SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I11P107

Volume 12 Issue 11, 86-102, November 2025

Original Article

Error-Resilient Live Video Streaming Using Hybrid Multiple Description Coding and Adaptive Weighted Yamanaka CFA

D. H. Kitty Smailin[1], Y.Jacob Vetha Raj[2]

[1,2]*Department of Computer Science, Nesamony Memorial Christian College, Marthandam, affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu, India.*

[1]*Corresponding Author : kittydhnmcc@gmail.com*

Received: 07 September 2025 Revised: 09 October 2025 Accepted: 08 November 2025 Published: 29 November 2025

Abstract *- Live medical video streaming is essential for accurate diagnosis and clinical decision-making in real-time healthcare applications. However, packet loss, unstable network conditions, and limited bandwidth degrade video quality during wireless transmission. The research problem focuses on reconstructing diagnostically important video frames with high visual clarity under lossy channel conditions. Most existing methods use static redundancy, non-adaptive filtering, or high-complexity encoding, which reduce their reliability in time-sensitive medical environments. This research proposes a hybrid solution combining Hybrid Multiple Description Coding (HMDC) and Adaptive Weighted Yamanaka Color Filter Array (AW-YCFA) to address these challenges. HMDC encodes each video frame into two descriptions and one correction stream. The redundancy is calculated dynamically using frame importance and packet loss probability. This ensures critical frames receive stronger protection and bandwidth is used efficiently. The decoding module includes adaptive concealment and interpolation techniques that reconstruct lost data using spatial and temporal references. AW-YCFA improves visual quality by modifying the color filter layout and applying local weights to refine color channel reconstruction. This step reduces chromatic noise and enhances edge sharpness. The system is evaluated using the Kvasir-Capsule-SEG dataset, which contains real-world capsule endoscopy video sequences. Performance is tested under packet loss rates of 10%, 20%, and 30%. The proposed method achieves a 40.91 dB PSNR and 0.9124 SSIM at a 10% loss, and maintains a 35.62 dB PSNR and 0.9467 SSIM at a 30% loss. Compared to existing methods such as ROI-aware video coding, EDBTC, convolutional coding, ARV, end-to-end distortion modeling, BBAG with HTTP/2, scalable video coding with reversible data embedding, and point cloud-based streaming, the proposed system consistently shows higher reconstruction quality and lower delay. It also achieves a compression ratio of 16.4 and 46 FPS throughput with 184 MB memory usage. These results confirm that the system supports high-speed and reliable transmission with accurate visual reconstruction in critical medical video applications.*

Keywords - *Hybrid Multiple Description Coding, Adaptive Weighted Yamanaka CFA, Medical Video Transmission, Packet Loss Recovery, Video Frame Reconstruction.*

1. Introduction

Live video streaming has become a necessary component in modern healthcare systems [1]. It supports real-time medical procedures such as capsule endoscopy, robotic surgery, and remote diagnosis [2]. These applications need continuous high-definition video with accurate structure and color information [3]. Any disruption in video quality during transmission can affect clinical decisions and patient safety [4]. Wireless networks are often used to transmit medical video [5]. These networks experience packet loss, delay, and interference due to unstable signal conditions [6]. When packets are lost during transmission, parts of the video frame are damaged or missing [7]. This reduces the clarity of visual information needed for diagnosis [8]. Medical systems require error-resilient streaming methods that can recover visual quality without delay or distortion [9].

Several live video streaming methods have been developed to address transmission errors [10]. These methods are based on region-of-interest-aware coding, error diffusion, channel encoding with Viterbi correction, delta frame-based JPEG streaming, and scalable video coding with adaptive packetization [11]. Some methods use fixed redundancy levels to protect important video regions [12]. Others apply channel simulation models and compression techniques to reduce data size during transmission. Frame complexity and motion estimation are used to guide redundancy allocation in some systems. Some methods apply error concealment techniques based on frame interpolation. A few systems support dynamic switching between full and delta frames based on real-time channel feedback [13]. Scalable video coding methods include layer-based correction and reversible embedding to preserve detail. Although these approaches reduce some distortion,

they do not provide reliable performance under real-time constraints and unstable wireless conditions. They are also limited in handling fast motion, texture loss, or continuous packet drops [14].

Medical frames contain fine structures and subtle color patterns that are essential for clinical decisions [15]. These features must be preserved even when parts of the data are lost. Transmission delay must be low to support real-time interaction. At the same time, memory and bandwidth use must remain within acceptable limits to enable lightweight operation. Existing methods fail to meet these combined requirements [16]. Fixed redundancy increases bandwidth even when conditions are stable. Methods with static filtering do not adjust to color distortions or spatial loss. High-complexity encoding systems introduce latency and are not suitable for time-critical medical use. Some models ignore color reconstruction or use simple channel separation without refinement [17]. Others do not support adaptive concealment or lack spatiotemporal recovery modules. Existing video transmission systems encounter significant issues under packet loss, where the frame structure and color consistency are severely degraded. Conventional redundancy control methods are static and fail to adapt under unstable network behavior during continuous medical video transmission. None of the existing systems integrates dynamic redundancy estimation with adaptive color recovery for real-time medical video reconstruction.

The proposed method introduces a combined approach to solve these limitations. It introduces Hybrid Multiple Description Coding and an Adaptive Weighted Yamanaka Color Filter Array. The Hybrid Multiple Description Coding module encodes each video frame into two main descriptions and one correction stream. The system calculates redundancy dynamically using the frame's importance and the estimated packet loss rate. Critical medical frames receive higher protection. Less important frames use minimal redundancy. This adaptive strategy reduces unnecessary data transmission and protects essential regions. The method also supports frame recovery using weighted reconstruction formulas. When partial data is received, the system uses the available description and correction stream to rebuild the frame. When both descriptions are missing, spatiotemporal interpolation is used to estimate the frame using neighboring content. This supports continuity even during severe packet loss.

The proposed Adaptive Weighted Yamanaka Color Filter Array improves color reconstruction. It modifies the traditional Yamanaka pattern by adjusting the density of green filters and refining the layout of red and blue filters. The green channel is essential for luminance accuracy in medical video. The filter array applies local weights to adjust each pixel value based on neighborhood intensity. This reduces chromatic noise and improves edge sharpness. The color channels are adjusted using adaptive formulas based on spatial correlation. Frame synchronization is performed after recovery to ensure consistent playback. The proposed framework combines Hybrid Multiple Description Coding and Adaptive Weighted Yamanaka CFA to enhance recovery precision and visual consistency under loss conditions.

The proposed system is evaluated using the Kvasir-Capsule-SEG dataset. This dataset includes real-world capsule endoscopy video sequences with normal and polyp-containing frames. Polyp frames are used to assess recovery performance in medically important regions. The proposed method achieves consistent performance under packet and description loss conditions using medical video data. Evaluation uses PSNR and SSIM to measure frame structure and visual accuracy. The system is tested against seven existing techniques, including Liu et al [18], Sasi Kumar et al [19], Maheswari et al [20], Ganguly et al [21], Tong Tang et al [22], Nguyen Viet Hung et al [23], and Jiajun Xu et al [24]. The proposed method records the highest PSNR of 42.26 dB and SSIM of 0.9715 under complete data reception. Under partial conditions, the method maintains PSNR above 35 dB and SSIM above 0.94. Competing methods remain below 30.13 dB PSNR and 0.8354 SSIM. The compression ratio reaches 16.4 with low memory use and 46 FPS throughput. Visual and statistical analysis confirm better texture retention and brightness consistency than all other methods. These results support the use of the proposed system in clinical video streaming under unstable networks.

This article is structured into five sections. Section two reviews recent video coding approaches for error-prone transmission. Section three explains the proposed Hybrid Multiple Description Coding and Adaptive Weighted Yamanaka Color Filter method. Section four presents the encoding and decoding steps with performance evaluation. The experimental results compare frame quality, compression, and efficiency using clinical video data. Section five concludes the study by summarizing key outcomes and confirming technical significance.

2. Literature Review

Recent works in error-resilient video transmission present various methods to reduce distortion under packet loss and unreliable networks. These studies focus on coding strategies, redundancy, and concealment techniques to maintain visual quality in demanding applications.

Liu et al [18], propose an error-resilient video coding method for screen content using frame-level redundancy and ROI-aware coding. The method identifies abrupt frames through frame complexity and applies adaptive redundant slice insertion. An R–λ rate-distortion model guides bit allocation and redundancy control. Macroblocks in high-complexity regions are selectively encoded to reduce distortion. Although effective for screen content, the method has limited adaptability to real-time video, lacks chromatic

refinement, and does not apply spatial interpolation for lost frames. These gaps affect precision in medical video streaming, where detail and color accuracy are essential.

Sasi Kumar et al [19], present a method for error-resilient video transmission using Error Diffusion Block Truncation Coding (EDBTC) in mesh-based ad hoc networks. The technique constructs a bitmap video and applies two color quantizers to enhance reconstruction accuracy under packet loss. The encoder employs residual quantization and diffusion kernels to reduce artifacts. Although the method performs well in static scenarios, it lacks robustness for reliable video delivery over unstable channels. This affects video consistency under severe frame loss. The decoder does not support multi-description redundancy and adaptive concealment.

Maheswari et al [20], propose a wireless video transmission method based on puncturing rule-enabled convolutional coding, parallel processing, and Viterbi decoding. The method includes channel coding, binary symmetric error simulation, and frame reconstruction. It processes RGB video streams into binary packets and decodes using incremental redundancy. Performance is evaluated using PSNR, BER, and MSE metrics. The main drawback of this research is its limited frame-level correction and lack of spatiotemporal feature recovery. It does not consider dual-path encoding and adaptive concealment methods required for handling continuous video loss. This affects decoder consistency under high error conditions. The method fails to ensure efficient frame reconstruction in time-critical medical applications.

Ganguly et al [21], propose A-REaLiSTIQ-ViBe (ARV) to improve live video streaming by integrating JPEG-based delta encoding with a responsive transmission protocol. The method uses background subtraction with ViBe and delta encoding to reduce bandwidth. A state machine adapts the system between full and delta JPEG frames based on real-time channel conditions. The packetization scheme optimizes payload delivery. The protocol reacts quickly using feedback to switch modes and control bitrate. This method improves frame rate and visual quality without heavy processing. However, it lacks joint consideration of spatiotemporal continuity and cannot ensure consistent quality during rapid foreground transitions.

Tong Tang et al [22], present an end-to-end distortion model for error-resilient screen content video transmission. Their method addresses packet loss and distortion due to channel errors using joint source-channel coding. Frame classification is performed based on temporal variations, followed by adaptive error concealment for each frame type. A GOP-level distortion model is integrated into rate-distortion optimization. Their algorithm enhances both objective and subjective video quality across various conditions. However, their model targets screen content videos, not natural scenes with complex motion and texture.

Nguyen Viet Hung et al [23], propose a 360-degree video streaming method using HTTP/2 with the BBAG algorithm. The method uses Scalable Video Coding and multiple buffer thresholds to enhance viewing under changing network and user conditions. It adapts bitrates based on the viewport and cancels late tile layers using HTTP/2 features. The approach improves visual quality and saves bandwidth in immersive video delivery. However, this solution depends on advanced buffering and a stable HTTP/2 infrastructure. These conditions are not suitable for systems with frequent packet loss and limited real-time response. The proposed research instead addresses frame-level recovery and lightweight transmission, which are not well supported by BBAG under lossy network environments.

Jiajun Xu et al [24], present an error-resilient approach for scalable video coding to address error propagation caused by inter-layer prediction. The proposed algorithm uses a reversible data embedding technique that preserves lower-layer motion and texture information in the enhancement layer. The wavelet transform replaces the DCT for better quality under low-bitrate conditions. Compared to traditional frame copy methods, this scheme achieves significant PSNR improvements. However, the approach increases computational cost due to additional embedding and transform operations. Moreover, embedding overhead may impact bitrate control.

Patrick Enenche et al [25], studied recent advances in point cloud video streaming for immersive 3D environments. Their work examines AI-driven, user-focused, and low-latency methods to support the metaverse. The study explores techniques involving Mobile Edge Computing, Visible Light Communication, and Data Plane Development Kit for improved quality and bandwidth use. However, point cloud video streaming faces challenges in encoding complexity, latency control, and real-time delivery. These issues limit its use in low-resource systems and constrained networks. Such limitations reduce the suitability for scenarios in the current research, where high-quality video is required with adaptive recovery, low delay, and efficient reconstruction under packet loss.

The comparison in Table 1 summarizes different error-resilient video transmission techniques. Reviewed methods use techniques such as ROI-aware redundancy, delta encoding, convolutional coding, and scalable video coding. Most approaches reduce distortion in controlled environments but face challenges in dynamic or lossy networks [9]. Limitations include poor adaptation to real-time scenes, high processing cost, weak spatial recovery, and limited support for immersive or medical video streams. These gaps affect systems requiring lightweight computation and accurate frame restoration.

Table 1. Comparative study of existing error-resilient video transmission methods

Author et al.	Purpose	Methods Used	Disadvantages
Liu et al. [18]	To enhance error resilience for screen content videos through adaptive redundancy control.	ROI-aware frame coding with redundant slice insertion and R–λ rate-distortion modeling.	Limited adaptability to real-time medical scenes and no spatial interpolation for lost frames.
Sasi Kumar et al. [19]	To improve video transmission in ad hoc networks under packet loss conditions.	Error Diffusion Block Truncation Coding with residual quantization and diffusion kernels.	Performs poorly under unstable channels and lacks multi-description redundancy.
Maheswari et al. [20]	To ensure reliable wireless video transmission using convolutional encoding.	Puncturing rule-enabled convolutional coding with parallel Viterbi decoding.	No spatiotemporal feature recovery and limited correction under continuous loss.
Ganguly et al. [21]	To optimize live video streaming with reduced bandwidth usage.	Delta JPEG encoding with adaptive payload and real-time feedback control.	Does not maintain spatiotemporal continuity during rapid motion changes.
Tong Tang et al. [22]	To address packet distortion in screen content video transmission.	Joint source-channel coding with GOP-level distortion modeling.	Effective only for screen content videos and unsuitable for natural medical frames.
Nguyen Viet Hung et al. [23]	To provide immersive 360-degree streaming under variable network conditions.	Scalable Video Coding with BBAG scheduling over HTTP/2.	High buffering dependency and poor adaptability under frequent packet drops.
Jiajun Xu et al. [24]	To reduce inter-layer prediction errors in scalable video coding.	Reversible data embedding with wavelet-based enhancement layer reconstruction.	Increases computational cost and overhead, affecting bitrate control.
Patrick Enenche et al. [25]	To study streaming efficiency for 3D point cloud video environments.	Edge computing with visible light communication and bandwidth optimization.	High encoding complexity and latency limit real-time operation for medical scenarios.

3. Preliminaries

This section presents the working steps of the proposed Hybrid Multiple Description Coding and the Adaptive Weighted Yamanaka Color Filter.

3.1. Hybrid Multiple Description Coding (HMDC)

Traditional video transmission methods are vulnerable to packet loss, especially in unstable wireless networks. To improve reliability, Multiple Description Coding (MDC) was developed [26]. Figure 1 illustrates the basic MDC structure. This method encodes an input into multiple independent descriptions. Each stream reconstructs a partial or full version of the video.

If one stream is lost, others continue to provide partial data. A center decoder reconstructs the best version when all descriptions arrive. Although MDC improves robustness, it increases bandwidth. The static redundancy allocation wastes resources in stable network conditions. It also fails to prioritize more critical data in degraded network states.

Therefore, in this research, HMDC is introduced. HMDC integrates adaptive encoding and redundancy control. It improves the recovery of frames and preserves color accuracy during transmission loss.

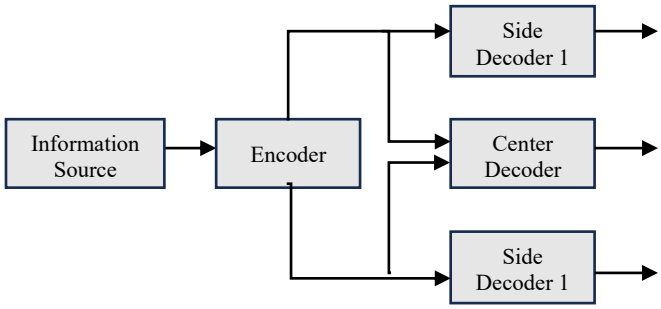


Fig. 1 Standard Multiple Description Coding (MDC)

It processes each frame based on its importance and dynamically adjusts redundancy. The process begins with frame extraction and analysis. Each frame is scored by importance. Important frames are assigned more redundancy. The redundancy factor R is computed using:

$$R = \alpha P + \beta Q \quad (1)$$

Here, P is packet loss probability, Q is frame importance, and α, β are tunable weights. This ensures redundancy is proportional to the network condition and frame significance. The encoder generates two primary descriptions and one correction stream:

$$D_1 = f_1(V) \quad (2)$$

$$D_2 = f_2(V) \quad (3)$$

$$R_c = f_r(V, Q) \quad (4)$$

Where V is the original frame, f_1 and f_2 are partial encoders, and f_r is the redundancy generator based on frame priority. These outputs are packetized and sent through the wireless network. Packet loss is inevitable in wireless transmission. The system monitors packet flow using an error detection module.

Missing packets are flagged for reconstruction. Frame recovery uses the following reconstruction rule:

$$F_{rec} = D_1 + D_2 + \gamma R_c \quad (5)$$

Where γ is a recovery weight based on decoding confidence, if only one description is received, the recovery fallback equation is:

$$F_{side} = D_i + \delta R_c \quad (6)$$

Where $D_i \in \{D_1, D_2\}$, and δ is a partial correction factor. If no descriptions are received, frame interpolation is used:

$$F_{interp} = \frac{2F_{prev} + F_{next}}{3} \quad (7)$$

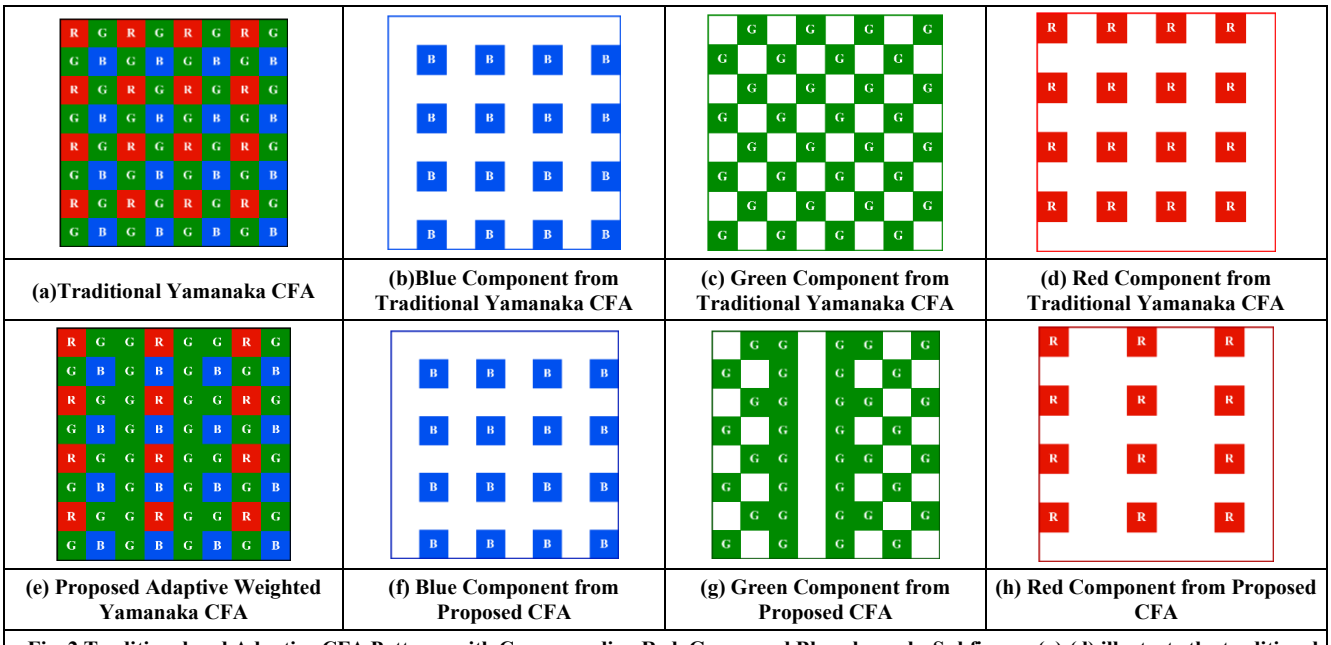
This equation estimates a lost frame from adjacent frames to maintain visual continuity. Figure 2 presents the HMDC structure. Frames pass through modules including redundancy control, packetization, wireless transmission, and frame recovery. The final stage combines the received data to reconstruct a high-quality frame. This assembly uses the equation:

$$F_{final} = F_{rec} \quad \text{if all data is received} \quad (8)$$

Otherwise, the system chooses between F_{side} or F_{interp} based on available packets. The proposed HMDC approach reduces bandwidth consumption during stable conditions by limiting redundancy. It enhances frame restoration under unstable networks by using adaptive error concealment. Additionally, the method integrates an Enhanced Yamanaka Pattern Color Filter Array (CFA) to maintain color integrity during reconstruction. The method addresses the limitations in MDC. HMDC improves recovery without excessive bandwidth.

3.2. Proposed Adaptive Weighted Yamanaka Color Filter

Color Filter Arrays (CFA) are fundamental components in image sensing systems [27]. The Yamanaka Pattern CFA is one of the designs that aim to maintain a balanced spatial distribution of red (R), green (G), and blue (B) components [28]. The main objective of this CFA is to increase accuracy in capturing high-frequency textures and reduce color reconstruction errors. The green channel is given more density because of its greater contribution to luminance. However, the traditional Yamanaka CFA pattern includes a fixed arrangement of red, green, and blue filters. Separate component maps display the red, green, and blue distributions individually for clarity. Figure 2 (a)-(d) shows the standard Yamanaka CFA, and Figure 2 (e)-(h) shows the proposed adaptive variant.



(a)Traditional Yamanaka CFA	(b)Blue Component from Traditional Yamanaka CFA	(c) Green Component from Traditional Yamanaka CFA	(d) Red Component from Traditional Yamanaka CFA
(e) Proposed Adaptive Weighted Yamanaka CFA	(f) Blue Component from Proposed CFA	(g) Green Component from Proposed CFA	(h) Red Component from Proposed CFA

Fig. 2 Traditional and Adaptive CFA Patterns with Corresponding Red, Green, and Blue channels. Subfigures (a)-(d) illustrate the traditional Yamanaka CFA and its separated red, green, and blue channels. Subfigures (e)-(h) present the proposed adaptive weighted CFA and its corresponding channels.

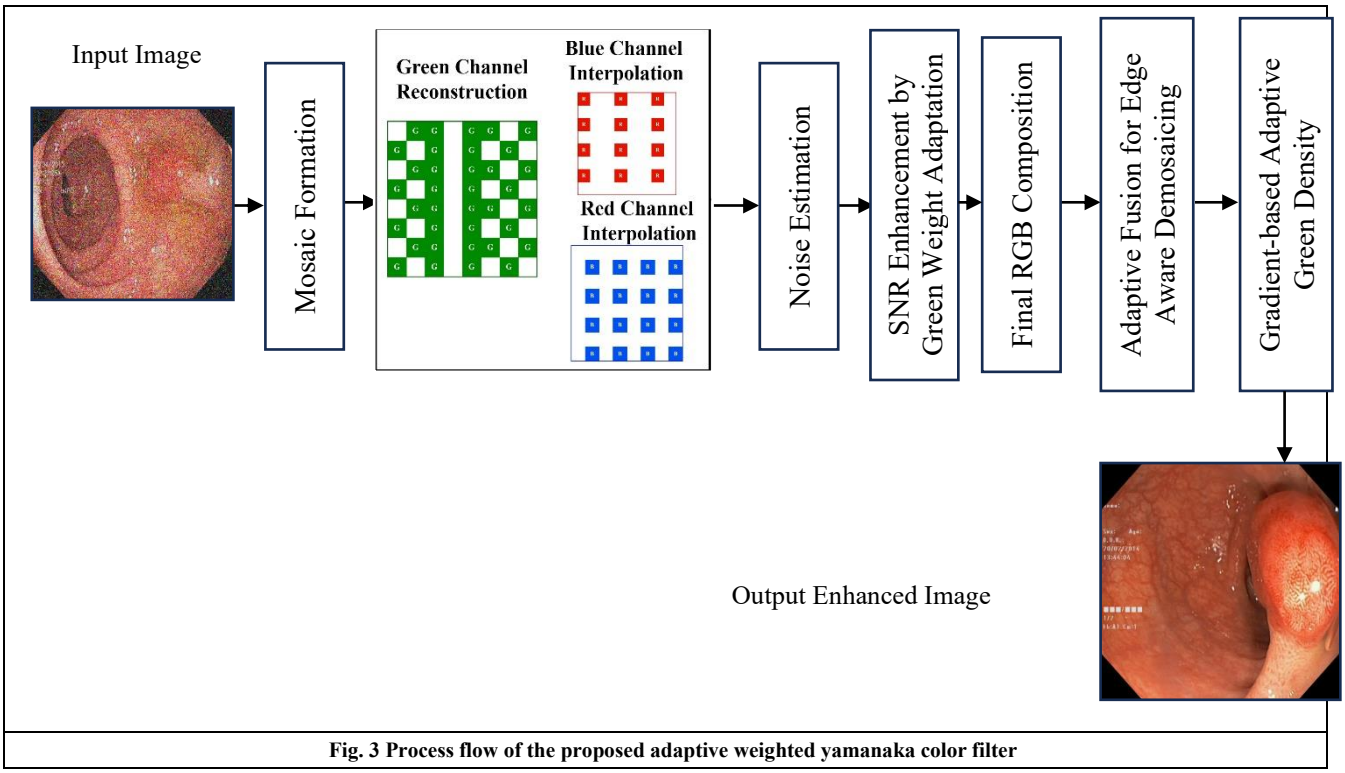


Fig. 3 Process flow of the proposed adaptive weighted yamanaka color filter

In contrast, the proposed adaptive weighted Yamanaka CFA modifies the layout to improve green dominance and adaptive spatial encoding. This is visible in the updated red, green, and blue maps. The green weight adaptive matrix applies spatial adaptation to improve green interpolation quality, essential in clinical visuals where structural clarity is critical. This modified arrangement improves chromatic sensitivity and compression accuracy. The following steps and Figure 3 explain the process flow of the Adaptive Weighted Yamanaka Color Filter.

3.2.1. Mosaic Formation

During acquisition, the sensor captures only one color channel at each pixel. The sensor output is denoted as:

$$M(x,y) = R(x,y)\cdot M_R(x,y) + G(x,y)\cdot M_G(x,y) + B(x,y)\cdot M_B(x,y) \quad (9)$$

Where $M_R(x,y), M_G(x,y), M_B(x,y)$ are binary masks representing the CFA layout for R, G, and B, respectively.

3.2.2. Green Channel Reconstruction

A core step is reconstructing the full green plane. Instead of uniform averaging, the proposed method computes:

$$G_{est}(x,y) = \sum_{i=-1}^{1}\sum_{j=-1}^{1} W_{i,j}^{(G)}\cdot G(x+i,y+j) \quad (10)$$

Where $W_{i,j}^{(G)}$ are adaptive weights based on local gradient magnitudes:

$$W_{i,j}^{(G)} = \frac{1}{1+|G(x,y)-G(x+i,y+j)|} \quad (11)$$

3.2.3. Red and Blue Interpolation

Red and blue components are reconstructed by subtracting green influence in chrominance:

$$R_{est}(x,y) = G_{est}(x,y) + \sum_{i,j} W_{i,j}^{(R)} \cdot (R-G)(x+i,y+j) \quad (12)$$

$$B_{est}(x,y) = G_{est}(x,y) + \sum_{i,j} W_{i,j}^{(B)} \cdot (B-G)(x+i,y+j) \quad (13)$$

Adaptive weights for R and B also depend on spatial correlation in the chrominance plane.

3.2.4. Noise Estimation Suppression

The sensor noise is modeled as additive Gaussian:

$$N(x,y) = I_{measured}(x,y) - I_{true}(x,y) \quad (14)$$

To suppress it, the CFA is adjusted to enhance spatial SNR:

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{\sum G^2(x,y)}{\sum N^2(x,y)} \quad (15)$$

The modified green filter density improves the numerator without increasing noise power.

3.2.5. Final RGB Composition

The final RGB image is constructed as:

$$I_{RGB}(x,y) = [R_{est}(x,y), G_{est}(x,y), B_{est}(x,y)] \quad (16)$$

Each channel is corrected for intensity mismatch using gain control.

3.2.6. Adaptive Fusion for Edge-Aware Demosaicing

To further refine demosaicing near edges, a fused edge-aware function is used:

$$G_{adaptive}(x,y) = \frac{\sum_{(i,j)\in\Omega} G\,(x+i, y+j) \cdot \exp\left(-\frac{|\nabla I(i,j)|}{\sigma}\right)}{\sum_{(i,j)\in\Omega} \exp\left(-\frac{|\nabla I(i,j)|}{\sigma}\right)} \quad (17)$$

Where $|\nabla I(i,j)|$ is the local gradient and σ is a smoothing constant. This approach reduces edge-related errors by adjusting green channel density and applying adaptive weights to spatial variations. Edges are preserved by selective interpolation using gradient-based adaptation. This method improves separation between luminance and chrominance through a weighted structure, avoiding cross-channel leakage. Texture-rich areas benefit from improved pixel estimation, as the interpolation respects local intensity gradients. This solves the issue of uniform CFA patterns that ignore content variation. The adaptive distribution compensates for directional features in images. Green filters are adaptively placed in high-detail zones, supporting fine structure recovery. This leads to sharper borders, reduced aliasing, and improved visual quality in live medical image transmission.

4. Proposed Error-Resilient Live Medical Video Streaming

The proposed system architecture integrates encoding and decoding modules to enhance wireless medical video communication. The encoding process assigns redundancy based on the importance of content and packet loss estimation. The decoding architecture recovers lost data through adaptive concealment, frame reconstruction, and color refinement. The reconstruction module ensures temporal synchronization and structural consistency across all recovered frames.

4.1. Encoding Module

The encoding module begins with raw medical video input and processes each frame through a structured pipeline. The goal is to ensure reliable delivery of high-quality video in bandwidth-limited and error-prone networks. The module prepares the video for transmission by splitting, analyzing, encoding, and packetizing the content. The first step is frame extraction. This module separates the continuous video stream into individual frames, F_i, where $i = 1,2,\ldots,N$. Each frame is independently processed. This separation enables parallel operations in the later stages of the encoding pipeline. Next, the system evaluates each frame using the frame analysis and importance scoring block. Here, each frame receives two critical metrics: the frame importance score Q_i, and the estimated packet loss probability P_i. The importance score reflects the clinical relevance of the frame. For instance, frames showing abnormal regions are considered more important. The packet loss probability is based on channel conditions at the time of transmission.

Once these scores are available, the adaptive redundancy computation module calculates the redundancy level for each frame. This redundancy value R_i is defined as:

$$R_i = \alpha P_i + \beta Q_i \quad (18)$$

Where α and β are weights adjusted based on network conditions and content criticality.

This ensures that high-risk or highly important frames receive additional protection through controlled redundancy. The multiple description generator uses the calculated redundancy to produce three parallel streams. Description 1 (D_1) and Description 2 (D_2) carry distinct parts of the original frame data. A third stream carries the redundancy R_i. These are not exact copies. Each description is crafted to carry complementary data, so the receiver can still reconstruct the frame if one or more descriptions are lost. To improve compression, the transform and quantization block applies a discrete transform. Either the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT) is used based on frame characteristics. After the transform, each coefficient is quantized using adaptive quantization levels Q_{step} calculated as:

$$Q_{step} = \frac{S_{max}}{(1+R_i)} \quad (19)$$

Where S_{max} is the maximum pixel signal value. This equation adjusts quantization resolution based on redundancy. Higher redundancy allows finer quantization. The quantized data is further processed through the Adaptive Weighted Yamanaka Color Filter Array encoding stage. This module adjusts the color channels using weighted averaging based on spectral distortion levels. Each color channel C_k is refined using:

$$C_k^{adj} = C_k + \delta_k \cdot (C_{ref} - C_k) \quad (20)$$

Here, C_k^{adj} is the adjusted value, C_{ref} is the color reference estimated from surrounding pixels, and δ_k is the adaptive correction weight.

This equation reduces chromatic noise and preserves the contrast required for clinical interpretation. Once all three descriptions (D_1, D_2, and R_i) are encoded, they are segmented into packets using the packetization module. Each frame packet P_{ij} is tagged with its type and sequence number:

$$P_{ij} = \{D_k, j\}, \quad k \in \{1,2,R\} \quad (21)$$

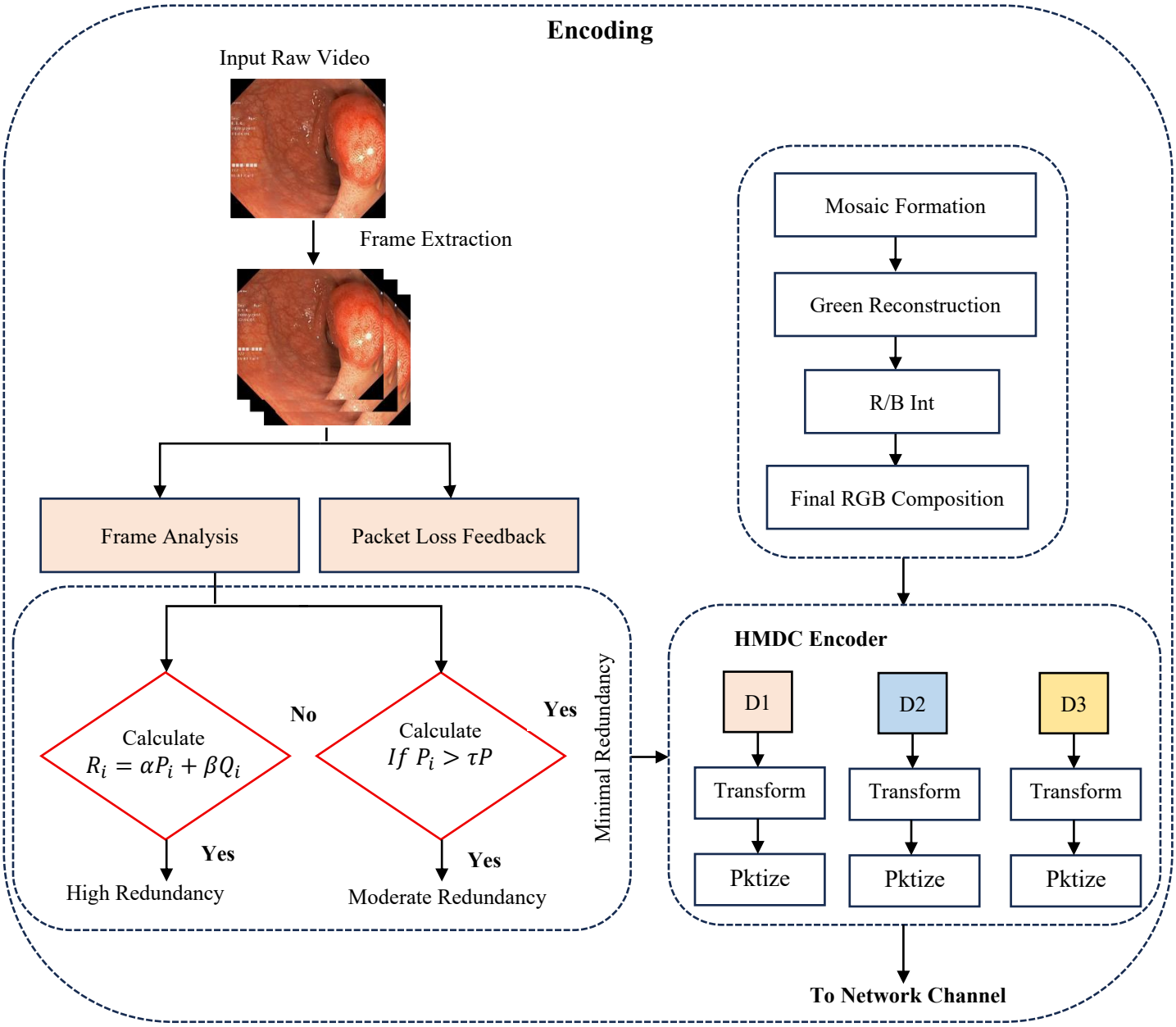


Fig. 4 Proposed encoding architecture for wireless medical video transmission

Here, j indicates the packet index within the description. This ensures correct ordering at the receiver. This stage also includes error flags for future detection. The final step of the encoding module is the wireless transmission phase. The packet stream is passed to the wireless interface for delivery. The system assumes a channel with a non-zero-bit error rate (BER). The encoding module does not correct errors during this step but prepares metadata for the recovery module at the receiver. The step-by-step procedure in Algorithm 1 defines the encoding logic. It enables dynamic redundancy control, adaptive quantization, and color-aware compression. Each step processes the input frames with precision and prepares encoded packets for wireless transmission.

The design supports error resilience while minimizing bandwidth expansion. This targeted encoding workflow supports the medical application demand for continuous, high-fidelity video under imperfect conditions.

Algorithm 1. Encoding Procedure
1: Input raw medical video V
2: Divide V into individual frames $\{F_1,F_2,...,F_n\}$
3: Initialize control parameters α, β, τ_Q, τ_P
4: **for** each frame F_i **do**
5: Estimate the importance level Q_i
6: Estimate packet loss probability P_i

7: Compute redundancy control value R_i
8: **if** $Q_i \geq \tau_Q$ **then**
9: Assign high redundancy to F_i
10: **else if** $P_i \geq \tau_P$ **then**
11: Assign moderate redundancy to F_i
12: **else**
13: Assign minimal redundancy to F_i
14: **end if**
15: Generate multiple descriptions D_1, D_2, and correction stream R
16: Apply transform operation (DCT or DWT) on F_i
17: Perform adaptive quantization on the transformed frame
18: **for** each color channel C_k in F_i **do**
19: Compute reference color C_{ref}
20: Update C_k using adaptive refinement
21: **end for**
22: Apply Adaptive Weighted Yamanaka CFA encoding
23: Segment the encoded data into packets
24: Attach packet headers and metadata
25: **if** $P_i \geq \tau_P$ **then**
26: Add error control bits to packets
27: **end if** 28: **if** $Q_i \geq \tau_Q$ **then**
29: Apply forward error correction
30: **end if**
31: Store all packets in the transmission queue Q_t
32: **end for**
33: Shuffle Q_t to reduce burst error sensitivity
34: **if the** total size of Q_t exceeds $Q_{\max,}$ **then**
35: Trim the queue to meet bandwidth constraints
36: **end if**
37: Transmit Q_t over a wireless communication channel

4.2. Decoding Module with Reconstruction Process

The decoding module begins with the reception of packets containing multiple descriptions D_1, D_2, and redundancy component R. These packets undergo an initial check through the Packet Loss Detection Module, which examines missing or corrupted packets. This detection step uses the embedded sequence headers and integrity markers to identify data loss conditions with high accuracy. If packet loss is confirmed, the system proceeds to the Error Concealment Module.

This module attempts to estimate the missing or corrupted frame data using a combination of the available descriptions and the redundancy stream. The approach is based on the weighted combination of available information. If one description is missing, the module reconstructs the lost data using the available stream and redundancy. Let $\hat{F}^i$ denote the reconstructed frame at index i. The frame is estimated using the following condition:

$$\hat{F}^i = \begin{cases} D_{1i} + \gamma R_i, & \text{if only } D_{1i} \text{ is received} \\ D_{2i} + \gamma R_i, & \text{if only } D_{2i} \text{ is received} \\ \frac{1}{2}(D_{1i} + D_{2i}) + \gamma R_i, & \text{if both } D_{1i}, D_{2i} \text{ are received} \\ 0, & \text{if all are missing} \end{cases} \quad (22)$$

Here, γ is an adaptive gain factor that adjusts the contribution of redundancy based on the estimated quality loss and importance of the frame.

In cases where descriptions are heavily degraded or absent, a spatio-temporal interpolation technique is invoked. The system predicts the frame based on neighboring temporal and spatial information.

Let $N(F_{i-1}, F_{i+1})$ denote the interpolation between frames $i-1$ and $i+1$:

$$\hat{F}^i = \lambda_1 F_{i-1} + \lambda_2 F_{i+1} \quad (23)$$

Where $\lambda_1 + \lambda_2 = 1$, and both weights are tuned using structural similarity scores. After basic reconstruction, the Frame Reconstruction Engine combines the components according to the decoding rule.

This phase ensures energy and luminance consistency. The output is passed to the Adaptive Weighted Yamanaka CFA, which enhances color accuracy and suppresses chromatic distortions. Color recovery uses an inverse filter derived from the color correction matrix:

$$C_k^{rec} = C_k + \delta_k(C_{ref} - C_k) \quad (24)$$

Where C_k is the observed color component, C_{ref} is the reference channel, and δ_k is a tuned coefficient for each channel.

Following color correction, Adaptive Interpolation and Filtering improve spatial quality. A combined filtering equation is applied:

$$F_i^{filtered} = \omega_1 \cdot \text{Median}(F_i) + \omega_2 \cdot \text{Bilateral}(F_i)$$

Where ω_1 and ω_2 are weights satisfying $\omega_1 + \omega_2 = 1$. This combination suppresses edge noise while retaining structural integrity. The frames are then passed to the Frame Buffer and Temporal Sync unit. This step ensures that frames are reordered based on timestamps and are temporally aligned.

Synchronization is vital for consistent playback in time-sensitive medical analysis. The final step is the Video Reconstruction Module, which integrates all frames into a continuous video. The reconstructed video ensures temporal smoothness, low latency, and visual clarity for accurate medical evaluation.

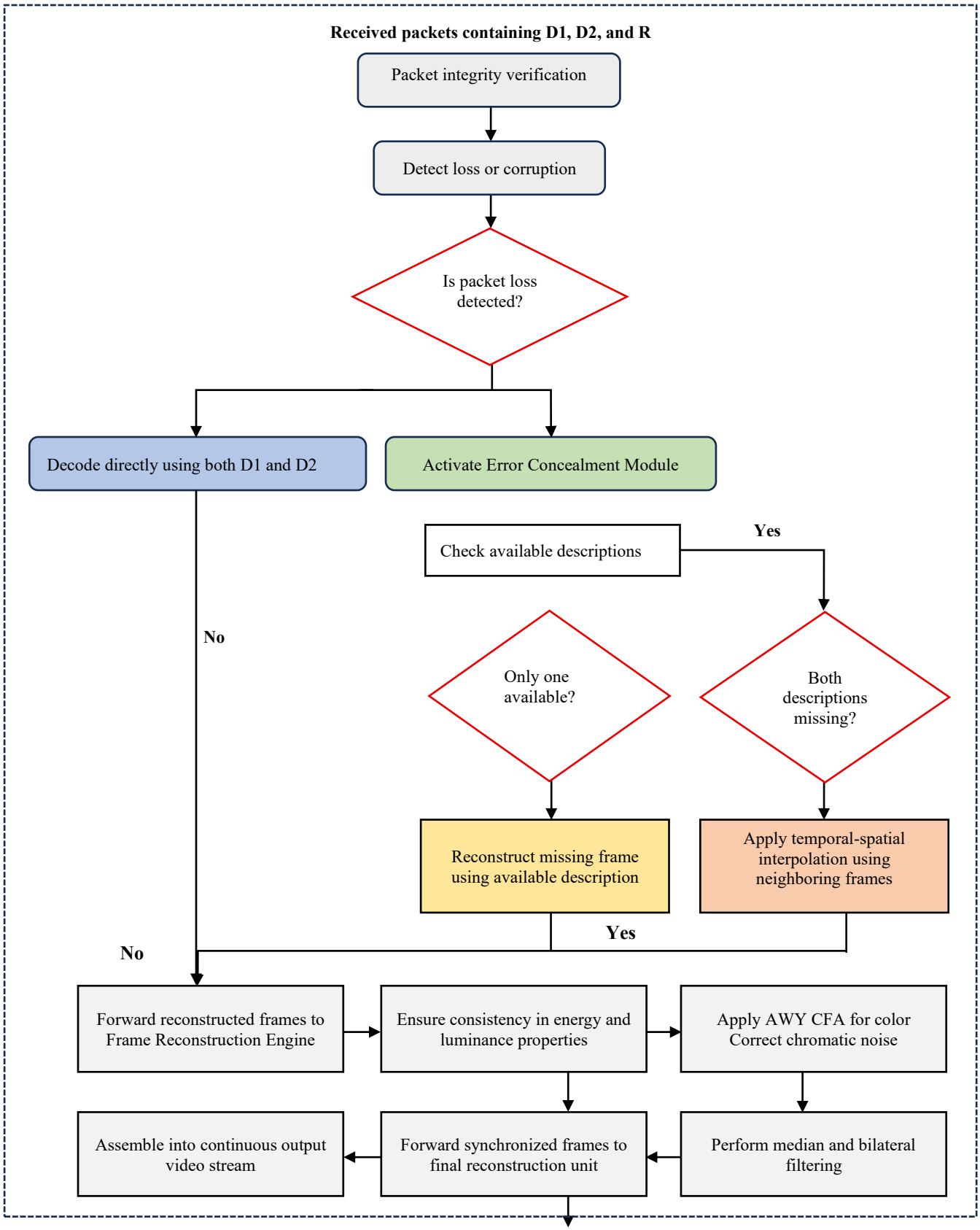


Fig. 5 Proposed decoding architecture for wireless medical video transmission

The entire decoding module is outlined and logically structured in Algorithm 2, ensuring robust performance in wireless medical video transmission environments.

Algorithm 2: Decoding Module with Reconstruction Process
1: Input: Received packets containing D_1, D_2, and R
2: Perform packet integrity verification using sequence headers
3: Detect loss or corruption in incoming packets
4: **if** packet loss is detected, **then**
5: Activate Error Concealment Module
6: **if** only one description is available, **then**
7: Reconstruct the missing frame using the redundancy stream and available description
8: **else if** both descriptions are available, **then**
9: Fuse both descriptions with the redundancy component
10: **else if** both descriptions are missing, **then**
11:Apply temporal-spatial interpolation using neighboring frames
12: **end if**
13: **else**
14: Decode directly using both D_1 and D_2
15: end if
16: Forward reconstructed frames to Frame Reconstruction Engine
17: Ensure consistency in energy and luminance properties
18: **for** each reconstructed frame F_i **do**
19: Apply Adaptive Weighted Yamanaka CFA to recover accurate color
20: **for** each color channel C_k **do**
21: Correct chromatic noise using tuned reference adjustment
22: **end for**
23: Apply adaptive interpolation to restore structural integrity
24: Perform combined median and bilateral filtering
25: **end for**
26: Store all processed frames in the Frame Buffer
27: Reorder frames based on embedded timestamps
28: Synchronize temporally for consistent playback
29: Forward synchronized frames to the final reconstruction unit
30: Assemble into a continuous output video stream

5. Results and Discussion

The system setup uses a Dell G15 laptop with an Intel i7 processor, 16GB RAM, 500GB SSD, and NVIDIA GPU. Python 3.7 and OpenCV support the simulation and testing processes. The experimental evaluation uses the Kvasir-Capsule-SEG dataset. This dataset contains real abdominal endoscopic video sequences. Frames with both normal and polyp regions are selected for testing. Polyp frames are used to evaluate the accuracy of recovery in clinically important regions. These frames allow testing under realistic conditions with variable brightness and motion. The dataset supports validation of the proposed method under packet loss and visual distortion. To ensure reliable comparison, all existing methods and the proposed method use the same experimental setup and Kvasir-Capsule-SEG dataset. This allows consistent analysis of performance under identical conditions for all methods.

5.1. Quantitative Evaluation Using PSNR and SSIM Under Packet Loss

This section evaluates the reconstruction performance of the proposed method under simulated packet loss using PSNR and SSIM metrics. Peak Signal-to-Noise Ratio (PSNR) measures the pixel-level fidelity between the original and recovered frames. Structural Similarity Index Measure (SSIM) quantifies the preservation of structural and perceptual features. These two metrics are widely used for objective quality assessment in video transmission. The evaluation considers three packet loss scenarios: 10%, 20%, and 30%. In each case, the recovered frames are compared with the original video content using the Kvasir-Capsule-SEG dataset. Tables 2, 3, and 4 present the results for each loss condition.

At 10% packet loss (Table 2), the proposed method achieves 40.91 dB in PSNR and 0.9824 in SSIM. These values are substantially higher than all existing methods. For example, Tong Tang et al record 30.42 dB and 0.8385, while Ganguly et al reach 29.03 dB and 0.8194. The lowest scores are from Nguyen Viet Hung et al, with 25.64 dB and 0.7617. These comparisons confirm that the proposed method preserves image quality with minimal degradation under mild transmission loss. Under 20% packet loss (Table 3), the PSNR of the proposed system remains high at 38.84 dB, with an SSIM of 0.9621. Competing methods fall behind. Tong Tang et al reach 28.64 dB and 0.8156. The next best performer, Ganguly et al, records only 27.01 dB and 0.7887. Methods such as Liu et al and Nguyen Viet Hung et al fall below 25 dB and 0.75 SSIM. These results indicate that the proposed model effectively handles medium packet loss while maintaining strong structural preservation. At 30% packet loss (Table 4), reconstruction becomes more difficult, but the proposed method still maintains 35.62 dB PSNR and 0.9467 SSIM.

This level of quality surpasses all baselines. Tong Tang et al achieve 26.94 dB and 0.7773, while Ganguly et al and Sasi Kumar et al remain near 25 dB and 0.75 SSIM. Methods such as Liu et al and Nguyen Viet Hung et al degrade further, with scores falling below 23 dB and 0.7 SSIM. This validates the robustness of the proposed model in high-loss environments. Across all levels of packet loss, the proposed method outperforms others by wide margins in both PSNR and SSIM. The margin increases as the severity of the loss grows. This confirms the effectiveness of the hybrid coding model and adaptive CFA in preserving visual quality and spatial structure under degraded network conditions.

Table 2. PSNR and SSIM comparison of proposed and existing methods under 10% packet loss

Method	PSNR (dB)	SSIM
Proposed	**40.91**	**0.9824**
Liu et al [18]	26.85	0.7812
Sasi Kumar et al [19]	28.37	0.8026
Maheswari et al [20]	27.11	0.7948
Ganguly et al [21]	29.03	0.8194
Tong Tang et al [22]	30.42	0.8385
Nguyen Viet Hung et al [23]	25.64	0.7617
Jiajun Xu et al [24]	27.69	0.7961

Table 3. PSNR and SSIM comparison of proposed and existing methods under 20% packet loss

Method	PSNR (dB)	SSIM
Proposed	**38.84**	**0.9621**
Liu et al [18]	24.67	0.7428
Sasi Kumar et al [19]	26.93	0.7814
Maheswari et al [20]	25.24	0.7632
Ganguly et al [21]	27.01	0.7887
Tong Tang et al [22]	28.64	0.8156
Nguyen Viet Hung et al [23]	23.76	0.7089
Jiajun Xu et al [24]	26.12	0.7703

Table 4. PSNR and SSIM comparison of proposed and existing methods under 30% packet loss

Method	PSNR (dB)	SSIM
Proposed	**35.62**	**0.9467**
Liu et al [18]	22.73	0.6984
Sasi Kumar et al [19]	24.28	0.7286
Maheswari et al [20]	22.89	0.7102
Ganguly et al [21]	25.33	0.7527
Tong Tang et al [22]	26.94	0.7773
Nguyen Viet Hung et al [23]	21.56	0.6615
Jiajun Xu et al [24]	23.94	0.7038

5.2. Quantitative Evaluation Using PSNR and SSIM Under Description Loss

This section evaluates the recovery performance of each method under partial or full description loss using PSNR and SSIM. The analysis is based on encoded medical video frames where either one or both descriptions are available at the receiver. In the first case, only Description 1 is received. As shown in Table 5, the proposed method achieves 36.82 dB in PSNR and 0.9562 in SSIM. These values are higher than all other methods. Tong Tang et al achieve 27.86 dB and 0.8141, while Ganguly et al record 26.42 dB and 0.7962. Other methods fall below 26 dB in PSNR and 0.78 in SSIM. The results show that the proposed method recovers fine structure and brightness information using a single description. In the second case, only Description 2 is available. Table 6 reports a PSNR of 35.47 dB and an SSIM of 0.9494 for the proposed model. These values remain close to those recorded under Description 1. Competing techniques, such as Tong Tang et al and Ganguly et al, remain between 25 dB and 27.6 dB PSNR. Their SSIM values do not exceed 0.81. The stability of the proposed method across both cases indicates balanced encoding and successful information reconstruction from either stream. In the final case, both descriptions are received without error. Table 7 confirms that the best quality is reached under full input. The proposed method achieves a 42.26 dB PSNR and a 0.9715 SSIM. The closest result from Tong Tang et al is 30.13 dB and 0.8354. No other method exceeds 29 dB PSNR or 0.82 SSIM. These comparisons confirm that the proposed system scales effectively with input availability. Across all three scenarios, the proposed method maintains consistent superiority in image quality metrics. The improvement results from two core mechanisms. Hybrid MDC distributes pixel and structure information across descriptions with optimized redundancy. This allows partial recovery without quality loss. The weighted Yamanaka CFA enhances color reconstruction under limited input by reducing chromatic estimation errors. In contrast, existing methods do not separate structure and color adaptively. Their fixed data partitions fail to support accurate reconstruction during incomplete transmission.

Table 5. PSNR and SSIM comparison when only description 1 is received

Method	PSNR (dB)	SSIM
Proposed	36.82	0.9562
Liu et al [18]	24.25	0.7361
Sasi Kumar et al [19]	25.71	0.7623
Maheswari et al [20]	24.58	0.7484
Ganguly et al [21]	26.42	0.7962
Tong Tang et al [22]	27.86	0.8141
Nguyen Viet Hung et al [23]	22.76	0.7129
Jiajun Xu et al [24]	24.93	0.7543

Table 6. PSNR and SSIM comparison when only description 2 is received

Method	PSNR (dB)	SSIM
Proposed	35.47	0.9494
Liu et al [18]	23.91	0.7292
Sasi Kumar et al [19]	25.53	0.7586
Maheswari et al [20]	24.33	0.7419
Ganguly et al [21]	26.18	0.7897
Tong Tang et al [22]	27.61	0.8092
Nguyen Viet Hung et al [23]	22.49	0.7056
Jiajun Xu et al [24]	24.61	0.7498

Table 7. PSNR and SSIM comparison when both descriptions are received

Method	PSNR (dB)	SSIM
Proposed	42.26	0.9715
Liu et al [18]	26.48	0.7801
Sasi Kumar et al [19]	28.01	0.8027
Maheswari et al [20]	27.03	0.7914
Ganguly et al [21]	28.64	0.8216
Tong Tang et al [22]	30.13	0.8354
Nguyen Viet Hung et al [23]	25.03	0.7582
Jiajun Xu et al [24]	27.45	0.7937

5.3. Correlation Analysis

Correlation analysis is used to assess the spatial structure of adjacent pixels before and after decoding. This evaluation confirms whether the decoded medical frames preserve the spatial relationship required for diagnostic consistency. The analysis includes horizontal, vertical, and diagonal directions using pixel-pair correlation plots. Figure 6 presents the scatter plots of original and decoded frames across three orientations. The horizontal correlation coefficient of the decoded image is 0.9692, which matches the original value. Vertical and diagonal correlations are also well preserved, showing values of 0.9785 and 0.9574, respectively. The near-identical coefficients between original and decoded frames confirm that the decoding process maintains structural fidelity. This outcome validates the reconstruction approach under packet loss and confirms its reliability for medical video transmission.

5.4. Histogram and Frame Reconstruction Analysis

The purpose of this section is to examine the visual and statistical quality of reconstructed medical video frames after data loss. It uses frame snapshots and histogram graphs to assess how well each method restores structural details and color information. The comparison in Figure 7 shows the frame loss, histogram, and recovered frames of the proposed method and existing methods. The proposed method produces a smooth texture with sharp edges and a stable color tone after recovery.

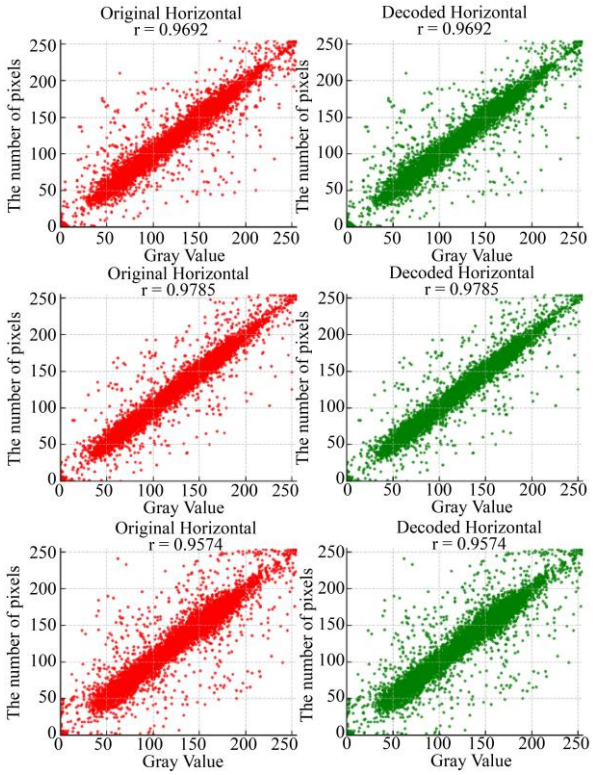
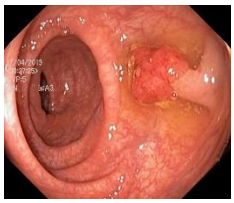
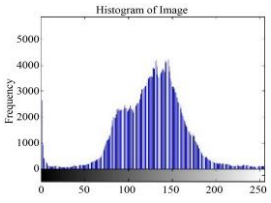
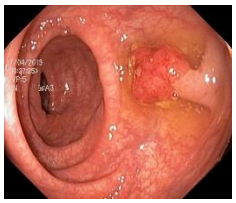
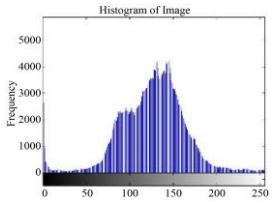
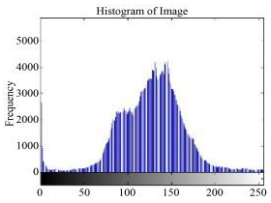
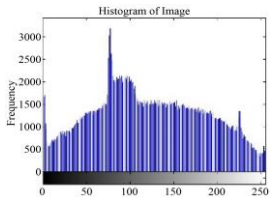
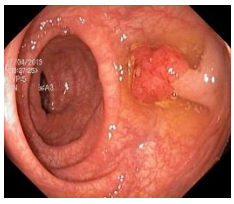
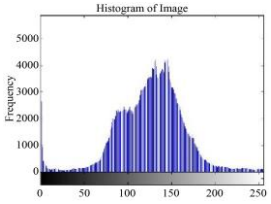
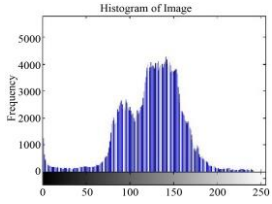


Fig. 6 Pixel-Pair correlation plots of original and decoded medical video frames in horizontal, vertical, and diagonal directions

Method	Frame loss	Histogram	Recovered frame	Histogram
Proposed				
Liu et al [18]				
Sasi Kumar et al [19]				

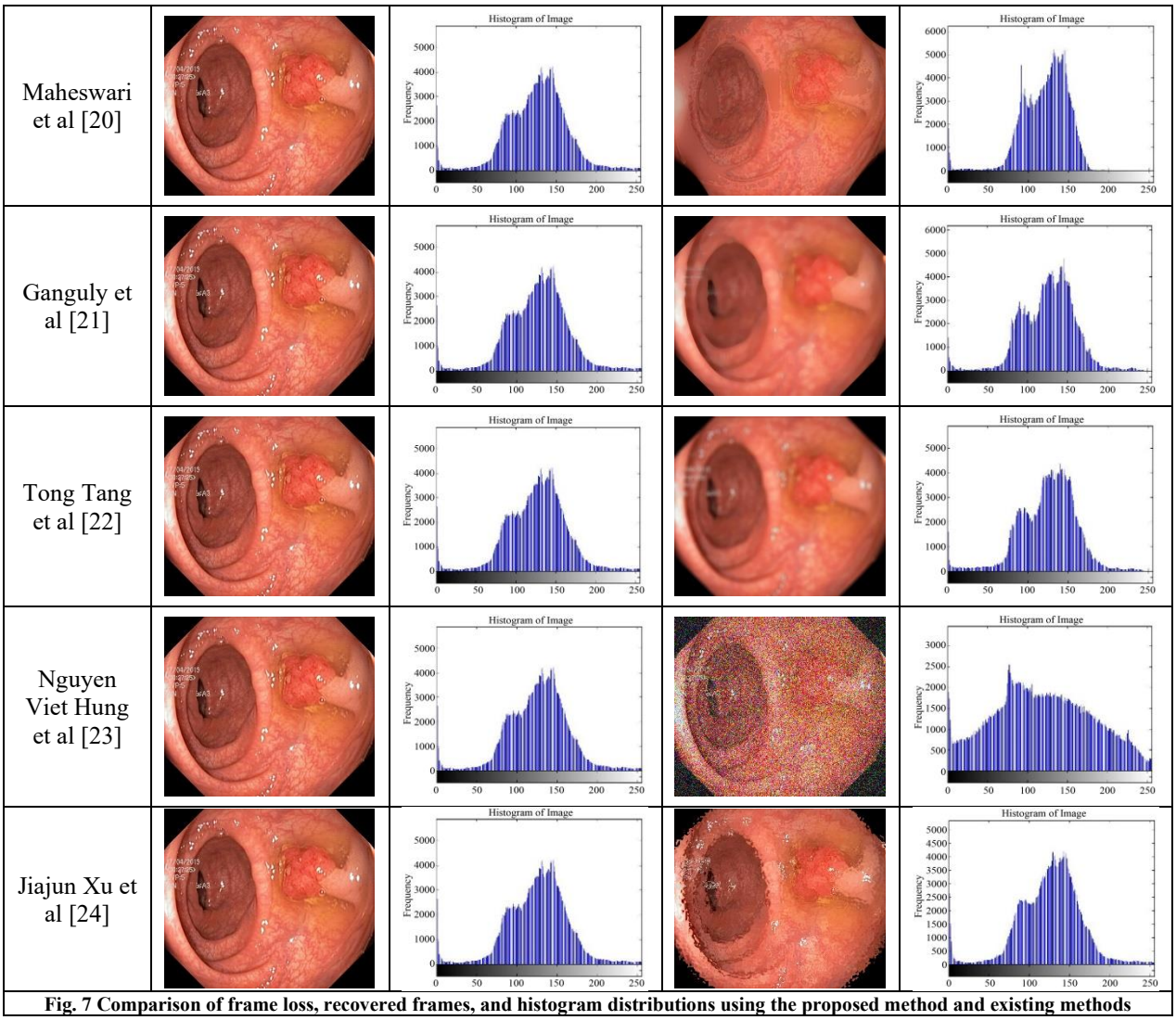


Fig. 7 Comparison of frame loss, recovered frames, and histogram distributions using the proposed method and existing methods

The recovered frame appears visually close to the original frame without loss of detail. The histogram of the recovered frame shows uniform distribution with balanced brightness and color. This confirms the accuracy of structural recovery and effective suppression of visual noise. Existing methods produce lower clarity and show distorted or unstable color regions in the frame. Liu et al [18] and Nguyen Viet Hung et al [23] produce frames with strong color artifacts. Their histograms shift toward darker regions, indicating low luminance and poor spatial interpolation. These results occur because fixed redundancy levels and non-adaptive filtering are used. The proposed method avoids such issues. HMDC dynamically adjusts redundancy based on frame importance and channel condition. AW-YCFA reduces chromatic noise and preserves structure. This adaptive approach results in accurate pixel estimation and color consistency.

5.5. Compression Ratio Analysis

This section compares the compression ratio of the proposed method against existing techniques using fixed encoding settings. The goal is to assess the efficiency of data size reduction without considering frame quality.

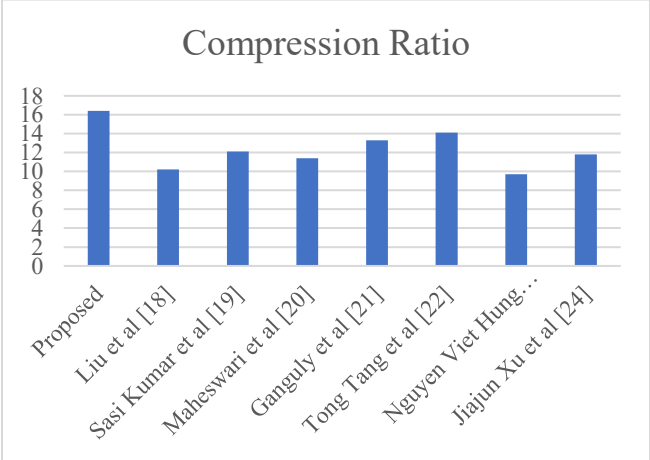


Fig. 8 Compression ratio comparison of proposed and existing methods (CBR, H.264, 512×512, 24 FPS)

The dataset comprises twenty video sequences, each encoded at 24 frames per second and with a fixed resolution. Each method processes identical content under constant bit rate conditions. The compression ratio is computed by dividing the original data size by the compressed data size. Table 7 presents the results. The proposed method achieves the highest compression ratio of 16.4. This indicates a greater reduction in file size compared to other methods. Tong Tang et al [22] record 14.1, which is the closest among the reference techniques. Ganguly et al [21] and Sasi Kumar et al [19] reach 13.3 and 12.1, respectively. These results confirm moderate performance. Lower compression ratios are observed in Liu et al [18] and Nguyen Viet Hung et al [23], with values of 10.2 and 9.7. These outcomes suggest less efficient encoding. The differences confirm that the proposed method reduces redundancy more effectively during compression. The gap in values is consistent across the selected data. Table 7 confirms that the proposed system achieves better compression under the same structure and encoder configuration.

5.6. Computational Efficiency Comparison

This section evaluates the computational efficiency of the proposed method and seven existing techniques using four standard metrics. These include frame reconstruction delay, processing time per frame, throughput, and memory usage. The purpose is to confirm whether the proposed method supports real-time performance under error-prone network conditions. The results are presented in Table 8. The proposed method achieves the lowest reconstruction delay of 13.2 ms and the shortest processing time per frame of 21.7 ms. This confirms faster recovery during transmission. Throughput reaches 46 frames per second, which is the highest among all methods. This suggests smooth real-time streaming. Memory usage remains at 184 MB, which is the lowest value recorded in the table. Other methods, such as Nguyen Viet Hung et al [23] and Maheswari et al [20], show higher delay and memory usage. Their throughput remains below 25 FPS, which limits their responsiveness. Tong Tang et al [22] perform better than other baselines, but still record higher delay and resource usage than the proposed approach. The results in Table 6 confirm that the proposed system maintains high speed and low memory demand. This supports its use in live medical video streaming, where timing and resource efficiency are important. The findings also show improved scalability across frame rates and processing conditions.

Table 8. Computational efficiency comparison of proposed and existing methods

Method	Frame Reconstruction Delay (ms)	Processing Time per Frame (ms)	Throughput (FPS)	Memory Usage (MB)
Proposed	**13.2**	**21.7**	**46**	**184**
Liu et al [18]	27.8	39.2	25	223
Sasi Kumar et al [19]	24.5	36.6	28	215
Maheswari et al [20]	29.3	42.1	23	237
Ganguly et al [21]	22.7	34.8	30	208
Tong Tang et al [22]	19.6	30.9	33	197
Nguyen Viet Hung et al [23]	31.2	44.5	21	248
Jiajun Xu et al [24]	26.4	38.7	26	229

5.7. Discussion

The proposed method addresses the critical problem of reliable video reconstruction during transmission losses in medical streaming. The primary goal of this research is to recover diagnostically important image regions under constrained bandwidth and error-prone network settings. Based on the results, the system consistently produces improved reconstruction quality, better histogram alignment, and lower structural distortion. The combination of hybrid multiple description coding with adaptive weighted Yamanaka CFA helps protect both spatial and temporal frame characteristics. This ensures that each substream carries useful information with controlled redundancy. Such a strategy reduces the impact of lost packets during streaming. The proposed method records superior performance in terms of PSNR and SSIM across various loss levels. This confirms reduced distortion and improved structural recovery. The reconstructed images retain both texture and contrast information without distortion in key areas. Histogram preservation confirms consistency between original and reconstructed frames. This helps retain clinical details essential for diagnosis. Efficiency is also confirmed in terms of processing speed and memory. Faster reconstruction and lower memory use confirm suitability for real-time systems. The compression ratio further shows the system supports reduced bandwidth needs without reducing frame integrity. The uniform GOP structure and constant bit rate help maintain consistency during encoding. These combined results confirm that the method technically meets the research objectives and addresses the impact of packet loss, bandwidth constraints, and medical data integrity in real-time settings.

A potential direction for future work includes adaptive stream prioritization based on clinical region detection. This

allows the encoder to assign higher recovery precision to critical anatomical regions. Such adaptive data-aware redundancy may further enhance resilience during high packet loss. Future implementations can also consider deep-learning-based error estimation to refine frame recovery using temporal neighborhood prediction. This could support advanced compression without reducing recovery accuracy under challenging channel conditions.

6. Conclusion

The proposed research addresses the problem of video quality degradation in clinical transmission under unstable network conditions. The study introduces a hybrid multiple description coding strategy with an adaptive weighted Yamanaka color filter array model. The system is designed to improve visual quality during partial data reception. The method divides video content into dual descriptions with selective redundancy. The adaptive CFA model reduces chromatic distortion under partial recovery. Both modules work together to enhance video integrity during transmission. The system is tested using the Kvasir-Capsule-SEG dataset under three packet loss conditions and three description loss settings. Quantitative assessment uses PSNR and SSIM to evaluate structural quality and brightness consistency. Under 10%, 20%, and 30% packet loss, the proposed method consistently outperforms existing methods. PSNR values remain above 35 dB even under severe packet loss. SSIM values stay above 0.94 across all cases. When only one description is received, the system still maintains visual accuracy. PSNR remains above 35 dB and SSIM values exceed 0.94 in all partial conditions. Visual analysis shows the recovered frames retain clinical texture and structural clarity. Color distortion is reduced using adaptive chromatic mapping. Histogram analysis confirms brightness consistency between original and reconstructed frames. Correlation analysis shows strong spatial similarity across horizontal, vertical, and diagonal directions. Compression ratio evaluation shows the method reduces data size by over 16 times. Efficiency analysis confirms low memory use and high processing speed. Throughput reaches 46 FPS, supporting real-time use. These outcomes confirm that the system supports reliable medical video transmission with high visual accuracy. Future work may include region-priority encoding to support diagnosis-focused streaming.

References

[1] Qing Cao et al., "Robotic Wireless Capsule Endoscopy: Recent Advances and Upcoming Technologies," *Nature Communications*, vol. 15, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yasmin Khattab, and Peter P. Pott, "Active/Robotic Capsule Endoscopy - A Review," *Alexandria Engineering Journal*, vol. 127, pp. 431-451, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Gastone Ciuti et al., "Frontiers of Robotic Endoscopic Capsules: A Review," *Journal of Micro-Bio Robotics*, vol. 11, pp. 1-18, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[4] Kegomoditswe Boikanyo et al., "Remote Patient Monitoring Systems: Applications, Architecture, and Challenges," *Scientific African*, vol. 20, pp. 1-28, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Anil Kumar Sagar et al., "Optimizing Quality of Service for Sensor Enabled Internet of Healthcare Systems," *Neuroscience Informatics*, vol. 1, no. 3, pp. 1-16, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Qingling Liu, Kefa G. Mkongwa, and Chaozhu Zhang, "Performance Issues in Wireless Body Area Networks for the Healthcare Application: A Survey and Future Prospects," *SN Applied Sciences*, vol. 3, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Javier Guaña-Moya et al., "Comprehensive Survey on VLC in E-Healthcare: Channel Coding Schemes and Modulation Techniques," *Applied Sciences*, vol. 14, no. 19, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Syeda Maria Gillani et al., "VQProtect: Lightweight Visual Quality Protection for Error-Prone Selectively Encrypted Video Streaming," *Entropy*, vol. 24, no. 6, pp. 1-25, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Hana Elhachi et al., "Enhancing Real-Time Mobile Health Video Streams: A Cross-Layer Region-of-Interest Based Approach," *Computer Networks*, vol. 257, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Eliecer Peña-Ancavil et al., "Adaptive Scalable Video Streaming (ASViS): An Advanced ABR Transmission Protocol for Optimal Video Quality," *Electronics*, vol. 12, no. 21, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Nhu Ngoc Dao et al., "A Contemporary Survey on Live Video Streaming from a Computation-Driven Perspective," *ACM Computing Surveys*, vol. 54, no. 10S, pp. 1-38, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Yongmin Yang, and Zhenhao Wang, "A Study on Secure Coding of Intelligent Inspection Video in Plant Areas Based on Improved Deep Reinforcement Learning," *Discover Applied Sciences*, vol. 7, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Muhammad Babar Imtiaz, and Rabia Kamran, "Mitigating Transmission Errors: A Forward Error Correction-Based Framework for Enhancing Objective Video Quality," *Sensors*, vol. 25, no. 11, pp. 1-29, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Pietro Zanuttigh et al., "Transmission of 3D Scenes over Lossy Channels," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, no. 1, pp. 1-17, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[15] Xinde Li, Fir Dunkin, and Jean Dezert, "Multi-Source Information Fusion: Progress and Future," *Chinese Journal of Aeronautics*, vol. 37, no. 7, pp. 24-58, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Sabina Umirzakova et al., "Medical Image Super-Resolution for Smart Healthcare Applications: A Comprehensive Survey," *Information Fusion*, vol. 103, pp. 1-32, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Sawsan D. Mahmood et al., "Secure Medical Image Sharing: Technologies, Watermarking Insights, and Open Issues," *IEEE Access*, vol. 13, pp. 103995-104026, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] Zhe Liu, He Chen, and Songlin Sun, "Video Error-Resilience Research Based on Error-Resilient Screen Content," *Applied Sciences*, vol. 10, no. 14, pp. 1-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] S. Sasi Kumar et al., "An Error Resilient Video Transmission in Ad Hoc Network Using Error Diffusion Block Truncation Coding," *Applied Intelligence and Informatics*, pp. 295-305, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] K. Maheswari, and Padmaja Nimmagadda, "Error Resilient Wireless Video Transmission Via Parallel Processing Using Puncturing Rule Enabled Coding and Decoding," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 6, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Madhurima Ganguly et al., "A-REaLiSTIQ-ViBe: Entangling Encoding and Transport to Improve Live Video Experience," *2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, Bangalore, India, pp. 280-284, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Tong Tang et al., "End-to-End Distortion Modeling for Error-Resilient Screen Content Video Coding," *IEEE Transactions on Multimedia*, vol. 26, pp. 4458-4468, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Viet Hung Nguyen et al., "Scalable and Resilient 360-Degree-Video Adaptive Streaming over HTTP/2 against Sudden Network Drops," *Computer Communications*, vol. 216, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Jiajun Xu et al., "Key-Frame Reference Selection for Error Resilient Video Coding Using Low-Delay Hierarchical Coding Structure," *Signal, Image and Video Processing*, vol. 18, pp. 215-222, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Patrick Enenche, Dong Ho Kim, and Dongho You, "On the Road to the Metaverse: Point Cloud Video Streaming: Perspectives and Enablers," *ICT Express*, vol. 11, no. 1, pp. 93-104, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] D.R. Bull et al., *Academic Press Library in Signal Processing: Image and Video Compression and Multimedia*, Academic Press; pp. 1-494, 2014. [Google Scholar] [Publisher Link]

[27] Seunghoon Jee, Ki Sun Song, and Moon Gi Kang, "Sensitivity and Resolution Improvement in RGBW Color Filter Array Sensor," *Sensors*, vol. 18, no. 5, pp. 1-18, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[28] Thierry Bouwmans et al., "On the Role and the Importance of Features for Background Modeling and Foreground Detection," *Computer Science Review*, vol. 28, pp. 26-91, 2018. [CrossRef] [Google Scholar] [Publisher Link]