioritizing different information layers based on their importance [8,9]. To further enhance the reliability of MVD transmission over TWRC, we propose integrating IL-FEC with Hierarchical QAM [10–12]. The IL-FEC technique provides targeted error protection within each hierarchical layer, effectively mitigating the impact of channel impairments and preserving the integrity of critical data. Previous works, such as those presented in [10,12], have addressed the challenges associated with layered video coding, which creates multiple layers of varying importance to progressively enhance video quality. Meanwhile, the study in [11] focuses on issues related to conventional stereoscopic video (CSV) transmission, where left and right video streams are transmitted with one stream prioritized to ensure 2D service continuity. In this paper, we consider the importance problem that arises in the MVD format, and the combined strategy harnesses the strengths of both techniques, aiming to deliver high-quality 3D video content with efficient bandwidth utilization and robust error resilience.

In this manuscript, we explore the theoretical foundations of our proposed scheme, evaluate its performance under various conditions, and demonstrate its effectiveness through simulations. Our approach addresses the unique challenges of MVD content exchange by optimizing error protection and modulation strategies, ultimately enhancing the robustness and quality of transmitted 3D video. This research contributes to the advancement of efficient and reliable 3D video communication.

The rest of this paper is organized as follows: Section 2 details the proposed system model. Section 3 describes the simulation setup and results, highlighting the performance improvements achieved by our approach. Finally, Section 4 concludes with a summary of our findings and suggestions for future research directions.

2. Related Works and the Key Contribution of This Paper

The related works referenced in this paper provide a broad foundation for the proposed method by integrating hierarchical modulation schemes, error correction techniques, and network coding strategies. The proposed approach aims to improve 3D video transmission quality and efficiency by combining these elements to handle the unique challenges of MVD content in TWRC settings. The details are as follows.

2.1. Layered Video Coding Techniques

This paper refers to works that have addressed layered video coding challenges, which involve creating multiple layers of varying importance to progressively enhance video quality. Prior studies, such as those in [13,14], have examined methods for improving the robustness of video transmission by encoding videos in hierarchical layers. These techniques aim to preserve critical data and enhance the overall quality of the transmitted video by prioritizing different information layers based on their importance.

2.2. Conventional Stereoscopic Video (CSV) Transmission

We also consider related works on CSV transmission, where left and right video streams are transmitted with one stream prioritized to ensure 2D service continuity [15].

This approach is relevant to the current study as it addresses issues related to preserving video quality in the presence of transmission errors. The paper builds on these concepts by applying them to the MVD format, which poses additional challenges due to its higher data requirements and complexity.

2.3. HQAM and IL-FEC

This paper integrates previous findings on hierarchical QAM schemes, such as those discussed in [16,17]. Hierarchical QAM offers a flexible approach to balancing data rate and robustness by organizing data hierarchically. This method is beneficial for transmitting MVD content, as it allows for prioritizing more critical information, such as color data, over less critical data, like depth maps. In addition, the proposed approach uses IL-FEC, which has been studied extensively in previous research [13,14]. IL-FEC provides targeted error protection within each hierarchical layer, thereby mitigating the impact of channel impairments and preserving the integrity of critical data. This technique has been used effectively to improve the reliability of video transmissions by reducing the BER in scenarios where multiple layers of video data need to be protected.

2.4. TWRC and Network Coding

The study builds on the concept of TWRC, which has been explored in prior works, such as [10–12], for improving spectral efficiency and reducing latency in high-bandwidth applications. These studies provide a foundation for understanding how TWRC can facilitate bidirectional communication between nodes exchanging MVD content. Furthermore, the paper references the use of network coding strategies that enhance video transmission over relay channels, aiming to optimize the balance between transmission quality and bandwidth efficiency.

2.5. Key Contributions of This Paper

1. Combination of HM and IL-FEC: This paper specifically integrates HQAM and IL-FEC to address the unique dual-layer data structure of 3D video in the MVD format. This combination provides differentiated protection levels for color and depth data, which is essential for maintaining the visual quality of 3D video under various transmission conditions.
2. Appropriate Hierarchical Value (α): The study uniquely explores the impact of a different α value in hierarchical modulation on video quality, providing a systematic way to adjust the value to achieve appropriate SSIM in different relay settings.
3. Efficient Use of Relay Resources: By reducing the number of relayed bitstreams by half, the proposed method improves relay efficiency without compromising video quality, which is a significant advancement over previous methods.
4. Comprehensive Performance Evaluation: The use of both SSIM and BER as performance metrics offers a holistic view of the method's effectiveness in ensuring high-quality 3D video transmission.

3. System Description

We focus on TWRC systems specifically designed for the transmission of 3D video in the MVD format. In this format, nodes A and B communicate via a relay node, R, utilizing network coding based on the Log-likelihood Ratio (LLR). Each node operates under a time-division half-duplex mode, meaning that they cannot transmit and receive simultaneously. Therefore, the proposed system requires three distinct time slots to complete the communication process.

As illustrated in Figure 1, nodes A and B exchange 3D video content formatted in the MVD format, which includes both multi-view video and corresponding depth maps. At node A, the multi-view RGB video and depth information are separately compressed into bitstreams u_1 and u_2, respectively, using multi-view video coding (MVC) [16]. To enhance transmission reliability, we apply the IL-FEC scheme as the channel coding method. In this

process, the bitstream u_1, which contains the critical 2D color video data, is first interleaved and then combined with the depth map bitstream u_2 (i.e., $\prod(u_1) \oplus u_2 = u_{12}$), where $\prod(\cdot)$ represents the Universal Mobile Telecommunication System (UMTS) interleaver, and $\oplus$ denotes the XOR (exclusive OR) operation. The primary purpose of combining these streams (i.e., the XOR operation) is to introduce redundancy, ensuring the important 2D color video data are more robust against errors, interleaving further dispersed errors, and improving overall transmission reliability.
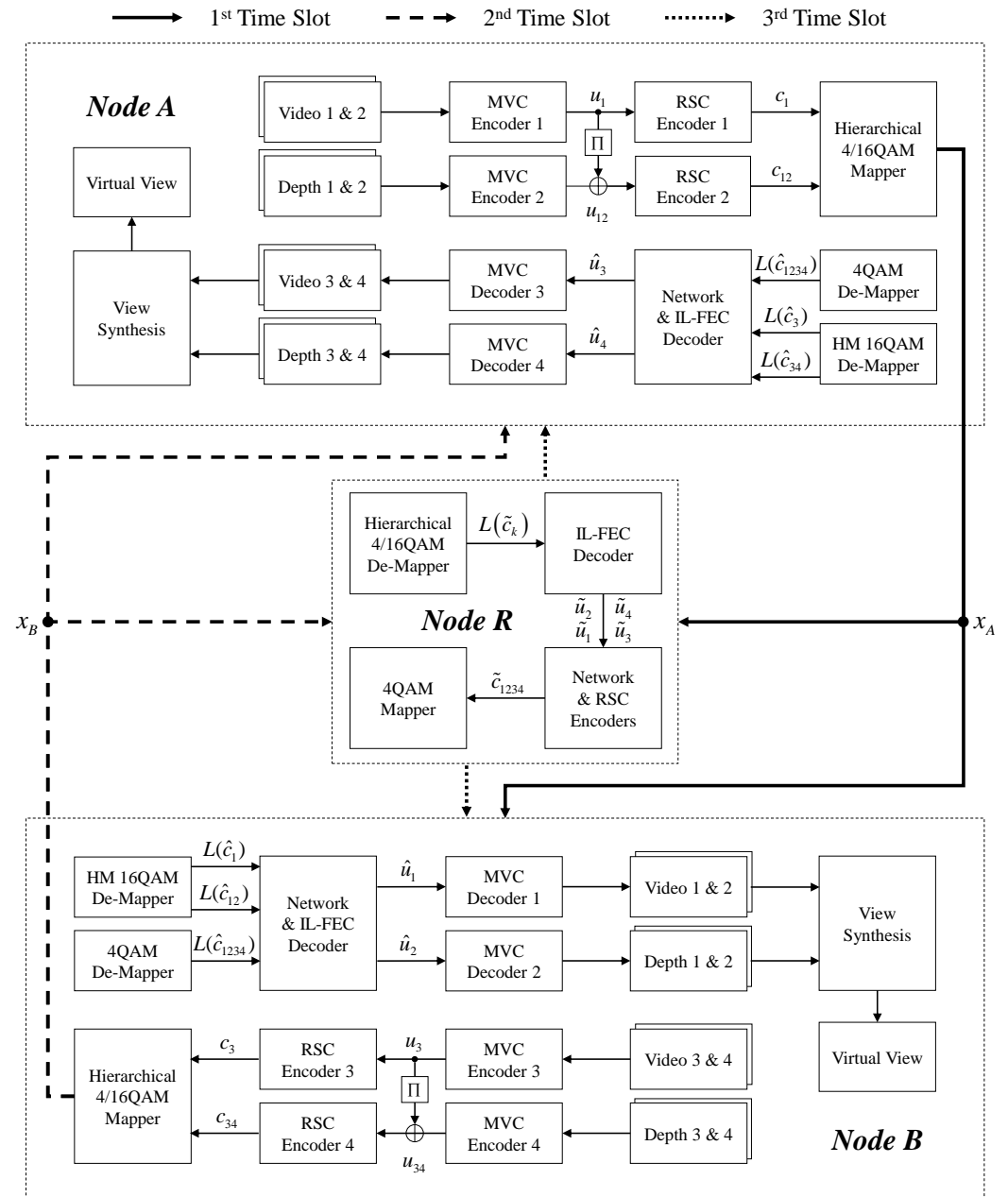

Figure 1. The proposed system model.

Following this, the crucial bitstream u_1 and the combined bitstream u_{12} are encoded into codewords c_1 and c_{12}, respectively, using two Rate-1/2 Recursive Systematic Convolutional (RSC) encoders. These codeword streams are then mapped to different priority levels for modulation: c_1 to high-priority (HP) bits and c_{12} to low-priority (LP) bits. Our approach leverages Hierarchical Quadrature Amplitude Modulation (HQAM) to prioritize the transmission of different data layers—assigning higher protection to the more critical color data (c_1) while allowing the depth data (c_{12}) to receive weaker protection, which is

suitable for its role in depth rendering. This differentiation is crucial for maintaining the perceived quality of 3D video content, as even minor degradation in color data can significantly affect user experience, whereas the impact of errors in depth data is comparatively less perceptible. This hierarchical approach is specifically tailored to 3D video.

Figure 2 presents the constellation diagram for hierarchical 16QAM modulation. In this diagram, black points indicate virtual symbols, while white points represent the actual transmitted symbol x_A. The symbol x_A can be expressed mathematically as:

$$x_A = K \left\{ f_{(\alpha)}^{16} \left(c_1^{(2m-1)}, c_{12}^{(2m-1)}, c_1^{2m}, c_{12}^{2m} \right) \right\} \tag{1}$$

where $1 \le m \le M$, and m represents the m-th bit constituting the transmitted symbol x_A. The function $f_{(\alpha)}^{16}(\cdot)$ represents the hierarchical 16QAM mapper, where α is a hierarchical value that determines the balance of protection between critical and non-critical data by adjusting the distances between constellation points. The hierarchical value $\alpha = d_1'/d_2$, where d_1' is the minimum distance between two actual (white) symbols in adjacent quadrants. It is crucial to note that d_1' is related to d_1 and d_2 through the equation $d_1' = d_1 - d_2$ [17]. Here, $2d_1$ represents the distance between the virtual center points of the quadrants, and $2d_2$ represents the distance between the actual transmitted symbols within each quadrant. Additionally, K serves as a scaling factor to normalize the average symbol energy to unity, as detailed in [13].
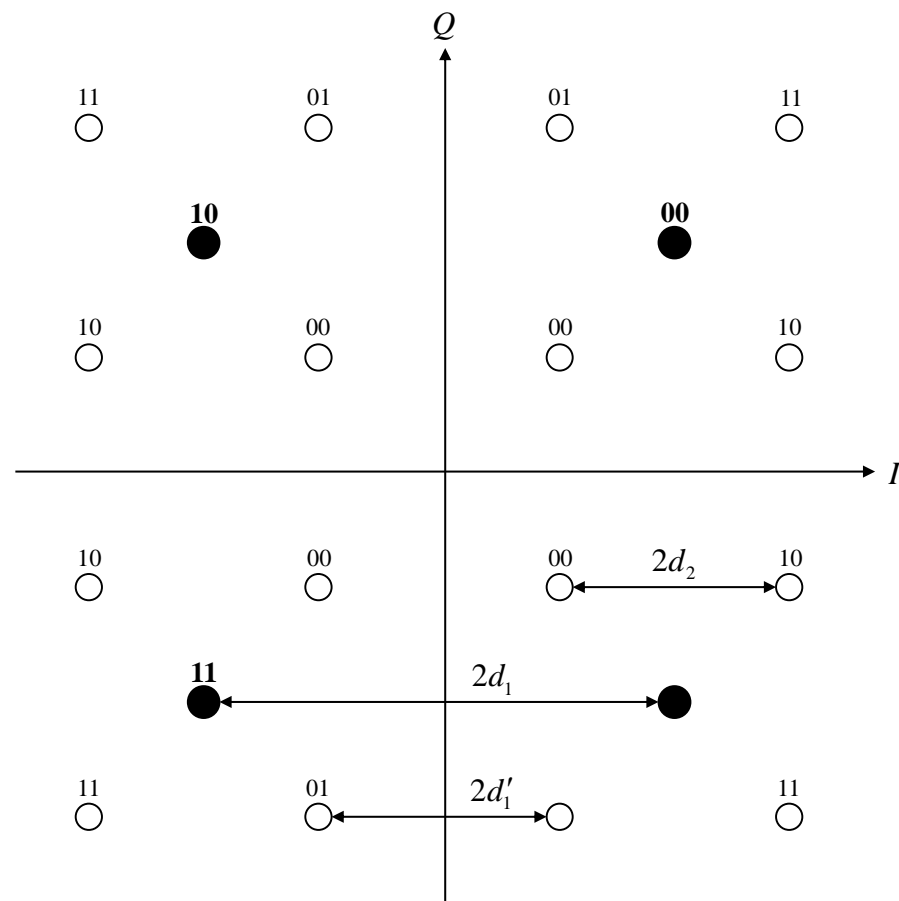

Figure 2. Hierarchical 16QAM constellations: Black points represent virtual symbols, enhancing error protection for critical data, while white points indicate actual transmitted symbols.

In hierarchical 16QAM, the first and third bits represent HP bits, while the second and fourth bits represent LP bits. Generally, increasing the value of α enhances the protection of the important codeword $c1$ but weakens the less important codeword $c12$, and vice versa.

This means that a high value of α does not guarantee high quality because the degradation of c_{12} affects both the RGB and depth videos. Therefore, finding a suitable value of α is crucial for improving the quality of the stereoscopic 3D video. The detailed formulation for determining the optimal value of α is discussed later. The process for generating the symbol x_B at node B is identical to the process for x_A, and thus, it is omitted for brevity.

The transmitted symbols x_A and x_B are directly exchanged between nodes A and B during the first and second time slots and are also received at the relay node R due to the wireless broadcast nature of the transmission. At the relay node, the received symbols are demodulated and decoded into bitstreams $\tilde{u}_1$ through $\tilde{u}_4$. These decoded bitstreams are then re-encoded using network and channel coding methods. This paper explores three such methods, as illustrated in Figure 3. Figure 3a shows the proposed method, while Figure 3b,c depict the first and second reference methods, respectively. In the proposed method shown in Figure 3a, the decoded bitstreams containing 2D color videos ($\tilde{u}_1$ and $\tilde{u}_3$) and the corresponding depth maps ($\tilde{u}_2$ and $\tilde{u}_4$) are first combined into $\tilde{u}_{13}$ and $\tilde{u}_{24}$, respectively, using an XOR operation. These combined streams are then separately encoded using a 1/2 RSC encoder. Following this, $\tilde{u}_{13}$ is interleaved and combined with $\tilde{u}_{24}$ to form $\tilde{u}_{1234}$, and the parity bitstreams $\tilde{p}_{13}$ and $\tilde{p}_{24}$ are punctured using an odd–even puncturing technique. As a result, the length of the concatenated codeword stream $\tilde{c}_{1234}$ is halved, allowing it to be modulated using low-order modulation schemes such as 4QAM or QPSK. In the first reference method (reference 1) shown in Figure 3b, the process begins similarly to the proposed method, with the decoded bitstreams containing 2D color videos ($\tilde{u}_1$ and $\tilde{u}_3$) and the corresponding depth maps ($\tilde{u}_2$ and $\tilde{u}_4$) are combined into $\tilde{u}_{13}$ and $\tilde{u}_{24}$, respectively, using an XOR operation. The channel coding then follows the same process as the existing IL-FEC method used at nodes A and B. Consequently, the length of the relayed codeword stream $\tilde{c}_{1234}$ is identical to that of the transmitted codeword streams from nodes A and B, and this stream is modulated using hierarchical 16QAM. This approach has been proposed in previous studies [6,11] In the second reference method (Reference 2) shown in Figure 3c, the XOR-based network encoding and channel coding using 1/2 RSC encoders follow the same procedure as in the first reference method shown in Figure 3. However, in this method, only the parity streams $\tilde{p}_{13}$ and $\tilde{p}_{24}$ are modulated using QPSK, without the source streams. This approach of retransmitting only parity bitstreams has been proposed in previous studies [14,15].

In the third time slot, nodes A and B demodulate the directly received symbols y_{BA} and x_{AB}, as well as the relayed symbols y_{RA} and y_{RB}, respectively. For this demodulation, each node utilizes a soft-bit demodulator, which is detailed in the appendix. The received symbols can be expressed as follows:

$$y_{ik} = \sqrt{G_{ik}} h_{il} x_i + n_k \tag{2}$$

where G_{ik} and h_{ik} represent the path gain and the complex-valued channel gain, respectively, between transmitter i and receiver k. Additionally, x_i denotes the transmitted symbol from i, and n_k represents additive white Gaussian noise (AWGN) with a variance of $N_0/2$. While more complex channel models (such as fading or shadowing models) could offer a more detailed representation of real-world conditions, the AWGN model provides a fundamental baseline to evaluate the effectiveness of our proposed techniques. This model simplifies the analysis, allowing for a focus on the impact of modulation and error correction strategies on the transmitted 3D video quality. According to [18,19], if the path gain between nodes A and B is assumed to be one, then the path gain between the relay node R and either node A or B is given by:

$$G_{Rk} = \left(\frac{\lambda_{AB}}{\lambda_{Rk}} \right)^{\beta} \tag{3}$$

where λ denotes the distance between nodes, k represents either node A or B, and β is the path-loss exponent. In this study, we consider only the free-space path loss scenario (i.e., $\beta = 2$), as used in [18,19].
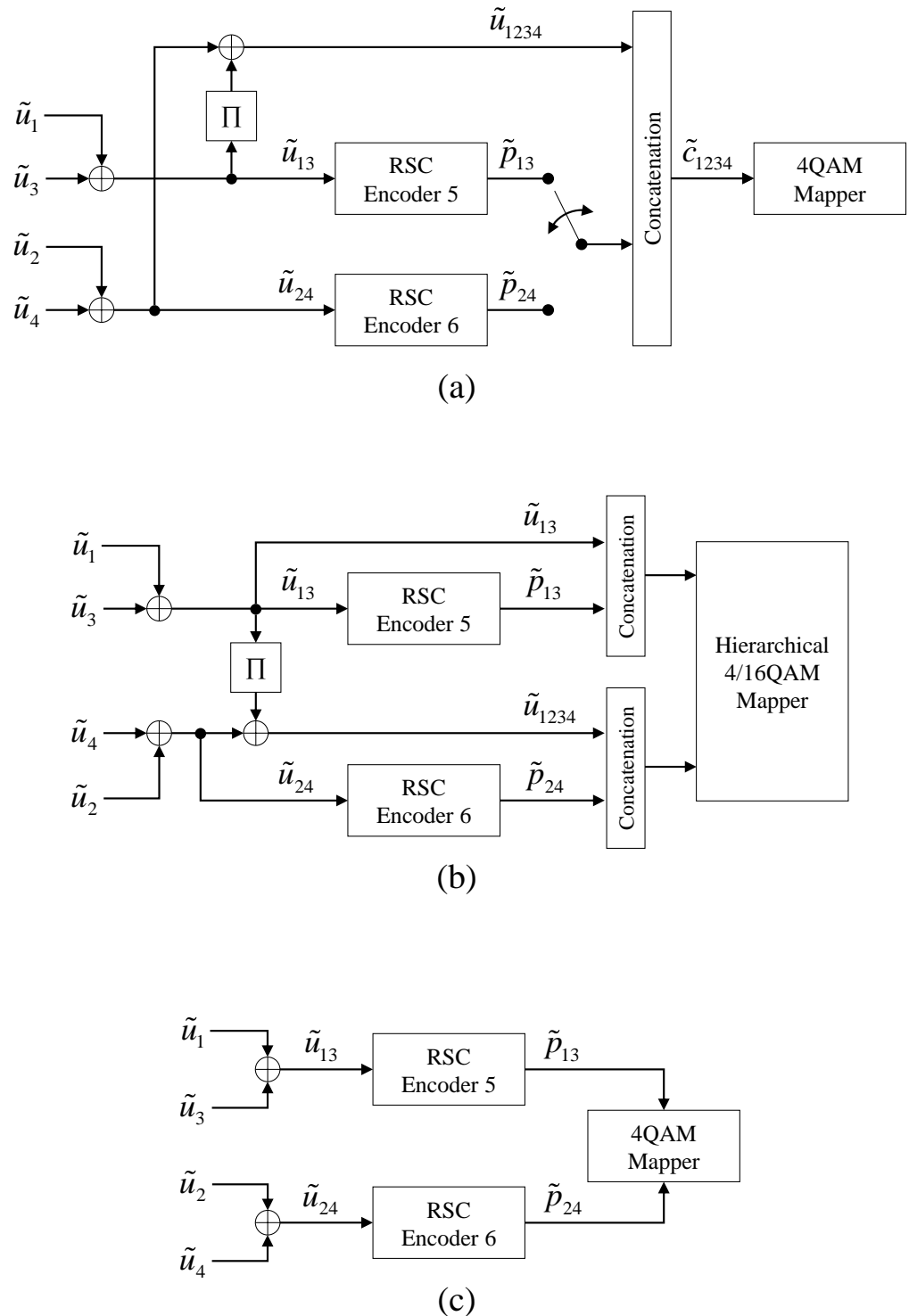

(a)

(b)

(c)

Figure 3. A joint network and channel encoder in node R: (**a**) proposed, (**b**) reference 1, (**c**) reference 2.

After demodulation, the soft-bit codewords are decoded by the iterative network and IL-FEC decoder, as shown in Figure 4, which illustrates the decoder for node A. The same principle applies to the decoder for node B, so it is not separately depicted. The soft-bit streams $L(\hat{u}_3)$, $L(\hat{u}_{34})$, $L(\hat{p}_3)$, and $L(\hat{p}_4)$ are directly received from node B, while

$L(\tilde{u}_{1234})$, $L(\hat{p}_{13})$, and $L(\hat{p}_{24})$ are received from node R. The proposed network and IL-FEC decoder provide a priority information (API) before the first half-iteration by using the received soft-bit streams from node R and the soft-bit streams already available at node A (i.e., $L(u_{12})$, $L(p_1)$, and $L(p_2)$). This approach differs from the conventional IL-FEC decoding process shown in Figure 4. The initial APIs are extracted using the $\boxplus$ operation, defined as $L(\tilde{u}_{1234}) \boxplus L(u_{12}) \overset{\Delta}{=} L(\tilde{u}_{1234} \oplus u_{12})$ [20]. Through the extracted $L(\tilde{u}_{34})$, $L(\tilde{p}_3)$, and $L(\tilde{p}_4)$, the directly received soft-bit streams $L(\hat{u}_{34})$, $L(\hat{p}_3)$, and $L(\hat{p}_4)$ can be improved before the iterative decoding process begins.
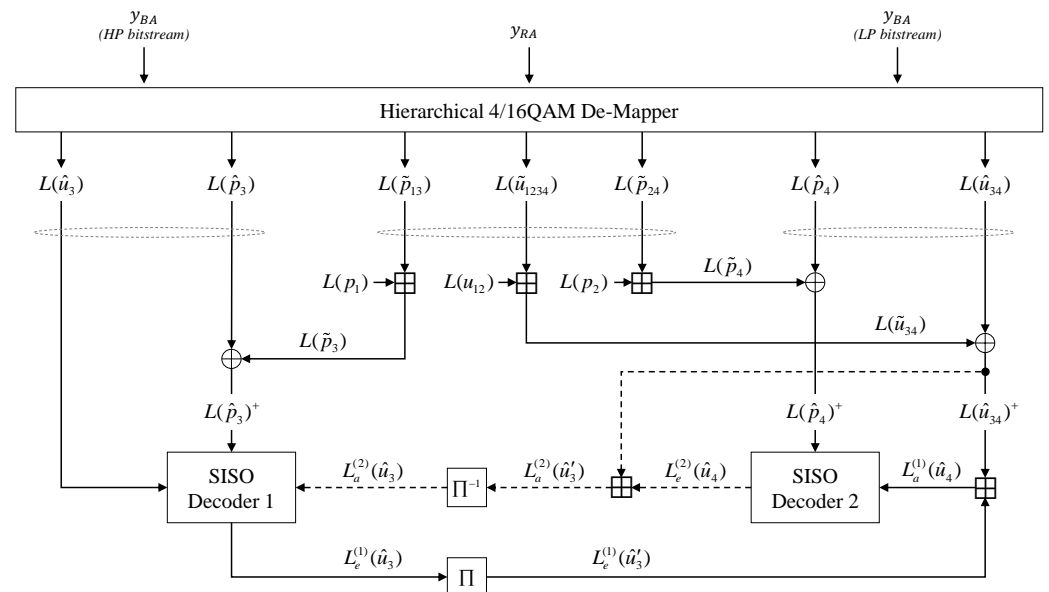

Figure 4. The iterative network and IL-FEC decoder for Node A.

The same demodulation and decoding process applies to $L(\hat{u}_1)$ and $L(\hat{u}_2)$ at node B, so the details for node B are omitted. Once decoding is complete, $\hat{u}_3$ and $\hat{u}_4$ are decompressed into multi-view videos 3 and 4 and the corresponding depth maps. If node A wishes to view an intermediate virtual viewpoint that was not directly received, it can be created using the received MVD content in conjunction with the depth image-based rendering (DIBR) technique [21–23].

4. Performance Evaluation

In this section, we evaluate the performance of our approach in the context of 3D video transmission, assessing the quality of three viewpoint videos (the received left and right views, along with the virtually generated center view) using the SSIM index. This evaluation demonstrates the performance improvement of the proposed network and channel coding method in comparison to the referenced methods. For this assessment, *Champagne Tower* and *Pantomime* MVD sequences are used, each with a resolution of 1280×960 and a frame rate of 29.411 fps [24]. Additionally, MPEG's View Synthesis Reference Software (VSRS) 4.1 is employed to generate the virtual center view [25]. The VSRS implements DIBR to perform view synthesis. Essentially, the reference software uses depth images to render new views, showcasing how DIBR is applied in practical scenarios. The overall experimental setup is illustrated in Figure 5.
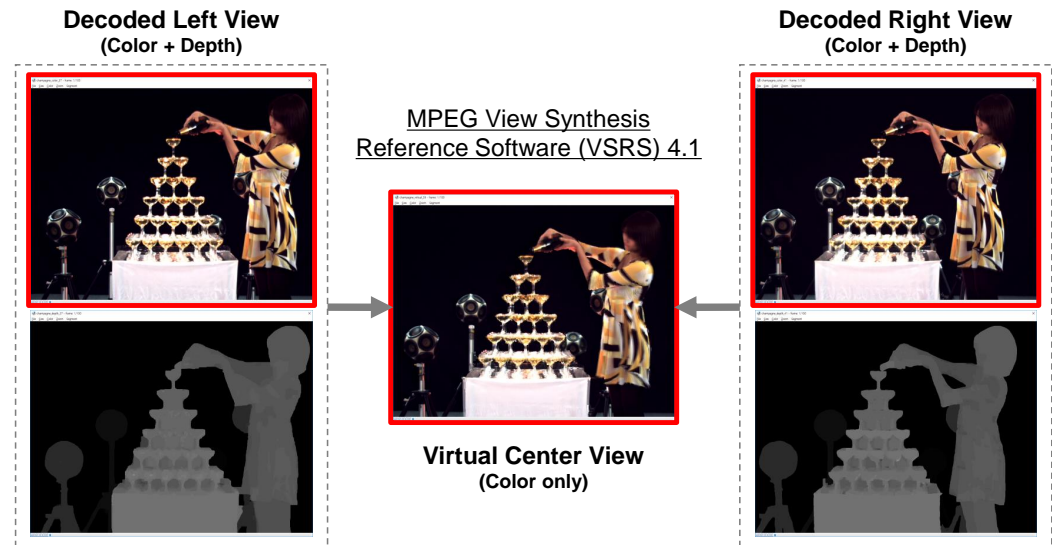
Figure 5. The example of view synthesis using MPEG VSRS.

4.1. Bit Error Rate Performance

The figures from Figures 6–8 present the Bit Error Rate (BER) performances of the proposed and referenced network and channel coding methods, as illustrated in Figure 3. In each figure, subfigures (a) and (b) depict the BER performances for the color video and depth map streams, respectively. The X-axis represents the hierarchical value (α), the Y-axis represents the relay location (λ), and the Z-axis represents the BER. For the theoretical analysis of LLR and the exact BER calculation used in this paper, please refer to the appendix of [11] and [17], respectively.
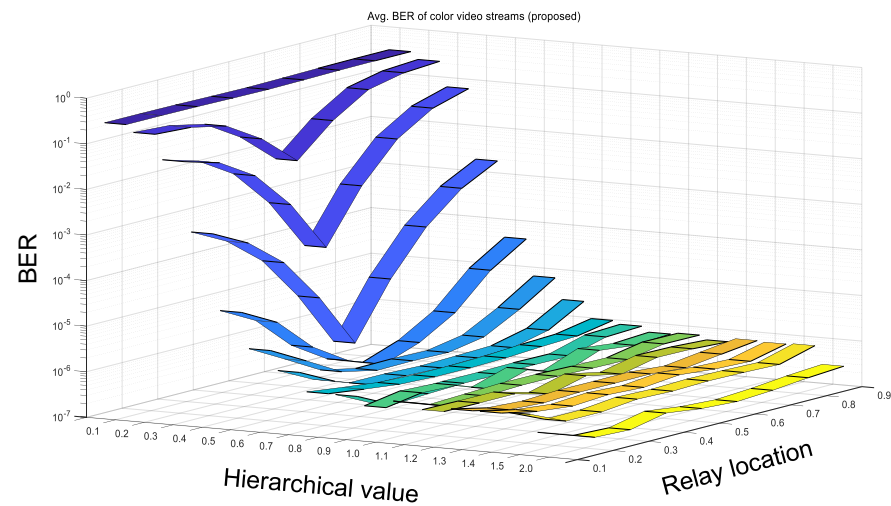
In the case of the proposed method shown in Figure 6, the BER performance for the color video stream improves as α increases, while the BER for the depth map stream deteriorates. This trend is consistent across different relay locations, except when α is between 0.1 and 0.4. Additionally, the target BER is set to 10^{-6} for the more important stream (color video) and 10^{-5} for the less important stream (depth map). As a result, the color stream meets the target BER when α is greater than 0.6, while the depth stream meets the target BER only when α is exactly 0.6.

For the first referenced method shown in Figure 7, the BER performance of the color video stream improves with increasing α, whereas the BER for the depth map stream worsens. However, this method shows suboptimal performance unless the relay node is positioned near the center. Consequently, the BERs for both color and depth streams meet their respective target BERs when α is greater than 0.4 and 0.3, respectively, but only if the relay is located near the center (i.e., λ is between 0.4 and 0.6).

Finally, for the second referenced method shown in Figure 8, the BER performance exhibits a similar pattern to the proposed method in Figure 6, but it can be observed that the second referenced method has a lower BER performance than the proposed method. Based on these BER performances, the following conclusions can be intuitively drawn.

- The qualities of left and right views may be improved as the hierarchical value α increases because the increased α makes the 2D color streams stronger. However, the quality improvement becomes saturated at any moment.
- Relay positioning significantly affected the performance, with central locations providing balanced protection to both streams due to equal path gains, while asymmetric placements favored the nearer stream.
- Increasing the α does not necessarily improve the quality of the virtual center view. This is because the virtual view is typically generated by synthesizing the color views and the related depth maps. In other words, since the quality of the depth maps, which are degraded due to the increased α, also affects the quality of the virtual center view, increasing the α may not always be good for the quality.

- It can be found that the higher modulation level (e.g., 16QAM) is more sensitive to the effect of path loss. Therefore, the proposed and the second referenced methods may show improved qualities of left and right views compared to that of the first reference method, regardless of the relay location.
- However, in the case of the virtual center view, perhaps only certain α values in the proposed method are expected to show improved quality. This is because only a certain α value of the proposed method reaches the target BERs of the 2D color and depth streams, respectively, regardless of the relay location (e.g., $\alpha = 0.6$ in Figure 6).
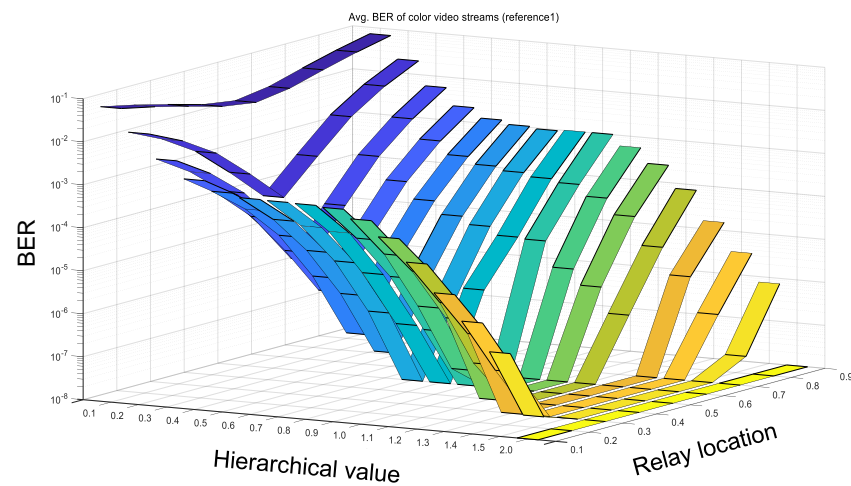

(a) Color video stream
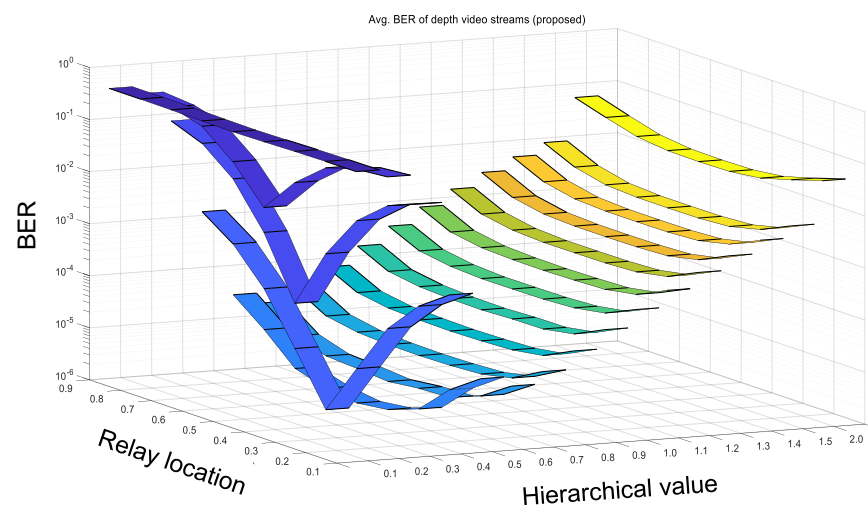
(b) Depth map stream

Figure 6. BER performances of the proposed network and channel coding method: as α increases, the BER for the color video stream improves, while the BER for the depth stream deteriorates, highlighting the trade-off in protection levels between streams.

Avg. BER of color video streams (reference1)

(a) Color video stream
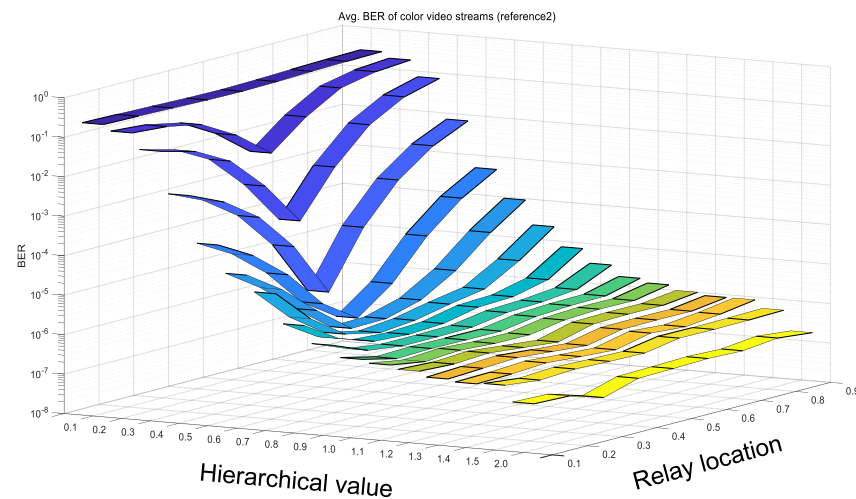
Avg. BER of depth video streams (reference1)
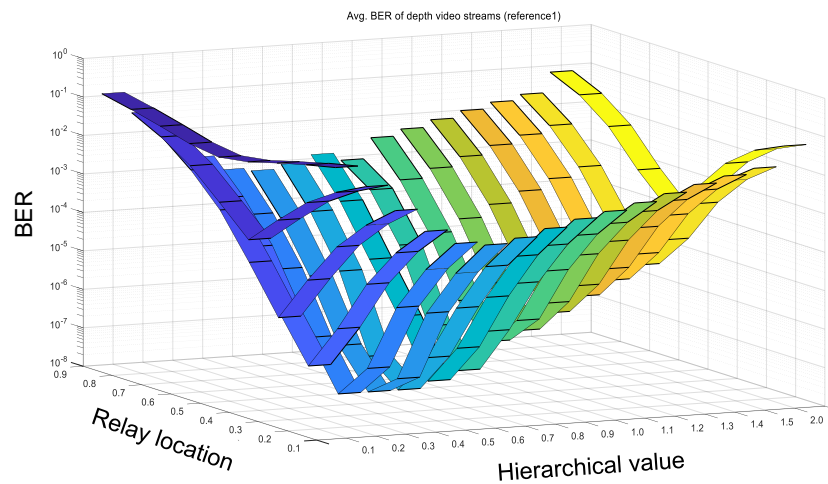
(b) Depth map stream

Figure 7. The BER performances of the first referenced network and channel coding method.
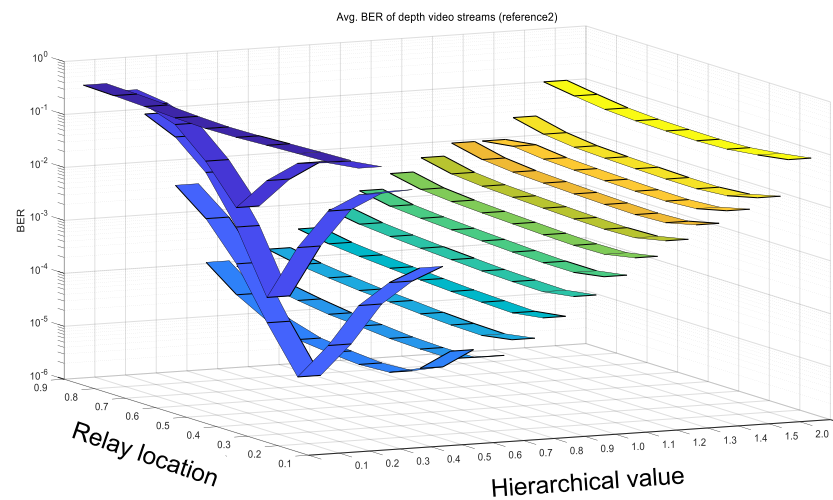
4.2. SSIM Performance

In this subsection, we examine how the BER performances influence the SSIM performances shown in Figures 9 and 10, which correspond to the *Pantomime* and *Champagne Tower* sequences, respectively. The target SSIM value is set to 0.9, as this is generally considered to indicate good quality by human observers [26]. Since the final output displayed to viewers consists of color videos rather than depth videos, the evaluation focuses solely on the quality of the left and right color videos, as well as the virtually generated center color video.

A key interest of this study is to ensure that all receivers obtain high-quality video, regardless of the relay node's location. To this end, the SSIM values obtained from each relay node position are averaged for each hierarchical value (α). This approach means that the overall average SSIM can be impacted if the performance deteriorates at any specific relay position.

(a) Color video stream

(b) Depth map stream

**Figure 8.** The BER performances of the second referenced network and channel coding method.

For the proposed method, the target SSIM is achieved for the reconstructed left and right view videos when α is greater than 0.6, regardless of the test sequences used. In contrast, the first and seocond reference methods require α to be greater than 1.4 and 0.8, respectively, to meet the target SSIM. Both the proposed and second reference methods, which use a low modulation order like 4QAM, demonstrate improved SSIM performances across various α values, regardless of relay location. However, when α exceeds 1.4, all network and channel coding methods show saturated SSIM performances, irrespective of the relay position, indicating that the proposed method does not significantly enhance SSIM for the reconstructed left and right view videos in these conditions. In other words, the proposed method consistently outperformed the referenced methods in terms of BER and SSIM, especially for critical color data. However, the first referenced method showed advantages in scenarios where relay positioning was less central, indicating its suitability in environments with less ideal relay placements.
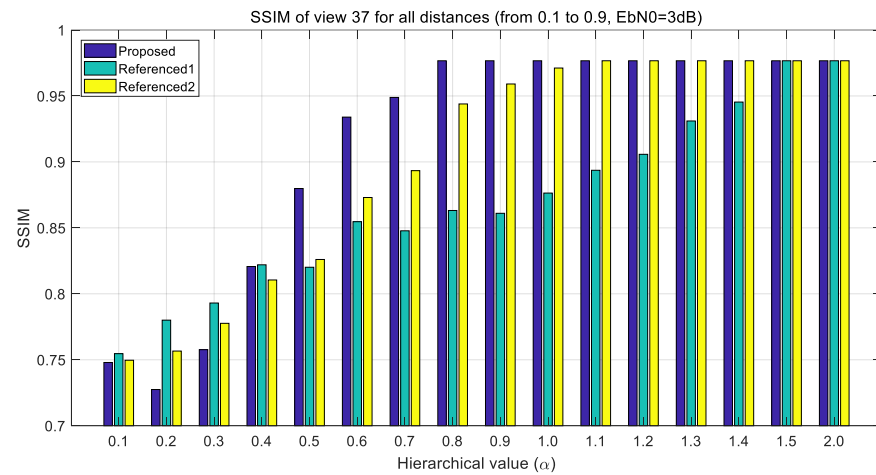
(a) Left view
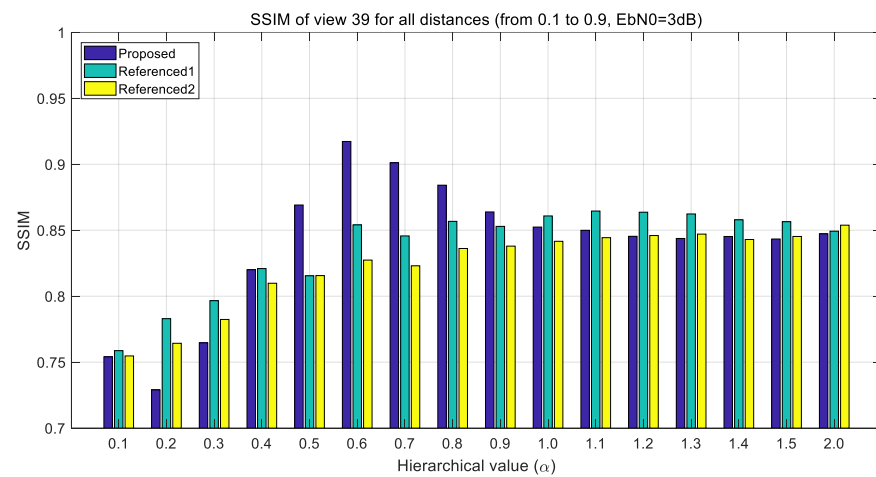
(b) Virtual center view
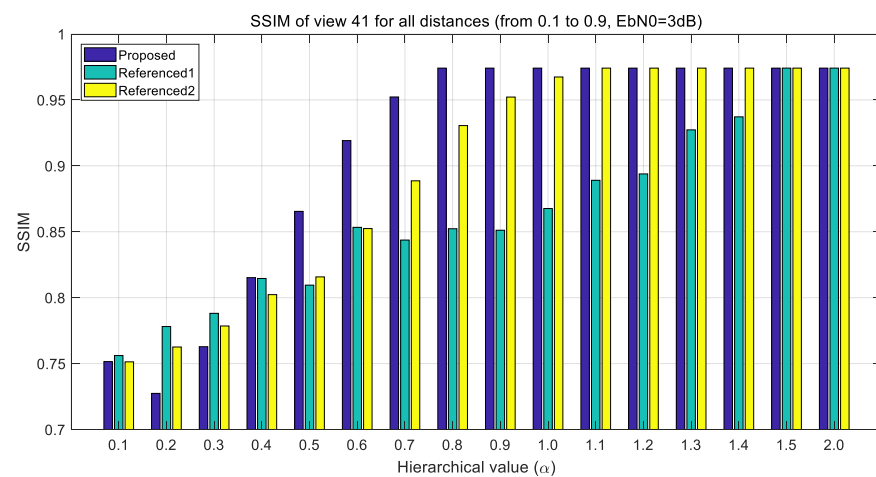
(c) Right view

Figure 9. The SSIM performances of the *Pantomime* test sequence.

(a) Left view

(b) Virtual center view

(c) Right view

Figure 10. The SSIM performances of the *Champagne Tower* test sequence.

The virtual center view presents a different scenario, as shown in Figures 9b and 10b. For the proposed method, the target SSIM is met when α is set to 0.6. While the virtual view for the *Champagne Tower* sequence slightly falls short of the target SSIM, this is not a major issue due to the complexity of the subject in this sequence. To illustrate this subjectively,

Figure 11 presents four video frames. Figure 11a shows the original view (SSIM = 1), and Figure 11b shows a virtually generated center frame (SSIM = 0.9025), generated by VSRS under error-free conditions. For *Champagne Tower*, the target SSIM is not met even without transmission errors because the complexity of the champagne tower subject causes a reduction in synthesis quality. However, as shown in Figure 11c, when the SSIM drops to about 0.85, the quality is nearly indistinguishable from the error-free case. If the SSIM drops further to 0.8, the degradation becomes visible to the naked eye.
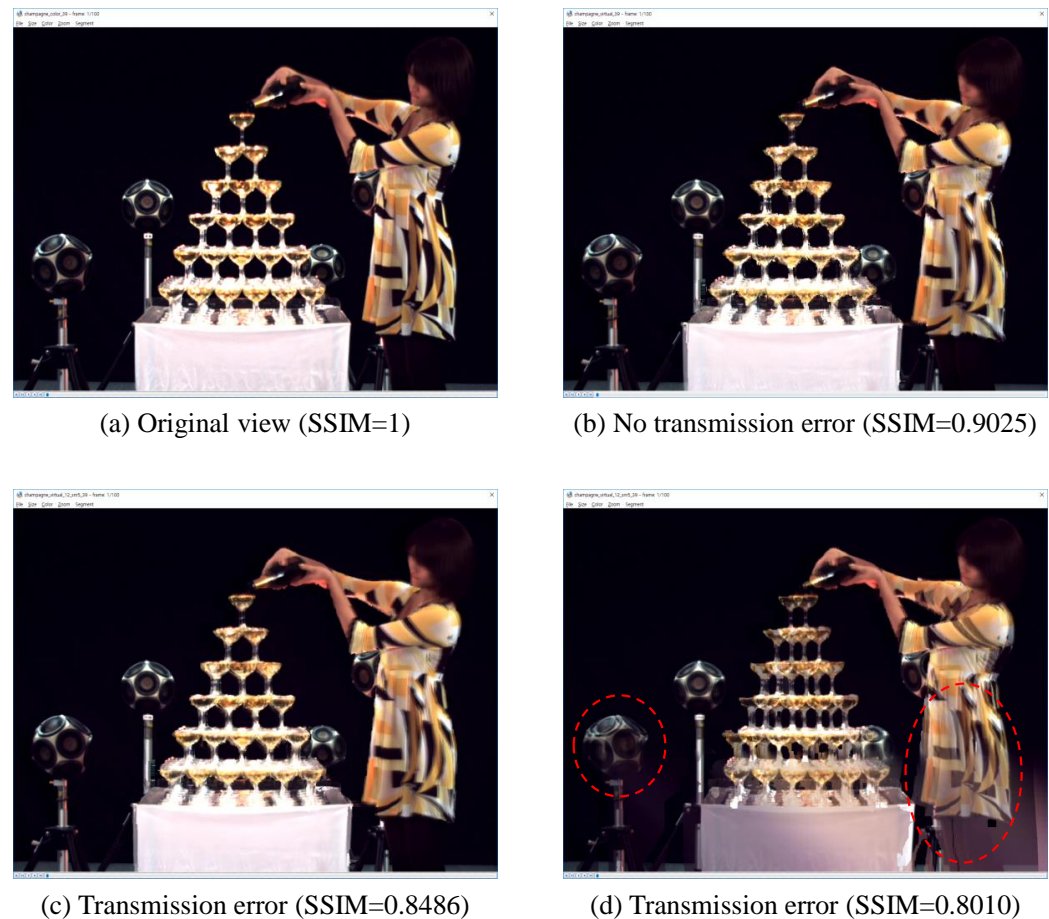

(a) Original view (SSIM=1)

(b) No transmission error (SSIM=0.9025)

(c) Transmission error (SSIM=0.8486)

(d) Transmission error (SSIM=0.8010)

Figure 11. The subjective assessment of the *Champagne Tower* test sequence.

The key reason the proposed method with $\alpha = 0.6$ excels in performance for the virtually generated center view is that both the 2D color video streams and the associated depth map streams meet the target BERs simultaneously, regardless of relay location, as illustrated in Figure 6. In summary, our evaluation demonstrates that the proposed method significantly outperforms the referenced methods in terms of BER performance, particularly for color video streams. This directly translates into better video quality, as evidenced by the SSIM results.

5. Conclusions

In this paper, we explore the use of hierarchical 16QAM with IL-FEC for transmitting 3D video in MVD format over a Two-Way Relay Channel (TWRC), where two users communicate with the assistance of a single relay node. Additionally, we propose a novel network and channel coding method for DF relaying and compare it with two other referenced methods. The proposed method not only reduces the number of relayed bitstreams by half but also provides initial API information to both receiving ends before the iterative decoding process begins. These enhancements significantly improve BER performance, which directly impacts the quality of the reconstructed videos.

For performance evaluation, we assess the SSIM performance based on the quality of the reconstructed left and right view videos, as well as the virtually generated center view video. For the left and right view videos, both the proposed and referenced methods demonstrate improved SSIM performance, with minor variations depending on the hierarchical value α, irrespective of the relay location. However, for the virtually generated center view video, only the proposed method with $\alpha = 0.6$ consistently meets the target SSIM of 0.9, regardless of the relay location.

The main contribution of this paper is the introduction of a network and channel coding method suitable for use in the relay node, and the identification of the optimal α value to enhance SSIM performance for reconstructed left and right view videos, as well as the virtually generated center view video when transmitting MVD-formatted 3D video in a TWRC setting. This study not only contributes to the field of 3D video transmission but also lays the groundwork for future research in optimizing transmission strategies for other high-dimensional media formats.

In addition, there are several potential directions for future research to expand upon our findings. These include exploring higher-order modulation schemes (e.g., 64QAM, 256QAM) to enhance data rates, developing adaptive coding and modulation techniques to dynamically adjust based on real-time channel conditions, and extending the approach to MIMO systems (e.g., movable antennas [27,28]) to leverage spatial diversity for improved transmission reliability. Additionally, applying the method to other 3D formats, such as Point Cloud Compression (PCC) [29] or Light Field (LF) video, testing it under more realistic channel models and network conditions, and pursuing cross-layer optimization strategies could further enhance its applicability and performance. Evaluating our approach against other advanced error correction techniques, like LDPC or polar codes, and developing a prototype for real-world testing are also promising areas for future exploration.

Author Contributions: Conceptualization, D.Y.; methodology, D.Y.; software, D.Y.; validation, D.Y. and D.H.K.; formal analysis, D.Y.; investigation, D.Y.; resources, S.-H.K.; data curation, D.Y.; writing—original draft preparation, D.Y.; writing—review and editing, D.Y., S.-H.K. and D.H.K.; visualization, D.Y.; supervision, D.H.K.; project administration, S.-H.K.; funding acquisition, D.Y. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Merkle, P.; Smolic, A.; Muller, K.; Wiegand, T. Multi-View Video Plus Depth Representation and Coding. In Proceedings of the 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 16 September–19 October 2007; Volume 1, pp. I-201–I-204. [CrossRef]
2. Chan, Y.L.; Fu, C.H.; Chen, H.; Tsang, S.H. Overview of current development in depth map coding of 3D video and its future. *IET Signal Process.* **2020**, *14*, 1–14. [CrossRef]
3. Zhang, X.; Zhao, Y.; Tillo, T.; Lin, C.; Xiao, J.; Wang, A. Packetization strategies for MVD-based 3D video transmission. In Proceedings of the 2016 Visual Communications and Image Processing (VCIP), Chengdu, China, 27–30 November 2016; pp. 1–4. [CrossRef]
4. Zhang, X.; Zhao, Y.; Tillo, T.; Lin, C. A packetization strategy for interactive multiview video streaming over lossy networks. *Signal Process.* **2018**, *145*, 285–294. [CrossRef]
5. Bi, C.; Liang, J. Joint Source-Channel Coding of JPEG 2000 Image Transmission Over Two-Way Multi-Relay Networks. *IEEE Trans. Image Process.* **2017**, *26*, 3594–3608. [CrossRef] [PubMed]

6.   You, D.; Kim, D.H.; Salah, H.; Fitzek, F.H.P. Hierarchical FEC and Modulation in Network-coded Two-way Relay Channel. In Proceedings of the European Wireless 2019; 25th European Wireless Conference, Aarhus, Denmark, 2–4 May 2019; pp. 1–6.

7.   Bouteggui, M.; Merazka, F.; Kurt, G.K. Effective capacity of two way relay channels under retransmission schemes. *AEU Int. J. Electron. Commun.* **2020**, *124*, 153321. [CrossRef]

8.   Hu, Z.; Xu, Y.; Deng, Y.; Zhang, Z. Distance-Enhanced Hybrid Hierarchical Modulation and QAM Modulation Schemes for UAV Terahertz Communications. *Drones* **2024**, *8*, 300. [CrossRef]

9.   Zhu, J.; Zhou, S.; Zhao, M. QAMA: Hierarchical QAM based Downlink Multiple Access with a Simple Receiver. *IEEE Trans. Wirel. Commun.* **2024**, 1. [CrossRef]

10.   Huo, Y.; Hellge, C.; Wiegand, T.; Hanzo, L. A Tutorial and Review on Inter-Layer FEC Coded Layered Video Streaming. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 1166–1207. [CrossRef]

11.   You, D.; Kim, D.H. Combined Inter-layer FEC and Hierarchical QAM for Stereoscopic 3D Video Transmission. *Wirel. Pers. Commun.* **2020**, *110*, 1619–1636. [CrossRef]

12.   Zhang, Y. Adaptive Unequal Error Protection for Scalable Video Streaming. Ph.D. Thesis, University of Southampton, Southampton, UK, 2020.

13.   You, D.; Kim, D.H. Normalization Factor for Three-Level Hierarchical 64QAM Scheme. *J. Korean Inst. Commun. Inf. Sci.* **2016**, *41*, 77–79. [CrossRef]

14.   Hausl, C.; Hagenauer, J. Iterative Network and Channel Decoding for the Two-Way Relay Channel. In Proceedings of the 2006 IEEE International Conference on Communications, Istanbul, Turkey, 11–15 June 2006; Volume 4, pp. 1568–1573. [CrossRef]

15.   Hausl, C.; Dupraz, P. Joint Network-Channel Coding for the Multiple-Access Relay Channel. In Proceedings of the 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, Reston, VA, USA, 28 September 2006; Volume 3, pp. 817–822. [CrossRef]

16.   Vetro, A.; Wiegand, T.; Sullivan, G.J. Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard. *Proc. IEEE* **2011**, *99*, 626–642. [CrossRef]

17.   Vitthaladevuni, P.; Alouini, M.S. BER computation of 4/M-QAM hierarchical constellations. *IEEE Trans. Broadcast.* **2001**, *47*, 228–239. [CrossRef]

18.   Ochiai, H.; Mitran, P.; Tarokh, V. Design and Analysis of Collaborative Diversity Protocols for Wireless Sensor Networks. In Proceedings of the IEEE Vehicular Technology Conference (VTC) Fall, Los Angeles, CA, USA, 26–29 September 2004. [CrossRef]

19.   Ng, S.X.; Li, Y.; Hanzo, L. Distributed Turbo Trellis Coded Modulation for Cooperative Communications. In Proceedings of the IEEE International Conference on Communications (ICC), Dresden, Germany, 14–18 June 2009. [CrossRef]

20.   Hagenauer, J.; Offer, E.; Papke, L. Iterative Decoding of Binary Block and Convolutional Codes. *IEEE Trans. Inf. Theory* **1996**, *42*, 429–445. [CrossRef]

21.   Fehn, C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In Proceedings of the Stereoscopic Displays and Virtual Reality Systems XI, San Jose, CA, USA, 19–22 January 2004; Bolas, M.T., Woods, A.J., Merritt, J.O., Benton, S.A., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 2004; Volume 5291, pp. 93–104. [CrossRef]

22.   Zinger, S.; Do, L.; de With, P. Free-viewpoint depth image based rendering. *J. Vis. Commun. Image Represent.* **2010**, *21*, 533–541. [CrossRef]

23.   Chen, X.; Liang, H.; Xu, H.; Ren, S.; Cai, H.; Wang, Y. Virtual View Synthesis Based on Asymmetric Bidirectional DIBR for 3D Video and Free Viewpoint Video. *Appl. Sci.* **2020**, *10*, 1562. [CrossRef]

24.   Tanimoto, M.; Fujii, T.; Fukushima, N. 1D Parallel Test Sequences for MPEG-FTV. ISO/IEC JTC1/SC29/WG11, Doc. M15378, April 2008.

25.   Stankiewicz, O.; Wegner, K.; Tanimoto, M.; Domanski, M. Enhanced View Synthesis Reference Software (VSRS) for Free-Viewpoint Television. ISO/IEC JTC 1/SC 29/WG 11, Doc. M31520, October 2013.

26.   Chang, T.C.; Xu, S.S.D.; Su, S.F. SSIM-Based Quality-on-Demand Energy-Saving Schemes for OLED Displays. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *46*, 623–635. [CrossRef]

27.   Zhu, L.; Ma, W.; Zhang, R. Movable Antennas for Wireless Communication: Opportunities and Challenges. *IEEE Commun. Mag.* **2024**, *62*, 114–120. [CrossRef]

28.   Zhang, R.; Cheng, L.; Zhang, W.; Guan, X.; Cai, Y.; Wu, W.; Zhang, R. Channel Estimation for Movable-Antenna MIMO Systems Via Tensor Decomposition. *IEEE Wirel. Commun. Lett.* **2024**, 1. [CrossRef]

29.   Schwarz, S.; Preda, M.; Baroncini, V.; Budagavi, M.; Cesar, P.; Chou, P.A.; Cohen, R.A.; Krivokuća, M.; Lasserre, S.; Li, Z.; et al. Emerging MPEG Standards for Point Cloud Compression. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 133–148. [CrossRef]