EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Wyner-Ziv video coding for wireless lightweight multimedia applications

Nikos Deligiannis[1,2*], Frederik Verbist[1,2], Athanassios C Iossifides[3], Jürgen Slowack[2,4], Rik Van de Walle[2,4], Peter Schelkens[1,2] and Adrian Munteanu[1,2]

Abstract

Wireless video communications promote promising opportunities involving commercial applications on a grand scale as well as highly specialized niche markets. In this regard, the design of efficient video coding systems, meeting such key requirements as low power, mobility and low complexity, is a challenging problem. The solution can be found in fundamental information theoretic results, which gave rise to the distributed video coding (DVC) paradigm, under which lightweight video encoding schemes can be engineered. This article presents a new hash-based DVC architecture incorporating a novel motion-compensated multi-hypothesis prediction technique. The presented method is able to adapt to the regional variations in temporal correlation in a frame. The proposed codec enables scalable Wyner-Ziv video coding and provides state-of-the-art distributed video compression performance. The key novelty of this article is the expansion of the application domain of DVC from conventional video material to medical imaging. Wireless capsule endoscopy in particular, which is essentially wireless video recording in a pill, is proven to be an important application field. The low complexity encoding characteristics, the ability of the novel motion-compensated multi-hypothesis prediction technique to adapt to regional degrees of temporal correlation (which is of crucial importance in the context of endoscopic video content), and the high compression performance make the proposed distributed video codec a strong candidate for future lightweight (medical) imaging applications.

Keywords: Wyner-Ziv coding, distributed video coding, hash-based motion estimation, wireless lightweight multimedia applications

1. Introduction

Traditional video coding architectures, like the H.26x [1] recommendations, mainly target broadcast applications, where video content is distributed to multiple users, and focus on optimizing the compression performance. The source redundancy is exploited at the encoder by means of predictive coding. In this way, traditional video coding implies joint encoding and decoding of video. Namely, the encoder produces a prediction of the source and then codes the difference between the source and its prediction. Motion-compensated prediction in particular, a key algorithm to achieve high compression performance by removing the temporal correlation between successive frames in a sequence, is very effective but computationally demanding.

The need for highly efficient video compression architectures maintaining lightweight encoding remains challenging in the context of wireless video capturing devices that have only modest computational capacity or operate on limited battery life. The solution to reduce the encoding complexity can be found in the fundamentals of information theory, which constitute an original coding perspective, known as distributed source coding (DSC). The latter stems from the theory of Slepian and Wolf [2] on lossless separate encoding and joint decoding of correlated sources. Subsequently, Wyner and Ziv [3] extended the DSC problem to the lossy case, deriving the rate distortion function with side information at the decoder. Driven by these principles, the distributed, alias Wyner-Ziv, video coding paradigm has arisen [4,5].

* Correspondence: ndeligia@etro.vub.ac.be
[1]Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
Full list of author information is available at the end of the article

Unlike traditional video coding, in distributed video coding (DVC), the source redundancies are exploited at the decoder side, implying separate encoding and joint decoding. Specifically, a prediction of the source, named side information, is generated at the decoder by using the already decoded information. By expressing the statistical dependency between the source and the side information in the form of a virtual correlation channel, e.g. [4-8], compression can be achieved by transmitting parity or syndrome bits of a channel code, which are used to decode the source with the aid of the side information. Hence, computationally expensive tasks, like motion estimation, could be relocated to the decoder, allowing for a flexible sharing of the computational complexity between the encoder and the decoder and enabling the design of lightweight encoding architectures.

DVC has been recognized as a potential strategic component for a wide range of lightweight video encoding applications, including visual sensor networks and wireless low-power surveillance [9,10]. A unique application of particular interest in this article is wireless capsule endoscopy[a]. Conventional endoscopy, like colonoscopy or gastroscopy, has proven to be an indispensable tool in the diagnosis and remedy of various diseases of the digestive track. Significant advances in miniaturization have led to the emergence of wireless capsule endoscopy [11]. At the size of a large pill, a wireless capsule endoscope comprises a light source, an integrated chip video camera, a radio telemetry transmitter and a limited lifespan battery. The small-scale nature of the recording device forces severe constraints on the required video coding technology, in terms of computational complexity, operating time, and power consumption. Moreover, since the recorded video is used for medical diagnosis, high-quality decoded video at an efficient compression ratio is of paramount importance.

Generating high-quality side information plays a vital role in the compression performance of a DVC system. Conversely to traditional predictive coding, in DVC the original frame is not available during motion estimation, since this is performed at the decoder. Producing accurate motion-compensated predictions at the decoder for a wide range of video content, while at the same time constraining the encoding complexity and guaranteeing high compression performance, poses a major challenge. This problem becomes even more intricate in the largely unexplored application of DVC in wireless capsule endoscopy, in which the recorded video material contains extremely irregular motion, due to low frame acquisition rates and the erratic movement of the capsule along the gastrointestinal track. Towards tackling this challenge, this study presents a novel hash-based DVC architecture.

First and foremost, this study paves the road for the application of DVC systems in lightweight medical imaging where the proposed codec achieves high compression efficiency with the additional benefit of low computational encoding complexity. Second, the proposed Wyner-Ziv video codec incorporates a novel motion-compensated multi-hypothesis prediction scheme, that supports online tuning to the spatial variations in temporal correlation in a frame by obtaining information from the coded hash in case temporal prediction is unreliable. Third, this article includes a thorough experimental evaluation of the proposed hash-based DVC scheme on (i) conventional test sequences, numerous (ii) traditional endoscopic as well as (iii) wireless capsule endoscopic video content. The experimental results show that the proposed DVC outperforms alternative DVC schemes, including DISCOVER, the hash-based DVC from [12] and our previous study [13], as well as conventional codecs, namely, Motion JPEG and H.264/AVC Intra [1]. Four, this article incorporates a detailed analysis of the encoding complexity and buffer size requirements of the proposed system.

The rest of the article is structured as follows. Section 2 covers an overview of Slepian-Wolf and Wyner-Ziv coding and their instantiation in DVC. Section 3 describes two application scenarios, both relevant to DVC in general and the proposed video codec in particular. Our novel DVC codec is explained in Section 4 and experimentally evaluated in Section 5, using conventional test sequences as well as endoscopic test video. Section 6 draws the conclusions of this study.

2. Background and contributions

2.1. Slepian-Wolf coding

Consider the compression of two correlated, discrete, identically and independently distributed (i.i.d.) random sources X and Y. According to Shannon's source coding theory [14], the achievable lower rate bound for lossless joint encoding and decoding is given by the joint entropy $H(X, Y)$ of the sources. Slepian and Wolf [2] studied the lossless compression scenario in which the sources are independently encoded and jointly decoded. According to their theory, the achievable rate region for decoding X and Y with an arbitrarily small error probability is given by $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$, $R_X + R_Y \geq H(X, Y)$, where $H(X|Y)$ and $H(Y|X)$ are the conditional entropies of the considered sources, and R_X, R_Y are the respective rates at which the sources X and Y are coded, i.e., the Slepian-Wolf theorem states that even when correlated sources are encoded independently, a total rate close to the joint entropy suffices to achieve lossless compression.

The Slepian-Wolf theorem constructs a random binning argument, in which the employed code generation is asymptotic and non-constructive. In [15], Wyner pointed out the strong relation between random binning and channel coding, suggesting the use of linear channel codes as a practical solution for Slepian-Wolf coding. Wyner's methodology was recently used by Pradhan and Ramchandran [16], in the context of practical Slepian-Wolf code design based on conventional channel codes like block and trellis codes. In the particular case of binary symmetric correlation between the sources, Wyner's scheme can be extended to state-of-the-art binary linear codes, such as Turbo [5,17], and low-density parity-check (LDPC) codes [18], approaching the Slepian-Wolf limit. A turbo scheme with structured component codes was used in [17] while parity bits instead of syndrome bits were sent in [5]. Although breaking the close link with channel coding, characterized by syndromes and coset codes, the latter solutions offer inherent robustness against transmission errors.

2.2. Wyner-Ziv coding

Wyner-Ziv coding [3] refers to the problem of lossy compression with decoder side information. Suppose X and Y are two statistically dependent i.i.d. random sources, where X is independently encoded and decoded using Y as side information. The reconstructed source $\hat{X}$ has an expected distortion $D = Ed\left(x, \hat{x}\right)$. According to the Wyner-Ziv theorem [3], a rate loss is sustained when the encoder is ignorant of the side information, namely $R^{*}_{X|Y}\ (D) \geq R_{X|Y}\ (D)$, where $R^{*}_{X|Y}\ (D)$ is the Wyner-Ziv rate and $R_{X|Y}(D)$ is the rate when the side information is available to the encoder as well. However, Wyner and Ziv further showed that equality holds for the quadratic Gaussian case, namely the case where X and Y are jointly Gaussian and a mean-square distortion metric $d(\bullet,\bullet)$ is used.

Initial practical Wyner-Ziv code design focused on finding good nested codes among lattice [19] and trellis-based codes [16] for the quadratic Gaussian case. However, as the dimensionality increases, lattice source codes approach the source coding limit much faster than lattice channel codes approach capacity. This observation has induced the second wave of Wyner-Ziv code design which is based on nested lattice codes followed by binning [20]. The third practical approach to Wyner-Ziv coding considers non-nested quantization followed by efficient binning, realized by a high-dimensional channel code [5]. Other constructions in the literature propose turbo-trellis Wyner-Ziv codes, in which trellis coded quantization is concatenated with a Turbo [21] or an LDPC [22] code.

2.3. DVC

One of the applications of DSC that has received a substantial amount of research attention is DVC. Except for providing low-complexity encoding solutions for video, Wyner-Ziv coding has been shown to provide error resilient video coding by means of distributed joint-source channel coding [23], or systematic forward error protection [24]. Moreover, layered Wyner-Ziv code [25] constructions support scalable video coding [23].

An early practical DVC implementation was the PRISM codec [4], combining Bose-Chaudhuri-Hocquenghem channel codes with efficient entropy coding and performing block-based joint decoding and motion estimation. An additional CRC check was sent to the decoder to select between many decoded versions of a block, each version in fact corresponding to a different motion vector. An alternative DVC architecture, that implemented Wyner-Ziv coding as quantization followed by turbo coding using a feedback channel to enable decoder-driven optimal rate control, was presented in [5]. In this architecture, side information was generated at the decoder using motion-compensated interpolation (MCI). The architecture was further improved upon, resulting in the DISCOVER codec [26], which included superior MCI [27] through block-based bidirectional motion estimation and compensation combined with spatial smoothing. The DISCOVER codec is a well-established reference in DVC, delivering state-of-the-art compression performance.

In sequences with highly irregular motion content, blind motion estimation at the decoder, by means of MCI for example, fails to deliver adequate prediction quality. One technique to overcome this problem is to perform hash-based motion estimation at the decoder. Aaron et al. [28] proposed a hash code consisting of a coarsely sub-sampled and quantized version of each block in a Wyner-Ziv frame. The encoder performed a block-based decision whether to transmit the hash. For the blocks for which a hash code was sent, hash-based motion estimation was carried out at the decoder, while for the rest of the blocks, for which no hash was sent, the co-located block in the previous reconstructed frame was used as side information. In [29], several hash generation approaches–either in the pixel or in the transform domain–were investigated. It was shown that hash information formed by a quantized selection of low-frequency DCT bands per block was outperforming the other methods [29]. In [12], a block-based selection, based on the current frame to be coded and its future and past frames in hierarchical order, was performed at the encoder. Blocks for which MCI was foreseen to fail were low-quality H.264/AVC Intra encoded and transmitted to the decoder to assist MCI. The residual frame, given by the difference between all reconstructed intra coded blocks or the central luminance value (for non-hash blocks) and the corresponding blocks in the Wyner-Ziv frame, was formed and Wyner-Ziv encoded. In our previous study [30], we have introduced a hash-based DVC, where the auxiliary information conveyed to the decoder comprised a number of most significant bit-planes of the original Wyner-Ziv frames. Such a bit-plane-based hash facilitates accurate decoder-side motion estimation and advanced probabilistic motion compensation [31]. Transform-domain Wyner-Ziv encoding was applied on the remaining least significant bit-planes, defined as the difference of the original frame and the hash [31]. In [32], hash-based motion estimation was combined with side information refinement to further improve the compression performance at the expense of minimal structural decoding delay.

Driven by the requirements of niche applications like wireless capsule endoscopy, this study proposes a novel hash-based DVC architecture introducing the following novelties. First, in contrast to our previous DVC architectures [30,31], which employed a bit-plane hash, the presented system generates the hash as a downscaled and subsequently conventionally intra coded version of the original frames. Second, unlike our previous study [30-32], the hash is exploited in the design of a novel motion-compensated multi-hypothesis prediction scheme, which is able to adapt to the regional variations in temporal correlation in a frame by extracting information from the hash when temporal prediction is untrustworthy. Compared to alternative techniques in the literature, i.e., [12,13,26,27], the proposed methodology delivers superior performance under strenuous conditions, namely, when irregular motion content is encountered as in for example endoscopic video material, where gastrointestinal contractions can generate severe morphological distortions in conjunction with extreme camera panning. Third, the way the hash is constructed and utilized to generate side information in the proposed codec also differs from the approaches in [28,29]. Fourth, conversely to alternative hash-based DVC systems [12,31], the proposed architecture codes the entire frames using powerful channel codes instead of coding only the difference between the original frames and the hash. Fifth, unlike existing works in the literature, this article experimentally shows the state-of-the-art compression performance of the proposed DVC not only on conventional test sequences, but also on traditional and wireless capsule endoscopic video content, while low-cost encoding is guaranteed.

3. Application scenarios for DVC

3.1. Wireless lightweight many-to-many video communication

Wyner-Ziv video coding can be a key component to realize many-to-many video streaming over wireless networks. Such a setting demands optimal video streams, tailored to specific requirements in terms of quality, frame-rate, resolution, and computational capabilities imposed by a set of recorders and receivers. Consider a network of wireless visual sensors that is deployed to monitor specific scenes, providing security and surveillance. The acquired information is gathered by a central node for decoding and processing. Wireless network surveillance applications are characterized by a wide variety of scene content, ranging from complex motion sequences, e.g., crowd or traffic monitoring, to surveillance of scenes mostly devoid of significant motion, e.g., fire and home monitoring.

In such scenarios, wireless visual sensors are understood to be cheap, battery powered and modest in terms of complexity. In this concept, Wyner-Ziv video coding facilitates communications from the sensors to the central base station, by maintaining low computational requirements at the recording sensor, while simultaneously ensuring fast, highly efficient, and scalable coding. From a complementary perspective, a conventional predictive video coding format with low-complexity decoding characteristics provides a broadcast oriented one-to-many video stream for further dissemination from the base station. Such a video communications' scenario centralizes the computational complexity in the fixed network infrastructure, which would be responsible for transcoding the Wyner-Ziv video coding streams to a conventional format.

3.2. Wireless capsule endoscopy

Although the history of ingestible capsules for sensing purposes goes surprisingly back to 1957, it was the semiconductor revolution of the 1990s that created a rush in the development of miniaturized devises performing detailed sensing and signal processing inside the body [33]. Among the latest achievements in this regard is wireless capsule endoscopy, which aims at providing visual recordings of the human digestive track. From a technological perspective, capsule endoscopic video transmission poses an interesting engineering challenge. Encapsulating the appropriate system components comprising a camera, light source, power supply, CPU, or memory in a biocompatible robust ingestible housing–see Figure 1, resistant to the gatrointestinal's hostile environment, is no easy task. The reward however is great. Capsule endoscopy has been shown to have a superior positive diagnosis rate compared to other methods, including push enteroscopy, barium contrast studies, computed tomographic enteroclysis, and magnetic resonance imaging [11]. The principal drawback of contemporary capsule endoscopes is that they only detect and record but are unable to take biopsies or perform therapy. In case a pathology is diagnosed, a more uncomfortable or even surgical therapeutic procedure is necessary. Nevertheless, because of its valuable diagnostic potential, the clinical use of capsule endoscopy has a bright future. Namely, wireless endoscopy offers the only non-invasive means to examine areas of the small intestine that cannot be reached by other types of endoscopy such as colonoscopy or esophagogastroduodenoscopy [11]. In addition to this, capsule endoscopy offers a less unpleasant alternative to traditional endoscopy, lowering the

Figure 1 The Pill-Cam ESO2, a wireless capsule endoscope, relative to a one euro coin.

threshold for preventive periodic screening procedures, where the large majority of patients are actually healthy.

Focussing on the video coding technology part, it is apparent that wireless endoscopy is subjected to severe constraints in terms of available computational capacity and power consumption. Contemporary capsule video chips employ conventional coding schemes operating in a low-complexity, intra-frame mode, i.e., Motion JPEG [34], or even no compression at all. Current capsule endoscopic video systems operate at modest frame resolutions, e.g., 256 × 256 pixels, and frame rates, e.g., 2-5 Hz, on a battery life time of approximately 7 h. Future generations of capsule endoscopes are intended to transmit at increased resolution, frame rate, and battery life time and will therefore require efficient video compression at a computational cost as low as possible. In addition, a video coding solution supporting temporal scalability has an attractive edge, enabling increased focus during the relevant stages of the capsules bodily journey. DVC is a strong candidate to fulfil the technical demands imposed by wireless capsule endoscopy, offering low-cost encoding, scalability, and high compression efficiency [10].

4. Proposed DVC architecture

A graphical overview of our DVC architecture, which targets the aforementioned application scenarios, is given in Figure 2.

4.1. The encoder

Every incoming frame is categorized as a key or a Wyner-Ziv frame, denoted by K and W, respectively, as to construct groups of pictures (GOP) of the form $KW \dots W$. The key frames are coded separately using a conventional intra codec, e.g., H.264/AVC intra [1] or Motion JPEG.[b] The Wyner-Ziv frames on the other hand are encoded in two stages. For every Wyner-Ziv frame, the encoder first generates and codes a hash, which will assist the decoder during the motion estimation process. In the second stage, every Wyner-Ziv frame undergoes a discrete cosine transform (DCT) and is subsequently coded in the transform domain using powerful channel codes, thus generating a Wyner-Ziv bit stream.

4.1.1. Hash formation and coding

Our Wyner-Ziv video encoder creates an efficient hash that consists of a low-quality version of the downsized original Wyner-Ziv frames. In contrast to our previous hash-based DVC architectures [30,31], where the dimensions of the hash were equal to the dimensions of the original input frames, coding a hash-based on the downsampled Wyner-Ziv frames reduces the computational complexity. In particular, every Wyner-Ziv frame undergoes a downscaling operation by a factor, $d \in \mathbb{Z}_+$. To limit the involved operations, straightforward downsampling is applied. Foregoing a low-pass filter to bandlimit, the signal prior to downsampling runs the risk of introducing undesirable aliasing artefacts. However, experimental experience has shown that the impact on the overall rate-distortion (RD) performance of the entire system does not outweigh the computational complexity incurred by the use of state-of-the-art downsampling filters, e.g., Lanczos filers [35].

After the dimensions of the original Wyner-Ziv frames have been reduced, the result is coded using a conventional intra video codec, exploiting spatial correlation within the hash frame only. The quality at which the hash is coded has experimentally been selected and constitutes a trade-off between (i) obtaining a constant quality of the decoded frames, which is of particular interest in medical applications, (ii) achieving high RD performance for the proposed system and (iii) maintaining a low hash rate overhead. We notice that constraining the hash overhead comes with the additional benefit of minimizing the hash encoding complexity. On the other hand, ensuring sufficient hash quality so that the accuracy of the hash-based motion estimation at the decoder is not compromised or so that even pixels in the hash itself could serve as predictors is important. Afterwards, the resulting hash bit stream is multiplexed with the key frame bit stream and sent to the decoder.

We wish to highlight that, apart from assisting motion estimation at the decoder as in contemporary hash-based systems, the proposed hash code is designed to also act as a candidate predictor for pixels for which the temporal correlation is low. This feature is of particular significance especially when difficult-to-capture endoscopic video content is coded. To this end, the presented hash generation approach was chosen over existing methods in which the hash consists of a number of most significant Wyner-Ziv frame bit-planes [30,31], of coarsely subsampled and quantized versions of blocks [28], or of

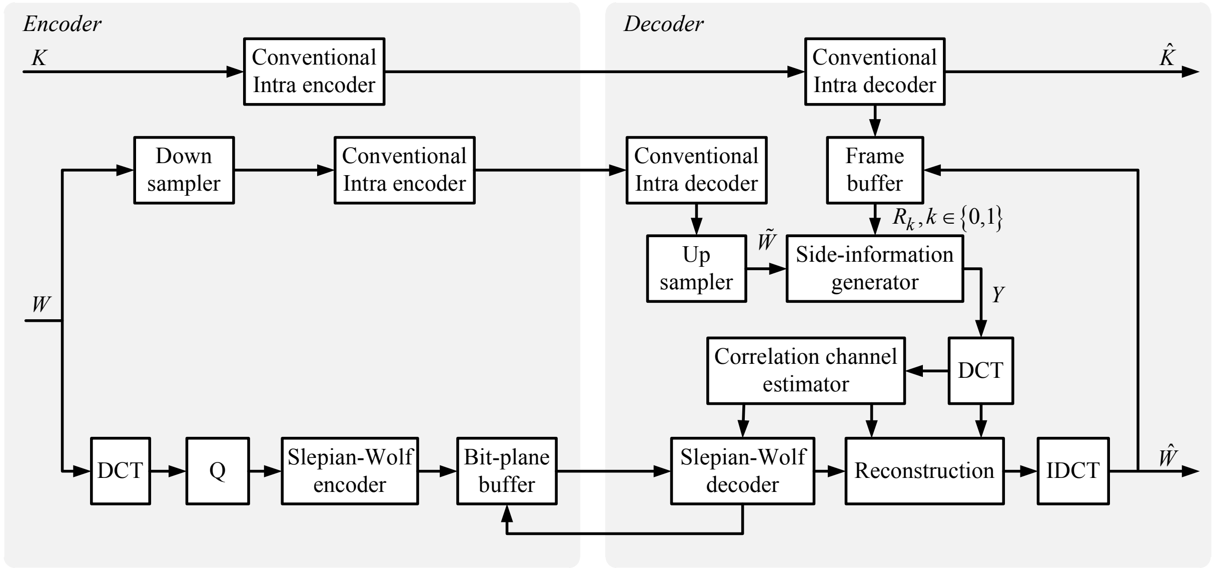

Encoder
Decoder
K
Conventional Intra encoder
Conventional Intra decoder
K̂
Down sampler
Conventional Intra encoder
Conventional Intra decoder
Frame buffer
$R_k, k \in \{0,1\}$
Up sampler
W̃
Side-information generator
Y
W
Correlation channel estimator
DCT
DCT
Q
Slepian-Wolf encoder
Bit-plane buffer
Slepian-Wolf decoder
Reconstruction
IDCT
Ŵ

Figure 2 Schematic overview of the proposed Wyner-Ziv video codec.

quantized low frequency DCT bands [29] in the Wyner-Ziv frames.

Furthermore, we note that, in contrast to other hash-based DVC solutions [12,28], the proposed architecture avoids block-based decisions on the transmission of the hash at the encoder side. Although this can increase the hash rate overhead when easy-to-predict motion content is coded, it comes at the benefit of constraining the encoding complexity, in the sense that the encoder is not burdened by expensive block-based comparisons or memory requirements necessary for such mode decision. An additional key advantage of the presented hash code is that it facilitates accurate side information creation using pixel-based multi-hypothesis compensation at the decoder, as explained in Section 4.2.2. In this way, the presented hash code enhances the RD performance of the proposed system especially for irregular motion content, e.g., endoscopic video material.

4.1.2. Wyner-Ziv encoding

In addition to the coded hash, a Wyner-Ziv layer is created for every Wyner-Ziv frame, providing efficient compression [5] and scalable coding [25]. In line with the DVC architecture introduced in [5], the Wyner-Ziv frames are first transformed with a 4 × 4 integer approximation of the DCT [1] and the obtained coefficients are subsequently assembled in frequency bands. Each DCT band is independently quantized using a collection of predefined quantization matrices (QMs) [26], where the DC and the AC bands are quantized with a uniform and double-deadzone scalar quantizer, respectively. The quantized symbols are translated into binary codewords and passed to a LDPC Accumulate (LDPCA) encoder [36], assuming the role of Slepian-Wolf encoder.

The LDPCA [36] encoder realizes Slepian and Wolf's random binning argument [15] through linear channel code syndrome binning. In detail, let **b** be a binary M-tuple containing a bit-plane of a coded DCT band β of a Wyner-Ziv frame, where M is the number of coefficients in the band. To compress **b**, the encoder employs an (M, k) LDPC channel code C constructed by the generator matrix $\mathbf{G}_{k\times M} = \left[\mathbf{I}_k \quad \mathbf{P}_{k\times(M-k)}\right] \cdot \mathrm{c}^{\mathrm{c}}$. The corresponding parity check matrix of C is $\mathbf{H}_{(M-k)\times M} = \left[\mathbf{P}^T_{k\times(M-k)} \quad \mathbf{I}_{M-k}\right]$. Thereafter, the encoder forms the syndrome vector as $\mathbf{s} = \mathbf{b}\mathbf{H}^T$. In order to achieve various puncturing rates, the LDPC syndrome-based scheme is concatenated with an accumulator [36]. Namely, the derived syndrome bits **s** are in turn mod-2 accumulated, producing the accumulated syndrome tuple $\boldsymbol{\alpha}$. The encoder stores the accumulated syndrome bits in a buffer and transmits them incrementally upon the decoder's request using a feedback channel, as explained in Section 4.2.3. Note that contemporary wireless (implantable) sensors–including capsule endoscopes–support bidirectional communication [33,37,38]. That is, a feedback channel from the encoder to the decoder is a viable solution for the pursued applications. The effect of the employed feedback channel on the decoding delay, and in turn on the buffer requirements at the encoder of a wireless capsule endoscope, is studied in Section 5.3.

Note that the focus of this study is to successfully target various lightweight applications by improving the compression efficiency of Wyner-Ziv video coding while maintaining low computational cost at the encoder. Hence, in order to accurately evaluate the impact of the proposed techniques on the RD performance, the proposed system employs LDPCA codes which are also used in the state-of-the-art codecs of [13,26]. Observe that for distributed compression under a noiseless transmission scenario the syndrome-based Slepian-Wolf scheme [15] is optimal since it can achieve the information theoretical bound with the shortest channel codeword length [23]. Nevertheless, in order to address distributed joint source-channel coding (DJSCC) in a noisy transmission scenario the parity-based [23] Slepian-Wolf scheme needs to be deployed. In the latter, parity-check bits are employed to indicate the Slepian-Wolf bins, thereby achieving equivalent Slepian-Wolf compression performance at the cost of an increased codeword length [23].

It is important to mention that, conversely to other hash-driven Wyner-Ziv schemes operating in the transform domain, e.g., [12,31], the presented Wyner-Ziv encoder encodes the entire original Wyner-Ziv frame, instead of coding the difference between the original frame and the reconstructed hash. The motivation for this decision is twofold. The first reason stems from the nature of the hash. Namely, coding the difference between the Wyner-Ziv frame and the reconstructed hash would require decoding and interpolating the hash at the encoder, an operation which is computationally demanding and would pose an additional strain on the encoder's memory demands. Second, compressing the entire Wyner-Ziv frame with linear channel codes enables the extension of the scheme to the DJSCC case [23], thereby providing error-resilience for the entire Wyner-Ziv frame if a parity based Slepian-Wolf approach is followed.

4.2. The decoder

The main components of the presented DVC architecture's decoding process are treated separately, namely dealing with the hash, side information generation and Wyner-Ziv decoding. The decoder first conventionally intra decodes the key frame bit stream and stores the reconstructed frame in the reference frame buffer. In the following phase, the hash is handled, which is detailed next.

4.2.1. Hash decoding and reconstruction

The hash bit-stream is decoded with the appropriate conventional intra codec. The reconstructed hash is then upscaled to the original Wyner-Ziv frame's resolution. The ideal upscaling process consists of upsampling followed by ideal interpolation filtering. The ideal interpolation filter is a perfect low-pass filter with gain d and cut-off frequency π/d without transition band [39]. However, such a filter corresponds to an infinite length impulse response h_{ideal}, to be precise, a sinc function $h_{\text{ideal}}(n) = \text{sinc}(n/d)$ where $n \in \mathbb{Z}_+$, which cannot be implemented in practice.

Therefore our system employs a windowing method [39] to create a filter with finite impulse response $h(n)$, namely

$$h(n) = h_{\text{ideal}}(n) \cdot z(n), \quad |n| < 3 \cdot d, \tag{1}$$

where the window function $z(n)$ corresponds to samples taken from the central lobe of a sinc function, that is

$$z(n) = \text{sinc}\left(\frac{n}{3 \cdot d}\right), \quad |n| < 3 \cdot d. \tag{2}$$

Such interpolation filter is known in the literature as a Lanczos3 filter [35]. Following [40], the resulting filter taps are normalized to obtain unit DC gain while the input samples are preserved by the upscaling process since $h_0(n) = 1$.

4.2.2. Side information generation

After the hash has been restored to the same frame size as the original Wyner-Ziv frames, it is used to perform decoder-side motion estimation. The quality of the side information is an important factor on the overall compression performance of any Wyner-Ziv codec, since the higher the quality the less channel code rate is required for Wyner-Ziv decoding. The proposed side information generation algorithm performs bidirectional overlapped block motion estimation (OBME) using the available hash information and a past and a future reconstructed Wyner-Ziv and/or key frame as references.

Temporal prediction is carried out using a hierarchical frame organization, similar to the prediction structures used in [5,12,26]. It is important to note that conversely to our previous study [30], in which motion estimation was based on bit-planes, this study follows a different approach regarding the nature of the hash as well as the block matching process. Before motion estimation is initiated, the reference frames are preprocessed. Specifically, to improve the consistency of the resulting motion vectors, the reference frames are first subjected to the same downsampling and interpolation operation as the hash.

Figure 3 contains a graphical representation of the motion estimation algorithm. To offer a clear presentation of the proposed algorithm, we introduce the following notation. Let $\tilde{W}$ be the reconstructed hash of a Wyner-Ziv frame, let Y be the side information and let $\tilde{R}_k$, $k \in \{0,1\}$ be the preprocessed versions of the reference frames R_k, respectively. Also, denote by $Y_{\mathbf{m}}$, $R_{k,\mathbf{m}}$, $\tilde{R}_{k,\mathbf{m}}$, $\tilde{R}_{k,\mathbf{m}}$ the blocks of size $B \times B$ pixels with top-left coordinates $\mathbf{m} = (m_1, m_2)$ in Y, R_k, $\tilde{W}$ and $\tilde{R}_k$,

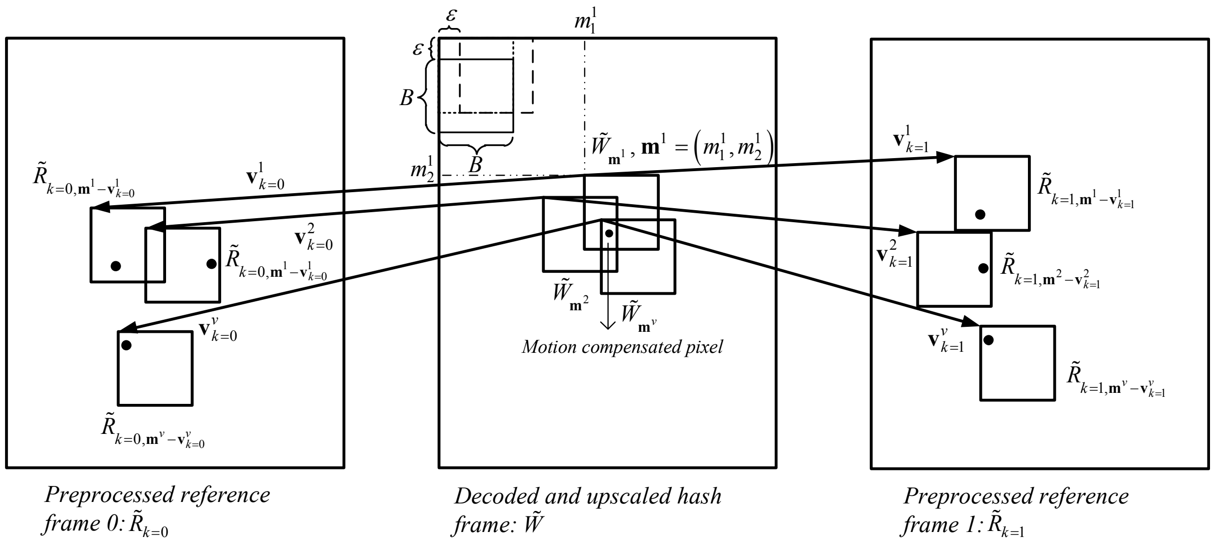

ε
m_1^1
ε
B
B
m_2^1
$\tilde{W}_{\mathbf{m}^1}, \mathbf{m}^1 = (m_1^1, m_2^1)$
$\tilde{R}_{k=0,\mathbf{m}^1-\mathbf{v}^1_{k=0}}$
$\mathbf{v}^1_{k=0}$
$\mathbf{v}^2_{k=0}$
$\tilde{R}_{k=0,\mathbf{m}^1-\mathbf{v}^1_{k=0}}$
$\mathbf{v}^v_{k=0}$
$\tilde{R}_{k=0,\mathbf{m}^v-\mathbf{v}^v_{k=0}}$
$\tilde{W}_{\mathbf{m}^2}$
$\tilde{W}_{\mathbf{m}^v}$
Motion compensated pixel
$\mathbf{v}^1_{k=1}$
$\tilde{R}_{k=1,\mathbf{m}^1-\mathbf{v}^1_{k=1}}$
$\mathbf{v}^2_{k=1}$
$\tilde{R}_{k=1,\mathbf{m}^2-\mathbf{v}^2_{k=1}}$
$\mathbf{v}^v_{k=1}$
$\tilde{R}_{k=1,\mathbf{m}^v-\mathbf{v}^v_{k=1}}$
Preprocessed reference frame 0: $\tilde{R}_{k=0}$
Decoded and upscaled hash frame: $\tilde{W}$
Preprocessed reference frame 1: $\tilde{R}_{k=1}$

Figure 3 Graphical representation of the motion estimation algorithm. All the overlapping $B \times B$ blocks, that contain the current motion-compensated pixel position in the hash frame $\tilde{W}$, are designated $\tilde{W}_{\mathbf{m}^i}$, where $\mathbf{m}^i = (m_1^i, m_2^i)$ are the top-left coordinates of the block. For every block $\tilde{W}_{\mathbf{m}^i}$ the best matching block in the preprocessed reference frames $\tilde{R}_{k=0}, \tilde{R}_{k=1}$ is found by minimizing the SAD, yielding the motion vectors $\mathbf{v}^i_{k=0}, \mathbf{v}^i_{k=1}$. The co-located pixels in the estimated blocks $\tilde{R}_{k=0,\mathbf{m}^i-\mathbf{v}^i_{k=0}}, \tilde{R}_{k=1,\mathbf{m}^i-\mathbf{v}^i_{k=1}}$ serve as potential temporal predictors for the current motion-compensated pixel position.

respectively. Finally, let $Y_{\mathbf{m}}(\mathbf{p})$ designate the sample at position $\mathbf{p} = (p_1, p_2)$ in the block $Y_{\mathbf{m}}$.

At the outset, the available hash frame is divided into overlapping spatial blocks, $\tilde{W}_{\mathbf{u}}$, with top-left coordinates $\mathbf{u} = (u_1, u_2)$, using an overlapping step size $\varepsilon \in \mathbb{Z}_+$, $1 \leq \varepsilon \leq B$. For each overlapping block $\tilde{W}_{\mathbf{u}}$, the best matching block within a specified search range ρ, is found in the reference frames $\tilde{R}_k$. In contrast to our earlier study [30], the proposed algorithm retains the motion vector $\mathbf{v} = (v_1, v_2)$, $-\rho < v_1, v_2 \leq \rho$, which minimizes the sum of absolute differences (SAD) between $\tilde{W}_{\mathbf{u}}$ and a block $\tilde{R}_{k,\mathbf{u}-\mathbf{v}}$, $\mathbf{v} = (v_1, v_2)$, in other words

$$\mathbf{v} = \arg\min_{\mathbf{v}} \sum_{\mathbf{p}} \left| \tilde{W}_{\mathbf{u}}(\mathbf{p}) - \tilde{R}_{k,\mathbf{u}-\mathbf{v}}(\mathbf{p}) \right|, \tag{3}$$

where $\mathbf{p}$ visits all the co-located pixel positions in the blocks $\tilde{W}_{\mathbf{u}}$ and $\tilde{R}_{k,\mathbf{u}-\mathbf{v}}$, respectively. The motion search is executed at integer-pel accuracy and the obtained motion field is extrapolated to the original reference frames R_k. By construction, every pixel $Y(\mathbf{p})$, $\mathbf{p} = (p_1, p_2)$ in the side information frame Y is located inside a number of overlapping blocks $Y_{\mathbf{u}_n}$ with $\mathbf{u}_n = (u_{n,1}, u_{n,2})$. After the execution of the OBME, a temporal predictor block $R_{k,\mathbf{u}_n}$ for every block $Y_{\mathbf{u}_n}$ has been identified in one reference frame. As a result, each pixel $Y(\mathbf{p})$ in the side information frame has a number of associated temporal predictors $r_{k,\mathbf{u}_n}$ in the blocks $R_{k,\mathbf{u}_n}$.

However, some temporal predictors may stem from rather unreliable motion vectors. Especially when the input sequence was recorded at low frame rates or when the motion content is highly irregular, as might be the case in endoscopic sequences, temporal prediction is not the preferred method for all blocks at all times. Therefore, to avoid quality degradation of the side information due to untrustworthy predictors, all obtained motion vectors are subjected to a reliability screening. Namely, when the SAD, based on which the motion vector associated with temporal predictor $r_{k,\mathbf{u}_n}$ was determined, is not smaller than a certain threshold T, the motion vector and associated temporal predictor is labeled as *unreliable*. In this case, a temporal predictor for the side information pixel $Y(\mathbf{p})$ is replaced by the co-located pixel of $Y(\mathbf{p})$ in the upsampled hash frame, that is $\tilde{W}(\mathbf{p})$. In other words, when motion estimation is considered not to be trusted, the hash itself is assumed to convey more dependable information. This feature of OBME is referred to as hash-predictor-selection (HPS).

During the motion compensation process, the obtained predictors per pixel, whether being temporal predictors or taken from the upsampled hash, are combined to perform multi-hypothesis pixel-based prediction. Specifically, every side information pixel $Y(\mathbf{p})$ is calculated as the mean value of the predictor values $g_{k,\mathbf{u}_n}$:

$$Y(\mathbf{p}) = \frac{1}{N_{k,\mathbf{u}_n}} \sum_{\mathbf{u}_n} g_{k,\mathbf{u}_n}, \tag{4}$$

where, $N_{k,\mathbf{u}_c}$ denotes the number of predictors for pixel $Y(\mathbf{p})$ and $g_{k,\mathbf{u}_n} = r_{k,\mathbf{u}_n}$ when $r_{k,\mathbf{u}_n}$ is *reliable* or $g_{k,\mathbf{u}_n} = \tilde{W}(\mathbf{p})$ when $r_{k,\mathbf{u}_n}$ is *unreliable*. The derived multi-hypothesis motion field is employed in an analogous manner to estimate the chroma components of the side information frame from the chroma components of the reference frames R_k or the upsampled hash.

4.2.3. Wyner-Ziv decoding

The derived motion-compensated frame is first DCT transformed to serve as side information Y for decoding the Wyner-Ziv bit stream in the transform domain. Then, online transform domain correlation channel estimation [7] is carried out to model the correlation channel between the side information Y and the original Wyner-Ziv frame samples W in the DCT domain. As in [7], the correlation is expressed by an additive input-dependent noise model, $W = Y + N$, where the correlation noise $N \sim \mathcal{L}(0, \sigma(y))$ is zero-mean Laplacian with standard-deviation $\sigma(y)$, which varies depending on the realization y of the input of the channel, i.e., the side information, namely [7],

$$f_{N|Y}\left(n\,|y\right) = \frac{1}{\sigma\left(y\right)\sqrt{2}} e^{-\frac{\sqrt{2}\,|n|}{\sigma\left(y\right)}} \tag{5}$$

Thereafter, the estimated correlation channel statistics per coded DCT band bit-plane are interpreted into soft estimates, i.e., log-likelihood ratios (LLRs). These LLRs, which provide *a priori* information about the probability of each bit to be 0 or 1, are passed to the variable nodes of the LDPCA decoder. Then, the message passing algorithm [41] is used for iterative LDPC decoding, in which the received syndrome bits correspond to the check nodes on the bipartite graph.

Notice that the scheme follows a layered Wyner-Ziv coding approach to provide quality scalability without experiencing a performance loss [25]. Namely, in the formulation of the LLRs, infor-mation given by the side information and the already decoded source bit-planes is taken into account. In detail, let b_l denote a bit of the lth bit-plane of the source and b_1, ..., b_{l-1} be the already decoded bits in the previous l -1 bit-planes. Then the estimated LLR at the corresponding variable node of the LDPCA decoder is given by

$$\text{LLR} = \log\frac{p\left(b_l = 0\,|y, b_1, \ldots, b_{l-1}\right)}{p\left(b_l = 1\,|y, b_1, \ldots, b_{l-1}\right)} = \log\frac{p\left(b_1, \ldots, b_{l-1}, b_l = 0\,|y\right)}{p\left(b_1, \ldots, b_{l-1}, b_l = 1\,|y\right)} \tag{6}$$

where equality in (6) stems from: $p(b_l|y, b_1, ..., b_{l-1}) = p(b_1, ..., b_{l-1}, \mathrm{b}_l|y)/p(b_1, ..., b_{l-1}|y)$. Hence, in (6) the nominator and the denominator are calculated by integrating the conditional probability density function of the correlation channel, i.e., $f_{X|Y}(x|y)$, over the quantization bin indexed by b_1, ..., b_l.

Remark that the LDPCA decoder achieves various rates by altering the decoding graph upon reception of an additional increment of the accumulated syndrome [36]. Initially, the decoder receives a short syndrome based on an aggressive code and the decoder tries to decode [36]. If decoding falls short, the encoder receives a request to augment the previously received syndrome with extra bits. The process loops until the syndrome is sufficient for successful decoding.

Once all the L bit-planes of a DCT band of a Wyner-Ziv frame are LDPCA decoded, the obtained L binary M-tuples $\mathbf{b}_1$, $\mathbf{b}_2$, ..., $\mathbf{b}_L$ are combined to form the decoded quantization indices of the coefficients of the band. Subsequently, the decoded quantization indices are fed to the reconstruction module which performs inverse quantization using the side information and the correlation channel statistics. Since the mean square error distortion measure is employed, the optimal reconstruction of a Wyner-Ziv coefficient w is obtained as the centroid of the random variable W given the corresponding side information coefficient y and the decoded quantization index q [25]. Namely

$$E\left[w\,|y, q\right] = \frac{\int_{q_L}^{q_H} w f_{W|y}\left(w\,|y\right)}{\int_{q_L}^{q_H} f_{W|y}\left(w\,|y\right)} \tag{7}$$

where, q_{L}, q_{H} denote the lower and upper bound of the quantization bin q. Finally, the inverse DCT transform provides the reconstructed frame $\hat{W}$ in the spatial domain. The reconstructed frame is now ready for display and is stored in the reference frame buffer, serving as a reference for future temporal prediction.

5. Evaluation

The experimental results have been divided into three distinct parts. Namely, first the proposed system is compared against a set of relevant alternative video coding solutions using traditional test sequences. The second part comprises the experimental validation of our system in the application of wireless capsule endoscopy, comparing its performance against coding solutions currently used for the compression of endoscopic video. The third part elaborates on the encoding complexity of the proposed architecture.

We begin by defining the configuration elements of the proposed system, which are common to both types of input video. Namely, the motion estimation algorithm was configured with an overlap step size ε = 4, the size of the overlapping blocks was set to B = 16 and the threshold was chosen T = 400. The motion search was executed in an exhaustive manner at integer-pel accuracy within a search range of ± 16 pixels. The downscaling factor to create the hash was fixed at d = 2.

5. 1. Evaluation on conventional test sequences

Regarding the performance evaluation of the proposed hash-based DVC on conventional video sequences, comparisons were conducted against three state-of-the-art reference codecs, namely, DISCOVER[d] [26], the hash-based scheme in [12] and H.264/AVC Intra. Comparative tests were carried out on the complete Foreman, Soccer, Silent, and Ice sequences, at QCIF resolution and at a frame rate of 15 Hz. To assess the RD performance of the codec in a small and a large GOP, results are depicted for GOP sizes of two and eight frames. To express the difference in the coding performance in terms of the Bjøntegaard Delta (BD) metric [42], four RD points have been drawn corresponding to QMs 1, 5, 7, and 8 of [26]. In this experimental setting, the hash and the key frames of our proposed system were coded with the H.264/AVC Intra codec (Main profile), since the assessed codecs employ H.264/AVC Intra as well. For a fair comparison, the employed quality parameters (QPs) (per RD point and per sequence) of the key frames are exactly the same with the ones employed in the reference Wyner-Ziv codecs [26]. Similar to the method used to find the key frames' QPs in [26], an off-line iterative scheme has been employed to determine the hash QPs in our codec. The process[e] was carried out on the first 15 frames of a sequence and on a GOP of 2 (this GOP size was also used in [26] to determine the QPs for the key frames). The relative standard deviation (RSD) of the PSNR values was used as a metric of the quality fluctuation of the decoded sequence. The parameters used and the resulting RSD per sequence and RD point are reported in Table 1. Although the proposed codec supports chroma (YUV) encoding –see Section 4.2.2, the experimental results presented in this section are only obtained for the luma (Y) component to allow a meaningful comparison with prior art [12,26].

The experimental results in Figures 4 and 5 show that the proposed hash-based DVC regularly outperforms the DISCOVER [26] codec. Notice that when the size of the GOP and the amount of motion in the sequence increases, the overall compression performance of the DISCOVER codec notably decreases with respect to the proposed DVC, which is mainly due to the quality degradation of DISCOVER's MCI-based side information generation. Hence, the proposed system consistently outperforms DISCOVER in Foreman, Ice, and Soccer, all of which contain rather complex motion patterns (above all Soccer), and this for both GOP sizes. Especially for a GOP size of 8, the recorded gains are significant with BD rate savings [42] of 24.77, 26.30, and 32.13%, in Foreman, Ice, and Soccer, respectively. Comparing the compression performance of both DVC systems on the Silent sequence, which contains a low amount of motion activity, the MCI-based DISCOVER slightly surpasses the proposed DVC. This is due to the fact that in low-motion sequences a hash is not required to accurately capture the motion pattern at the

Table 1 Employed quantization parameters for the key, the hash and the Wyner-Ziv frames as well as the resulting RSD for the entire sequence

	RD point 1 (QM1)	RD point 2 (QM4)	RD point 3 (QM7)	RD point 4 (QM8)
Ice				
Key frame QP	40	34	29	25
Hash QP	41	40	39	38
RSD(%)	2.25	2.26	1.90	1.28
Foreman				
Key frame QP	40	34	29	25
Hash QP	41	40	39	38
RSD(%)	2.92	2.97	2.58	1.96
Silent				
Key frame QP	37	33	29	24
Hash QP	40	39	38	37
RSD(%)	2.29	1.02	0.54	2.38
Soccer				
Key frame QP	44	36	31	25
Hash QP	45	42	41	38
RSD(%)	4.41	3.29	2.96	2.73

The RSD is given by RSD(%) = 100 × σ_{PSNR}/μ_{PSNR}, where σ_{PSNR} and μ_{PSNR} are the standard deviation and the mean of the PSNR values

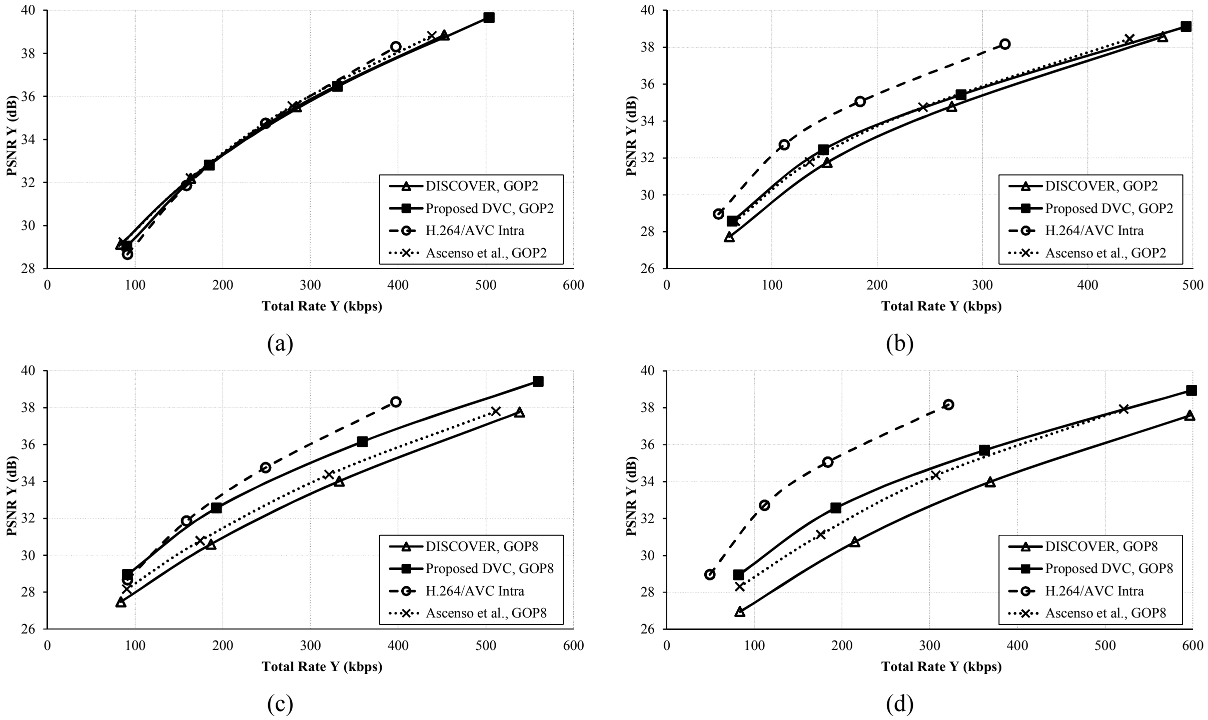


Figure 4 Experimental results obtained on traditional test sequences. The proposed hash-based DVC is compared against DISCOVER, the system in [12] and H.264/AVC Intra. The figure shows the RD performance corresponding to Foreman (left), and Soccer (right) at QCIF resolution, a frame rate of 15 Hz, and a GOP of **(a b)** 2 and **(c, d)** 8. Only the *Y* component is coded.

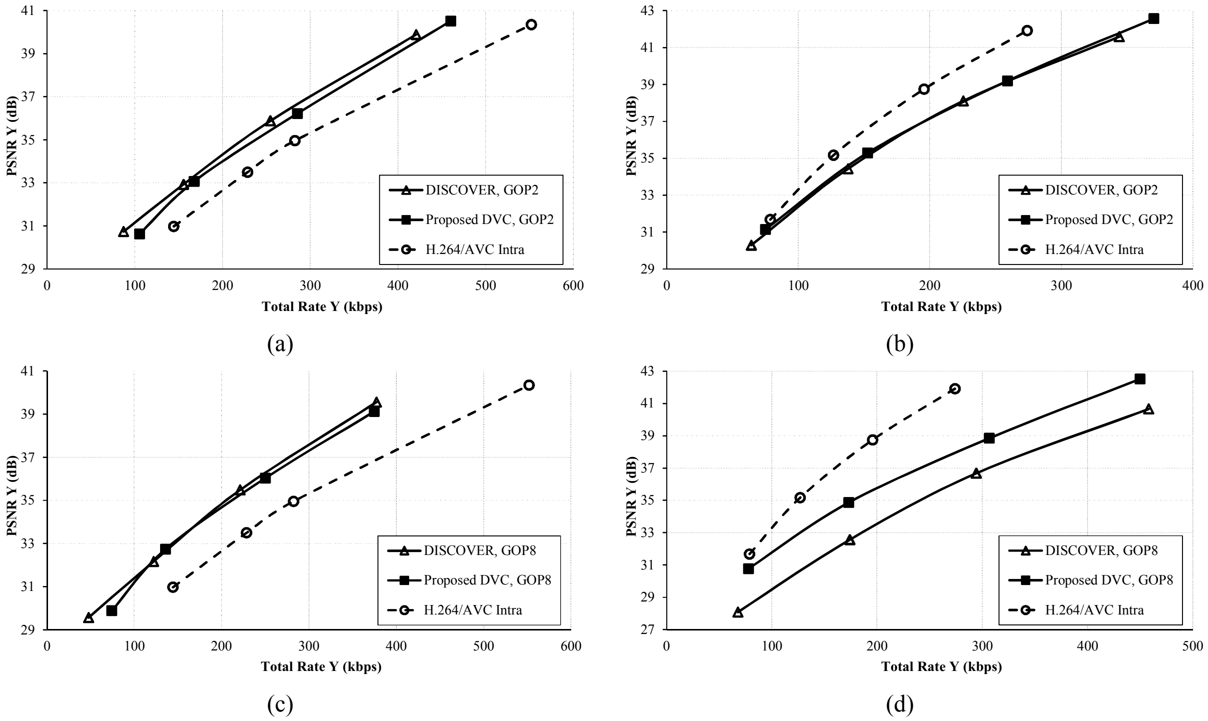


Figure 5 Experimental results obtained on traditional test sequences. The proposed hash-based DVC is compared against DISCOVER and H.264/AVC Intra. The figure shows the RD performance corresponding to Silent (left), and Ice (right) at QCIF resolution, a frame rate of 15 Hz, and a GOP of **(a, b)** 2 and **(c, d)** 8. Only the *Y* component is coded.

decoder, as this can be simply achieved via interpolation. The incurred loss in RD per-formance is albeit reasonable, at the level of 7.9% for GOP2, and decreasing with growing GOP size to 5.4% for GOP8.

To further evaluate the performance of our proposed scheme, the coding results of [12] are included in Figure 4. The hash-based Wyner-Ziv video codec of [12] combines MCI with hash-driven motion estimation using low quality H.264/AVC Intra coded Wyner-Ziv blocks to generate side information. Even though the codec of [12] advances over DISCOVER, our proposed hash-based solution generally exhibits higher performance bringing BD rate savings of 17.68 and 12.18% in Foreman and Soccer, in GOP8, respectively.

Lastly, the proposed DVC is compared with H.264/AVC Intra, which represents the low-complexity configuration of the state-of-the-art traditional coding paradigm. One can observe from Figure 5 that in low-motion sequences the proposed codec is superior to H.264/AVC Intra, bringing BD rate savings of up to 26.7% in Silent, GOP8. However, under difficult motion conditions like in Ice or Soccer H.264/AVC Intra is very efficient compared to DVC systems, which is in agreement with the results shown in Figures 4 and 5. We emphasize that the encoding complexity of H.264/AVC Intra is much higher than any of the presented DVC solutions, as discussed in Section 5.3.

In Figure 6, we schematically depict the contribution of the LPDCA, the hash, and the key-frame rate to the total rate of the proposed coding system. The results show that as the GOP size increases, namely as more Wyner-Ziv frames are coded, the hash and the LDPCA rates increase, whereas the key-frame rate decreases. Notice also that, for a given sequence and GOP size, as the total rate increases from RD points 1 to 4, the contribution of the hash rate diminishes in favor of the LDPCA rate. Furthermore, the relative contribution of the hash rate to the total rate becomes smaller when high-motion sequences are coded–see Figure 6b, since relatively more LDPCA rate is spent.

5.2. Evaluation on endoscopic video sequences

A major contribution of this article is the assessment of Wyner-Ziv coding for endoscopic video data, characterized by its unique content. In the proposed codec, the quantization parameters of the Wyner-Ziv frames, the key frames, and the hash are meticulously selected so as to retain high and quasi-constant decoded frames' quality, as demanded by medical applications. Furthermore, in order to deliver high-quality decoding under the strenuous conditions of highly irregular motion content and low frame acquisition rates, the proposed codec employs a GOP size of 2.

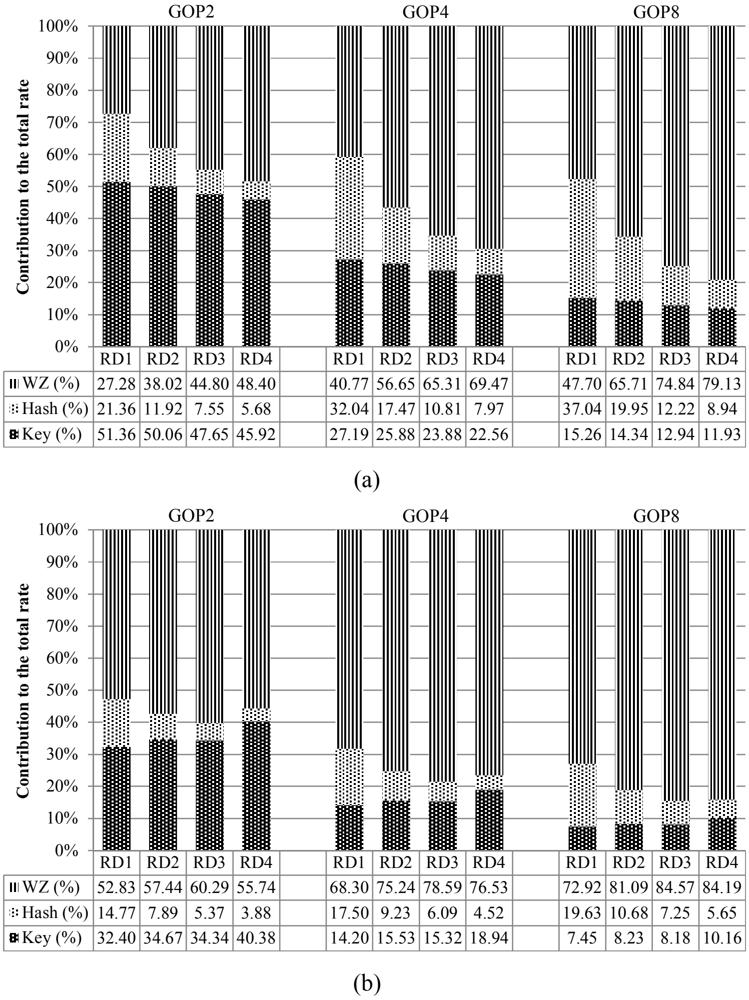
Initially, in order to prove the potential of its application in contemporary wireless capsule endoscopic technology, the proposed codec has been appraised using four capsule endoscopic test video sequences visualizing diverse areas of the gastrointestinal track. These sequences were extracted from extensive capsule endoscopic video material of two capsule examinations from two random volunteers[f] performed at the Gastroenterology Clinic of the Universitair Ziekenhuis Brussels, Belgium. In the aforementioned clinical examinations, the capsule acquisition rate was two frames per second with a frame resolution of 256 × 256 pixels. The obtained test video sequences[g] are termed "Capsule Test Video 1" to "Capsule Test Video 4" in the remainder of the article.

In the set of experiments comprising capsule endoscopic video content, Motion JPEG has been set as benchmark, since this technology is commonly employed in up-to-date capsule endoscopes [34]. To enable a fair comparison, Motion JPEG has also been employed to code the key and the hash frames in the proposed codec. We note that in this experimental setting the luma (Y) and the chroma (U and V) components were coded. The results, which are illustrated in Figure 7, show that the proposed codec generally outperforms Motion JPEG for the capsule endoscopic sequences. In particular, in "Capsule Test Video 1" and "Capsule Test Video 2" the proposed codec brings average Bjøntegaard rate savings of, respectively, 6.16 and 9.33% against Motion JPEG.

Figure 7 also evaluates the impact of the flexible scheme that enables the proposed OBME method to identify erroneous motion vectors and to replace the temporal predictor pixel with the decoded and interpolated hash. The results show that the proposed system with the HPS module remarkably advances over its equivalent that solely retains predictors from the reference frames. Specifically, in "Capsule Test Video 1" to "Capsule Test Video 4" adding the HPS functionality results in BD [42] rate improvements of 21.1, 16.02, 12.93, and 12.06%, respectively.

The visual assessment of the proposed codec (with HPS) compared to Motion JPEG for a Wyner-Ziv frame of "Capsule Test Video 1" and "Capsule Test Video 2" is depicted in Figures 8 and 9, respectively.

Future generations of capsule endoscopic technology aim at diminishing the quality difference with respect to conventional endoscopy by increasing the frame rate and resolution. Therefore, to confirm its capability under these conditions, the proposed Wyner-Ziv video codec is evaluated using conventional endoscopic video sequences monitoring diverse parts of the digestive track of several patients. The endoscopic test video



	GOP2					GOP4					GOP8			
	RD1	RD2	RD3	RD4		RD1	RD2	RD3	RD4		RD1	RD2	RD3	RD4
▥ WZ (%)	27.28	38.02	44.80	48.40		40.77	56.65	65.31	69.47		47.70	65.71	74.84	79.13
⁞ Hash (%)	21.36	11.92	7.55	5.68		32.04	17.47	10.81	7.97		37.04	19.95	12.22	8.94
▩ Key (%)	51.36	50.06	47.65	45.92		27.19	25.88	23.88	22.56		15.26	14.34	12.94	11.93

(a)

	GOP2					GOP4					GOP8			
	RD1	RD2	RD3	RD4		RD1	RD2	RD3	RD4		RD1	RD2	RD3	RD4
▥ WZ (%)	52.83	57.44	60.29	55.74		68.30	75.24	78.59	76.53		72.92	81.09	84.57	84.19
⁞ Hash (%)	14.77	7.89	5.37	3.88		17.50	9.23	6.09	4.52		19.63	10.68	7.25	5.65
▩ Key (%)	32.40	34.67	34.34	40.38		14.20	15.53	15.32	18.94		7.45	8.23	8.18	10.16

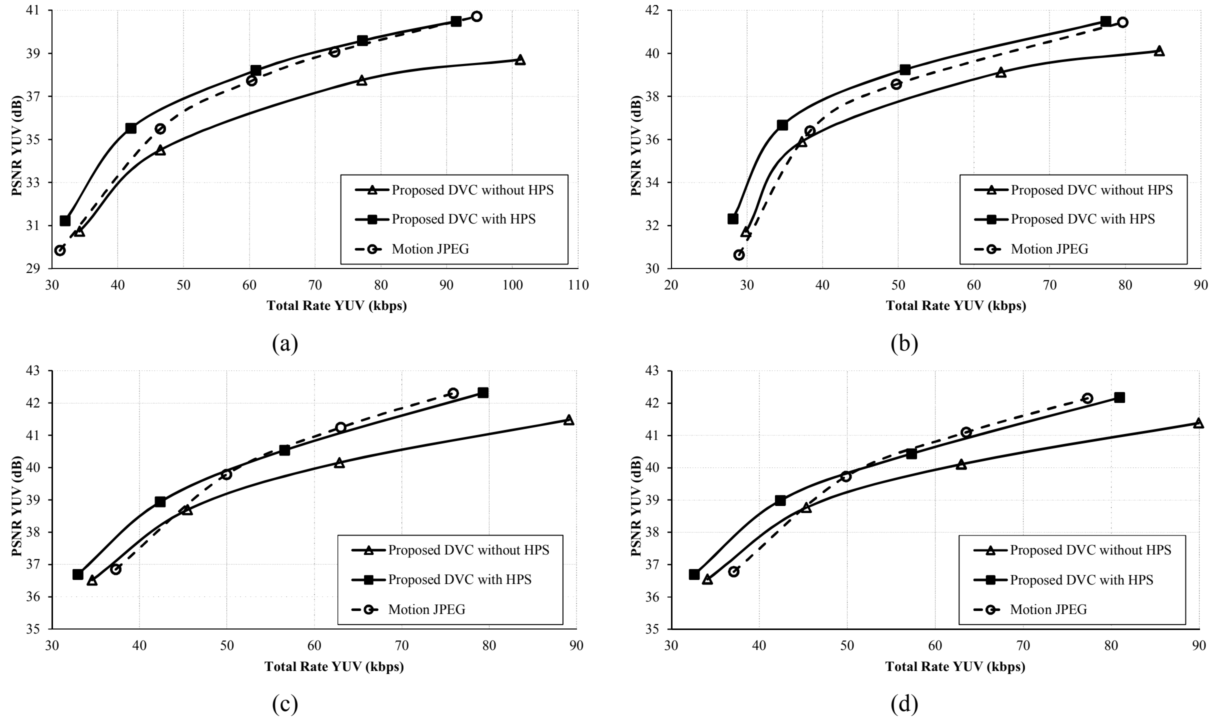
(b)

Figure 6 Contribution of the LDPCA (WZ), the hash and the key-frame rate to the total rate of the proposed system for (a) Foreman and (b) Soccer. The results are provided for GOP sizes of 2, 4, and 8 frames and the four rate points.

sequences considered in this experimental setting have a frame rate of 30 Hz and a frame resolution of 480 × 320 pixels. These endoscopic test video sequences are further referred to as “Endoscopic Test Video 1” to “Endoscopic Test Video 6”. In this experiment, the proposed codec employs H.264/AVC Intra (Main profile) to code the key and the hash frames. Notice that the H.264/AVC Intra codec constitutes a recognized reference for medical video compression, e.g. [43].

In Figure 10, the proposed DVC system (with and without HPS) is evaluated against the H.264/AVC Intra and our previous TDWZ codec of [13]. The latter features an MCI framework comprising overlapped block motion compensation to generate side information at the decoder. The TDWZ codec of [13] provides state-of-the-art MCI-based DVC performance, outperforming the DISCOVER [26] codec. Remark that the DISCOVER codec could not be included in the comparison since it does not support the frame resolution of the video data. The results are provided only for the luma component of the sequences “Endoscopic Test Video 1” to “Endoscopic Test Video 4”, since the codec in [13] does not support chroma encoding. The experimental results depicted in Figure 10 show that the proposed codec (with HPS) delivers significant compression gains over the state-of-the-art TDWZ codec of [13]. Specifically, in “Endoscopic Test Video 1” and “Endoscopic Test Video 2” the proposed codec introduces average BD [42] rate savings of 43.14 and 43.37%, respectively. These remarkable compression gains clearly motivate the proposed hash-based Wyner-Ziv architecture comprising our novel motion-compensated multi-hypothesis prediction scheme over MCI-based solutions.

Compared to H.264/AVC Intra, the experimental results in Figure 10 show that the proposed codec delivers BD rate savings of 4.1% in “Endoscopic Test Video



(a)

(b)

(c)

(d)

Figure 7 Experimental results on data acquired from a wireless capsule endoscope: (a) “Capsule Test Video 1”, (b) “Capsule Test Video 2”, (c) “Capsule Test Video 3”, and (d) “Capsule Test Video 4”. The RD performance of the proposed system (with and without HPS) is compared against that of Motion JPEG. All three *Y, U*, and *V* components are coded. The average YUV PSNR is given by $\mathrm{PSNR}_{YUV} = (4 \times \mathrm{PSNR}_Y + \mathrm{PSNR}_U + \mathrm{PSNR}_V)/6$.

2”. In “Endoscopic Test Video 1” and “Endoscopic Test Video 3” the proposed codec falls behind H.264/AVC Intra, incurring a BD rate loss of 3.84 and 0.20%, respectively. Only in “Endoscopic Test Video 4”, which comprises highly irregular motion, the experienced Bjøntegaard rate overhead is notable amounting to 15.68%. Notice that the benefit of the HPS functionality of the proposed codec is reduced in case of conventional endoscopic video with respect to the capsule endoscopic sequences. This is due to the fact that the former sequences were recorded at a much higher frame rate and contain more temporal correlation. Nevertheless, in “Endoscopic Test Video 4” the HPS module brings BD rate savings of 4.63%.

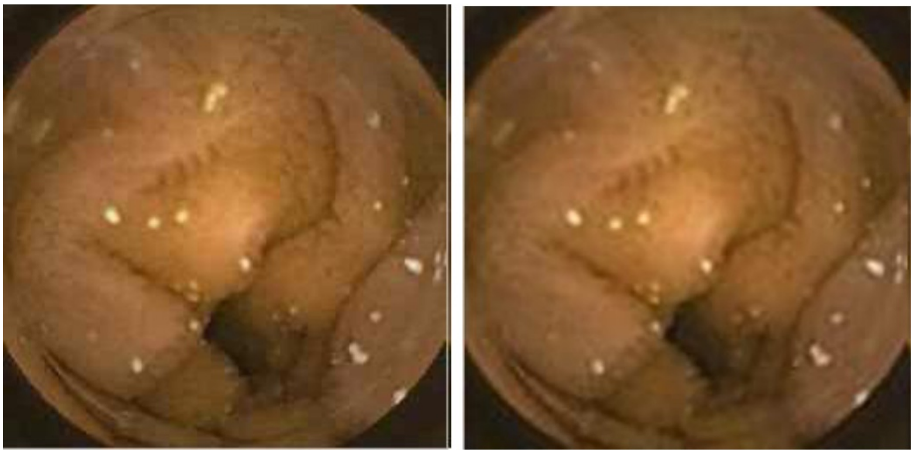


Figure 8 Visual snapshots from the decoded “Capsule Test Video 1” sequence. (left) Motion JPEG (72.94 kbps, 39.07 dB), (right) the proposed hash-based Wyner-Ziv video codec (77.1 kbps, 39.59 dB).

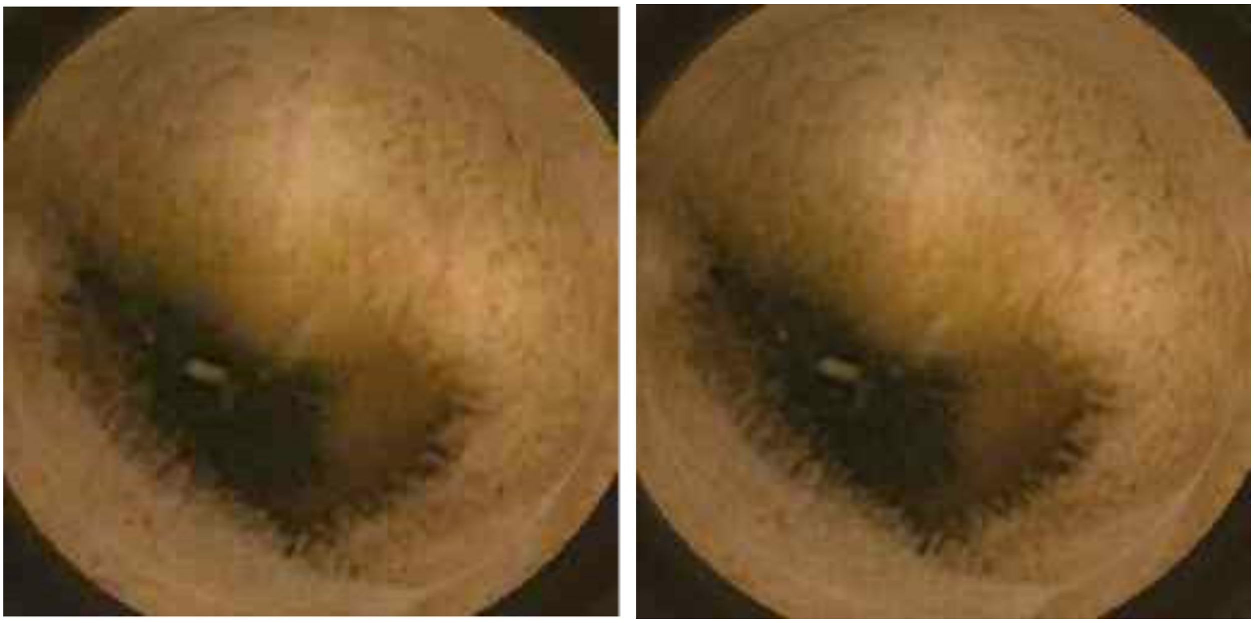


Figure 9 Visual snapshots from the decoded "Capsule Test Video 2" sequence. (left) Motion JPEG (49.74 kbps, 38.56 dB), (right) the proposed hash-based Wyner-Ziv video codec (50.88 kbps, 39.25 dB).

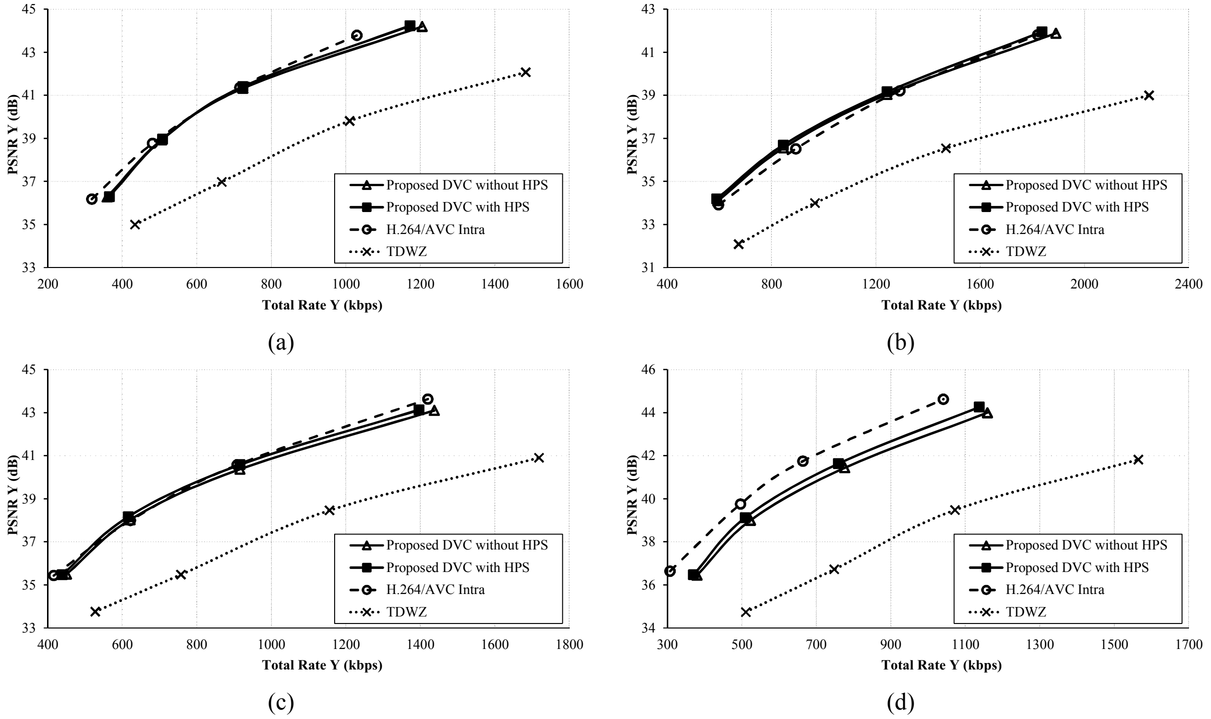
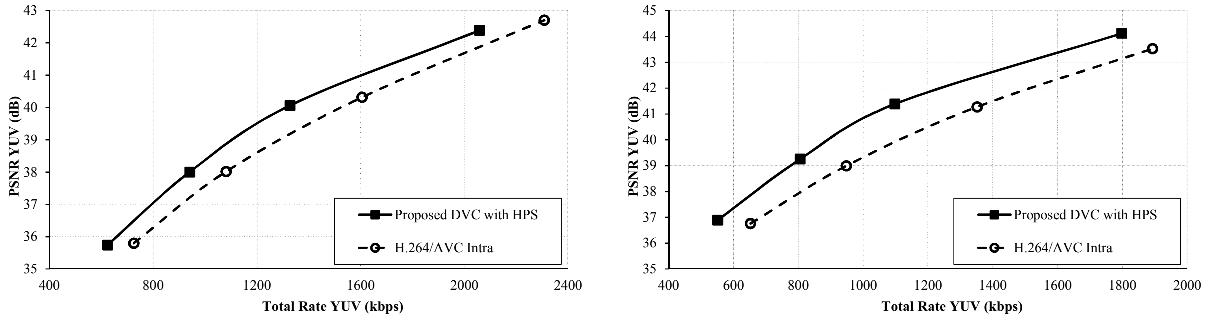


Figure 10 Experimental results obtained on data acquired from conventional endoscopy: (a) "Endoscopy Test Video 1", (b) "Endoscopy Test Video 2", (c) "Endoscopy Test Video 3", and (d) "Endoscopy Test Video 4". The RD performance of the proposed system (with and without HPS) is compared against that of H.264/AVC Intra and that of the TDWZ codec of [13]. Only the *Y* component is coded.



(a)

(b)

Figure 11 Experimental results obtained on data acquired from conventional endoscopy: (a) "Endoscopy Test Video 5" and (b) "Endoscopy Test Video 6". The RD performance of the proposed system (with HPS) is compared against that of H.264/AVC Intra. All three *Y, U*, and *V* components are coded. The average *YUV* PSNR is given by $\text{PSNR}_{YUV} = (4 \times \text{PSNR}_Y + \text{PSNR}_U + \text{PSNR}_V)/6$.

To benchmark the performance of the presented codec (with HPS) against H.264/AVC when all three *Y, U*, and *V* components are coded, both systems are also tested on "Endoscopic Test Video 5" and "Endoscopic Test Video 6". The results, see Figure 11, show that the proposed codec outperforms the competition. Specifically, the proposed codec delivers a significant BD rate reduction of 12.41 and 18.61% in "Endoscopic Test Video 5" and "Endoscopic Test Video 6", respectively.

Visual comparisons between TDWZ [13] and the proposed codec are given in Figures 12 and 13. The proposed codec yields significantly better visual quality and does not suffer from blocking artefacts, typically affecting TDWZ [13] at this rate. The superior visual quality delivered by the proposed system compared to the state-of-the-art TDWZ codec of [13] confirms the potential of the former in medical imaging applications, where high visual quality is a fundamental demand.

5.3. Encoding complexity

Low-cost encoding is a key aspect of distributed video compression. During the evaluation of the DISCOVER [26] codec, it was shown that the Wyner-Ziv frames' encoding complexity is very low compared to the complexity associated with the intra encoding of the key frames. Therefore, the lower the number of key frames, i.e., the longer the GOP, the higher the gain in complexity reduction offered by DVC over H.264/AVC Intra frame coding. Execution time measurements under controlled conditions, as established by the DISCOVER group [26], have shown that our codec (using H.264/AVC Intra to code the hash and the key frames) brings a reduction in average encoding time of approximately 30, 50, and 60% for a GOP size of 2, 4, and 8, respectively, compared to H.264/AVC Intra.

In contrast to hash-less Wyner-Ziv codecs, e.g. [26], our proposed codec has a higher encoding complexity caused by the additional hash formation and coding. However, the hash-related complexity overhead is kept low, since the hash dimensions were reduced to one fourth of the original frame resolution prior to coarse H.264/AVC Intra frame coding. When compared to Motion JPEG, the proposed codec (although currently not optimized for speed) exhibits similar encoding time

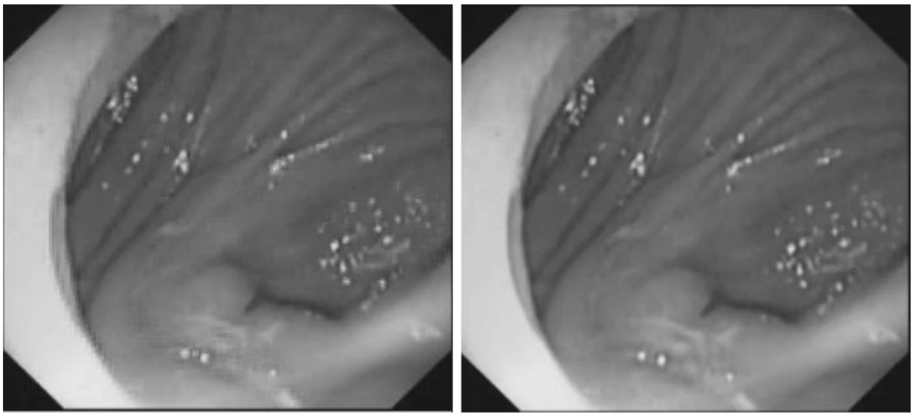


Figure 12 Visual snapshots from the decoded "Endoscopy Test Video 1" sequence. (left) The TDWZ codec in [13] (1206 kbps, 40.3 dB), (right) the proposed DVC (1204.7 kbps, 44.2 dB).

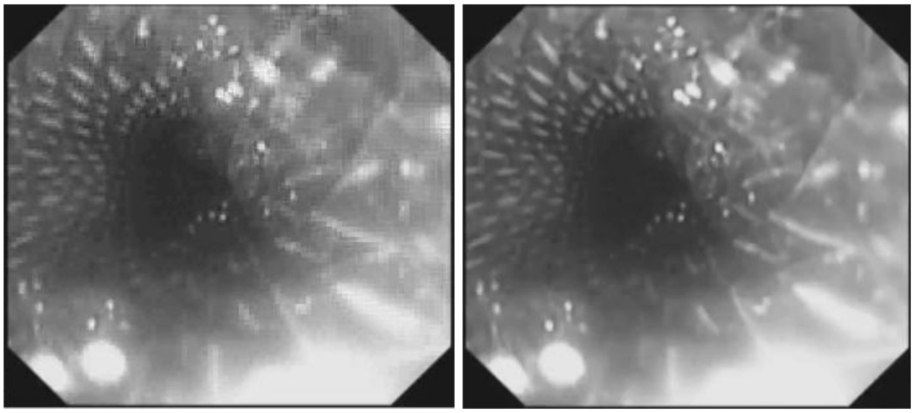


Figure 13 Visual snapshots from the decoded "Endoscopy Test Video 2" sequence. (left) The TDWZ codec in [13] (1281 kbps, 35.43 dB), (right) the proposed DVC (1242 kbps, 39.15 dB).

but offers superior compression performance. We remark that compared to DISCOVER or Motion JPEG, the proposed codec offers a significant reduction of the encoding rate for a given distortion level. Such a notable rate reduction induces an important decrease in power consumption by the transmission part of wireless video recording devices, e.g., wireless capsule endoscopes.

The proposed system links the encoder to the decoder via a feedback channel. Such a reverse channel implies that the encoder is forced to store Wyner-Ziv data in a buffer pending the decoder's directives. Based on our prior work [44], we analyze of the buffer size requirements imposed on the presented system's encoder due to the decoding delay for the capsule endoscopy application scenario. Recall that the GOP size in this scenario is restricted to 2 frames (see Section 5.2). The prime factors determining the decoding delay are the frame acquisition period t_F, the time t_{SI} to generate a side information frame, the transmission time (time-of-flight) t_{TOF} between encoder and decoder, and the LDPC soft-input soft-output decoding time, denoted by t_{SISO}.

For simplicity, the combined intra encoding and decoding time is at most t_F [44]. Given the fact that the presented system applies bidirectional motion estimation, the decoding of a Wyner-Ziv frame can commence only after the next key frame (in a GOP of 2) has been decoded. This induces a structural latency, starting from receiving the Wyner-Ziv frame, of $3 \times t_F$, which corresponds to the acquisition time of two frames, that is, the Wyner-Ziv frame proper and the next key frame, as well as the encoding and decoding of the latter. Adding the time to generate the side information and to perform Wyner-Ziv decoding yields the total time delay. Hence, the total time that the Wyner-Ziv frame bits need to be stored at the encoder, measured from the time the frame is received, is given by $3 \times t_F + t_{\mathrm{SI}} + 2 \times F \times t_{\mathrm{TOF}} + F \times t_{\mathrm{SISO}}$, where F represents the number of feedback requests, soliciting additional syndrome bits and inducing another LDPC decoding attempt. Since in a GOP of 2, the encoder receives a Wyner-Ziv frame every $2 \times t_F$, the size of the encoder's buffer, expressed in number of frames L, is given by

$$L = \text{ceiling}\left[\frac{3 \times t_F + t_{\mathrm{SI}} + 2 \times F \times t_{\mathrm{TOF}} + F \times t_{\mathrm{SISO}}}{2 \times t_F}\right] \quad (8)$$

Continuing our analysis, the reported capsule and the conventional endoscopic sequences were recorded using a camera with an acquisition rate of 2 and 30 Hz, respectively, corresponding to an acquisition period t_F of 500 and 33.33 ms.

An estimation of the transmission time t_{TOF} through the body can be made by calculating the velocity of a uniform plane in a lossy medium [45], characterized by its dielectric properties, i.e. the conductivity and permittivity. These values can be calculated based on [46,47] for a wide range of body tissues and frequencies. It can be verified that at a frequency of 433 MHz the velocity is always greater than 10% of the speed of light through all body tissue cases included in [47], leading to a time-of-flight t_{TOF} in the order of 15 ns through 0.5 m of tissue.

It is clear that the time t_{SI} to generate a side information frame is dominated by OBME. Fortunately, several VLSI designs for hardware implementation of block motion estimation have been proposed. Considering the state-of-the-art architecture of [48], full integer-pel motion search can be executed at $4\rho^2 + B$-1 cycles per macroblock (MB), where ρ and B are the search range and MB size, respectively. However, our presented scheme employs bidirectional OBME. Specifically, the total number of overlapping blocks per frame is $(H{\cdot}V)/$

ε^2, where H and V are the horizontal and vertical frame dimensions and ε is the overlap size. Hence, based on the VLSI architecture in [48], the total number of cycles per frame is given by $2\cdot[(4\cdot\rho^2 + B\text{-}1)\cdot H\cdot V]/\varepsilon^2$, where the factor 2 stems from the bidirectionality. Considering a simplified decoding device with a single core CPU running at 800 MHz with a 1DIMPS/MHz/core[h] and instantiating the OBME parameters for $\rho = 16$, $\varepsilon = 4$, and $B = 16$, yields a delay of 10.63 and 24.9 ms per frame for the capsule endoscopic ($H = V = 256$) and the endoscopic ($H = 480$, $V = 320$) sequences, respectively.

LDPC decoding of a Wyner-Ziv frame is performed in t_{SISO}. If decoding is unsuccessful, a request through the feedback channel is issued. Supposing that the encoder responds immediately, the decoder receives the next chunk of syndrome bits in $2 \times t_{\text{TOF}}$ after which a new LDPC decoding attempt is made. Table 2 tabulates the average number of feedback channel requests per Wyner-Ziv frame for the four considered RD points for the capsule endoscopic sequences. Taking recent advances in LDPC decoder implementations into account, decoding latencies[i] in the order of 6 μs have been reported for a codeword length of 2304 bits (802.16e standard) and 25 iterations or of 270 μs for a codeword length of 64800 bits (DVB-S2 standard) and 50 iterations [49]. For parallel architectures a latency of 280 ns per iteration can be reached [50] for a rate 1/2 code of 2304 bit codewords. Our system employs a LDPC codeword of 6336 bits and 50 decoding iterations. Considering a worst-case scenario, namely decoding codewords as big as in the DVB-S2 standard and with $F = 100$ feedback requests (see Table 2), the total LDPC decoding latency $F \times t_{\text{SISO}}$ would be roughly 27 ms.

Based on the above approximations, Equation (8) yields an estimated buffer size of $L = 2$ and $L = 3$ frames for the capsule ($t_F = 500$ ms) and the conventional endoscopic ($t_F = 33.33$ ms) sequences, respectively, thus confirming the applicability of the proposed scheme. However, the encoder buffer size can be further restrained. An elegant solution is to constrain the number of feedback requests to a fixed number of requests for an entire Wyner-Ziv frame as proposed in our previous study [44], where we show that the loss of compression efficiency compared to unconstrained feedback is less than 5% when at most $F = 5$ requests per Wyner-Ziv frame are allowed. In addition to this, the structural latency induced by bidirectional temporal prediction could be reduced by employing unidirectional prediction.

Table 2 Average feedback channel requests per Wyner-Ziv frame for the capsule endoscopic video sequences

	RD point 1	RD point 2	RD point 3	RD point 4
Capsule Test Video 1	35.4	41.2	59.3	86.6
Capsule Test Video 2	33.7	39.7	59.6	85.4
Capsule Test Video 3	34.1	39.4	55.8	80.9
Capsule Test Video 4	32.7	40.5	59.0	88.3

6. Conclusions

Motivated by the strict prerequisites of wireless lightweight multimedia applications, such as wireless capsule endoscopy, this article has introduced a novel video codec based on the principles of Wyner-Ziv coding. The proposed codec maintains low encoding complexity and facilitates quality and temporal scalability. Intrinsically, the proposed codec achieves high compression performance by embracing a novel hash-driven motion estimation technique, which generates accurate side information at the decoder. The presented technique performs motion-compensated multi-hypothesis prediction, enabling adaptation to the regional variations in temporal correlation in a frame. Concrete experimentation using various conventional and endoscopic test video sequences has confirmed the superior compression performance of the proposed codec against several state-of-the-art traditional and Wyner-Ziv video codecs. In effect, in conventional and endoscopic test video material significant Bjøntegaard rate savings of up to 32.13 and 43.37% over the state-of-the-art have been obtained.

Endnotes

[a] This paper has been presented in part at the IEEE International Conference on Image Processing, Brussels, Belgium, September 2011 [51]. [b] Motion JPEG is based on the JPEG coding standard [52] and includes a file format that can handle multiple JPEG images. Unlike the Motion JPEG 2000 standard [53], no standard specification has been defined for Motion JPEG, and hence only proprietary solutions are available (e.g., support in Microsoft AVI files, Apple Quicktime format or the RFC 2435 spec that describes how Motion JPEG can be supported by an RTP stream). [c] To simplify the presentation, the LDPC code is assumed systematic. [d] The experimental results of DISCOVER [26] have been obtained using the executable of the DISCOVER codec which is available on the projects website [26]. [e] Given an RD point and a number of iterations, the process starts from a specific hash QP value (QP_hash = QP_key+1), and calculates the total and the hash rate, and the resulting RSD of the decoded frames (both key and WZ frames). If the RSD is lower that a strict threshold, the QP and the rate values are stored;

otherwise they are discarded. Next, the hash QP is increased and the algorithm continuous till it reaches a given number of iterations. Out of the retained QPs, the one which minimizes the total rate is chosen as the best for the specific rate point. In case of equal total rates the highest QP value is selected. [f]These volunteers presented no evidence of gastrointestinal pathologies. [g]These sequences were transformed to the YUV 4:2:0 format supported by the proposed codec. [h]Nowadays more powerful processors exist to be deployed in devices at the size of the decoder of a capsule endoscope. For instance, the Apple A5 processor of iPhone 4S has an ARM Cortex-A9 MPCore 32-bit multicore processor at 800 MHz, 2.5DIMPS/MHz/core. [i]All these figures correspond to Soft-Input Soft Output (SISO) decoders.

Abbreviations

AVC: advanced video coding; CPU: central processing unit; CRC: cyclic redundancy check; DCT: discrete cosine transform; DISCOVER: distributed coding for video services; DJSCC: distributed joint source-channel coding; DVC: distributed video coding; DSC: distributed source coding; GOP: group of pictures; HPS: hash-predictor-selection; JPEG: joint photographic experts group; LDPCA: low-density parity-check accumulate; LLR: log likelihood ratio; MCI: motion-compensated interpolation; OBME: overlapped block motion estimation; PRISM: power-efficient robust high-compression syndrome-based multimedia coding; PSNR: peak signal-to-noise ratio; RSD: relative standard deviation; QCIF: quarter common intermediate format; QM: quantization matrix; QP: quality parameter; RD: rate-distortion; SAD: sum of the absolute differences; TDWZ: transform-domain Wyner-Ziv.

Acknowledgements

This study was supported by the FWO Flanders projects G.0391.07, G.0146.10 and the postdoctoral fellowship of Peter Schelkens. The authors would like to thank Prof. Dr. Daniel Urbain, head of the Gastroenterology clinic of the Universitair Ziekenhuis Brussel for providing numerous endoscopic video sequences.

Author details

[1]
Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium [2]
Interdisciplinary Institute for Broadband Technology, Gaston Crommenlaan 8, B-9050 Ghent, Belgium [3]
Department of Electronics, Alexander Technological Educational Institute of Thessaloniki, P.O. Box 141, GR-57400 Thessaloniki, Greece [4]
ELIS Department, Multimedia Lab, Ghent University, Gaston Crommenlaan 8, B-9050 Ghent, Belgium

Competing interests

Parts of the research presented in this article have been filled by IBBT under patent applications EP07120604.9 (T Clerckx, A Munteanu, Motion estimation and compensation process and device. November 2007), and PCT/EP2011/071296 (N Deligiannis, A Munteanu, J Barbarien, Method and device for correlation channel estimation, November 2011).

Received: 10 October 2011 Accepted: 14 March 2012
Published: 14 March 2012

References

1. T Wiegand, GJ Sullivan, G Bjøntegaard, A Luthra, Overview of the H.264/AVC video coding standard. IEEE Trans Circ Syst Video Technol. **13**(7), 560–576 (2003)
2. D Slepian, JK Wolf, Noiseless coding of correlated information sources. IEEE Trans Inf Theory. **19**(4), 471–480 (1973). doi:10.1109/TIT.1973.1055037
3. AD Wyner, J Ziv, The rate-distortion function for source coding with side information at the decoder. IEEE Trans Inf Theory. **22**(1), 1–10 (1976). doi:10.1109/TIT.1976.1055508
4. R Puri, K Ramchandran, PRISM: a new robust video coding architecture based on distributed compression principles, in *40th Allerton Conference on Communication, Control, and Computing*, vol. 40. Allerton, IL, pp. 586–595 (October 2002)
5. B Girod, A Aaron, S Rane, D Rebollo-Monedero, Distributed video coding. Proc IEEE. **93**(1), 71–83 (2005)
6. GR Esmaili, PC Cosman, Wyner-Ziv video coding with classified correlation noise estimation and key frame coding mode selection. IEEE Trans Image Process. **20**(9), 2463–2474 (2011)
7. N Deligiannis, J Barbarien, M Jacobs, A Munteanu, A Skodras, P Schelkens, Side-information-dependent correlation channel estimation in hash-based distributed video coding. IEEE Trans Image Process (2012, in press)
8. V Toto-Zarasoa, A Roumy, C Guillemot, Non-uniform source modeling for distributed video coding, European Signal Processing Conference, Aalborg, Denmark, pp. 1889–1893 (August 2010)
9. F Pereira, L Torres, C Guillemot, T Ebrahimi, R Leonardi, S Klomp, Distributed video coding: selecting the most promising application scenarios. Signal Process Image Commun. **23**(5), 339–352 (2008). doi:10.1016/j.image.2008.04.002
10. L Stankovic, V Stankovic, S Cheng, Distributed compression: overview of current and emerging multimedia applications, in *IEEE International Conference on Image Processing*, Brussels, Belgium, pp. 1841–1844 (September 2011)
11. R Sidhu, D Sanders, M McAlindon, Gastrointestinal capsule endoscopy: from tertiary centres to primary care. BMJ. **332**, 528–531 (2006). doi:10.1136/bmj.332.7540.528
12. J Ascenso, C Brites, F Pereira, A flexible side information generation framework for distributed video coding. Multimedia Tools Appl. **48**(3), 381–409 (2010). doi:10.1007/s11042-009-0316-6
13. N Deligiannis, M Jacobs, J Barbarien, F Verbist, J Škorupa, R Van de Walle, A Skodras, P Schelkens, A Munteanu, Joint DC coefficient band decoding and motion estimation in Wyner-Ziv video coding, in *International Conference on Digital Signal Processing*, Corfu, Greece, pp. 1–6 (July 2011)
14. CE Shannon, Coding theorems for a discrete source with a fidelity criterion. IRE National Convention Record 142–163 (1959)
15. A Wyner, Recent results in the Shannon theory. IEEE Trans Inf Theory. **20**(1), 2–10 (1974). doi:10.1109/TIT.1974.1055171
16. S Pradhan, K Ramchandran, Distributed source coding using syndromes (DISCUS): design and construction. IEEE Trans Inf Theory. **49**, 626–643 (2003). doi:10.1109/TIT.2002.808103
17. P Mitran, J Bajscy, Coding for the Wyner-Ziv problem with turbo-like codes, in *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, p. 91 (June 2002)
18. A Liveris, Z Xiong, C Georghiades, Compression of binary sources with side information at the decoder using LDPC codes. IEEE Commun Lett. **6**(10), 440–442 (2002)
19. R Zamir, S Shamai, U Erez, Nested linear/lattice codes for structured multiterminal binning. IEEE Trans Inf Theory. **48**(6), 1250–1276 (2002). doi:10.1109/TIT.2002.1003821
20. Z Liu, S Cheng, A Liveris, Z Xiong, Slepian-Wolf coded nested quantization (SWC-NQ) for Wyner-Ziv coding: performance analysis and code design, in *IEEE Data Compression Conference*, Snowbird, UT, pp. 322–331 (March 2004)
21. J Chou, S Pradhan, K Ramchandran, Turbo and trellis-based costructions for source coding with side information, in *IEEE Data Compression Conference*, Snowbird, UT, pp. 33–42 (March 2003)
22. Y Yang, S Cheng, Z Xiong, W Zhao, Wyner-Ziv coding based on TCQ and LDPC codes. IEEE Trans Commun. **57**(2), 376–387 (2009)
23. Q Xu, V Stankovic, Z Xiong, Layered Wyner-Ziv video coding for transmission over unreliable channels. Signal Process Elsevier. **86**, 3212–3225 (2006)
24. A Aaron, S Rane, D Rebollo-Monedero, B Girod, Systematic lossy forward error protection for video waveforms, in *International Conference on Image Processing*, vol. I. Barcelona, Spain, pp. 609–612 (September 2003)
25. S Cheng, Z Xiong, Successive refinement for the Wyner-Ziv problem and layered code design. IEEE Trans Signal Process. **53**(8), 3269–3281 (2005)
26. X Artigas, J Ascenso, M Dalai, S Klomp, D Kubasov, M Quaret, The DISCOVER codec: architecture, techniques and evaluation. *Picture Coding Symposium* (Lisboa, Portugal, 2007), ([http://www.discoverdvc.org], 2007)

27. J Ascenso, C Brites, F Pereira, Content adaptive Wyner-Ziv video coding driven by motion activity, in *IEEE International Conference on Image Processing*, Atlanta, GA, pp. 605–608 (October 2006)
28. A Aaron, S Rane, B Girod, Wyner-Ziv video coding with hash-based motion compensation at the receiver, in *IEEE International Conference on Image Processing*, vol. 5. Singapore, pp. 3097–3100 (October 2004)
29. TN Dinh, G Lee, J-Y Chang, H-J Cho, Side information generation using extra information in distributed video coding, in *IEEE International Symposium on Signal Processing and Information Technology*, Cairo, Egypt, pp. 138–143 (December 2007)
30. N Deligiannis, A Munteanu, T Clerckx, J Cornelis, P Schelkens, Overlapped block motion estimation and probabilistic compensation with application in distributed video coding. IEEE Signal Process Lett. **16**(9), 743–746 (2009)
31. F Verbist, N Deligiannis, M Jacobs, J Barbarien, P Schelkens, A Munteanu, A statistical approach to create side information in distributed video coding, in *ACM/IEEE International Conference on Distributed Smart Cameras* (August 2011)
32. N Deligiannis, M Jacobs, F Verbist, J Slowack, J Barbarien, R Van de Walle, P Schelkens, A Munteanu, Efficient hash-driven Wyner-ZIv video coding for visual sensors. *ACM/IEEE International Conference on Distributed Smart Cameras* (Ghent, Belgium, 2011)
33. C McCaffrey, O Chevalerias, C O'Mathuna, K Twomey, Swallowable-capsule technology. IEEE Pervasive Comput. **7**(1), 23–29 (2008)
34. D Turgis, R Puers, Image compression in video radio transmission for capsule endoscopy. Sensors Actuat A: Physical. **123-124**, 129–136 (2005)
35. CE Duchon, Lanczos filtering in one and two dimensions. J Appl Meteorol. **18**(8), 1016–1022 (1979). doi:10.1175/1520-0450(1979)0182.0.CO;2
36. D Varodayan, A Aaron, B Girod, Rate-adaptive codes for distributed source coding. Signal Process Elsevier. **86**, 3123–3130 (2006)
37. J Abouei, JD Brown, KN Plataniotis, S Pasupathy, Energy efficiency and reliability in wireless biomedical implant systems. IEEE Trans Inf Technol Biomed. **15**(3), 456–466 (2011)
38. PD Bradley, An ultra low power, high performance Medical Implant Communication System (MICS) transceiver for implantable devices, in *IEEE Biomedical Circuits and Systems Conference*, London, UK), pp. 158–161 (August 2008)
39. AV Oppenheim, RW Schafer, JR Buck, *Discrete-Time Signal Processing*, 2nd edn. (Prentice Hall, Upper Saddle River, 1999)
40. K Turkowski, S Gabriel, Filters for common resampling tasks, in *Graphics Gems I*, ed. by Glassner AS (Academic Press, San Diego, 1990)
41. S Chung, T Richardson, R Urbanke, Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation. IEEE Trans Inf Theory. **47**(2), 657–670 (2001). doi:10.1109/18.910580
42. G Bjøntegaard, Calculation of average PSNR differences between RD-curves. (2001) Technical Report, VCEG, April 2001, Contribution VCEG-M33
43. H Yu, Z Lin, F Pan, Applications and improvement of H.264 in medical video compression. IEEE Trans Circ Syst I. **52**(12), 2707–2716 (2005)
44. Namely: J Slowack, J Skorupa, N Deligiannis, P Lambert, A Munteanu, R Van de Walle, Distributed video coding with feedback channel constraints. in IEEE Trans Circ Syst Video Technol (2012)
45. CA Balanis, *Advanced Engineering Electromagnetics*, (John Wiley and Sons, Inc., New York, 1989)
46. C Gabriel, Compilation of the dielectric properties of body tissues at RF and microwave frequencies, (Occupational and environmental health directorate, Radiofrequency Radiation Division, Brooks Air Force Base, TX, 1996)
47. http://niremf.ifac.cnr.it/tissprop/htmlclie/htmlclie.htm
48. C-Y Chen, S-Y Chien, Y-W Huang, T-C Chen, T-C Wang, L-G Chen, Analysis and architecture design of variable block-size motion estimation for H.264/AVC. IEEE Trans Circ Syst I. **53**(2), 578–592 (2006)
49. T Brack, M Alles, T Lehnigk-Emden, F Kienle, N When, NE L'Insalata, F Rossi, M Rovini, L Fanucci, Low complexity LDPC code decoders for next generation standards, in *Design, Automation, and Test in Europe*, Nice, France, pp. 1–6 (April 2007)
50. Y Sun, JR Cavallaro, T Ly, Scalable and low power LDPC decoder design using high level algorithmic synthesis, in *IEEE International System-on-Chip Conference*, Belfast, Northern Ireland, UK, pp. 267–270 (September 2009)
51. N Deligiannis, F Verbist, J Barbarien, J Slowack, R Van de Walle, P Schelkens, A Munteanu, Distributed coding of endoscopic video, in *IEEE International Conference on Image Processing*, Brussels, Belgium, pp. 1853–1856 (September 2011)
52. WB Pennebaker, JL Mitchell, *JPEG Still Image Data Compression Standard*, (Van Nostrand Reinhold, New York, 1993)
53. P Schelkens, A Skodras, T Ebrahimi, *The JPEG 2000 Suite*, (Wiley, Chichester, 2009)

doi:10.1186/1687-1499-2012-106
Cite this article as: Deligiannis *et al.*: **Wyner-Ziv video coding for wireless lightweight multimedia applications.** *EURASIP Journal on Wireless Communications
and Networking* 2012 **2012**:106.