

Spatio-Temporal Iterative Source-Channel Decoding Aided Video Transmission

Yongkai Huo, Chuan Zhu and Lajos Hanzo, *Fellow, IEEE*

School of ECS, University of Southampton, UK.

Email: {yh3g09,cz12g09,lh}@ecs.soton.ac.uk, <http://www-mobile.ecs.soton.ac.uk>

Abstract—Low-complexity uncompressed video transmission meets the requirements of home networking and quality/delay-sensitive medical applications. Hence it attracted research-attention in recent years. The redundancy inherent in the uncompressed video signals may be exploited by joint source-channel decoding for improving the attainable error resilience. Hence in this treatise we study the application of iterative joint source-channel decoding aided uncompressed video transmission, where correlation inherent in the video signals is modelled by a first-order Markov process. Firstly, we propose a spatio-temporal joint source-channel decoding system using a recursive systematic convolutional codec, where both the horizontal and the vertical intra-frame correlations as well as the inter-frame correlations are exploited by the receiver, hence relying on three-dimensional (3D) information exchange. This scheme may be combined with arbitrary channel codecs. Then we analyze the three-stage decoder's convergence behavior using 3D EXIT charts. Finally, we benchmark the attainable system performance against a couple of video communication systems, including our previously proposed 2D error concealment scheme, where only intra-frame correlations were exploited without invoking a channel codec. Our simulation results show that substantial E_b/N_0 improvements are attainable by the proposed technique.

I. INTRODUCTION

Shannon's source- and channel-coding separation theorem [1] states that reliable transmission may be accomplished by separate source coding using lossless entropy codes and channel coding under the idealized assumption of Gaussian channels and potentially infinite encoding/decoding delay and complexity. However, there are restrictions on the source encoder in many practical applications, where the transmitters fail to remove all the redundancy residing in the source signals. Hence, joint source-channel coding (JSCC) [2] was proposed for jointly exploiting the source's correlation and the channel decoder's error correction capability for error concealment. Fingscheidt and Vary proposed Softbit source decoding (SBSD) [3] for error concealment of speech signals by modelling the speech source signals using a first-order Markov process. Görtz [4], [5], Adrat and Vary [6], [7] developed the iterative source channel decoding (ISCD) philosophy, where turbo-like iterative decoding was performed by exchanging extrinsic information between the source encoder and decoder. A novel double low-density parity-check (LDPC) [8] code was proposed for JSCC, which was decoded using the standard belief propagation (BP). In the LDPC code, two traditional concatenated low-density parity-check (LDPC) [9] codes were employed as the source LDPC and channel LDPC processing blocks, respectively. At the receiver, the source LDPC and channel LDPC schemes performed joint decoding by exchanging their extrinsic information. The authors of [10] presented a novel system that invokes jointly optimized iterative source and channel decoding for enhancing the error resilience of the adaptive multirate wideband (AMR-WB) speech codec. The resultant AMR-WB-coded speech signal was protected by a recursive systematic convolutional (RSC) code and transmitted using a non-coherently detected multiple-input-multiple-output (MIMO) differential space-time spreading (DSTS) scheme. The same principles are also applicable to video signals.

Similar to the context of audio signals discussed above, a number of investigations have been conducted by applying the JSCC principle for improving the achievable video/image quality in visual multimedia delivery. A review of JSCC aided scalable image streaming was conducted in [11]. In [12], a joint source decoder and maximum *a-posteriori* probability (MAP) channel decoder was applied for decoding the motion vectors of H.264 [13] coded video streams. Kliewer, Görtz and Mertins [14], [15] modelled images using a Markov Random Field (MRF) for the sake of exploiting the correlation among adjacent pixels. In [16], Raptor codes were utilized for both the Slepian-Wolf (SW) [17] coding and the channel coding component for the sake of transmitting distributed video coded [18] streams over networks inflicting packet loss events. This philosophy leads to a cross-layer design, where video compression and error protection are performed jointly. In practical finite-delay, finite-complexity video systems, different source bits tend to inflict different perceptual video-quality degradations, hence the more vulnerable bits are protected by stronger channel codecs [19]. A JSCC scheme conceived for 3D video coding was proposed in [20], where different source and channel coding rates were considered, in order to find the optimum configuration for a WiMAX based communication channel. In [21], the authors applied JSCC to a Slepian-Wolf codec, which exploited both the channel statistics and the correlation between video frames. The authors of [22] imposed artificial redundancy on the H.264-encoded bitstream, which was then further encoded by a recursive systematic convolutional (RSC) codec. At the receiver, the source decoder which exploited the artificially imposed redundancy performed iterative decoding by exchanging extrinsic information with the RSC decoder. In order to increase the efficiency of wireless video sensor networks (WVSN), a joint source/channel coding rate control strategy was proposed in [23], where the channel codec's rate adaptation operated in unison with the network's resource allocation, which additionally also relied on source-rate optimization. The unified treatment of the topic of near-capacity multimedia communication systems using iterative detection aided JSCC employing sophisticated transmission techniques was studied in [24].

The traditional lossy video coding methods of the MPEG and the ITU-T H.26x codecs have been researched for decades [25]. However, they may be inappropriate for some applications. Firstly, they impose a high encoder complexity by the discrete cosine transform (DCT) transform, motion compensation, etc which may become excessive in video sensor networks, mobile camera phones and wireless personal area networks (WPAN) [26], [27], for example. Secondly, the processing time generates an intrinsic latency, which may violate the delay budget of delay-sensitive applications, such as interactive gaming [28]. Thirdly, some video quality degradation is inevitable and remains unrecoverable at the receiver, which may be unacceptable in high quality medical applications [29], [30]. Last but not the least, compressed video streaming is limited to devices, where matching encoding/decoding techniques are employed. A transcodec converting between compressed video formats is required, when a device has to relay the received video stream to another device employing a different compression technique, which may increase both the cost and complexity. On the other hand, the emerging 60 GHz wireless personal area networks (WPAN) within the IEEE 802.15.3c standard family [31], [32] were designed for short-range (<10 m) transmission of very-high-speed (>2 Gb/s) multimedia information

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The financial support of the RC-UK under the auspices of the India-UK Advanced Technology Centre (IU-ATC) and that of the EU under the CONCERTO project is gratefully acknowledged. The fiscal support of the ERC's Advanced Fellow grant is also thankfully acknowledged.

to both computer terminals and to consumer appliances centered around an individual person's workspace, such as offices, residential rooms, etc. The WirelessHD specification [33], [34], which is another WPAN standard, increases the maximum data rate to 28 Gb/s. Hence it is capable of supporting the transmission of either compressed or uncompressed digital high definition (HD) multimedia signals. Using the 60-GHz band as detailed by the WirelessHD specification, SiBEAM's chipset was designed for supporting uncompressed HD video transmission [28]. Hence uncompressed video transmission may be used both for home networking [26] and for other high-quality applications, such as lossless medical video communications [35], [36].

In recent years, a number of investigations have been conducted in uncompressed video transmissions. Singh *et al.* [26], [37], [38] developed a system, where both an unequal error protection (UEP) and automatic repeat request (ARQ) protocols were conceived for achieving an improved video quality. The authors of [39] investigated the specific technical challenges imposed by mm-wave systems supporting reliable video streaming using multi-beam transmissions. A flexible UEP method was proposed for the uncompressed video context in [40], which offers an improved visual quality and resource efficiency over both conventional UEP and equal error protection (EEP). An error correction scheme was conceived in [41] for mitigating the bit error effects imposed on the video, where the spatial redundancy present in the uncompressed HD video signal was exploited.

Since in uncompressed video transmission substantial spatio-temporal correlation is exhibited, it is beneficial to exploit this redundancy at the receiver, either for the sake of achieving an improved video quality or for reducing the transmission power. On the other hand, most of the state-of-art JSCC techniques were not conceived for uncompressed video communications. To name a few, the solutions proposed in [6], [7] were conceived for one-dimensional audio signals, where artificially generated correlated one-dimensional signals - rather than true 2D video signals were employed for performance evaluation. In [24], artificial redundancy was imposed by a so-called short block code for the sake of approaching the capacity of the system in the scenario of H.264-compressed video streaming. Motivated by the congenial principle of iterative JSCC, in [42] we proposed the so-called Iterative Horizontal-Vertical Scanline Model (IHVSM) using a bit-based iterative error concealment technique, where the intra-frame redundancy was exploited. More specifically, we proposed an iterative Error Concealment (EC) technique for low-complexity uplink video communications, where the correlation of the video signal is exploited by a first-order Markov process aided decoder. Firstly, we derived reduced-complexity rules for our first-order Markov modelling aided source decoder. Then we proposed a bit-based two-dimensional iterative EC algorithm, where a horizontal and a vertical source decoder were employed for exchanging their soft extrinsic information using the turbo-like iterative decoding philosophy. This scheme may be combined with low-complexity video codecs, provided that some residual redundancy resides in the video signals and the video decoders are capable of estimating the softbit information of the video pixels. We applied our proposed two-dimensional iterative EC in two design examples, namely in both distributed video coding [18] and in uncompressed video transmission scenarios. By contrast, in this treatise, we propose the novel concept of first-order Markov process aided Three-Dimensional Iterative Source-Channel decoding using a Recursive Systematic Convolutional (M3DISC-RSC) codec, where both the horizontal and vertical intra-frame correlations as well as the inter-frame correlations are exploited by the receiver. To elaborate a little further, we have three source-channel decoder-pairs, which perform iterative decoding by exchanging extrinsic information for the sake of improved video quality. Furthermore, novel three-dimensional EXIT charts [43]–[45]

will be employed for analyzing the convergence behavior of the system. Note that our proposed system may be utilized in scenarios that when the receiver can afford the associated computational complexity. Furthermore, only modest changes have to be imposed on the wireless transmitter for the sake of applying our proposed techniques.

Against this background, our novel contributions are:

1. We conceive the novel M3DISC-RSC system, which exchanges extrinsic information both among the rows and columns of video frames, as well as between the consecutive frames.
2. A single systematic channel codec is combined with three independent source decoders, effectively forming three source-channel decoder pairs for three-stage decoding, where the systematic RSC codec generates gradually improved extrinsic information, leading to a near-unimpaired video quality.

The paper is organized as follows. In Section II, we briefly review the decoding technique of first-order Markov processes. In Section III, we detail the design of the proposed scheme by employing a RSC channel codec. Then we analyze its convergence behavior by using three-dimensional EXIT charts in Section IV. The simulation results characterizing our system are provided in Section IV-D. Finally, our conclusions are offered in Section V.

II. MARKOV MODELLED VIDEO SCANLINE BASED SOFTBIT SOURCE DECODING

The *a-posteriori* probability determination technique conceived for first-order Markov processes was detailed in [7], [42]. In this section, we will briefly detail the technique of first-order Markov process decoding. Firstly the representation and decoding of the first-order Markov source-process trellis is presented in Section II-A. Then, the technique of incorporating the proposed decoding algorithm into our turbo-like iterative decoder is detailed in Section II-B. Let us commence by stipulating the following assumptions:

- x_i : an m -bit pattern of pixels scanned from the original video pixels at time instant i , which is expressed as $\{x_i(0), \dots, x_i(m-1)\} = x_i \binom{m-1}{0}$;
- m : the number of bits in each m -bit pattern x_i of pixels;
- $X_m = \{0, 1, \dots, 2^m - 1\}$: the set of all possible values in an m -bit pattern x_i ;
- $x_0^t = x_0, \dots, x_t$: the bit patterns of the 1^{st} frame of the original video consisting of $(t+1)$ m -bit patterns during the time interval spanning from 0 to t ;
- $y_0^t = y_0, \dots, y_t$: potentially error-infested bit pattern of the 1^{st} frame;

A. BCJR Decoding of First-Order Markov Chain

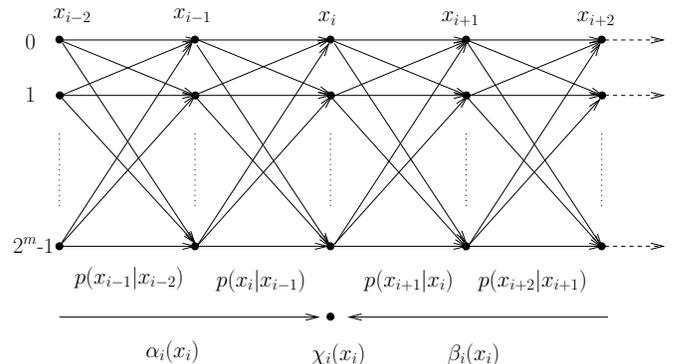


Fig. 1. Trellis of first-order Markov process for BCJR decoding, where $p(x_{i+1}|x_i)$ is the Markov transition probability.

The corresponding trellis of the first-order Markov process is displayed in Fig. 1, where the m -bit pattern x_i indicates the trellis state at time instant i and the probability $p(x_{i+1}|x_i)$ indicates the

transition from state x_i to state x_{i+1} . For the sake of decoding the trellis, let us initially follow the process of the classic BCJR [46] based determination rule of the *maximum a-posteriori probability*. At the receiver, the *a-posteriori* probability of the m -bit pattern $x_i, x_i \in X_m$ conditioned on the specific received frame of m -bit patterns y_0, \dots, y_t may be expressed as

$$p(x_i | y_0^t) = \frac{p(x_i \wedge y_0^t)}{p(y_0^t)}, \quad (1)$$

where the joint probability $p(x_i \wedge y_0^t)$ of the m -bit pattern x_i and of the received frame y_0^t may be further formulated as [42]

$$p(x_i \wedge y_0^t) = \beta_i(x_i) \cdot \chi_i(x_i) \cdot \alpha_i(x_i). \quad (2)$$

In Eq. (2), the components α, β, χ may be readily derived as in [42]. In Eq. (2), the symbol-based channel information $\chi_i(x_i) = p(y_i | x_i)$ may be calculated from the bit-based channel information as

$$\chi_i(x_i) = C_{\chi_i} \cdot \exp \sum_{k=0}^{m-1} \frac{x_i(k)}{2} \cdot L[y_i(k) | x_i(k)], \quad (3)$$

where C_{χ_i} is the normalization factor, which solely depends on y_i . Furthermore, similar to the forward recursion calculation of the BCJR algorithm, the component $\alpha_i(x_i)$ in Eq. (2) may be formulated as

$$\alpha_i(x_i) = \sum_{x_{i-1} \in X_m} \chi_{i-1}(x_{i-1}) \cdot p(x_i | x_{i-1}) \cdot \alpha_{i-1}(x_{i-1}).$$

Similarly, the backward recursion calculation of the component $\beta_i(x_i)$ in Eq. (2) is given by:

$$\beta_i(x_i) = \sum_{x_{i+1} \in X_m} \beta_{i+1}(x_{i+1}) \cdot \chi_{i+1}(x_{i+1}) \cdot p(x_{i+1} | x_i).$$

The determination of the bit-based *a-posteriori* LLRs from the symbol-based *a-posteriori* probability $p(x_i | y_0^t)$ was presented in [7]. Similarly, the bit-based *a-posteriori* LLR $L[x_i(k) | y_0^t]$ may be formulated as

$$L[x_i(k) | y_0^t] = \ln \frac{\sum_{\substack{x_i(k)=0 \\ x_i \in X_m}} \beta_i(x_i) \cdot \chi_i(x_i) \cdot \alpha_i(x_i)}{\sum_{\substack{x_i(k)=1 \\ x_i \in X_m}} \beta_i(x_i) \cdot \chi_i(x_i) \cdot \alpha_i(x_i)}. \quad (4)$$

B. Extrinsic Information Exchange for Iterative Decoding

A limitation of the formulas provided in Section II-A is that they cannot be directly used for iterative decoding, since they cannot exploit the *a-priori* LLR information $L[x_i(k)]$, which was generated from the *extrinsic* information gleaned from the other decoder involved in the turbo-like iterative decoding process [47]. The rules of iterative source and channel decoding were derived by Vary and his team in [6], [7]. To make use of the *a-priori* LLR information $L[x_i(k)]$, the combined bit-based log-likelihood information may be utilized as [7]

$$\gamma_i(x_i) = \exp \sum_{k=0}^{m-1} \frac{\bar{x}_i(k)}{2} \cdot \{L[x_i(k)] + L[y_i(k) | x_i(k)]\}, \quad (5)$$

where the symbol-based m -bit information γ is the combination of the bit-based log-likelihood *a-priori* information $L[x_i(k)]$ and of the channel information $L[y_i(k) | x_i(k)]$. We note in this context that γ of Eq. (5) contains more valuable information than the channel information χ . By replacing χ with γ in Eq. (5) we have the following formula:

$$L[x_i(k) | y_0^t] = \ln \frac{\sum_{\substack{x_i \in X_m \\ x_i(k)=0}} \beta_i(x_i) \cdot \gamma_i(x_i) \cdot \alpha_i(x_i)}{\sum_{\substack{x_i \in X_m \\ x_i(k)=1}} \beta_i(x_i) \cdot \gamma_i(x_i) \cdot \alpha_i(x_i)}. \quad (6)$$

Similar to the BCJR decoding technique of classic turbo codes [48], the bit-based *a-posteriori* LLR $L[x_i(k) | y_0^t]$ may be split into three components, namely the *a-priori* information $L[x_i(k)]$, the channel information $L[y_i(k) | x_i(k)]$ and the extrinsic information $L_e[x_i(k)]$. Specifically, the extrinsic information $L_e[x_i(k)]$ may be formulated as

$$L_e[x_i(k)] = \ln \frac{\sum_{\substack{x_i \in X_m \\ x_i(k)=0}} \beta_i(x_i) \cdot \gamma_i^{[ext]}[x_i(k)] \cdot \alpha_i(x_i)}{\sum_{\substack{x_i \in X_m \\ x_i(k)=1}} \beta_i(x_i) \cdot \gamma_i^{[ext]}[x_i(k)] \cdot \alpha_i(x_i)}, \quad (7)$$

where the extrinsic information component $\gamma_i^{[ext]}[x_i(k)]$ may be expressed as

$$\gamma_i^{[ext]}[x_i(k)] = \exp \sum_{l=0, l \neq k}^{m-1} \frac{\bar{x}_i(l)}{2} \cdot \{L[x_i(l)] + L[y_i(l) | x_i(l)]\}.$$

III. SYSTEM OVERVIEW

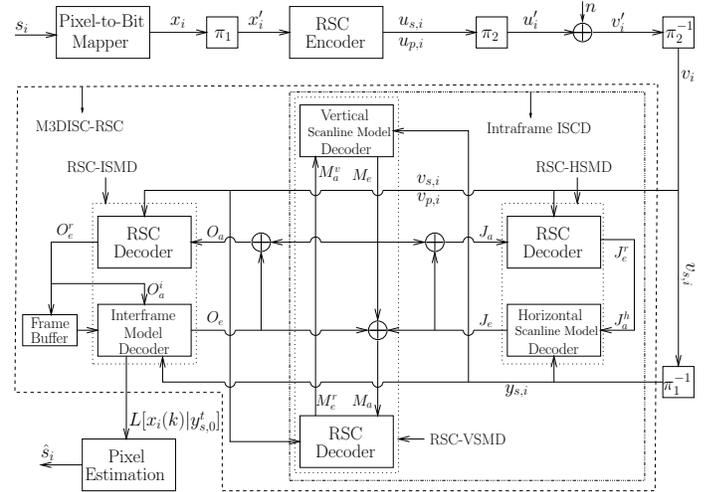


Fig. 2. System architecture of the M3DISC-RSC, where R represents reordering of the video pixels, while π represents the bit-interleaver.

A one-dimensional iterative system model was proposed and analyzed in [3], [5]–[7] in the context of audio signals. In [42], we conceived a system for iterative error concealment in two-dimensional video frames, which exploited the intra-frame correlation of practical video signals. In this section, we will detail our system model designed for ISCD exploiting the 3D correlation inherent in uncompressed video streaming. The system model of 3D iterative ISCD is displayed in Fig. 2, where R represents reordering of the video pixels, while π denotes the bit-interleaver. Assuming that our algorithm performs soft decoding on a (8×8) -pixel macroblock, all the soft pixels are ordered into 8 horizontal soft Markov processes, each of which will be input into a horizontal scanline model decoder. Then these 8 horizontal soft Markov processes will be reordered into 8 vertical soft Markov processes, each of which consists of 8 soft pixels and will be input into a vertical scanline model decoder. Note that the dimension of the video is the only parameter of the reordering process, hence given this parameter, the receiver can readily carry out the reordering. More details of the reordering are provided in [42]. The subscripts “s,” “p” denote the systematic and parity bits of a coded symbol, respectively. For instance, a systematic convolutional coded (RSC) symbol v_i consists of the systematic component $u_{s,i}$ and the parity component $u_{p,i}$. The subscripts “a,” “e” represent the *a-priori* extrinsic information, respectively, whereas “J,” “M,” and “O” indicate their relevance to the inner, middle and outer decoders, respectively. Furthermore, the superscripts “h,” “v,” “r,” and “i” indicate that the information is related to the horizontal scanline model decoder, vertical scanline model decoder, the RSC and the

inter-frame model decoder, respectively. We will further detail the system model below.

A. Transmitter

Since we consider uncompressed video communication in this treatise, we do not employ any video encoder at the transmitter. Hence, the uncompressed video pixels are simply converted to bits by the pixel-to-bit mapper shown in Fig. 2. Then the uncompressed bits are encoded by a RSC code. Specifically, at time instant i , the transmitter has to convey a video pixel s_i , which is mapped into the bit pattern x_i . This pixel-to-bit mapper may include the classic quantization operation [25]. Let us now assume that the m -bit pattern x_i is expressed as $\{x_i(0), \dots, x_i(m-1)\} = x_i \binom{m-1}{0}$ as well as that N consecutive and hence highly correlated m -bit patterns, such as x_0, \dots, x_{N-1} may be treated as a frame. Consider the first 2D video frame for example, which is mapped to a bit sequence $x'_0 \binom{m-1}{0}, \dots, x'_{N-1} \binom{m-1}{0}$ by a bit-based interleaver of length $N \cdot m$. Then the interleaved bit sequence is encoded by a channel codec, where a RSC is employed in our case. After this stage, the RSC coded bitstream consisting of the systematic bit sequence $u_{s,i}$ and the parity bit sequence $u_{p,i}$, will be interleaved by another interleaver before transmission over the channel. Then the signals are transmitted to the receiver through a Rayleigh channel using BPSK modulation. Note that the transmitter scans and transmits each video frame on a block by block basis, and similarly the receiver will reconstruct each video frame on a block basis.

B. Receiver

At the receiver, the softbit source decoding [3] principle is employed for mitigating the effects of the error-infested bit sequence $v_0 \binom{m-1}{0}, \dots, v_{N-1} \binom{m-1}{0}$, which is deinterleaved from the received bit sequence $v'_0 \binom{m-1}{0}, \dots, v'_{N-1} \binom{m-1}{0}$. Two decoding stages are involved in the softbit source channel decoding process, namely the M3DISC-RSC and the related pixel estimation. Firstly the received signal $v'_{s,i}$ will be deinterleaved at the receiver to generate the bit sequence $v_{s,i}$ and $v_{p,i}$, where $v_{s,i}$ indicates the systematic part of the deinterleaved bit sequence, while $v_{p,i}$ indicates the parity part. Then the systematic information $v_{s,i}$ will be further deinterleaved by interleaver π_1 ¹, hence the error-infested version of signal x_i namely $y_{s,i}$ can be obtained, as seen in Fig. 2.

At the first decoding stage of Fig. 2 four decoders are employed, namely the RSC channel decoder as well as the three source decoders, the Horizontal Scanline Model Decoder (HSMD) operating in the horizontal direction, the Vertical Scanline Model Decoder (VSMD) proceeding in the vertical direction and the Inter-frame Scanline Model Decoder (ISMD). However, each of the source decoders is paired with the RSC thereby forming three decoder pairs, namely the RSC-HSMD, the RSC-VSMD and the RSC-ISMD decoder pairs as seen in Fig. 2. The reason for this design is that the three source decoders jointly improve the attainable video peak signal-to-noise ratio (PSNR) owing to the improved bit error ratio (BER), while this was not possible in the previously proposed iterative EC system using the Iterative Horizontal-Vertical Scanline Model (IHVSM) of [42]. Hence the RSC codec is employed for providing increasingly improved extrinsic information from one source decoder to another for reducing the BER and hence improving the PSNR. Furthermore, we choose the RSC codec as the channel decoder, since the systematic information bits namely $v_{s,i}$ of Fig. 2 may be readily exploited by the source decoders. However, arbitrary channel codecs may also be employed. In our system, the three decoder pairs of Fig. 2 are treated as three amalgamated decoders for the sake of performing the three-stage decoding [24], while a certain number of iterations

is performed between the two decoders within each decoder pair to generate the extrinsic information during the integrated three-stage decoding process. Note that in our scenario, both the systematic part $v_{s,i}$ and the parity part $v_{p,i}$ can be directly exploited by the RSC decoder, while the three source decoders can only directly utilize the deinterleaved systematic information $y_{s,i}$.

Let us now continue by detailing the decoding process at the receiver by assuming the following scenarios:

- The f^{th} frame is being transmitted, which implies that the $(f-1)$ th frame has already been received.
- H horizontal scanlines and V vertical scanlines of the current frame have been received. This is equivalent to saying that a $(H \times V)$ -line block of the f^{th} frame has been received, which is represented by the $(H \cdot V)$ -bit patterns $v_{s,i}/y_{s,i}, v_{p,i}$.

Two different types of iterative decoding processes are invoked at the receiver, namely the iterative decoding process within the amalgamated three decoder pairs and the iterative decoding process exchanging extrinsic information among the three integrated decoder pairs. Below we now detail them separately.

1) *Iterative Decoding within the Decoder Pairs:* Again there are three amalgamated decoder-pairs in Fig. 2, where the RSC components accept *a-priori* information from the other two decoder-pairs, while the source decoders are responsible for generating the extrinsic information. The inner RSC-HSMD decoder pair of Fig. 2 performs decoding on a block of $(H \cdot V)$ bit-patterns, namely H horizontal and V vertical scanlines. The RSC codec accepts both the systematic information $v_{s,i}$ and the parity information $v_{p,i}$ of the $(H \cdot V)$ -bit patterns, which are deinterleaved by the interleaver π_2 of Fig. 2. The HSMD source decoder accepts the systematic information $y_{s,i}$ as its input, which is deinterleaved from $v_{s,i}$ by the interleaver π_1 . For each decoding iteration within a specific decoder-pair, the RSC decoder will take both the channel information v_i and the *a-priori* information J_a of Fig. 2 as its input to generate the extrinsic information J_e^r , which will be deinterleaved by interleaver π_1 . Then the deinterleaved extrinsic information will be reordered by R_1 to generate H horizontal scanlines of *a-priori* LLR information J_a^h , which can be exploited by the HSMD. Conversely, the extrinsic information J_e generated by the HSMD in the format of the H horizontal scanlines will be reordered and deinterleaved for further exploitation by the RSC, as a part of the *a-priori* information. Finally, after a preset number of iterations exchanging extrinsic information between the RSC and the HSMD, the extrinsic information J_e generated by the HSMD will be output as the extrinsic information of the inner decoder-pair as seen in Fig. 2. Similarly, the middle RSC-VSMD decoder pair of Fig. 2 performs decoding on a block of $(H \cdot V)$ bit-patterns. However, reordering is employed for the extrinsic information between the RSC and the VSMD. In contrast to the inner and intermediate decoder pairs, the RSC of the outer decoder pair of Fig. 2 performs decoding on a block of $(H \cdot V)$ bit-patterns, while the ISMD performs decoding on f -bit patterns, which are from the same frame position of f consecutive video frames. Firstly, the RSC takes the extrinsic information O_a from the ISMD and the intermediate decoder pair of Fig. 2 as *a-priori* information to generate the extrinsic information O_e^r , which will be deinterleaved by interleaver π_1 and reordered by R_3 . Then the reordered information will be stored in the “frame buffer” of Fig. 2, which will output $H \cdot V$ independent scanlines for the f consecutive frames. Each of the scanlines carries f -bit patterns, which obey a Markov process and this is exploited by the ISMD in order to perform decoding. Note that f is a flexible system parameter, which depends on the particular application. Specifically, for non-realtime video streaming applications, f may be set to a higher number, which will induce a maximal delay of $(f-1)$ video frames. However, for realtime applications, only a more limited range of the previously received frames may be utilized for decoding the current frame. Finally,

¹In Fig. 2, the reordering and interleaving operations are ignored for the sake of simplifying our system architecture, which are straightforward to add.

after a certain number of iterative decoding iterations, the extrinsic information generated by the ISMD will be output as the extrinsic information O_e generated by this particular decoder-pair.

2) *Iterative Decoding Exchanging Information Among Decoder Pairs*: When considering iterative decoding exchanging information among the three decoder-pairs of Fig. 2, each source-channel decoder pair will be treated as an amalgamated decoder. Generally, the extrinsic information exchange rules follow the three stage decoding rules stated in [24]. The inner source-channel decoder iterations exploit the intra-frame correlation within a video frame, while the outer decoder exploits the inter-frame correlation. As seen in Fig. 2 there are two inputs to the inner RSC-HSMD decoder pair, namely the channel information v_i and the *a-priori* information $J_a = J_e + M_e$, where M_e is the extrinsic information generated from the intermediate RSC-VSMD decoder pair, while J_e is generated by the inner decoder-pair itself. By contrast, the *a-priori* information forwarded to the intermediate RSC-VSMD decoder pair can be expressed as $M_a = M_e + J_e + O_e$, where O_e is the extrinsic information generated from the outer RSC-ISMD decoder pair. Similarly, the *a-priori* information provided for the outer RSC-ISMD decoder-pair of Fig. 2 may be expressed as $O_a = O_e + M_e$. Note that the extrinsic information must be appropriately deinterleaved and reordered to be exploited as the *a-priori* information by another decoder. However, in order to simplify our discussions, we ignored this interleaver and reordering operation in the above expressions. The final *a-posteriori* information may be generated by a hard-decision at the output of the outer decoder-pair of Fig. 2.

As a matter of fact, more extrinsic information may be gleaned from any of the six decoders of Fig. 2. For example, J_e^r can also be utilized by VSMD as part of the *a-priori* information. However, since M_e^r generated by the RSC of the intermediate decoder pair may be viewed as a more reliable version of J_e^r , J_e^r was excluded from the *a-priori* information provided for the VSMD.

In the discussions of Section III-B.1 and Section III-B.2, a number of extrinsic information exchanges are involved, namely $J_a, J_e, J_e^r, J_a^h, M_a, M_e, M_e^r, M_a^v, O_a, O_e, O_e^r, O_a^i$. Among them, we have $J_a^h = J_e^r, M_a^v = M_e^r$ and $O_a^i = O_e^r$, which are generated by the RSC decoder [46]. The extrinsic information terms J_e, M_e and O_e are generated by the horizontal, vertical and inter-frame model decoder, respectively, whose extrinsic information derivation rule is given by Eq. (7). After the first stage decoding, the relevant *a-posteriori* LLR information $L[x_i(k)|y_{s,0}^t]$ is generated, which may be exploited by either the bit-based MMSE or the bit-based MAP estimator for estimating the m -bit pattern x_i as well as for outputting the original pixel \hat{s}_i at the parameter estimation stage, which may be formulated as [3], [42]

- MAP estimator

$$\hat{x}_i = \arg \max_{\forall x_i \in X_m} \prod_{k=0}^{m-1} p[x_i(k)|y_0^t]; \quad (8)$$

- MMSE estimator

$$\hat{x}_i = \sum_{k=0}^{m-1} 2^k \cdot p[x_i(k) = 1|y_0^t]. \quad (9)$$

Finally, the original video source pixel \hat{s}_i may be obtained from the estimated m -bit pattern \hat{x}_i by using the inverse operations of the source encoder.

C. Parameter Training for Markov Processes

At the receiver, the first-order Markov process is utilized for modelling the correlation within multiple video frames. However, the Markov Model's State Transition Table (MMSTT) must be appropriately trained for the sake of reflecting the correlations inherent in a specific video sequence. Let us now detail the training process

using a gray-scale video sequence. The same process may be readily applied to color sequences. Let us commence by assuming that the training video sequence contains f frames, each of which carries $(W \times H)$ m -bit pixels, where W is the width and H is the height of a figure. To train the MMSTT parameters for the horizontal and the vertical Markov processes, we firstly initialize the $(2^m \times 2^m)$ -element MMSTT $T[0 : 2^m - 1, 0 : 2^m - 1]$ to zero values. Then we scan all the H horizontal scanlines from left to right and the W vertical scanlines from top to bottom in each frame of the training video sequence. For all scanlines, when pixel s_{i-1} and pixel s_i are scanned, the corresponding element $T[s_{i-1}, s_i]$ in the MMSTT is increased by 1. Finally, by normalizing the summation of all rows in the MMSTT $T[0 : 2^m - 1, 0 : 2^m - 1]$, the first-order Markov transition probabilities $p(s_j|s_i)$ can be obtained, where we have $s_i, s_j \in [0, 2^m)$. Similarly, a total of $W \cdot H$ scanlines along the time axis constituted by the consecutive frames are used to train the parameters of the inter-frame Markov processes. Each of the sequence contains f pixels, which are from the same position of f consecutive video frames.

Hence there are two MMSTTs constructed after the training process. They may be approximated by the Laplace distribution and used at the receiver. The parameter training process employed in this paper simply requires the evaluation, which is a low-complexity off-line process. Alternatively, the correlation may be evaluated from the stored previously transmitted video, which is more representative of the signal transmitted. The same process may be invoked at the receiver, provided that error probability is low. Furthermore, as we stated in the manuscript, the Markov transition table can be readily approximated using the analytical Laplace distribution. Hence, the parameter training process does not increase the decoding time and does not require a high signaling overhead. Alternatively, they may be trained on a long and sufficiently diverse video sequence. In this treatise, we do not focus on this issue in more depth. Note that the accuracy of transition probability tables substantially affects the performance of the system.

IV. PERFORMANCE ANALYSIS

In this section, we will analyze the performance of our proposed system introduced in Section III. Firstly, in Section IV-A we will introduce the scenario considered in our experiments. Then our EXIT-chart analysis will be presented in Section IV-B, followed by a couple of benchmarks discussed in Section IV-C. Finally, we will benchmark the performance of our system.

A. Scenario

	Akiyo	Foreman	Coastguard
Representation	YUV 4:2:0	YUV 4:2:0	YUV 4:2:0
Format	QCIF	QCIF	QCIF
Bits Per Pixel	8	8	8
FPS	30	15	15
Number of Frames	30	30	30
Bitrate	524 kbps	1579 kbps	1924 kbps
"Natural" Code Rate	1/8.7	1/2.89	1/2.37

TABLE I
FEATURES OF THE VIDEO SEQUENCES, AKIYO, FOREMAN AND COASTGUARD.

Generator of RSC	[11,13,13,15]	Modulation	BPSK
Channel Code Rate	1/2	Pair Iteration	2
Channel	Unc-Ray	Inner Iteration	2

TABLE II
TABLE OF PARAMETERS EMPLOYED FOR EXPERIMENT. *Unc-Ray* STANDS FOR UNCORRELATED RAYLEIGH.

In this section, we present our experimental parameters used for characterizing the convergence behavior of and benchmarking the

proposed M3DISC-RSC scheme introduced in Section III. Three 30-frame video sequences, Akiyo, Foreman and Coastguard, represented in (176×144) -pixel quarter common intermediate format (QCIF) and 4:2:0 YUV representation are employed. Moreover, the iterative intra-frame EC scheme of Fig. 2 operates on the basis of (8×8) -pixel blocks. Each QCIF luminance frame is divided into $(22 \times 18) 8 \times 8$ -pixel blocks and each QCIF chroma frame is divided into $(11 \times 9) 8 \times 8$ -pixel blocks. The uncompressed video bits are transmitted through an uncorrelated non-dispersive Rayleigh channel using BPSK modulation. We employ a RSC encoder having a rate of $R = 1/2$ and generator polynomials of $g_1 = 1011$, $g_2 = 1101$, $g_3 = 1101$, $g_4 = 1111$, which are represented as $G = [1, g_2/g_1, g_3/g_1, g_4/g_1]$, where g_1 is a feedback input and g_2, g_3, g_4 are the feedforward outputs. Moreover, the puncturing matrix $[1 0; 0 1; 1 0; 0 1]$ is employed, where the four rows correspond to the output of the systematic bit of the g_2, g_3 and g_4 , respectively. For the 3D iterative decoding, two iterations are employed for iterative decoding within each decoder-pair is employed, as well as for the inner/intra-frame iterative information exchange between decoder pairs. The parameters employed are listed in Table I and Table II. The parameters of the first-order Markov model MMSTTs were trained using the original video sequences according to the process detailed in Section III-C, which are utilized by the HSMD, VSMD and ISMD for improving the achievable error resilience.

Shannon's channel capacity theorem [1] was proposed for the transmission of i.i.d source. Hence, to be in line with the channel capacity theory, we have to consider the true entropy of the video sequence, when calculating the energy efficiency per bit. More explicitly, any redundancy inherent in the encoded sequence has to be taken into account by shifting the BER vs the channel SNR per bit curves, namely the E_b/N_0 curves to the right, regardless, whether the redundancy is natural inherent source redundancy or whether it was artificially imposed by channel coding. Similar to our previous work [42], assuming that the total uncompressed size of a video file is S_r bits and the entropy of this video source file is S_e , we might interpret the raw video file as being "naturally" losslessly encoded from the lowest possible number of S_e i.i.d bits, to generate an increased number of S_r bits, where the 'natural' code rate is $r = S_e/S_r$. In our simulations the E_b/N_0 (dB) value is calculated as $E_b/N_0 = 10 \log_{10} \frac{E_b S_r}{N_0 S_e}$. Since the true entropy of the video source cannot be readily evaluated, as a tool, the near-lossless coding mode of the H.264 codec [13], [25] was utilized for encoding the source video for the sake of approximating its entropy S_e . The "natural" code rates of the three video sequences used in our simulations are listed in Table I. Quantitatively, we found that the "natural" code rates (NCR) of the Akiyo, Foreman and Coastguard clips were $1/8.7$, $1/2.89$ and $1/2.37$, respectively for the scenario considered, which corresponds to the maximum lossless compression ratios of 8.7, 2.89 and 2.37.

B. Three-Dimensional EXIT Charts

In this section, we characterize the convergence behavior of the proposed M3DISC-RSC scheme introduced in Section III using the Akiyo sequence. For the ISMD, we divide the 30 frames of Akiyo into two $f = 15$ -frame groups for decoding, which imposes a maximal delay of $(f - 1)/FPS = 933$ millisecond (ms)³ in video transmission.

For the sake of analyzing the iterative decoding convergence of the M3DISC-RSC scheme seen in Fig. 2, each of the three

²We refer to $r = S_e/S_r$ as the 'natural' code rate, because we interpret the redundancy naturally inherent in the video source signal as being equivalent to an identical-rate channel code and hence conceive a receiver, which is capable of exploiting it.

³This delay of ~ 1 sec is unsuitable for lip-synchronized interactive video applications. However in Section IV-E, we will optimize our system and characterize its performance for reduced delays.

source-channel decoder-pairs is treated as an integrated decoder component. Again, the channel information v_i can be exploited by all of the three decoder-pairs. Hence all the EXIT functions of the inner, intermediate and outer decoder-pairs depend on E_b/N_0 , which can be expressed as $I(J_e) = f_j [I(\tilde{J}_e), I(M_e), E_b/N_0]$, $I(M_e) = f_m [I(J_e), I(\tilde{M}_e), I(O_e), E_b/N_0]$ and $I(O_e) = f_o [I(M_e), I(\tilde{O}_e), E_b/N_0]$ [43], [45], respectively. Furthermore, \tilde{J}_e, \tilde{M}_e and \tilde{O}_e indicate the relevant extrinsic information previously generated by the inner, intermediate and outer decoder-pairs, respectively. However, in order to draw the EXIT functions in the three-dimensional (3D) space, we approximate the EXIT functions as $I(J_e) = f_j [0, I(M_e), E_b/N_0]$, $I(M_e) = f_m [I(J_e), 0, I(O_e), E_b/N_0]$ and $I(O_e) = f_o [I(M_e), 0, E_b/N_0]$ in the simulations.

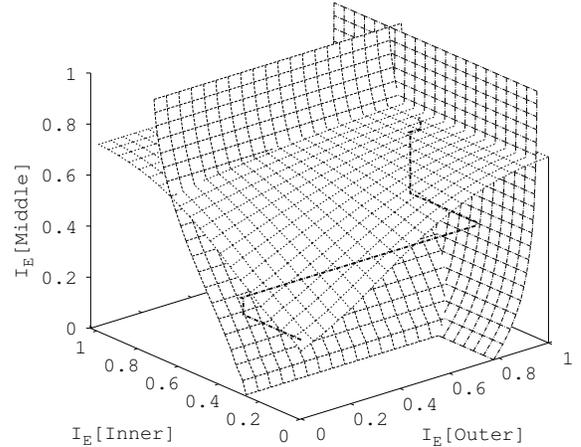


Fig. 3. Three dimensional EXIT chart of the M3DISC-RSC, when communicating over uncorrelated Rayleigh channel with E_b/N_0 of 5.4 dB.

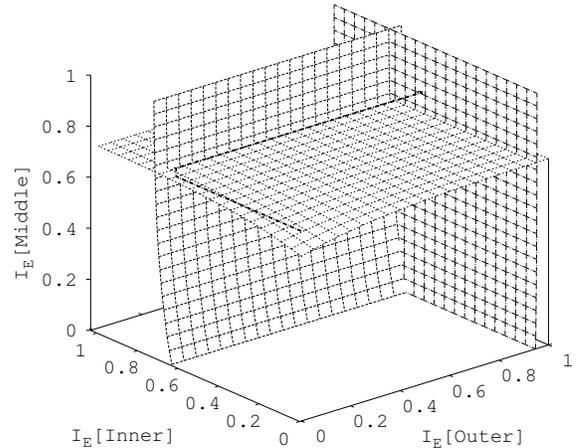


Fig. 4. Three dimensional EXIT chart of M3DISC-RSC, when communicating over an uncorrelated Rayleigh channel at $E_b/N_0 = 9.4$ dB.

We present the 3D EXIT charts recorded at the E_b/N_0 values of 5.4 dB and 9.4 dB, which are shown in Fig. 3 and Fig. 4, respectively. Observe from the figures that the inner and intermediate decoders generate similar mutual information (MI) with the aid of the same *a-priori* MI, which is due to the fact that the horizontal and vertical scanlines carry similar amount of correlations. The outer decoder generates substantially higher MI than the inner and the middle decoders at the same *a-priori* MI, since a higher amount of correlation is associated with the consecutive frames within a intra-frame. Observe from Fig. 3 and Fig. 4 that the three EXIT surfaces intersect at the points $(0.94, 0.62, 0.73)$ and $(0.95, 0.69, 0.73)$, respectively. Hence, the Monte-Carlo simulation based decoding trajectory is unable to reach the point $(1, 1, 1)$ at an E_b/N_0 of 5.4

or 9.4 dB. However, we will demonstrate in Section IV-D that we can still attain a high video quality at $E_b/N_0 = 9.4$ dB despite having a non-negligible BER.

C. Benchmarkers

In order to provide sufficiently deep insights on the performance of our proposed system, let us now describe the benchmarkers.

Firstly, we benchmark our M3DISC-RSC scheme against the RSC aided uncompressed video transmission system, where no source correlation is exploited at the receiver. We refer to this as RSC scheme for simplicity.

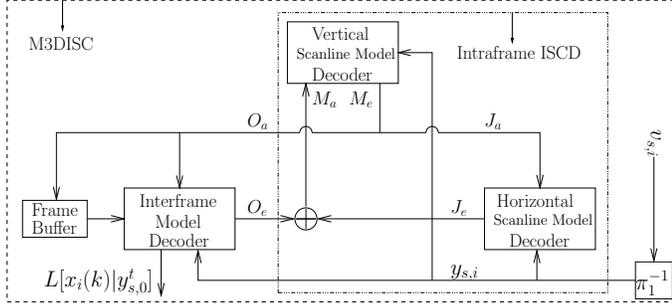


Fig. 5. Architecture of the M3DISC, where no-channel encoder is employed.

Then, to analyze the benefits of the RSC codec in the M3DISC-RSC scheme, we will benchmark it against the M3DISC arrangement, which is a non-channel-encoded version of the M3DISC-RSC regime. The architecture of the M3DISC scheme is portrayed in Fig. 5. We will also benchmark the performance recorded for the three video sequences against that of the MMSE-based hard decoder (MMSE-HD) [42], where no source correlation is exploited by the receiver. As further benchmarkers, both the first-order Markov modeling based SBS (FOMM-SBS) relying on a one-dimensional Markov process is employed at the receiver and the IHVSM of [42] were also invoked.

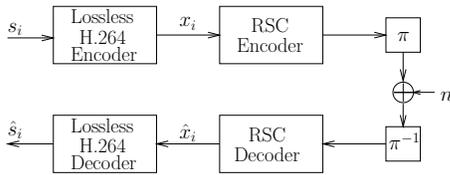


Fig. 6. Architecture of the Lossless-H.264-RSC system, where the H.264 codec operates in the near-lossless encoding mode.

Finally, the system employing the near-lossless H.264 codec of Fig. 6 is invoked. Specifically, the H.264 codec [13] is configured using the smallest quantization index. Furthermore, both predicted (P) and bidirectionally predicted (B) frames are enabled. More specifically, the 30-frame sequences were encoded into an intra-coded frame (I), followed by the periodically repeated PBBBBBBB frames. Again, this enables the H.264 codec to generate a near-lossless video bitstream. However, a delay of 8 frames delay was introduced by the employment of B frames. As shown in Fig. 6, the SC codec of the M3DISC-RSC scheme was utilized as our FEC codec for protecting the losslessly encoded bitstream. We refer to this system as the Lossless-H.264-RSC arrangement for simplicity. Note that the Lossless-H.264-RSC system imposes a high complexity at the transmitter, but a low complexity at the receiver. By contrast, our system imposes low complexity at the transmitter and a high complexity at the receiver.

Finally, to show the beneficial effects of our proposed system, the benchmark system of Fig. 7 was also considered. At the transmitter, the original video signals are encoded by a Huffman code. Then the compressed bitstream is encoded by the same RSC codec as that employed by the M3DISC-RSC system. At the receiver, the three

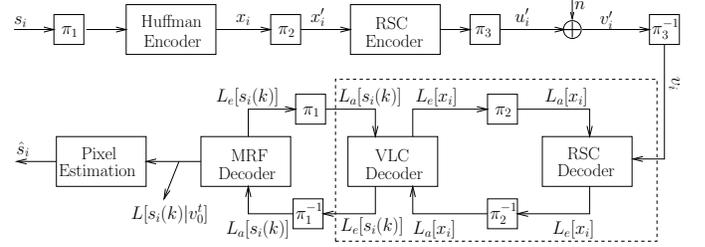


Fig. 7. Architecture of the three-stage VLC-MRF-RSC system, where the soft VLC decoder and RSC decoder consist the inner decoding stage. π_1 is a pixel-level interleaver [49].

decoders, namely the soft variable length codec (VLC) decoder [50], the soft MRF decoder [15] and the RSC decoder, perform three-stage decoding relying on joint source-channel decoding, where the MRF decoder was proposed for exploiting the correlation among adjacent pixels. In this system, the three Huffman codebooks (CB) designed for the YUV components have to be signaled to the receiver by the transmitter. Here we chose this system as a benchmark, since this system imposes similar complexity characteristics to those of our system. We refer to this system as VLC-MRF-RSC for simplicity.

A brief comparison of all the benchmarking schemes is shown in Table IV-C, where L-H.264-RSC represents the Lossless-H.264-RSC scheme. The Row Col (Trans./Rec.) compares the complexity imposed at the transmitter and receiver, respectively, while f represents the number of buffered frames invoked for decoding.

D. Numerical Results

In this section, we present our simulation results for benchmarking the scheme introduced in Section III using Akiyo, Foreman and Coastguard sequences. We rely on two types of curves for characterizing the attainable video quality, namely the PSNR versus E_b/N_0 curves and the bit error ratio (BER) versus E_b/N_0 curves. Since the MMSE-based estimator outperforms the MAP-based estimator in terms of the PSNR video quality [42], we only present the simulation results, where the MMSE-based estimator is employed for pixel estimation. Note that to avoid having infinite PSNR values when a video frame is perfectly reconstructed, we artificially set the total averaged mean squared error (MSE) value between the reconstructed and the original frame to a minimum value of 1. This is justified, since the same technique is employed in the H.264 reference software JM. Hence the maximum unimpaired video PSNR that may be obtained at the receiver is about 48.1 dB.

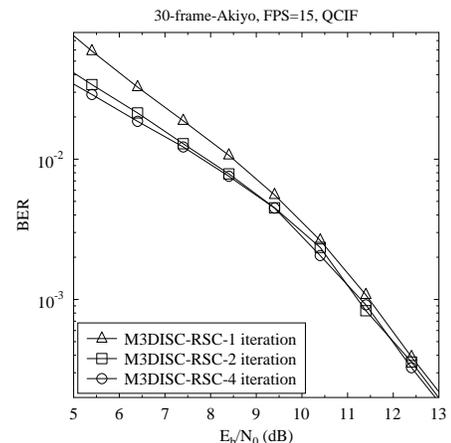


Fig. 8. BER vs E_b/N_0 for a Rayleigh channel, when the MMSE-based pixel estimation is employed.

Firstly, we present the BER versus E_b/N_0 performance of the M3DISC-RSC scheme of Fig. 2 in Fig. 8 using the Akiyo sequence when tolerating a maximal delay of $(f - 1)/FPS = 933$ ms

	MMSE-HD	RSC	L-H.264-RSC	FOMM-SBSD	IHVSM	M3DISC	VLC-MRF-RSC	M3DISC-RSC
Dimension	1	1	1	1	2	3	2	3
Bits Num to decode	8	10000	10000	64	512	$512 \sim f \times 512$	512	$512 \sim f \times 512$
Side Information	None	None	None	$1 \times \text{MMSTT}$	$1 \times \text{MMSTT}$	$2 \times \text{MMSTT}$	$3 \times \text{CB}$	$2 \times \text{MMSTT}$
Code Rate	NCR	$\frac{1}{2} \times \text{NCR}$	$\frac{1}{2}$	NCR	NCR	NCR	$(\frac{1}{2} \times \text{NCR}, \frac{1}{2})$	$\frac{1}{2} \times \text{NCR}$
Delay (frames)	0	0	8	0	0	$0 \sim f - 1$	0	$0 \sim f - 1$
Col (Trans./Rec.)	low/low	low/low	high/low	low/high	low/high	low/high	low/high	low/high

TABLE III

COMPARISON OF M3DISC-RSC AND THE BENCHMARKERS: MMSE-HD [3], RSC, LOSSLESS-H.264-RSC, FOMM-SBSD [7], IHVSM [42], M3DISC, VLC-MRF-RSC [15].

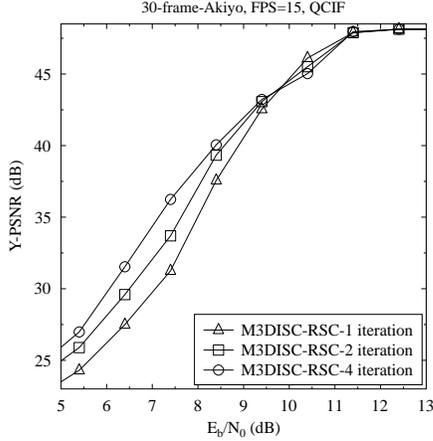


Fig. 9. Y-PSNR vs E_b/N_0 for a Rayleigh channel when the MMSE is employed for pixel estimation.

by setting $f = 15$, while the relevant Y-PSNR versus E_b/N_0 performance is displayed in Fig. 9. Observe from the two figures that we can achieve a BER of about 8×10^{-3} and a Y-PSNR of about 40 dB at a E_b/N_0 of 8.4 dB using 4 iterations. Furthermore, we observe in Fig. 9 that at a E_b/N_0 of 10.4 dB the M3DISC-RSC using 4 iterations performs slightly worse than after a single iteration. This may be due to the fact that the parameters of Markov processes trained using the Akiyo video sequence does not exactly match the distribution of some of the blocks in specific frames, as exemplified by the boundaries of objects, where the pixel values may change drastically. Another reason for this phenomenon is that we employ a short interleaver of only 512 bits, which cannot entirely prohibit error propagation during the decoding process.

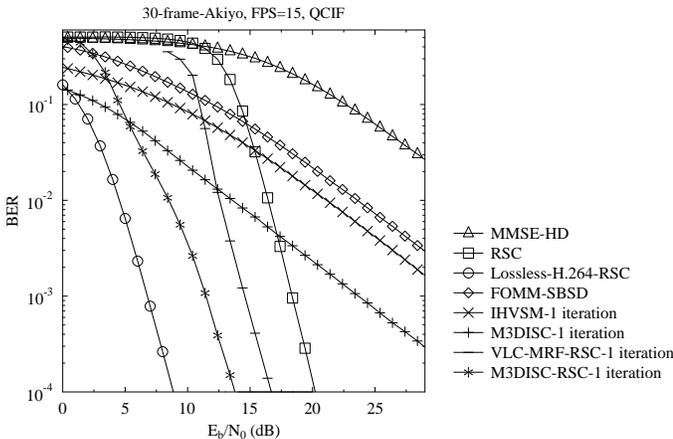


Fig. 10. BER comparison of M3DISC-RSC and the benchmarkers: MMSE-HD [3], RSC, Lossless-H.264-RSC, FOMM-SBSD [7], IHVSM [42], M3DISC, VLC-MRF-RSC [15]. Akiyo sequence.

Let us now present our performance comparison of the M3DISC-RSC scheme with a delay of $(f - 1) = 14$ frames and contrast it to the benchmarks, where $I_{outer} = 1$ (outer) iteration is employed for all schemes. Since the BER metric is less relevant than the PSNR metric in reflecting the perceptual video quality, here we present the

BER vs E_b/N_0 curves in Fig. 10 only for the Akiyo sequence. More specifically, $I_{inner} = 1$ inner iteration is employed for the M3DISC, since it outperforms the ones with more iterations [42]. Observe in Fig. 10 that at a BER of 5×10^{-3} , the M3DISC-RSC scheme of Fig. 2 outperforms the IHVSM, FOMM-SBSD and VLC-MRF-RSC schemes by about 14.7 dB, 17.1 dB and 3.5 dB, respectively, while the M3DISC scheme achieves a power reduction of 7.5 dB compared to the IHVSM. Even though the Lossless-H.264-RSC achieves the best BER performance, its bits are extremely sensitive to bit errors. Moreover, the PSNR vs E_b/N_0 curves are recorded in Fig. 11 for the Akiyo, Foreman and Coastguard sequences. As seen in Fig. 11 for the Akiyo sequence at a Y-PSNR of 46 dB⁴, the M3DISC-RSC scheme outperforms the IHVSM, FOMM-SBSD, Lossless-H.264-RSC and VLC-MRF-RSC arrangements by about 12.4 dB, 14.8 dB, 3 dB and 5.1 dB in terms of the required transmission power, respectively, while the M3DISC scheme attains a power reduction of 8.6 dB compared to the IHVSM. In other words, the M3DISC-RSC scheme outperforms the IHVSM, FOMM-SBSD, Lossless-H.264-RSC and VLC-MRF-RSC arrangements in terms of its reconstructed video quality by 12.9 dB, 15 dB, more than 20 dB and more than 20 dB of Y-PSNR at an E_b/N_0 level of 9.4 dB, respectively. Viewing Fig. 11 from a different perspective, we observe for the Foreman sequence, that at a Y-PSNR of 46 dB, the M3DISC-RSC scheme outperforms the IHVSM, FOMM-SBSD, Lossless-H.264-RSC and VLC-MRF-RSC arrangements by about 11.8 dB, 14.8 dB, 6.7 dB and 3.4 dB in terms of the required transmission power, respectively, while the M3DISC scheme achieves a power reduction of 3 dB compared to the IHVSM. In other words, it becomes explicit from Fig. 11 that the M3DISC-RSC scheme attains a Y-PSNR improvement of about 13.4 dB and 17.2 dB at a E_b/N_0 of 7.5 dB compared to the IHVSM and the FOMM-SBSD, respectively. As seen in Fig. 11 for the Coastguard sequence, when considering a Y-PSNR of 46 dB, the M3DISC-RSC scheme achieves a power reduction of about 13.2 dB, 7.8 dB and 3.3 dB compared to the IHVSM, the Lossless-H.264-RSC and the VLC-MRF-RSC, while the M3DISC scheme outperforms the IHVSM by about 3.1 dB. Alternatively, Fig. 11 suggests that the M3DISC-RSC scheme outperforms the IHVSM and the FOMM-SBSD by about 14 dB and 15.2 dB in terms of the attainable Y-PSNR at an E_b/N_0 of 7.5 dB, respectively.

From the above discussions, we may conclude that our proposed M3DISC-RSC system substantially outperforms the IHVSM, FOMM-SBSD, Lossless-H.264-RSC and VLC-MRF-RSC schemes in terms of the Y-PSNR video quality achieved. Even though the Lossless-H.264-RSC has the best BER performance, its error-sensitive bits reduce the robustness of the streamed video signals. Furthermore, according to the Y-PSNR results of Fig. 11, we may attain an improved power reduction by employing the M3DISC-RSC scheme for the video sequences exhibiting dynamic motions compared to the Lossless-H.264-RSC system, since the Lossless-H.264 codec substantially reduces the robustness of the system while only achieving a modest compression ratio.

A subjective comparison of the decoded Akiyo sequence at $E_b/N_0 = 9.4$ dB is displayed in Fig. 12, where $I_{outer} = 1$ iteration

⁴Here we are interested in this high video quality, since this treatise considers the quality-sensitive applications.

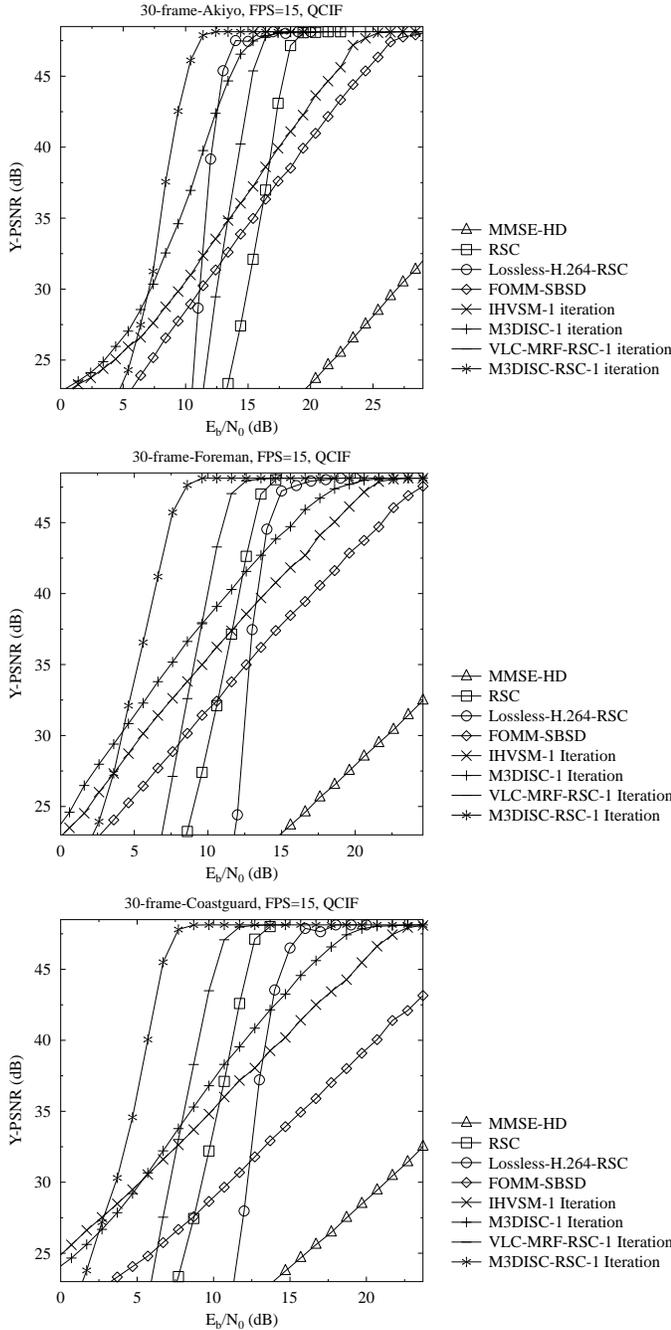


Fig. 11. Reconstructed video quality of M3DISC-RSC and of the benchmarks: MMSE-HD [3], RSC, Lossless-H.264-RSC, FOMM-SBSD [7], IHVSM [42], M3DISC, VLC-MRF-RSC [15]. Akiyo, Foreman and Coastguard sequences.

is employed for all the benchmarks. Observe from Fig. 12 that the proposed M3DISC-RSC scheme is capable of recovering the error-infested video substantially better than the benchmarks.

E. System Optimization

In this section, we characterize the M3DISC-RSC scheme associated with different delays using the Akiyo sequence. In the simulations of Section IV-D, we always buffered f frames for joint decoding, which induces a delay of $(f - 1)$ frames. However, in practical scenarios, we may buffer a reduced number of $d + 1$ ($\leq f$) frames and utilize the preceding $(f - d - 1)$ frames previously reconstructed at the receiver. In this case, we can perform decoding using f frames, which consist of $(d + 1)$ newly buffered softbit frames and $(f - d - 1)$ reconstructed hardbit frames. Correspondingly, we impose a delay of d frames. The PSNR vs E_b/N_0 curves using $f = 15$ and variable value of d is shown in Fig. 13. Observe

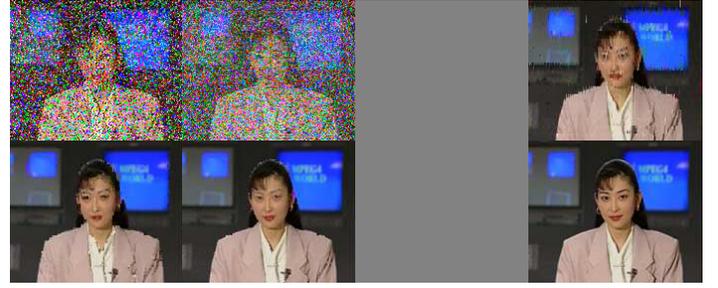


Fig. 12. A frame comparison of the decoded Akiyo sequence at $E_b/N_0 = 9.4$ dB. The frames are reconstructed by MMSE-HD [3], RSC, Lossless-H.264-RSC, FOMM-SBSD [7], IHVSM [42], M3DISC, VLC-MRF-RSC [15], M3DISC-RSC, respectively.

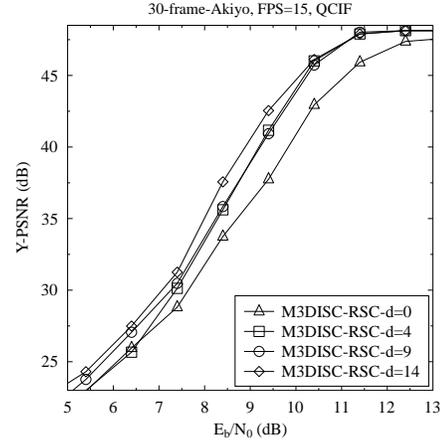


Fig. 13. Performance of M3DISC-RSC with different delay tolerance, where d represents the delay expressed in terms of number of frames.

that only a moderate gain of about $E_b/N_0 = 1$ dB can be achieved upon increasing the delay from $d = 0$ to 14 frames, which suggests that we may decrease the delay of our system by appropriately tuning the decoder at an acceptable PSNR performance degradation.

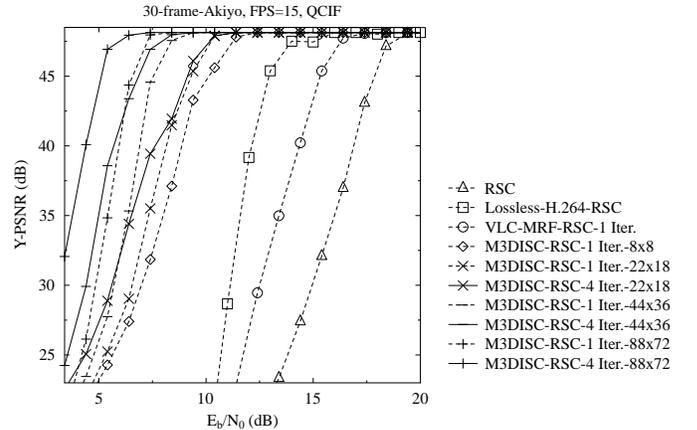


Fig. 14. Performance of M3DISC-RSC with different block size and constant $f = 15$, where 8×8 , 22×18 etc. represent size of block. The performance of RSC, Lossless-H.264-RSC and VLC-MRF-RSC are included for benchmarking.

Again, our system operates on a block-by-block basis. In Section IV-D, a constant block size of (8×8) -pixels was employed. Below, we present the performance of the M3DISC-RSC scheme configured for variable block sizes, ranging from (8×8) to (88×72) , using $f = 15$, as well as different number of iterations. The corresponding PSNR vs E_b/N_0 performance is displayed in Fig. 14 using the Akiyo sequence, where the system's performance substantially improved upon increasing the size of the block. This may be attributed to the fact that both the source decoders and the RSC decoder may benefit from increasing the length of the interleavers.

V. CONCLUSIONS

In this paper, we proposed the first-order Markov process aided three-dimensional iterative source-channel decoding concept relying on an RSC codec for uncompressed video transmissions, where both the horizontal and vertical intra-frame correlations as well as the inter-frame correlations were exploited by relying on first-order Markov processes. The proposed technique is capable of exploiting both the intra-frame and inter-frame correlations for iterative source-channel decoding. Furthermore, a single RSC codec was combined with three independent source decoders for forming three decoder-pairs, for three-stage decoding, where the RSC was utilized for improving the source decoder's convergence behavior. Our simulation results demonstrated that the proposed M3DISC-RSC scheme may facilitate a substantial power reduction compared to the benchmarks, including the IHVSM scheme, the Lossless-H.264-RSC system and the VLC-MRF-RSC system.

Our future work will focus on iterative decoding exchanging extrinsic information among the source decoder and channel decoder in a stereoscopic video context.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, June and October 1948.
- [2] K. Sayood and J. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Transactions on Communications*, vol. 39, pp. 838–846, June 1991.
- [3] T. Fingscheidt and P. Vary, "Softbit speech decoding: A new approach to error concealment," *IEEE Transaction on Speech and Audio Processing*, vol. 9, pp. 240–251, March 2001.
- [4] N. Görtz, "Joint source channel decoding using bit-reliability information and source statistics," *International Symposium on Information Theory*, p. 9, August 1998.
- [5] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1662–1670, September 2001.
- [6] M. Adrat, R. Vary, and J. Spittka, "Iterative source-channel decoder using extrinsic information from softbit-source decoding," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2653–2656, May 2001.
- [7] M. Adrat and P. Vary, "Iterative source-channel decoding: Improved system design using Exit charts," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 928–941, January 2005.
- [8] M. Fresia, F. Pérez-Cruz, H. Poor, and S. Verdú, "Joint source and channel coding," *IEEE Signal Processing Magazine*, vol. 27, pp. 104–113, November 2010.
- [9] R. Gallager, "Low-density parity-check codes," *IEEE Transactions on Information Theory*, pp. 21–28, 1962.
- [10] N. Othman, M. El-Hajjar, O. Alamri, and L. Hanzo, "Iterative AMR-WB source and channel-decoding using differential space-time spreading assisted sphere packing modulation," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 484–490, January 2009.
- [11] R. Hamzaoui, V. Stanković, and Z. Xiong, "Optimized error protection of scalable image bit streams [advances in joint source-channel coding for images]," *IEEE Signal Processing Magazine*, vol. 22, pp. 91–107, November 2005.
- [12] Y. Wang and S. Yu, "Joint source-channel decoding for H.264 coded video stream," *IEEE Transactions on Consumer Electronics*, vol. 51, pp. 1273–1276, November 2005.
- [13] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *ITU-T Rec. H.264/ISO/IEC 14496-10 AVC: Advanced Video Coding for Generic Audiovisual Services*, March 2010.
- [14] J. Kliewer, N. Görtz, and A. Mertins, "On iterative source-channel image decoding with Markov random field source models," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. iv–661–iv–664, August 2004.
- [15] J. Kliewer, N. Görtz, and A. Mertins, "Iterative source-channel decoding with Markov random field source models," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3688–3701, October 2006.
- [16] Q. Xu, V. Stankovic, and Z. Xiong, "Distributed joint source-channel coding of video using raptor codes," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 851–861, May 2007.
- [17] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.
- [18] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, pp. 71–83, January 2005.
- [19] H. Nguyen, H. Nguyen, and T. Le-Ngoc, "Signal transmission with unequal error protection in wireless relay networks," *IEEE Transactions on Vehicular Technology*, vol. 59, pp. 2166–2178, June 2010.
- [20] B. Kamolrat, W. Fernando, M. Mrak, and A. Kondoz, "Joint source and channel coding for 3D video with depth image - based rendering," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 887–894, May 2008.
- [21] Y. Zhang, C. Zhu, and K.-H. Yap, "A joint source-channel video coding scheme based on distributed source coding," *IEEE Transactions on Multimedia*, vol. 10, pp. 1648–1656, December 2008.
- [22] Nasruminallah and L. Hanzo, "EXIT-chart optimized short block codes for iterative joint source and channel decoding in H.264 video telephony," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 4306–4315, October 2009.
- [23] J. Zou, H. Xiong, C. Li, R. Zhang, and Z. He, "Lifetime and distortion optimization with joint source/channel rate adaptation and network coding-based error control in wireless video sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 1182–1194, March 2011.
- [24] Nasruminallah and L. Hanzo, "Near-capacity H.264 multimedia communications using iterative joint source-channel decoding," *IEEE Communications Surveys and Tutorials*, vol. 14, pp. 538–564, Second Quarter 2012.
- [25] L. Hanzo, P. Cherriman, and J. Streit, *Video Compression and Communications: From Basics to H.261, H.263, H.264, MPEG2, MPEG4 for DVB and HSDPA-Style Adaptive Turbo-Transceivers*. New York: John Wiley, 2007.
- [26] H. Singh, J. Oh, C. Kweon, X. Qin, H.-R. Shao, and C. Ngo, "A 60 GHz wireless network for enabling uncompressed video communication," *IEEE Communications Magazine*, vol. 46, pp. 71–78, December 2008.
- [27] K. Liu, X. Ling, X. Shen, and J. Mark, "Performance analysis of prioritized MAC in UWB WPAN with bursty multimedia traffic," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 2462–2473, July 2008.
- [28] J. Gilbert, C. Doan, S. Emami, and C. Shung, "A 4-Gbps uncompressed wireless HD A/V transceiver chipset," *IEEE Micro*, vol. 28, pp. 56–64, March-April 2008.
- [29] A. Hutanu, R. Paruchuri, D. Eiland, M. Liska, P. Holub, S. Thorpe, and Y. Xin, "Uncompressed HD video for collaborative teaching - an experiment," in *International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2007*, pp. 253–261, November 2007.
- [30] S.-T. Wei, C.-W. Tien, B.-D. Liu, and J.-F. Yang, "Adaptive truncation algorithm for Hadamard-transformed H.264/AVC lossless video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, pp. 538–549, May 2011.
- [31] R. Fisher, "60 GHz WPAN standardization within IEEE 802.15.3c," in *International Symposium on Signals, Systems and Electronics, 2007. ISSSE '07*, pp. 103–105, July 30–August 2 2007.
- [32] C. Park and T. Rappaport, "Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee," *IEEE Wireless Communications*, vol. 14, pp. 70–78, August 2007.
- [33] <http://www.wirelessHD.org>, *WirelessHD Specification Overview*, October 2007.
- [34] D. Pepe and D. Zito, "60-GHz transceivers for wireless HD uncompressed video communication in nano-era CMOS technology," in *15th IEEE Mediterranean Electrotechnical Conference (MELECON)*, pp. 1237–1240, April 2010.
- [35] S. Shimizu, N. Nakashima, and K. Okamura et al., "International transmission of uncompressed endoscopic surgery images via superfast broadband internet connections," *Surgical Endoscopy*, vol. 20, pp. 167–170, 2006.
- [36] V. Sanchez, P. Nasiopoulos, and R. Abugarbich, "Efficient lossless compression of 4-D medical images based on the advanced video coding scheme," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, pp. 442–446, July 2008.
- [37] H. Singh, X. Qin, H. Shao, C. Ngo, C. Kwon, and S. S. Kim, "Support of uncompressed video streaming over 60GHz wireless networks," in *Proceedings of 5th IEEE Consumer Communications and Networking Conference, CCNC 2008*, pp. 243–248, January 2008.
- [38] H. Singh, H. Niu, X. Qin, H. Shao, C. Y. Kwon, G. Fan, S. S. Kim, and C. Ngo, "Supporting uncompressed HD video streaming without retransmissions over 60GHz wireless networks," in *IEEE Wireless Communications and Networking Conference, WCNC 2008*, pp. 1939–1944, March 31-April 3 2008.
- [39] H.-R. Shao, C. Ngo, H. Singh, S. Qin, C. Kweon, G. Fan, and S. Kim, "Adaptive multi-beam transmission of uncompressed video over

60GHz wireless systems,” in *Future Generation Communication and Networking, FGCN 2007*, vol. 1, pp. 430–435, December 2007.

- [40] S.-E. Hong and W. Y. Lee, “Flexible unequal error protection scheme for uncompressed video transmission over 60GHz multi-Gigabit wireless system,” in *Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–6, July 31–August 4 2011.
- [41] M. Manohara, R. Mudumbai, J. Gibson, and U. Madhow, “Error correction scheme for uncompressed HD video over wireless,” in *IEEE International Conference on Multimedia and Expo, ICME 2009*, pp. 802–805, June 28–July 3 2009.
- [42] Y. Huo, T. Wang, R. G. Maunder, and L. Hanzo, “Iterative two-dimensional error concealment for low-complexity wireless video uplink transmitters,” *IEEE Transactions on Multimedia*, submitted for publication. Available at http://eprints.soton.ac.uk/339126/1/2DEC_Video.pdf.
- [43] S. ten Brink, “Convergence of iterative decoding,” *Electronics Letters*, vol. 35, pp. 806–808, May 1999.
- [44] M. Butt, R. Riaz, S. X. Ng, and L. Hanzo, “Near-capacity iterative decoding of binary self-concatenated codes using soft decision demapping and 3-D EXIT charts,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 1608–1616, May 2010.
- [45] R. Maunder and L. Hanzo, “Extrinsic information transfer analysis and design of block-based intermediate codes,” *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 762–770, March 2011.
- [46] L.R. Bahl and J. Cocke and F. Jelinek and J. Raviv, “Optimal decoding of linear codes for minimising symbol error rate,” *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.
- [47] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near shannon limit error-correcting coding and decoding: Turbo codes,” in *Proceedings of the International Conference on Communications*, (Geneva, Switzerland), pp. 1064–1070, May 1993.
- [48] L. Hanzo, T.H.Liew, B.L.Yeap, R. Tee, and S. Ng, *Turbo Coding, Turbo Equalisation and Space-Time Coding*. New York: John Wiley, 2011.
- [49] Y.-J. Wu and H. Ogiwara, “Symbol-interleaver design for turbo trellis-coded modulation,” *IEEE Communications Letters*, vol. 8, pp. 632–634, October 2004.
- [50] R. Bauer and J. Hagenauer, “Symbol-by-symbol MAP decoding of variable length codes,” in *Proc. 3rd ITG Conf. Source Channel Coding*, pp. 111–116, January 2000.



Yongkai Huo received the B.Eng. degree with distinction in computer science and technology from Hefei University of Technology, Hefei, China, in 2006 and the M.Eng. degree in computer software and theory from University of Science and Technology of China, Hefei, China, in 2009. He is currently working toward the Ph.D. degree with the Communications, Signal Processing and Control Group, School of Electronics and Computer Science, University of Southampton,

Southampton, U.K. He received a scholarship under the China-U.K. Scholarships for Excellence Programme. His research interests include distributed video coding, multiview video coding, robust wireless video streaming and joint source-channel decoding.



Chuan Zhu received the B.Eng. degree from Southeast University, Nanjing, China. In 2010, he obtained a M.Sc. degree with distinction in radio frequency communication systems from the University of Southampton, Southampton, UK. He was awarded the student case award towards his Ph.D. study from British Telecom, UK, and is currently working towards the Ph.D. degree with the Communications Research Group, School of Electronics and Computer Science, University

of Southampton, UK. His research interests include joint source-

channel decoding, video compression and transmission, and EXIT-chart-aided turbo detection, as well as cooperative communications.



Lajos Hanzo (<http://www-mobile.ecs.soton.ac.uk>) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate “Doctor Honoris Causa” by the Technical University of Budapest. During his 35-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since

1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 80 PhD students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1300 research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and since 2009 he has been a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to <http://www-mobile.ecs.soton.ac.uk>