A Low-Complexity Joint Source-Channel Video Coding for 3-D DWT Codec

Evgeny Belyaev[1], Karen Egiazarian[1], Moncef Gabbouj[1], and Kai Liu[2]
[1] Tampere University of Technology, Tampere 33720, Finland
[2] Xidian University, 710071 Xian, China
Email: {evgeny.belyaev,karen.egiazarian,moncef.gabbouj}@tut.fi; kailiu@mail.xidian.edu.cn

Abstract—In this paper we propose a low-complexity error-resilient and joint source-channel video coding algorithms for video codec based on three-dimensional discrete wavelet transform (3-D DWT). First, we show that the video stream, generated by 3-D DWT codec with the proposed error-resilient, is significantly more robust to packet losses than video stream generated by scalable extension of the H.264/AVC standard (H.264/SVC). Second, in comparison with H.264/SVC the proposed joint source-channel video coding algorithm demonstrates better performance for high packet loss rates. Finally, the computational complexity of 3-D DWT encoder with error-resilient and joint source-channel video coding is 3-6 times less than H.264/SVC encoder. Therefore, it can be more preferable for video transmission over highly unreliable channels and (or) in systems in which the video encoding computational complexity or power consumption play a critical role.

Index Terms—unequal loss protection, 3-D DWT, H.264/SVC

I. Introduction

Video coding and transmission are core technologies used in numerous applications such as streaming, surveillance, broadcasting and so on. Taking into account packet losses and time-varying bandwidth, the scalable video coding (SVC) is the most preferable compression method for video transmission over packet networks. During the last decade, in a number of papers it was demonstrated that in a combination with unequal loss protection (ULP) of different video stream layers, SVC provides robust transmission. Taking into account that bit errors are usually corrected at the lower network levels, only packet losses (or erasures) can happen at the application layer. Therefore, ULP can be easily implemented at the application layer using inter-packet Reed-Solomon (RS) codes without any modification of other network layers. In this case, the base video stream layer is protected using RS codes with a high redundancy level while the remaining layers are protected with a lower redundancy level or not protected at all.

Manuscript received September 9, 2013; revised November 21, 2013.
This work was supported by Tampere Doctoral Program in Information Science and Engineering (TISE), Finland and the project of National Natural Science Foundation of China (NSFC) for International Young Scientists.
Corresponding author email: evgeny.belyaev@tut.fi.
doi:10.12720/jcm.8.12.893-901

The most widespread ULP techniques, analyzed in the literature, are based on a scalable extension of the H.264/AVC standard (see, for example, [1], [2]) which includes temporal, spatial and quality scalability. H.264/SVC provides high compression efficiency due to motion compensation and inter-layer prediction. However, combined with the end-to-end distortion estimation caused by packet losses, erasure-correction coding and ULP optimization, the full video transmission system based on H.264/SVC has a high computational complexity. Finally, without ULP the video bit stream generated by H.264/SVC is very sensitive to packet losses and even 1% packet loss rate is enough for significant visual quality degradation when reconstructed video becomes not authentic (new objects, which are not present in original video, can appear).

As an alternative to H.264/SVC, scalable video coding based on three-dimensional discrete wavelet transform (3-D DWT) can be used [3]–[8]. Due to intensive investigation of Motion Compensated Temporal Filtering, currently 3-D DWT schemes have a comparable rate-distortion performance with the H.264/SVC. However, as in the previous case, combined with ULP these schemes also require high computational resources. For example, in [9] the encoding speed of 6 frames per second (fps) at CIF/SIF resolution was achieved on a Pentium M 2-GHz processor, which is not enough for real-time transmission (up to the authors knowledge there is no later work in the literature which shows complexity performance of joint compression and loss protection). But it is well known that there are a lot of applications such as video transmission for space missions [10], vehicle-to-infrastructure video delivery [11], underwater wireless video transmission [12], video transmission for wireless sensor networks [13] and so on, where the channel can be very unreliable, all video transmission parts such as video capturing, video encoding, decoding, packet loss protection and playback should operate in real-time, and transmitter computational resources (or power consumption) is very restricted. Therefore, a development of efficient low-complexity scalable video coders with ULP and rate control, which is a particular case of joint source-channel coding problem, is an important practical problem.

In this paper we extend our previous results [14] which were related to low-complexity video compression (without considering any transmission problems, like

packet loss protection, packetization and so on) and introduce a novel low-complexity error-resilient and joint source-channel video coding algorithms for video codec based on three-dimensional discrete wavelet transform. First, we show that the video stream, generated by 3-D DWT codec with the proposed error-resilient, is significantly more robust to packet losses than video stream generated by scalable extension of the H.264/AVC standard (H.264/SVC). Second, in comparison with H.264/SVC the proposed joint source-channel video coding algorithm demonstrates better performance for high packet loss rates. Finally, the computational complexity of 3-D DWT encoder with error-resilient and joint source-channel video coding is 3–6 times less than H.264/SVC encoder.

The rest of the paper is organized as follows. Section II introduces the low-complexity joint source-channel video coding algorithm based on 3-D DWT: Section II-A shows the general scheme of the 3-D DWT codec, Section II-B describes the packetization and packet loss protection based on inter-packet Reed-Solomon codes, Section II-C formulates the optimization task for joint source-channel video coding for considered case and describes how this task can be solved by high complex Lagrange relaxation method, Sections II-D and II-E propose how the complexity can be significantly reduced. Section III presents error concealment and error-resilient coding for the 3-D DWT coder as well as comparisons of the proposed codec with scalable extension of H.264/AVC standard. Finally, conclusions are drawn in Section IV.

II. Proposed Algorithm

A. General Video Coding Scheme Description

The considered video coding scheme is shown in Fig. 1. First, a group of frames (GOF) of length N are accumulated in the input frame buffer. Then, a one-dimensional multilevel DWT of length N in the temporal direction is applied. All frames in GOF are processed starting from low-frequency to high-frequency frames. For each frame, the spatial subbands are also processed from low-frequency to high-frequency spatial subbands. Depending on the channel rate and motion level in a current GOF, the rate controller chooses one of the following options for each spatial subband:

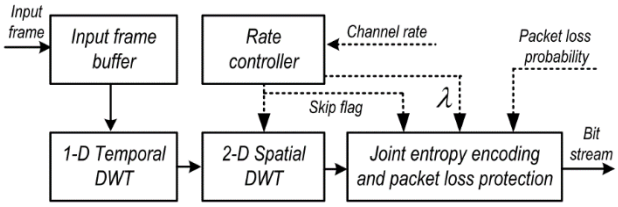

Fig. 1. The considered video coding scheme.

TABLE I. Packet Loss Protection Modes

Protection mode	1	2	3	…	11	12
Protection level	Four-fold repetition	RS (6,2)	RS (7,3)	…	RS (15,11)	No protection

- It permits calculation of 2-D spatial transform, selects the Lagrange multiplier λ and then permits minimization of a Lagrangian sum by selection of the ULP mode and the lowest bit-plane number (truncation point) which is Included into a bit stream;
- It prohibits the calculation of the 2-D spatial DWT, the entropy coding and ULP mode selection for the subband. At the decoder side, the corresponding transform coefficients in the subband are considered to be zero.

B. Packetization and Packet Loss Protection

The proposed entropy encoding and ULP scheme is shown in Fig. 2. Each wavelet subband is independently compressed using bit-plane entropy coding with a simplified JPEG2000 context modeling. As it is proposed in [14], the entropy encoder consists of two cores: Levenstein zero-run coder for low-entropy binary contexts and adaptive binary range coder for the remaining binary contexts. Each core has its own output buffer. Comparing to the traditional bit-plane encoders it allows to decrease the complexity of the wavelet encoder up to 2 times [14], without any significant rate-distortion performance degradation.

Compression is starting from the highest bit-plane t_{max} and, when a subband is truncated at the bit-plane t^*, each buffer contains $t_{max} - t^* + 1$ subpackets (each bit-plane fits into two subpackets). ULP is achieved due to the selection of different protection levels for each subband. The list of used packet loss protection levels is illustrated in Table I.

Levels 2-11 are based on inter-packet $RS(n,k)$code in finite field $GF(2^8)$ with the generator polynomial

$$g(x) = x^4 + 30x^3 + 216x^2 + 231x + 166.$$

In this approach k source bytes with the same index are used to form of the source polynomial $m(x)$ and corresponding r parity bytes are generated as:

$$r(x) = -x^r \cdot m(x) \; mod \; g(x).$$

If the source byte with the current index does not exist (stuffing byte), we use zero byte instead of it. Finally,

Real-time Transport Protocol (RTP) headers are added to each packet, and transmission is performed.

If k or more packets from n are delivered, then the RS decoder is able to recover all lost source packets. Let p denote the packet loss probability, then the probability that a source packet will be lost and not recovered by the RS decoder is

$$p_{dec}(p,n,k) = \sum_{i=n-k+1}^{n} \frac{i}{n} \cdot \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}. \quad (1)$$

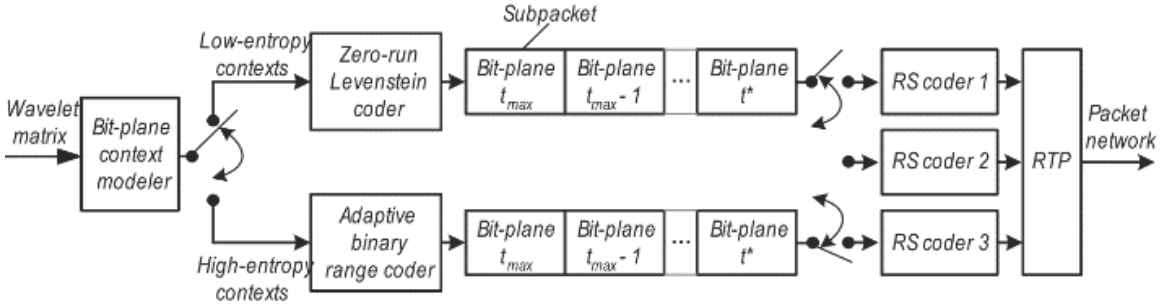

Fig. 2. Proposed joint entropy encoding and unequal packet loss protection scheme.

If the RS code (n_i, k_i) for the protection of two subpackets, containing a bit stream of bit plane i, is used, then the probability that these two subpackets will not be received and not recovered is

$$\pi_i = 1 - \left(1 - p_{dec}(p, n_i, k_i)\right)^2, \quad (2)$$

when these subpackets are placed into the different source packets and

$$\pi_i = p_{dec}(p, n_i, k_i), \quad (3)$$

when the subpackets are placed into the same source packet. It is easy to see, that in the second case, the probability that the subpackets are lost is always less than or equal to that of the first case. Therefore, we are using the second type of packetization.

Let $\psi = \{t^*, n^*, k^*\}$ denote a decision vector for a subband, which is truncated at a bit-plane $t^*$and for each bit-plane RS code (n^*, k^*) is used. Denote by d_i the subband distortion when the bit-planes $t_{max}, .., i$ are received and the bit-plane $i-1$ is not received. Then the expected distortion after truncation at bit-plane t^* is

$$E[d(\psi)] = \sum_{i=t^*+1}^{t_{max}} \prod_{j=i}^{t_{max}} (1-\pi_i) \cdot \pi_{i-1} \cdot d_i + \\ + \pi_{t_{max}} \cdot d_{skip} + \prod_{i=t^*}^{t_{max}} (1-\pi_i) d_{t^*}, \quad (4)$$

where d_{skip} is the distortion, when the subband is not received.

C. Joint Source-Channel Coding by Lagrangian-Relaxation

Let $\mathbf{\Psi} = \{\psi_i\}$ be the set of decision vectors, where ψ_i is the decision vector for the i'th subband in a GOF. The overall expected GOF distortion is

$$E[D(\mathbf{\Psi})] = \sum_i E[d(\psi_i)], \quad (5)$$

and the resulting GOF bit stream size is

$$R(\mathbf{\Psi}) = \sum_i r(\psi_i), \quad (6)$$

where $r(\psi_i)$ is the bit stream size of source and parity packets for the i'th subband when the decision vector ψ_i is applied.

Fig. 3 illustrates the expected visual quality for different loss protection redundancies $\frac{n-k}{n}$ when source and parity packets bit stream size $R(\mathbf{\Psi})$ is equivalently to the channel rate 3000 kbps and packet loss probability is 20%. One can see, that if the redundancy is zero (no any protection is used), then the expected visual quality has a minimum value. With increase of the redundancy the distortion caused by packet loss protection decreases and the expected visual quality increases until some maximum value. A further increase of the redundancy does not lead to increase of the quality, because in order to keep the source and parity packets bit rate equal to the channel rate, it is needed to compress the wavelet subbands with higher and higher compression ratio. In this case the distortion caused by compression increases and distortion caused by losses is not significant, because the video bit stream is overprotected. Therefore, the visual quality decreases.

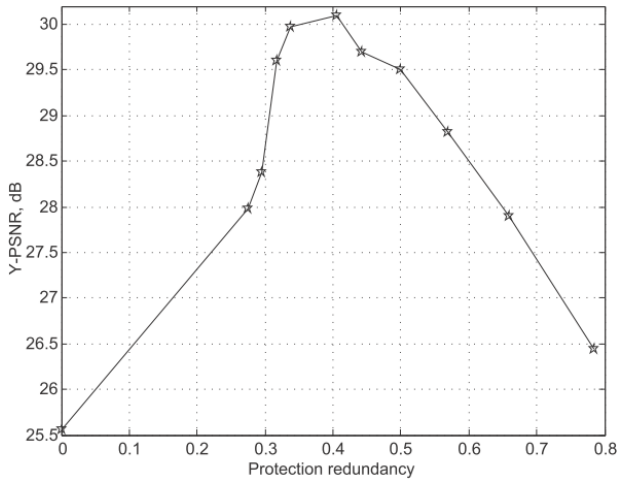

Fig. 3. Expected visual quality in case of protection by Reed-Solomon codes for packet loss probability 20% and channel rate 3000 kbps for "Football" (640x480).

In order to find the minimum overall expected distortion for a given video source, channel rate and packet loss probability the following joint source-channel video coding optimization task can be formulated. For each GOF we should determine the set of decision vectors $\mathbf{\Psi}^*$, such that

$$\begin{cases} \mathbf{\Psi}^* = \arg\min\limits_{\{\mathbf{\Psi}\}} E[D(\mathbf{\Psi})] \\ R(\mathbf{\Psi}^*) \le R_{max}, \end{cases} \quad (7)$$

where $R_{max} = \frac{N \cdot C}{f}$ is the bit budget for each GOF, C is the channel rate, and f is the video source frame rate. The optimization task can be solved using the Lagrangian relaxation method [15]. To use it the following statements should be proved.

Statement 1. *For each* $\lambda \ge 0$, *the set of decision vectors* $\mathbf{\Psi}_\lambda^* \in \{\mathbf{\Psi}\}$, *which minimizes*

$$E[D(\mathbf{\Psi})] + \lambda \cdot R(\mathbf{\Psi}), \quad (8)$$

is optimal solution of the optimization task, if $R_{max} = R(\mathbf{\Psi}_\lambda^*)$.

Proof. Let us assume that the statement is not hold, and there exist the set of decision vectors $\mathbf{\Psi} \in \{\mathbf{\Psi}\}$, so that $E[D(\mathbf{\Psi})] < E[D(\mathbf{\Psi}_\lambda^*)]$ and $R(\mathbf{\Psi}) \le R(\mathbf{\Psi}_\lambda^*)$. Then inequality

$$E[D(\mathbf{\Psi})] + \lambda \cdot R(\mathbf{\Psi}) < E\left[D\left(\mathbf{\Psi}_\lambda^*\right)\right] + \lambda \cdot R\left(\mathbf{\Psi}_\lambda^*\right)$$

means that the set of decision vectors $\mathbf{\Psi}_\lambda^*$ is not minimizes (8) that contradicts with the statement formulation.

From Statement 1 it follows that in order to solve (7) one should find λ, such that $R(\Psi_\lambda^*) = R_{max}$.

Statement 2. *Let us assume that for some* λ_1 *and* λ_2, *values* $\mathbf{\Psi}_{\lambda_1}^*$ *and* $\mathbf{\Psi}_{\lambda_2}^*$, *which minimizes (8) are found.Then, if* $R\left(\mathbf{\Psi}_{\lambda_1}^*\right) > R(\mathbf{\Psi}_{\lambda_2}^*)$, *then the following inequality holds:*

$$\lambda_2 \ge -\frac{E\left[D\left(\mathbf{\Psi}_{\lambda_1}^*\right)\right] - E\left[D\left(\mathbf{\Psi}_{\lambda_2}^*\right)\right]}{R\left(\mathbf{\Psi}_{\lambda_1}^*\right) - R\left(\mathbf{\Psi}_{\lambda_2}^*\right)} \ge \lambda_1. \quad (9)$$

Proof. From Statement 1 it follows that

$$E\left[D\left(\mathbf{\Psi}_{\lambda_1}^*\right)\right] + \lambda_1 \cdot R\left(\mathbf{\Psi}_{\lambda_1}^*\right) < E\left[D\left(\mathbf{\Psi}_{\lambda_2}^*\right)\right] + \lambda_1 \cdot R\left(\mathbf{\Psi}_{\lambda_2}^*\right). \quad (10)$$

From (10) and $R\left(\mathbf{\Psi}_{\lambda_1}^*\right) > R\left(\mathbf{\Psi}_{\lambda_2}^*\right)$ follows, that

$$-\frac{E\left[D\left(\mathbf{\Psi}_{\lambda_1}^*\right)\right] - E\left[D\left(\mathbf{\Psi}_{\lambda_2}^*\right)\right]}{R\left(\mathbf{\Psi}_{\lambda_1}^*\right) - R\left(\mathbf{\Psi}_{\lambda_2}^*\right)} \ge \lambda_1, \quad (11)$$

what proves the right side of (9). Analogously, from Statement 1 it follows that

$$E\left[D\left(\mathbf{\Psi}_{\lambda_2}^*\right)\right] + \lambda_2 \cdot R\left(\mathbf{\Psi}_{\lambda_2}^*\right) < E\left[D\left(\mathbf{\Psi}_{\lambda_1}^*\right)\right] + \lambda_2 \cdot R\left(\mathbf{\Psi}_{\lambda_1}^*\right). \quad (12)$$

From (12) and $R\left(\mathbf{\Psi}_{\lambda_1}^*\right) > R\left(\mathbf{\Psi}_{\lambda_2}^*\right)$ follows, that

$$\lambda_2 \ge -\frac{E\left[D\left(\mathbf{\Psi}_{\lambda_1}^*\right)\right] - E\left[D\left(\mathbf{\Psi}_{\lambda_2}^*\right)\right]}{R\left(\mathbf{\Psi}_{\lambda_1}^*\right) - R\left(\mathbf{\Psi}_{\lambda_2}^*\right)}. \quad (13)$$

From Statement 2 it follows, that function $R(\mathbf{\Psi}_\lambda^*)$ is a non-increasing function of λ. It means that needed λ value can be found by the bisection method [15].

For the considered codec, the rate-distortion function for the subband i is independent of the truncation and RS code parameters selected for other subbands. Therefore,

$$\begin{aligned} &\min_{\mathbf{\Psi}}\{E[D(\mathbf{\Psi})] + \lambda \cdot R(\mathbf{\Psi})\} = \\ &= \min_{\mathbf{\Psi}}\left\{\sum_i E[d(\psi_i)] + \lambda \cdot r(\psi_i)\right\} = \\ &= \sum_i \min_{\psi_i}(E[d(\psi_i)] + r(\psi_i)). \end{aligned} \quad (14)$$

Therefore, a solution of the task (7) can be written as $\mathbf{\Psi}^* = \{\psi_i^*\}$, where for each subband i the value $\psi_i^* = \{t_i^*, n_i^*, k_i^*\}$ can be found independently as

$$\psi_i^* = \arg\min_{\{\psi_i\}}\{E[d(\psi_i)] + \lambda \cdot r(\psi_i)\}. \quad (15)$$

D. Skipping of Spatial Subbands for Complexity Reduction

Algorithm 1 Wavelet subband truncation and RS code parameters selection

Input: *Subband*$\{x[i,j]\}, \lambda$

$t_{max} \leftarrow \left\lfloor \log_2(\max\limits_{(i,j)} x[i,j]) \right\rfloor$
$E \leftarrow \sum\limits_{(i,j)} x^2[i,j]$
$\Omega^* \leftarrow E$
for $t = t_{max},..,0$ **do**
  Encode bit-plane t as in [14]
  $r \leftarrow$ *get_subband_bit_stream_size*()
  $(n_t, k_t) \leftarrow \arg\min\limits_{(n,k)} E[d(t,n,k)] + \lambda \cdot r \cdot \frac{n}{k}$
  $\Omega(t, n_t, k_t) = E[d(t, n_t, k_t)] + \lambda \cdot r \cdot \frac{n_t}{k_t}$
  if $\Omega(t, n_t, k_t) \ge \Omega^*$ **then**
    if $\Omega^* = E$ **then**
      Skip subband
    else
      Truncate subband at bit-plane t^*
      Use Reed-Solomon (k^*, n^*) *code*
    end if
    exit
  else
    $\Omega^* \leftarrow \Omega$
    $t^* = t, n^* = n_t, k^* = k_t$
  end if
end for

In this paper we propose to use the following Algorithm 1 for ψ_i^* selection. First, all bit planes are compressed from the highest to the lowest bit plane as it is proposed in our previous work [14]. Then, after encoding of each bit plane with number t we calculate the bit stream size r for bit planes $t_{max},..,t$ of the subband. After that the Reed-Solomon (k^*, n^*) code from Table I which minimizes the Lagrange sum (15) is choosen taking into account that the expected distortion is calculated as in (4) and the full bit stream size includes the compressed subband bit stream size of r bits as well as the bit size of parity packets and calculated as $r \cdot \frac{n}{k}$. If the current Lagrange sum for bit-plane t is less than for bit plane $t-1$, then the bit plane $t+1$ is encoded. Otherwise, the encoding process is stopped, and the

subband is truncated at the current bit plane and protected by Reed-Solomon (k^*, n^*) code. If, it is more efficient to not transmit even the highest bit plane t_{max}, then the subband is skipped and at the decoder side all wavelet coefficients of the subband are considered to be zero. One can see, that Algorithm 1 corresponds to (15) with the assumption that $E[d(\psi_i) + \lambda \cdot r(\psi_i)$ is a convex function.

From Algorithm 1, it follows that if

$$\Omega(t_{max}, n_{t_{max}}, k_{t_{max}}) > \sum_{(i,j)} x^2[i,j], \tag{16}$$

then this subband is skipped. In this situation the encoder wastes computational resources for calculating the 2-D DWT and for forming the subband bit stream which is not included in the output bit stream. Therefore, it is important to find a criterion which can guarantee with ahigh probability that the current subband will be skippedbefore processing of this subband.

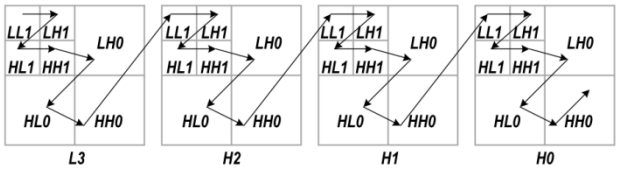

Fig. 4. Example of spatial subbands encoding order.

In our previous work [14] the authors have proposed the following parent-child tree based subband skip criterion, which can be used also for complexity reduction of joint compression and protection scheme considered in this work. It is well known that the value of the wavelet coefficient in the child subband is correlated with the corresponding coefficient in the parent subband. This property is widely used in image and video compression algorithms based on inter-subband correlation (see, e.g. [16]). Using the same reasoning, we first process subbands from low-frequency to high-frequency temporal subbands and from low-frequency to high-frequency spatial subbands (see Fig. 4). Second, we make the following *coding assumption*: if for any subband, inequality (16) holds, then for all temporal-spatial child-subbands, the same inequality also holds. In this case, all child-subbands will not be processed and hence spatial transform calculation, entropy coding and ULP parameters selection are skipped.

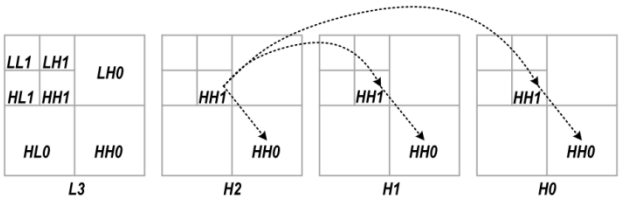

Fig. 5. Spatial subbands skipping.

Fig. 5 illustrates the proposed approach. Frames after temporal wavelet decomposition are denoted as L3, H2, H1, H0, where L3 is the low-frequency frame. If the spatial subband HH1 in frame H2 is skipped then the corresponding child-subbands HH0 in frame H2, HH1 and HH0 in frame H1, and HH1 and HH0 in frame H0 are skipped without any processing.

E. Proposed Joint Source-Channel Coding Algorithm

To find the necessary λ value with the bisection method, one needs to calculate the 2-D DWT and losslessly compress each subband in order to determine all sets of truncation points, which requires significant computational resources. In this paper we propose a one-pass joint source-channel video coding algorithm which uses the *virtual buffer* concept to estimate λ without lossless coding. Let us define b_{virt} to be the number of bits in avirtual buffer and B_{max} the virtual buffer size determined as $B_{max} = L \cdot C$, where L is the required bufferinglatency at the decoder side for continuous video playback, C is the channel rate. The proposed algorithm is presented below.

At step 4, Algorithm 2 selects the Lagrange multiplier λ value in the proportional to the virtual buffer fullness, where γ defines the proportion degree. It allows increasing or decreasing the compression ratio for the current GOF depending on the buffer fullness. At step 5, all frames in the GOF are processed starting from low-frequency to high-frequency frames. For each frame, the spatial subbands are also processed from low-frequency to high-frequency spatial subbands. At step 8, each subband is compressed starting from the highest bit-plane, until the Lagrangian sum decreases. At step 9, if the highest bit-plane is not included into the output bit stream, this subband is skipped together with all of its child-subbands. At steps 10 and 12 the virtual buffer fullness is updated.

Algorithm 2 Proposed joint source-channel coding algorithm

1: $b_{virt} \leftarrow b_{virt}(0)$
2: **for** $j = 0,..,j_{max} - 1$ **do**
3: Calculate 1-D temporal DWT for GOF j.
4: $\lambda \leftarrow \lambda_{max} \left(\frac{b_{virt}}{B_{max}} \right)^{\gamma}$
5: **for** $i = 0,..,i_{max} - 1$ **do**
6: **if** parent subband was skipped **then** skip subband i;
7: Calculate 2-D spatial DWT for subband i;
8: Find $\psi_i^* = \arg\min_{\{\psi_i\}} \{E[d(\psi_i)] + \lambda \cdot r(\psi_i)\}$
9: **if** $r(\psi_i^*) = 0$ **then** skip subband i;
10: $b_{virt} \leftarrow b_{virt} + r(\psi_i^*)$
11: **end for**
12: $b_{virt} \leftarrow b_{virt} - R_{max}$
13: **end for**

Note that if during the encoding process of a video sequence the number of bits in the virtual buffer requires

$$0 \le b_{virt} \le B_{max}, \tag{17}$$

then the average bit stream size for each GOF $\bar{R}$ is bounded by

$$R_{max} \le \bar{R} \le R_{max} + \frac{B_{max}}{j_{max}}, \tag{18}$$

where j_{max} is the number of GOF's in a video sequence. From (18) it follows that if (17) holds during the

encoding process, then the average bit stream size for each GOF $\bar{R}$ tends to R_{max} with an increasing j_{max}.

To further explain Algorithm 2, let us suppose that the video encoder compresses the series of identical GOF's. Let us define λ^* as the Lagrange multiplier value computed using the bisection method described in Section II-C and consider the virtual buffer fullness when the initial buffer fullness is

$$b_{virt}(0) = b^*_{virt}(0) = B_{max}\left(\frac{\lambda^*}{\lambda_{max}}\right)^{\frac{1}{\gamma}}, \quad (19)$$

Taking into account that for the bisection method $R(\Psi^*_\lambda) = R_{max}$ (see Statement 1), after processing each GOF, the buffer fullness will be constant and equal to

$$b_{virt} = b^*_{virt}(0).$$

Now let us analyze the buffer fullness in the case when the initial buffer fullness $b_{virt}(0) \neq b^*_{virt}(0)$. Let us assume that after compression of the previous GOF the current buffer fullness is $b_{virt} < b^*_{virt}(0)$. It means that λ value calculated by line 4 of Algorithm 2 for the current GOF will be less than the optimal λ^*. Taking into account that $R(\mathbf{\Psi}^*_\lambda)$ is a non-increasing function of λ (see Statement 2), after compression of this GOF the difference $R(\mathbf{\Psi}^*_\lambda) - R_{max} > 0$. Therefore, a new buffer fullness calculated by line 10 will be higher than previous one. It means that after compression of the next GOF, the current value of λ will be closer to the optimal λ^* and so on, until the current value becomes equal or exceeds λ^*.

Analogically, if the current $b_{virt} > b^*_{virt}(0)$ then after compression of the next GOF the current value of λ will also be closer to the optimal λ^* and so on, until the current λ value becomes equal or smaller than λ^*.

Taking into account the reasoning described above, the value of λ calculated by the proposed algorithm will oscillate around the optimal value λ^* founded by bisection method.

III. Simulation Results

The proposed codec run with a GOF size $N = 16$ using the Haar transform in the temporal direction and the 5/3 spatial wavelet transform at three-levels of the decomposition. At the decoder side we use the following error concealment. First, bit stream of each subband is decoded in a progressive way until a loss is detected. In this case, any losses in stream corresponds to the higher quantization at the encoder side and does not lead to error propagation. Second, if main low-frequency subband, which contains brightness information for all GOF, is lost we copy corresponding subband from the previous GOF. In this case an error propagation can happens. To minimize probability of this event we use the following simple error-resilient coding. At the encoder side for the main subband we use repetition of the highest bit-planes which can be placed into the two additional source packets.

In case of H.264/SVC we have used JSVM 9.8 reference software [17] with two spatial and five temporal scalable layers. GOP size and intra-frame period are set to 16. For error-resilient coding we have used flexible macroblock ordering with two slice groups and loss-aware rate-distortion optimized macroblock mode decision. Frame copy error concealment method is used at the decoder side. For ULP we use Reed-Solomon codes from Table I, so different level of protection can be achieved for each spatial and temporal layer.

For 3-D DWT codec, joint source-channel video coding algorithm was implemented in real-time mode as it was described above, while ULP optimization for H.264/SVC was realized in off-line mode.

Simulation results were obtained for the video sequence "Football" (640 ×480, 360 frames, 30 Hz) [18]. For both codecs the packet length was set to 800 bytes and channel with independent packet losses was simulated. For visual quality measurement we use the following approach. During each measurement with an index i the sum of square errors D_i between luma component of original and reconstructed video is calculated. Then we repeat this experiment $K = 20$ times (7200 frames) and estimate the Expected Peak Signal-to-Noise Ratio ($E[Y - PSNR]$) which we use as the visual quality metric:

$$E[Y - PSNR] = 10\log_{10}\frac{255^2}{\frac{1}{K}\sum_{i=0}^{K-1} D_i}. \quad (20)$$

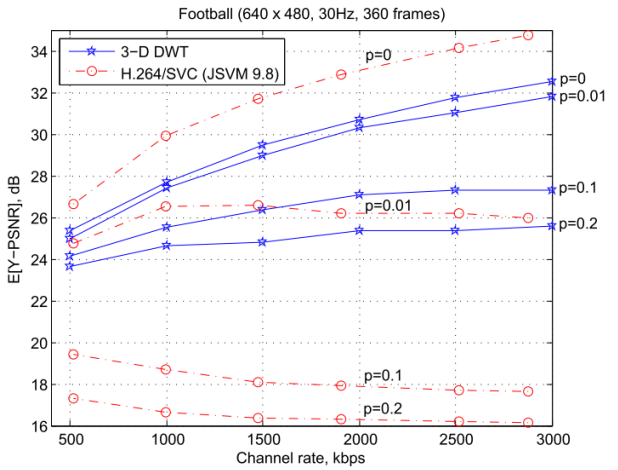

Fig. 6. Expected visual quality comparison for video sequence "Football" with no loss protection.

Fig. 6 and Fig. 7 show the expected PSNR for different packet loss probabilities and channel rates for the proposed codec and for H.264/SVC with and without loss protection (see detailed comparison at http://www.cs.tut.fi/~belyaev/ULP.htm). One can see that the proposed codec is significantly less sensitive to packet losses. Even if packet loss ratio is 5% it provides much better visual quality (see Fig. 8) while H.264/SVC has frames with unrecognizable objects. In case of ULP, H.264/SVC provides better performance (up to 1.5dB) for low packet loss probabilities, but with increasing of it the proposed codec provides the same or better performance (up to 1.5dB). It can be explained by the following reason.

Since H.264/SVC is more sensitive to packet losses it is needed to add more redundancy by Reed-Solomon coding than 3-D DWT codec, which has less compression performance, but requires less redundancy for protection (see Table II).

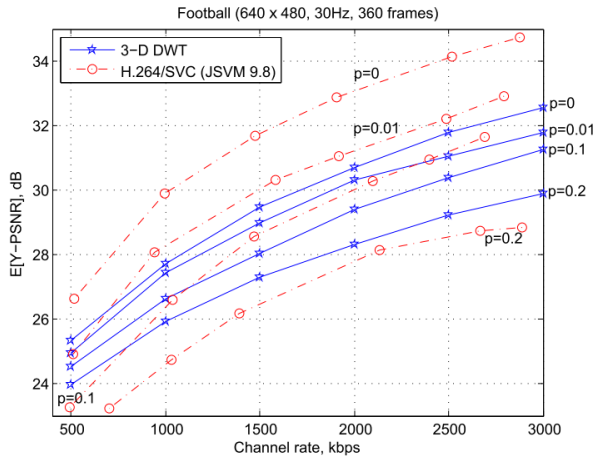

Fig. 7. Expected visual quality comparison for video sequence "Football" with unequal packet loss protection.

For computation complexity measurements we use the encoding speed which is defined as the number of frames which can be encoded in one second by given CPU. The complexity is considered as the inverse value of the encoding speed, which was measured without any use of assemblers, threads, or other program optimization techniques. Table III shows encoding speed for the proposed codec and H.264/SVC depending on the video bit rate. Taking into account that JSVM 9.8 reference software is not optimized for real-time operation we have also used fast implementation of the H.264/AVC standard in single-layer mode (x.264 codec in ultrafast profile [19]) to estimate possible encoding speed of H.264/SVC after optimization. Our estimation shows that the complexity of the proposed codec is 3–6 times less than H.264/SVC at least.

TABLE II: PROTECTION REDUNDANCY

Packet loss probability, p	0	0.01	0.1	0.2
H.264/SVC (JSVM 9.8)	0	0.276	0.464	0.617
3-D DWT	0	0	0.271	0.412

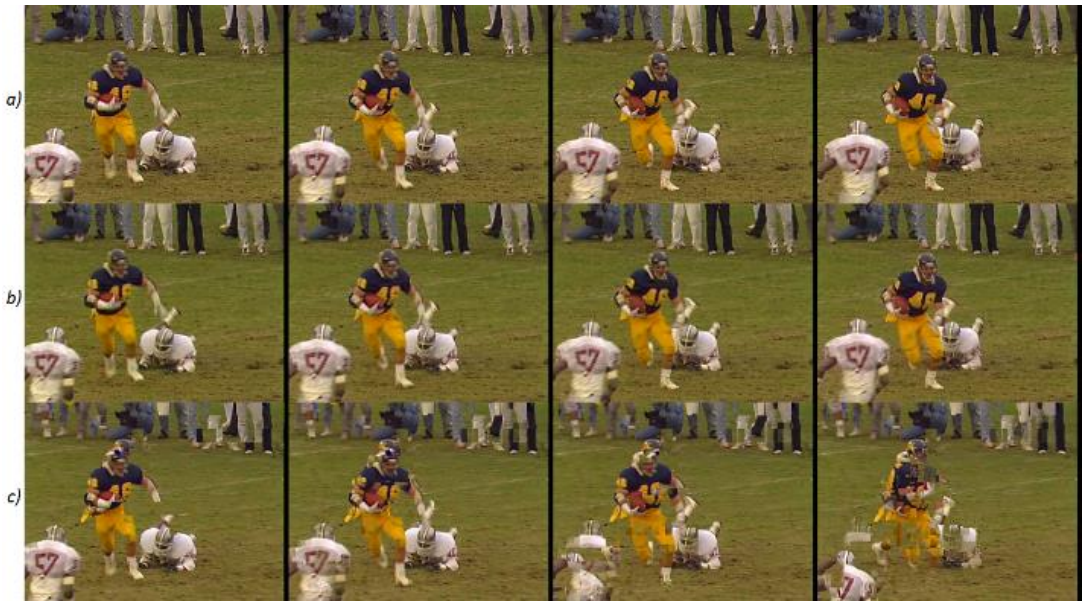

Fig. 8. Visual comparison for 5% packet loss. a) Original frames b) Proposed 3-D DWT c) H.264/SVC (JSVM 9.8)

TABLE III: ENCODING SPEED FOR CPU INTEL CORE 2.1 GHZ, FPS

Video bit rate, kbps	500	1000	2000	2500	3000
H.264/SVC (JSVM 9.8)	<1	<1	<1	<1	<1
x.264	49	44	40	37	36
3-D DWT	279	215	150	133	120
3-D DWT + ULP (p = 0.1)	288	185	160	137	130
3-D DWT + ULP (p = 0.2)	315	198	177	160	145

Additionally, we have added the full complexity for the proposed codec with ULP for different packet loss probabilities. One can see that the complexity of the proposed codec decreases with increasing of the packet loss probability or with decreasing of the bit rate. This can be explained as follows. An increase of a packet loss probability leads to an increase of the portion of parity packets in the output bit stream. In order to keep the source and parity data bit rate equal to the channel rate, the video compression ratio is increased, and condition $r(\psi_i^*) = 0$ in Algorithm 2 holds for more and more subbands which are skipped together with the corresponding child-subbands. As a result, 2-D spatial DWT, entropy coding and ULP mode selection are not performed for a significant portion of spatial subbands.

IV. CONCLUSIONS

A novel low-complexity error-resilient and joint source-channel video coding algorithms for video codec based on three-dimensional discrete wavelet transform were presented. We have shown that in comparison with

H.264/SVC, the proposed codec has the following advantages:

- It is much less sensitive to packet losses and provides robust video transmission and authentic reconstruction for 1–5% packet loss rates without any packet loss protection. In case of H.264/SVC even 1% packet loss rate is enough for significant visual quality degradation. Reconstructed video of H.264/SVC can be not authentic, i.e. new objects, which are not present in original video, can appear.
- The proposed joint source-channel video coding algorithm demonstrates better performance for high packet loss rates.
- The computational complexity of 3-D DWT encoder with error-resilient and joint source-channel video coding is 3–6 times less than H.264/SVC encoder. High complexity of H.264/SVC is caused by high complexity of motion estimation, backward loop at the encoder side, input frame down sampling, and end-to-end distortion estimation caused by packet losses. It is difficult to use of the H.264/SVC on mobile devices, personal computers and other systems, which compress one or more high resolution video sources in real-time. Therefore, H.264/SVC is a solution for systems based on pre-encoded video.

Therefore, the 3-D DWT codec with proposed error-resilient and joint source-channel video coding algorithms can be more preferable for video transmission over highly unreliable channels and (or) in systems in which the video encoding computational complexity or power consumption play a critical role.

REFERENCES

[1] E. Maani and A. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 20, no. 3, pp.407–416, 2010.

[2] Y. Huo, M. El-Hajjar, and L. Hanzo, "Inter-layer fec aided unequal error protection for multilayer video transmission in mobile TV," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 20, no. 3, pp. 1622–1634, 2013.

[3] M. van der Schaar and D. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Transaction on Multimedia*, vol. 9, no. 1, pp. 185–197, 2007.

[4] D. Taubman and A. Zakhor, "Multi-rate 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 572–588, 1994.

[5] J. Tham, S. Ranganath, and A. Kassim, "Highly scalable wavelet-based video codec for very low bit-rate environment," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 1, pp. 12–27, 1998.

[6] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. Van der Schaar, J. Cornelis, and P. Schelkens, "In-band motion compensated temporal filtering," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 653–673, 2004.

[7] D. Zhang, W. Zhang, J. Xu, F. Wu, and H. Xiong, "A cross-resolution leaky prediction scheme for in-band wavelet video coding with spatial scalability," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 516–521, 2008.

[8] N. Mehrseresht and D. Taubman, "A flexible structure for fully scalable motion-compensated 3-D DWT with emphasis on the impact of spatial scalability," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 740–753, 2006.

[9] J. Fowler, M. Tagliasacchi, and B. Pesquet-Popescu, "Video coding with wavelet-domain conditional replenishmentand unequal error protection," in *Proc. IEEE International Conference on Image Processing*, Atlanta, USA, 2006, pp. 1869-1872.

[10] R. Viola, L. Chiariglione, and R. Russo, "Video compression for manned space missions," in *Proc. IEEE Global Telecommunications Conference*, Dallas, USA, 1989.

[11] A. Vinel, E. Belyaev, O. Lamotte, M. Gabbouj, Y. Koucheryavy, and K. Egiazarian, "Video transmission over IEEE 802.11p: Real-world measurements," presented at the IEEE ICC-2013(Workshop on Emerging Vehicular Networks), Budapest, Hungary, June 9-13, 2013.

[12] J. Ribas, D. Sura, and M. Stojanovic, "Underwater wireless video transmission for supervisory control and inspection using acoustic OFDM," in *Proc. OCEANS*, Seattle, USA, 2010, pp. 1-9.

[13] I. Politis, M. Tsagkaropoulos, and S. Kotsopoulos, "Optimizing video transmission over wireless multimedia sensor networks," in *Proc. IEEE Global Telecommunications Conference*, New Orleans, USA, 2008, pp. 117-122.

[14] E. Belyaev, K. Egiazarian, and M. Gabbouj, "A low-complexity bit-plane entropy coding and rate control for 3-D DWT based video coding," *IEEE Transactions on Multimedia*, 2013.

[15] G. Schuster and A. Katsaggelos, "Rate-distortion based video compression, optimal video frame compression, and object boundary encoding," *Kluwer Academic Publisher*, Norwell, MA, USA, 1997.

[16] B. Kim, Z. Xiong, and W. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374–1378, 2000.

[17] JSVM 9.8 software package. [Online]. Available: http://iphome.hhi.de/

[18] Xiph.org test media. [Online]. Available: http://media.xiph.org/video/derf/

[19] x.264 Video Codec. [Online]. Available: http://x264.nl/

Evgeny Belyaev received the Engineer degree and the Ph.D. (candidate of science) degree in technical sciences from State University of Aerospace Instrumentation (SUAI), Saint-Petersburg, Russia, in 2005 and 2009, respectively. He is currently a Researcher with the Institute of Signal Processing, Tampere University of Technology, Finland.

His research interests include real-time video compression and transmission, video source rate control, scalable video coding, motion estimation and arithmetic encoding.

Karen Egiazarian received the M.Sc. degree in mathematics from Yerevan State University in 1981, the Ph.D. degree in physics and mathematics from Moscow State University, Moscow, Russia, in 1986, and the D.Tech. degree from the Tampere University of Technology (TUT), Tampere, Finland, in 1994.

He has been Senior Researcher with the Department of Digital Signal Processing, Institute of Information Problems and Automation, National Academy of Sciences of Armenia. Since 1996, he has been an Assistant Professor with the Institute of Signal Processing, TUT, and from 1999 a Professor, leading the Computational Imaging Group.

His research interests are in the areas of applied mathematics, image and video processing.

Moncef Gabbouj received the B.S. degree in electrical engineering from Oklahoma State University, Stillwater, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1986 and 1989, respectively.

He is currently an Academy of Finland Professor with the Department of Signal Processing, Tampere University of Technology, Tampere, Finland. He was the Head of the department from 2002 to 2007. He was on sabbatical leave at the American University of Sharjah, Sharjah, United Arab Emirates, from 2007 to 2008, and a Senior Research Fellow with the Academy of Finland, Helsinki, Finland, from 1997 to 1998 and in 2008. He is the Co-Founder and Past CEO with SuviSoftOy, Ltd., Tampere. From 1995 to 1998, he was a Professor with the Department of Information Technology, Pori School of Technology and Economics, Pori, Finland, and from 1997 to 1998 he was a Senior Research Scientist with the Academy of Finland. From 1994 to 1995, he was an Associate Professor with the Signal Processing Laboratory, Tampere University of Technology. From 1990 to 1993, he was a Senior Research Scientist with the Research Institute for Information Technology, Tampere. His current research interests include multimedia content-based analysis, indexing and retrieval, nonlinear signal and image processing and analysis, and video processing and coding.

Kai Liu received the B.S. and M.S. degrees in computer science and the Ph.D. degree in signal processing from Xidian University, Xian, China, in 1999, 2002, and 2005, respectively. Currently, he is an Associate Professor of computer science and technology with the Xidian University.

His major research interests include VLSI architecture design and image coding.