Downloaded from orbit.dtu.dk on: Apr 20, 2024

DTU Library

Error Concealment for 3-D DWT Based Video Codec Using Iterative Thresholding

Belyaev, Evgeny; Forchhammer, Søren; Codreanu, Marian

Published in:
I E E E Communications Letters

Link to article, DOI:
10.1109/LCOMM.2017.2697428

Publication date:
2017

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Belyaev, E., Forchhammer, S., & Codreanu, M. (2017). Error Concealment for 3-D DWT Based Video Codec Using Iterative Thresholding. *I E E E Communications Letters*, *21*(8), 1731-1734. https://doi.org/10.1109/LCOMM.2017.2697428

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Error Concealment for 3-D DWT Based Video Codec Using Iterative Thresholding

Evgeny Belyaev, Søren Forchhammer, *Member, IEEE*, and Marian Codreanu, *Member, IEEE*

***Abstract*— Error concealment for video coding based on a 3-D discrete wavelet transform (DWT) is considered. We assume that the video sequence has a sparse representation in a known basis different from the DWT, e.g., in a 2-D discrete cosine transform basis. Then, we formulate the concealment problem as l_1–norm minimization and solve it utilizing an iterative thresholding algorithm. Comparing different thresholding operators, we show that video block-matching and 3-D filtering provide the best reconstruction by utilizing spatial similarity within a frame and temporal similarity between neighbor frames. Experimental results show that the proposed error concealment provides up to 6-dB improvement in the peak signal-to-noise ratio in comparison with existing recovery methods.**

***Index Terms*— Video error concealment, iterative thresholding.**

I. Introduction

THERE exist many applications where video data is compressed and transmitted over unreliable wireless channels, e.g., video transmission for space missions, video delivery from drones, vehicle-to-vehicle video delivery, multimedia wireless sensor networks and so on. In such applications, a video coding algorithm should generate a bit stream, which is robust to packet losses. Video coding standards, such as H.264/AVC [1] and H.265/HEVC [2], can be utilized for the transmission. Due to the motion estimation and compensation they achieve very high compression efficiency at the expense of a high sensitivity of the video stream to packet loss. As an alternative, video coding based on a 3-D discrete wavelet transform (3-D DWT) [3]–[6] encodes each group of wavelet coefficients independently. It provides a good balance between compression efficiency and robustness to packet loss.

In a DWT-based video transmission, an inverse DWT can be performed assuming that all lost coefficients are zero. This approach gives relatively good reconstruction performance, especially if the lowest subbands coefficients are delivered using additional protection, such as duplication of lowest subband coefficients [3]–[5], [7], multiple description coding [8] and/or inter-packet forward error-correction [3]–[5]. If neighbor coefficients are delivered, then a lost coefficient can be interpolated [6], [9] utilizing local correlation of a subband coefficient. This recovery method can improve the reconstruction quality without significant increase of computational complexity, but it cannot be efficiently used when many or even all coefficients of a subband are lost.

The associate editor coordinating the review of this paper and approving it for publication was W. Xiang. *(Corresponding author: Evgeny Belyaev.)*

E. Belyaev and S. Forchhammer are with the Department of Photonics Engineering, Technical University of Denmark, 2800 Lyngby, Denmark (e-mail: fevbel@fotonik.dtu.dk; sofog@fotonik.dtu.dk).

M. Codreanu is with the Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland (e-mail: marian.codreanu@oulu.fi).

In this letter, we consider the error concealment (ERC) problem as a problem of signal recovery from incomplete measurements. The main contributions are the following:

1) We introduce a simplified video loss model where each frame is represented by 2-D DWT and transmitted over an unreliable channel, such that each coefficient is either delivered without any distortion or lost. Then we formulate the concealment problem as l_1–norm minimization and solve it via iterative thresholding [10]. Finally, we extend the simplified model to a realistic scenario of video streaming using the 3-D DWT codec applying 3-D transform, quantization of wavelet coefficients, bit stream packetization and packet loss.
2) We apply the proposed error concealment for 3-D DWT codecs and show that it provides up to 6 dB improvement in Peak Signal-to-Noise Ratio (PSNR) for a packet loss rate in the range from 0 to 40%.

The rest of the letter is organized as follows. In Section II we introduce the simplified video transmission model and recovery via iterative thresholding. In Section III we propose an extension of this model for a realistic scenario. The experimental results are presented in Section IV. Conclusions are drawn in Section V.

II. Simple Loss Model and Recovery

A. Wavelet Coefficients Loss Model

Let us assume the following image transmission model.[1] Consider an image $\mathbf{X} \in \mathbb{R}^{H \times W}$ as a 2-D signal of size $H \times W$ pixels. Each pixel is represented by an integer value in the range $0, \ldots, 255$. At the encoder side the image is represented via a l-level 2-D discrete wavelet orthogonal transform as

$$\mathbf{w} = \mathbf{W}\mathbf{x} = [\mathbf{w}^l_{LL}; \mathbf{w}^l_{LH}; \mathbf{w}^l_{HL}; \mathbf{w}^l_{HH}; \mathbf{w}^{l-1}_{LH}; \mathbf{w}^{l-1}_{HL}; \mathbf{w}^{l-1}_{HH}; \ldots; \times \ldots; \mathbf{w}^1_{LH}; \mathbf{w}^1_{HL}; \mathbf{w}^1_{HH}], \quad (1)$$

where $\mathbf{x} = \mathrm{vec}(\mathbf{X})$, $\mathbf{W}$ is the transform matrix, and subvectors $\mathbf{w}^{i+1}_{LL}$, $\mathbf{w}^{i+1}_{LH}$, $\mathbf{w}^{i+1}_{HL}$ and $\mathbf{w}^{i+1}_{HH}$ are called *wavelet subbands* at level $i+1$, which are calculated as

$$\begin{bmatrix} \mathbf{w}^{i+1}_{LL} \\ \mathbf{w}^{i+1}_{LH} \\ \mathbf{w}^{i+1}_{HL} \\ \mathbf{w}^{i+1}_{HH} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{i}_{LL} \\ \mathbf{W}^{i}_{LH} \\ \mathbf{W}^{i}_{HL} \\ \mathbf{W}^{i}_{HH} \end{bmatrix} \mathbf{w}^{i}_{LL}, \quad (2)$$

where $\mathbf{w}^0_{LL} = \mathbf{x}$, $\mathbf{W}^i_{LL}$ is a matrix corresponding to low-pass filtering and sub-sampling for rows and columns,

[1] We use the following notation. The column vectors and matrices are denoted by boldfaced lowercase and uppercase letters, respectively, e.g., $\mathbf{v}$ and $\mathbf{A}$. The superscript $(.)^T$ denotes the transpose operation, $\mathrm{vec}(\mathbf{A})$ concatenates columns of $\mathbf{A}$ into a vector, $[\mathbf{v}_1; \mathbf{v}_2]$ concatenates the column vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ into a vector, $\leftarrow$ means assignment operation, $\lfloor x \rfloor$ denotes mapping x to the nearest integer value not higher than x.

$\mathbf{W}^i_{HH}$ corresponds to high-pass filtering and sub-sampling for rows and columns, $\mathbf{W}^i_{LH}$ corresponds to low-pass filtering and sub-sampling for rows and high-pass filtering and sub-sampling for columns, $\mathbf{W}^i_{HL}$ corresponds to high-pass filtering and sub-sampling for rows and low-pass filtering and sub-sampling for columns. Then all coefficients are transmitted over an unreliable channel, so that each coefficient is either successfully received or lost. Let us define a coefficient loss matrix $\boldsymbol{C}$ as an identity matrix which does not have rows corresponding to the lost wavelet coefficients. Then the decoder received the following vector of wavelet coefficients

$$\mathbf{y} = \boldsymbol{C}\boldsymbol{W}\mathbf{x}. \tag{3}$$

B. Reconstruction via Iterative Soft Thresholding

The basic reconstruction can be performed assuming that all lost wavelet coefficients are zero. In this case the image is reconstructed as

$$\hat{\mathbf{x}} = \boldsymbol{W}^T\boldsymbol{C}^T\mathbf{y}, \tag{4}$$

Clearly, such a simplistic approach is suboptimal. To outperform it, let us consider the error-concealment problem as signal recovery from incomplete measurements, which is well-known in image super-resolution [12] and in compressive sensing [14]. In compressive sensing linear measurements are acquired for an image as $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{M\times WH}$, $M < WH$, denotes the measurement matrix composed by M rows which are randomly selected by a sensor from some matrix of size $WH \times WH$. Then it is assumed that the signal has a sparse representation in a known basis, i.e. $\mathbf{x} = \boldsymbol{\Psi}^{-1}\boldsymbol{\theta}$, where $\boldsymbol{\Psi}$ is a $WH \times WH$ representation matrix and $\boldsymbol{\theta}$ is the sparse vector of the transform coefficients. Under this assumption, the recovery is formulated as searching for the sparsest vector $\boldsymbol{\theta}$ which satisfies $\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}\boldsymbol{\theta}$. Herewith, the l_1–norm can be used as a sparsity metric. In our model, the measurement matrix is $\boldsymbol{\Phi} = \boldsymbol{C}\boldsymbol{W}$ and M rows are randomly selected from matrix $\boldsymbol{W}$ by a communication channel, not by a sensor, i.e., the sensing is performed by the channel. Building on this similarity, to reconstruct the video sequences we solve the l_1 minimization problem utilizing the *iterative soft thresholding* (IST) algorithm [10]. A frame at iteration k is estimated as:

$$\hat{\mathbf{x}}^k = soft\left(\hat{\mathbf{x}}^{k-1} + \Delta\hat{\mathbf{x}}^k, \sigma_k\right), \tag{5}$$

where $\Delta\hat{\mathbf{x}}^k = \boldsymbol{W}^T\boldsymbol{C}^T\left(\mathbf{y} - \boldsymbol{C}\boldsymbol{W}\hat{\mathbf{x}}^{k-1}\right)$, an initial estimate $\hat{\mathbf{x}}^0$ is zero-vector or a frame provided by other recovery method, and the operator $soft(\mathbf{x}, \sigma)$ includes three main steps:

1) A sparsifying transform with matrix $\boldsymbol{\Psi}$ is applied for a frame as $\boldsymbol{\theta} = \boldsymbol{\Psi}\mathbf{x}$.
2) Soft thresholded transform coefficients $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_i\}$ are calculated as

$$\tilde{\theta}_i = \begin{cases} 0, & \text{if } |\theta_i| < \sigma, \\ \left(1 - \dfrac{\sigma}{|\theta_i|}\right)\theta_i, & \text{otherwise.} \end{cases} \tag{6}$$

3) A soft thresholded frame is calculated as $\tilde{\mathbf{x}} = \boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{\theta}}$.

An optimal selection of σ_k and a number of iterations K for a given sparsifying transform is an open research problem. Experimental results show that increase of K provides better reconstruction performance of IST, but also increases computational complexity. At the same time, threshold value σ_k should be reduced from iteration to iteration. In this letter, we use a quadratic function [12], i.e., $\sigma_k = \sigma_0\left(\frac{K-k+1}{K}\right)^2$, where an initial threshold value σ_0 is experimentally selected to provide the maximum reconstruction quality for a given K.

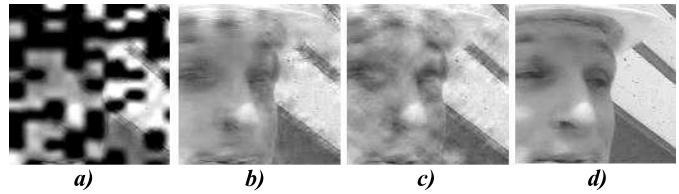

Fig. 1. *a)* Replacement by zeros (PSNR = 6.27 dB), *b)* Linear interpolation [9] (PSNR = 21.39 dB), *c)* IST utilizing $W \times H$ DCT (PSNR = 21.71 dB), *d)* IST utilizing VBM3D (PSNR = 33.20 dB).

Fig. 1 shows an example of reconstruction using the biorthogonal 4.4 wavelet transform with $l = 3$ levels at the encoder to represent 16 frames of a video sequence. The loss of wavelet coefficients is simulated in the following way. Each subband is divided into blocks of 2×2 coefficients. Then, each block is lost with say 40% probability. At the decoder side, we recover the video sequence utilizing a basic reconstruction where all lost coefficients are estimated as zeros (Fig. 1 *a)*), using linear interpolation [9] (Fig. 1 *b)*) and using IST with $\sigma_0 = 200$, $K = 40$ and $H \times W$ discrete cosine transform (DCT) as the sparsifying transform $\boldsymbol{\Psi}$ (Fig. 1 *c)*). One can see that both linear interpolation and IST provide much better visual quality than the basic reconstruction.

C. Thresholding via BM3D and VBM3D

A higher sparsity level of a signal representation achieved by Ψ in the iterative thresholding algorithm means better quality of reconstruction. The $H \times W$ DCT considered above achieves only a moderate sparsity level, because it only takes local similarity between pixels into account. However, it is well-known that many images have non-local self-similarity as well, i.e., there are similar blocks at different locations. In [13], an image denoising algorithm called Block-Matching and 3D filtering (BM3D) exploiting both local and non-local image similarities was proposed. First, BM3D achieves a highly sparse representation of an image using a transform which depends on an image. Then it utilizes hard thresholding and empirical Wiener shrinkage of transform coefficients depending on the noise level. As it is shown in [11] and [12], BM3D can also be used as a thresholding operator $soft(.)$ for other applications, such as image super-resolution and compressive sensing. Finally, we can apply the thresholding operator for a group of frames to achieve an even higher sparsity level utilizing temporal similarity between frames. For this purpose we use an extension of BM3D called Video Block-Matching and 3D filtering (VBM3D) [15] as the thresholding operator. As it is shown in Fig. 1 *d)*, VBM3D provides much higher visual quality than other approaches.

III. Error Concealment for 3-D DWT Video Codec

The loss model from Section II cannot be directly applied to the 3-D DWT codec [3]–[5] due to the following differences:

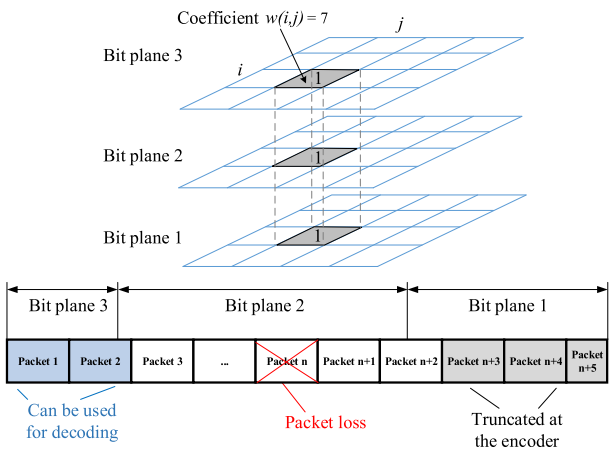

Fig. 2. Bit plane coding, packetization and packet loss.

1) 3-D DWT is used instead of 2-D DWT.
2) Each wavelet subband is quantized and compressed using context-adaptive binary arithmetic coding.
3) The bit stream generated by the arithmetic encoder is represented as a set of packets. Thereafter each packet is transmitted over the network, i.e., there are packet losses instead of coefficient losses.

Let us modify the loss model from Section II to take into account all differences listed above. In the 3-D DWT codec an input video sequence $\mathbf{X}_1, \ldots, \mathbf{X}_F$ is divided into a groups of frames (GOF) of length 2^G, where G is a positive integer number. Then 1-D wavelet transform with G levels of decomposition is applied in the temporal direction for each GOF. Finally, 2-D wavelet transform with l levels of decomposition is applied for each frame, i.e., 3-D DWT is written as

$$[\mathbf{y}_1; \ldots; \mathbf{y}_F] = \mathbf{W}_{3D}[\mathbf{x}_1; \ldots; \mathbf{x}_F], \tag{7}$$

where $\mathbf{W}_{3D}$ is the matrix of 3-D DWT.

Let us consider the quantization and packetization in the 3-D DWT codec. Each wavelet subband is represented as a set of bit planes and is compressed independently of other subbands. The bit at position n in binary representation of a wavelet coefficient belongs to bit plane n. As illustrated in Fig. 2, the binary representation of a wavelet coefficient $w(i, j) = 7$ with coordinates (i, j) is 111, i.e., this coefficient has 1 in the first bit plane, 1 in the second bit plane and 1 in the third bit plane. Each bit plane is processed from the highest (the most significant) bit plane to the lowest bit plane. Binary symbols within each bit plane is processed in a raster scan order and compressed by the arithmetic coder. After compression of the bit plane with index t, the encoder makes a decision to continue encoding bit plane $t-1$ or to stop the encoding, i.e., to truncate the bit stream at bit plane t. In the example, the encoder truncates the bit stream after bit plane 2, i.e., bit plane 1 is not transmitted. If there is no loss in the channel, the coefficient considered will be decoded as 6 (110), i.e., the truncation corresponds to uniform scalar quantization where the quantization step is a power of two.

Let us consider the packet losses. The compressed bit stream of a subband is represented as a set of packets and sent to the channel. Each packet contains the necessary headers (RTP, UDP headers etc.) and payload with video bit stream. Due to resynchronization in the arithmetic decoder, if packet n of a subband is not delivered, then all packets $n+1$, $n+2, \ldots$ belonging to the subband cannot be used for decoding (see Fig. 2). Therefore, the following three events can occur:

1) The first packet of a subband is not delivered. In this case all coefficients of the subband cannot be decoded and are marked as lost. Notice, that linear interpolation is not applicable for this model since all neighbor coefficients are also lost.
2) The encoder truncated a subband at the highest bit plane, i.e., the subband is not transmitted (skipped). In this case all coefficients of the subband are also marked as a lost coefficients.
3) One or more packets of a subband are lost, but the first packet is delivered. In this case we decode the subband until the first loss is detected, and then output the higher fully decoded bit planes. This situation is equivalent to a higher quantization level at the encoder side. In the example in Fig. 2, the original quantization step is 2^1, but because of packet loss, we can fully decode only bit plane 3. It is equivalent to the situation when the encoder uses quantization step 2^2 and loss does not occurr.

Thus, processing of events 1 and 2 for each subband give us coefficient loss matrix C_{3D}, and processing of event 3 gives us equivalent quantization steps for all delivered wavelet coefficients. Therefore, the model (3) is modified as

$$\begin{bmatrix} \mathbf{y}_1 \\ \ldots \\ \mathbf{y}_F \end{bmatrix} = \tilde{\mathbf{Q}}_{3D} \left\lfloor \mathbf{Q}_{3D}\mathbf{C}_{3D}\mathbf{W}_{3D} \begin{bmatrix} \mathbf{x}_1 \\ \ldots \\ \mathbf{x}_F \end{bmatrix} \right\rfloor, \tag{8}$$

where $\mathbf{Q}_{3D}$ and $\tilde{\mathbf{Q}}_{3D}$ are the forward and inverse equivalent quantization matrices. The basic recovery (4) is modified as

$$\left[\hat{\mathbf{x}}_1; \ldots; \hat{\mathbf{x}}_F\right] = \tilde{\mathbf{W}}_{3D}\mathbf{C}_{3D}^T\left[\mathbf{y}_1; \ldots; \mathbf{y}_F\right], \tag{9}$$

where $\tilde{\mathbf{W}}_{3D}$ is a matrix of inverse wavelet transform. Let us define $\mathbf{\Phi}_{3D} = \mathbf{C}_{3D}\mathbf{W}_{3D}$ and $\tilde{\mathbf{\Phi}}_{3D} = \tilde{\mathbf{W}}_{3D}\mathbf{C}_{3D}^T$. Then, utilizing (5), (8) and (9), the error concealment for the 3-D DWT codec is obtained by Algorithm 1.

IV. Experimental Results

Experimental results were obtained for the first 128 frames of the video sequence 'Foreman' with frame resolution 288×352, and frame rate 30 Hz. In the 3-D DWT codec [3]–[5], the input video sequence was divided into groups of 16 frames. Thereafter 1-D Haar wavelet transform with 4 levels of decomposition was applied in the temporal direction for each group, and 5/3 LeGall 2-D wavelet transform with 3 levels of decomposition was applied for each frame. Finally, bit plane coding and packetization were applied for each wavelet subband as it is described in Section III. At the encoder side, packets corresponding to the most important low-frequency subbands $\mathbf{w}_{LL}^3$ were duplicated similarly to [7]. The video bit rate was 820 kbps including duplicated packets.

Packet loss simulation was done in the following way. The video codec generated the video stream as a set of packets with length not higher than 800 bytes. Then, we randomly

Algorithm 1 Error Concealment for 3-D DWT Codec

Input: $\mathbf{y}_1, \ldots, \mathbf{y}_M, K, \sigma_0$

1: **for all** $k = 1, \ldots, K$ **do**

2: $$\begin{bmatrix} \Delta\mathbf{y}_1^k \\ \ldots \\ \Delta\mathbf{y}_F^k \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{y}_1 \\ \ldots \\ \mathbf{y}_F \end{bmatrix} - \tilde{\mathbf{Q}}_{3D} \left[\mathbf{Q}_{3D} \mathbf{\Phi}_{3D} \begin{bmatrix} \hat{\mathbf{x}}_1^{k-1} \\ \ldots \\ \hat{\mathbf{x}}_F^{k-1} \end{bmatrix} \right]$$

3: $$\begin{bmatrix} \Delta\mathbf{x}_1^k \\ \ldots \\ \Delta\mathbf{x}_F^k \end{bmatrix} \leftarrow \tilde{\mathbf{\Phi}}_{3D} \begin{bmatrix} \Delta\mathbf{y}_1^{k-1} \\ \ldots \\ \Delta\mathbf{y}_F^{k-1} \end{bmatrix}$$

4: $$\begin{bmatrix} \hat{\mathbf{x}}_1^k \\ \ldots \\ \hat{\mathbf{x}}_F^k \end{bmatrix} \leftarrow \begin{bmatrix} \hat{\mathbf{x}}_1^{k-1} \\ \ldots \\ \hat{\mathbf{x}}_F^{k-1} \end{bmatrix} + \begin{bmatrix} \Delta\mathbf{x}_1^k \\ \ldots \\ \Delta\mathbf{x}_F^k \end{bmatrix}$$

5: **if** $k \neq K$ **then**

6: $\sigma_k \leftarrow \sigma_0 \left(\frac{K-k+1}{K}\right)^2$

7: $[\hat{\mathbf{x}}_1^k, \ldots, \hat{\mathbf{x}}_F^k] \leftarrow \mathrm{VBM3D}([\hat{\mathbf{x}}_1^k, \ldots, \hat{\mathbf{x}}_F^k], \sigma_k)$

8: **end if**

9: **end for**

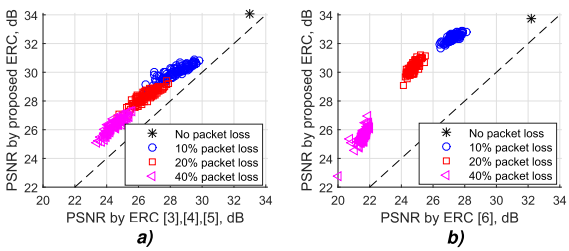

Fig. 3. Performance comparison of different ERC's.

removed packets from this set utilizing the independent packet loss model. The resulting bit stream was used for decoding and ERC by 3-D DWT decoder [3]–[5] to provide the recovered video sequence. Then this sequence was used as initial estimates $\hat{\mathbf{x}}_1^0, \ldots, \hat{\mathbf{x}}_F^0$ for ERC by Algorithm 1 with $\sigma_0 = 200$, $K = 40$. The reconstructed video quality was estimated utilizing Peak Signal-to-Noise Ratio (PSNR). Since the PSNR depends on packet loss realization, we simulated 100 realizations at loss rates 10, 20 and 40 percent.

Fig. 3 *a)* illustrates comparison of ERC performance for each packet loss realization. Here on the x-axis we show the PSNR value after ERC provided by the 3-D DWT codec [3]–[5], while the y-axis shows the corresponding PSNR achieved by Algorithm 1. The proposed algorithm increases the PSNR by 1.8 dB in average and from 1 to 3 dB depending on packet loss realization. Furthermore, 1.1 dB improvement is achieved even for the case when all packets are delivered. This is because in our model, when the encoder skips a subband, all coefficients of this subband are marked as lost and recovered by Algorithm 1 as well.

In order to provide comparison with ERC via linear interpolation as in [6], we applied biorthogonal 4.4 2-D DWT with 3 levels of decomposition for each input frame, then 1-D Haar DWT with 4 levels of decomposition for each group of 16 frames, and SPIHT [16] encoding for each frame. The packetization was simulated in the following way. Each packet contains 11 interleaved coefficients with an interval of 4 for the root subband and all corresponding spatio-temporal descendant coefficients. As result, at the video bit rate equal to 820 kbps, the packet size was approximately 700 bytes. Then we simulated packet loss using the independent loss model, recovered the video sequence utilizing root subband averaging [6] (Fig. 3 *b)*, x-axis) and futher used the recovered video as initial estimates $\hat{\mathbf{x}}_1^0, \ldots, \hat{\mathbf{x}}_F^0$ for Algorithm 1 with $\sigma_0 = 200$, $K = 40$ (Fig. 3 *b)*, y-axis). Notice, that in the packetization considered all delivered coefficients are not affected by packet losses, i.e., additional quantization as in [3], [4] does not occur. Therefore, in line 2 of Algorithm 1 we use identity matrices instead of $\mathbf{Q}_{3D}$ and $\tilde{\mathbf{Q}}_{3D}$. Comparing to the root subband averaging the proposed ERC increases the PSNR by 4.9 dB in average, from 3 to 6 dB depending on packet loss realization and by 1.5 dB when all packets are delivered.

V. Conclusion

In this letter we presented an error concealment algorithm based on iterative thresholding which can be utilized in 3-D DWT video codecs for both coding performance improvement and transmission over unreliable channels.

References

[1] *Advanced Video Coding for Generic Audiovisual Services*, document document ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), Feb. 2012.

[2] *High Efficiency Video Coding*, document ITU-T Rec. H.265 and ISO/IEC 23008-2, Oct. 2014.

[3] E. Belyaev, K. Egiazarian, M. Gabbouj, and K. Liu, "A low-complexity joint source-channel video coding for 3-D DWT codec," *J. Commun.*, vol. 8, no. 12, pp. 893–901, 2013.

[4] E. Belyaev, A. Vinel, A. Surak, M. Gabbouj, M. Jonsson, and K. Egiazarian, "Robust vehicle-to-infrastructure video transmission for road surveillance applications," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 2991–3003, Jul. 2015.

[5] *3-D DWT Codec, Version 2.0*, accessed on Apr. 26, 2017. [Online]. Available: http://www.cs.tut.fi/~belyaev/3d_dwt.htm

[6] S. Cho and W. A. Pearlman, "Error-resilient video coding with improved 3D SPIHT and error concealment," *Proc. SPIE*, vol. 5022, p. 125, May 2003.

[7] S. Kim, H. Tanaka, T. Ogawa, and M. Haseyama, "Error-resilient 3-D wavelet video coding with duplicated lowest sub-band coefficients and two-step error concealment method," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E93-A, no. 11, pp. 2173–2183, 2010.

[8] J. Zhu and R. M. Dansereau, "Error-resilient and error concealment 3-D SPIHT for multiple description video coding with added redundancy," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 855–868, Jun. 2012.

[9] J. Rombaut, A. Pizurica, and W. Philips, "Locally adaptive passive error concealment for wavelet coded images," *IEEE Signal Process. Lett.*, vol. 15, pp. 178–181, Jan. 2008.

[10] I. Daubechies, M. Fornasier, and I. Loris, "Accelerated projected gradient method for linear inverse problems with sparsity constraints," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 764–792, 2008.

[11] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, San Antonio, TX, USA, Sep./Oct. 2007, pp. I-549–I-552.

[12] K. Egiazarian and V. Katkovnik, "Single image super-resolution via BM3D sparse coding," in *Proc. 23rd Eur. Signal Process. Conf.*, 2015, pp. 2849–2853.

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[14] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 14–20, Mar. 2008.

[15] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th Eur. Signal Process. Conf.*, Poznań, Poland, 2007, pp. 145–149.

[16] *SPIHT Image Compression*, accessed on Apr. 26, 2017. [Online]. Available: http://www.cipr.rpi.edu/research/SPIHT/spiht3.html