A Study on Error Resilient Video Coding using 3D Wavelet Transform and its Error Concealment method

Sunmi KIM†  Hirokazu TANAKA†  Takahiro OGAWA†  and  Miki HASEYAMA†

† Graduate School of Information Science and Technology, Hokkaido University
Kita-14 Nishi-14, Kita-ku, Sapporo, 060-0814 Japap
E-mail: † {sunmikim, tanaka, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

Abstract In this paper, we propose an error resilient and error concealment method for 3D wavelet video coding over wireless transmission. The proposed method consists of the following two schemes; the first one is a 3D wavelet encoding with dispersive grouping, and the second one is an estimation scheme for dispersive grouped elementary streams using a Minimum Mean Square Error (MMSE) algorithm at the decoder. Experimental results indicate that the proposed algorithm provides better performance from aspects of PSNR (Peak Signal-to-Noise Ratio) and visual quality.

Keyword Wavelet video coding, Error Concealment, Dispersive Grouping, Minimum Mean Square Error (MMSE)

1. INTRODUCTION
The Scalable Video Coding (SVC) extension of the H.264/MPEG-4 Advanced Video Coding (AVC) standard (H.264/AVC) is the newest entry in the series of international video coding standards. It addresses coding schemes for reliably delivery of video to diverse clients over heterogeneous networks using available system resources, particularly in scenarios where the downstream client capabilities, system resources, and network conditions are not known in advance [1]. In recent years, extensive research activities in image coding have been focused on the discrete wavelet transform. While the good results obtained by wavelet coders (e.g., the embedded zerotree wavelet (EZW) coder [2] and the set partitioning in hierarchical trees (SPIHT) coder [3]) are partly attributable to the wavelet transform, we emphasize that much of the performance gain is obtained by using error resilient and error concealment method.

Wavelet transform generates highly compressed and scalable bit-streams, and is thus adequate for multimedia applications, video compression. However, wavelet based compressed sequences are very vulnerable to errors such as packet losses over wireless channel. Transmission errors not only corrupt the current frame, but also propagate to the successive frames [4]. Therefore, it is considerably important to develop an error resilient and concealment method to minimize the visual degradation.

Cresere divided the wavelet transform coefficients into several groups and independently processed each group. Thus, a bit error in one group does not affect the others [5].
Kim, et.al. proposed separating low-frequency sub-bands with high-frequency sub-bands and applied adaptive packetization to them for better error resilience [6].

Cho, et.al. presented a multilayered protection of embedded video bit-streams over bit errors and packet erasure channels using Error Resilient and Concealment 3-D SPIHT (ERC-SPIHT) algorithm [7]. He broke the wavelet transform into a number of spatial-temporal tree blocks which can be encoded and decoded independently by the 3-D SPIHT algorithm.

Bajic presented an adaptive maximum a posteriori (MAP) error concealment algorithm for dispersively packetized wavelet-coded images [8]. He proposed that dispersive packets spreads neighboring samples and proposed adaptive MAP concealment is based on modeling each sub-band of a still image as a Markov random field.

In this paper, we propose an efficient error resilient video coding using 3D wavelet transform and its error concealment method suitable for error prone channels. The proposed method consists of the following approaches; the first one is a 3D wavelet encoding with grouping and constructing scheme of an elementary stream format, the second one is an estimation algorithm for missing dispersive grouped elementary streams using a Minimum Mean Square Error (MMSE) algorithm at the decoder. The experimental results are shown to verify the performance of proposed scheme.

This paper is organized as follows: Section 2 presents the procedure of the error resilient video coding based on the dispersive grouping method. Furthermore we propose the error concealment method in section 3. In section 4, we show experimental results to verify the performance of our methods. Finally, concluding remarks are given in section 5.

2. Error Resilient with Dispersive Grouping

2.1. Error Resilient 3D Wavelet Video Coding

In this section, we propose a new error resilient coding using dispersive grouping method. In the first, we reviewed some of the basic concepts of the dispersive packetization method from Bajic's paper.

The proposed video coding system consists of spatial analysis, partitioning, temporal analysis, quantization and duplication as shown in Fig. 1.

Let \( w(u,v,l) \) denote the pixel value at location \((u,v)\) of \(l\)-th frame in the input video sequence of \((H \times V)\) pixels. The spatial analysis produces the wavelet coefficients \( w_p(u,v,l)(u = 1,\ldots,H; v = 1,\ldots,V) \).

The proposed video encoder employs a 3D discrete wavelet transform (DWT) to generate energy condensed coefficients. The Daubechies filters are considered in this paper because of their excellent performance. Daubechies showed how any discrete wavelet transform or two band sub-band filtering with finite filters can be decomposed into a finite sequence of simple filtering steps, which we call lifting steps but that are also known as ladder structures [9]. The implementation of a filter bank uses lifting coefficients like as:

5/3 tap wavelet filter

\[
\begin{bmatrix}
1 & 1/4(1+z^-) \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
-1/2(1+z^-) & 1
\end{bmatrix}
\]

9/7 tap wavelet filter

\[
\begin{bmatrix}
K_0 & 0 \\
0 & K_1
\end{bmatrix}
\begin{bmatrix}
\delta(1+z^-) & 1 \\
\gamma(1+z^-) & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

\(\alpha = -1.586134342059, \quad \beta = -0.5298011857, \quad \gamma = 0.882911075530, \quad \delta = 0.443506852043971, \quad K_0 = 0.8128930661, \quad K_1 = 1.230174104914001\)

Further these coefficients are fed into a partitioning module as shown in Fig. 2.

The proposed partitioning function is a remainder function to generate groups using unique index – location index of horizontal, vertical and time position in the frames – of video sequence. How to assign the wavelet coefficient to the group index \(n\) is calculated as follows:

\[ n = (u + v + l) \mod N, \]

where \(N\) is the number of divided groups.
The Institute of Image Information and Television Engineers
NII-Electronic Library Service

Fig. 3 The concept of the proposed constructing method.

The partitioning produces wavelet coefficients of $n$-th group $w_n(u', v', l) (u' = 1, \ldots, H/N; v' = 1, \ldots, V/N)$, $(u', v')$ is the location of coefficient in $n$-th group.

After the partitioning procedure, $p$-level temporal wavelet decomposition is processed in the group of frames (GOF) for each group. The GOF size allows maximum temporal composition level ($p_{\text{max}}$) as follows:

$$p_{\text{max}} = \log_2(\text{GOF}).$$

Therefore, we can get higher decomposition efficiency in longer GOF frames. After the wavelet transform, all coefficients are uniformly quantized by the following:

$$w_q^n = \text{sign}(w_{st}^n) \left[ \frac{w_{st}^n}{\delta} \right],$$

where $w_{st}^n$ is the spatial-temporal wavelet transform coefficient, $w_q^n$ is the quantization result, $\delta$ is the quantization step size, $\text{sign}(x)$ returns the sign of coefficient $x$, and $[x]$ operation obtains the largest integer that is less or equal than $x$.

2.2. Duplicated Method for Making Bit-stream

To minimize damage caused by the packet loss, a new packet duplication method is proposed while not significantly increasing the bit-rate. In the proposed method, the duplication is performed for only bit-stream of the lowest sub-band. This high priority data is critical to the decoded video quality but its size is small compared to the entire video data.

The lowest frequency sub-band is referred to as $\text{LL}_k$ band indicating that the signal has been low-pass filtered in both horizontal and vertical direction at the spatial decomposition level ($p = k$). Since the lowest frequency sub-band coefficients have the highest energy, we propose a duplication method for the constructing a bit stream as shown Fig. 3.

For example, the output bit stream of the conventional $k$-level spatial wavelet transmission encoder is shown as follows: $\text{LL}_k - \{ \text{LH}_k \cdots \text{LH}_l \} - \{ \text{HH}_k \cdots \text{HH}_l \} - \{ \text{HL}_k \cdots \text{HL}_l \}$ to send video sequence.

As shown in Fig.3, the output bit stream of the proposed encoder is like as $\text{LL}_k - \{ \text{LH}_k \cdots \text{LH}_l \} - \text{LL}_l - \{ \text{HH}_l \cdots \text{HH}_l \} - \{ \text{HL}_k \cdots \text{HL}_l \}$ in each $l$-th frame of $n$-th group.

The decoder can obtain a reasonable reconstruction quality from the multiple lowest sub-band information over packet loss environment.

3. Proposed Error Concealment Method

Previous works concealed the missing coefficients using the average value of surrounding coefficients [10]. We apply the wavelet transformation and error concealment method as shown in Fig. 4. Any early decoding failure or missing coefficients affect to the related spatial-temporal coefficients of same GOF region in the wavelet domain. Therefore, we apply the error concealment step in the lowest sub-band before the wavelet decomposition.

The general operations of the proposed error concealment method are as follows:

Step 1. Check the erroneous packet and error detection.
Step 2. Error concealment treats the lowest sub-band with duplicated coefficients.
Step 3. Temporal decomposition and inverse dispersive grouping and spatial decomposition.
Step 4. Error concealment treats all of video frames with MMSE estimation.

The lost coefficients of the lowest sub-band are replaced by their duplicated coefficients.

The concealed coefficient value $\tilde{w}_{st}^n(u', v', l)$ of a missing coefficient $w_{st}^n(u', v', l)$ in $n$-th group is derived from duplicated coefficients of other groups.

$$\tilde{w}_{st}^n(u', v', l) = \sum_{r=1}^{R} \frac{1}{R} w_{st}^n(u', v', l),$$

where $R$ is the number of duplication and $\tilde{w}_{st}^n(u', v', l)$ is the duplicated coefficient of $r$-th group.

After first concealment in the lowest sub-band as eq. (6), the temporal wavelet decomposition is applied to each group.

Fig. 4 The flowchart of the proposed concealment method.
Then we apply the inverse dispersive grouping and the spatial wavelet decomposition to reconstruct the original video sequence.

However, there are still some missing coefficient values which are not successfully recovered by the lowest spatial sub-band concealment in each frame. To recover the lost information, we proposed error concealment method by estimating the values of the missing coefficients using Minimum Mean Square Error (MMSE) algorithm.

The MMSE estimation turns out to be particularly simple when the observed values of missing coefficient and reference are jointly Gaussian distribution [11].

The concept of the proposed error concealment method in the decoder is shown in Fig.5, and the steps of the concealment are as follows.

\[ \hat{w}(u,v) = \frac{1}{N_{ref}} \sum \alpha_{i,j} x_{i,j}, \]

where \( N_{ref} \) is the number of neighbors for estimation.

In Fig. 5, \( t \) is the index of area of reference concealing coefficients and \( w(i,j) \) is neighbor of lost coefficient, \( u-t \leq i \leq u+t \) and \( v-t \leq j \leq v+t \). Hence, the total number of its neighbors for the estimation is

\[ N_{ref} = 2t+1 \ (t \geq 1), \]

To minimize the error of \( (\hat{w}(u,v) - w(u,v))^2 \), the optimal weighting factors \( \alpha_{i,j} \) are derived by using MMSE algorithm.

\[ a = \frac{1}{N_{ref}} \sum w(i,j) x_{i,j}, \]

where \( R_{YY} \) is the auto-correlation matrix of Y and \( R_{XY} \) is the cross-correlation matrix of X and Y. In practical implementation, an approximation of \( R_{YY} \) and \( R_{XY} \) can be calculated from the empirical data as

\[ \bar{R}_{YY} = \frac{1}{N_{ref}} \sum w(i,j) x_{i,j}, \]

\[ \bar{R}_{XY} = \frac{1}{N_{ref}} \sum w(i,j) x_{i,j}, \]

and

\[ X = \begin{bmatrix} w(u-t,v-t) & w(u-t+1,v-t) & \cdots & w(u+t,v-t) \\ w(u-t,v-t+1) & w(u-t+1,v-t+1) & \cdots & w(u+t,v-t+1) \\ \vdots & \vdots & \ddots & \vdots \\ w(u-t,v+t) & w(u-t+1,v+t) & \cdots & w(u+t,v+t) \end{bmatrix}, \]

\[ Y = \begin{bmatrix} w(i-j,t-j) & w(i-j+1,t-j) & \cdots & w(i-j+t,t-j) \\ w(i-j,t-j+1) & w(i-j+1,t-j+1) & \cdots & w(i-j+t,t-j+1) \\ \vdots & \vdots & \ddots & \vdots \\ w(i-j,t+j) & w(i-j+1,t+j) & \cdots & w(i-j+t,t+j) \end{bmatrix}, \]

\[ (12) \]

4. EXPERIMENTAL RESULTS

We tested the proposed error-concealment algorithm on Foreman (CIF resolution) YUV 4:2:2 color video sequence. The simulation parameters are listed in the table 1. In this simulation, it is assumed that one entire group is lost in the channel and the other groups are successfully received in the receiver. As for the concealment by MMSE, eight surrounding neighbor coefficients \((t=1)\) are referenced for interpolation.

We provide an example of visual comparison of the first frame of foreman with 3-level spatial compression and 4-level temporal compression. The number of dispersive group is 4 and the number of frames 32. Fig. 6 shows the improvement in reconstruction of missing coefficient using dispersive grouping and the proposed error concealment method. We observe that the PSNR of reconstructed frame with using dispersive higher than without grouping sequence. And the proposed error concealment algorithm provides better performance from aspects of PSNR and visual quality.

Experimentation of Fig. 7 is based on 3-level spatial compression and diverse temporal compression as follow GOF size while number of dispersive group is 8 and the number of frames is 32. Two downward lines show the 1/32 group error (=3%) and 2/32 groups error (=6%) cases.

As you can see in Fig. 7, the performance gain that can be obtained by the proposed algorithm is decrease in the GOF size aspect. On the other hand, the compression ratio of 16 GOF is increase by 150 percent compare to 2 GOF.

<table>
<thead>
<tr>
<th>Table 1 Simulation Parameter.</th>
</tr>
</thead>
<tbody>
<tr>
<td>file name</td>
</tr>
<tr>
<td>frame size</td>
</tr>
<tr>
<td>the number of frames</td>
</tr>
<tr>
<td>wavelet filter type</td>
</tr>
<tr>
<td>spatial compressed level</td>
</tr>
<tr>
<td>temporal compressed level</td>
</tr>
<tr>
<td>the size of GOF</td>
</tr>
</tbody>
</table>
Fig. 4. Experimented results (the number of groups: 4, GOF: 16, error rate: 6%) (a) frame : 0 without error (47.69dB) (b) one case without dispersive grouping in erroneous environment (12.57dB) (c) dispersive grouping in erroneous environment (14.52dB) (d) proposed algorithm (35.92dB).

Fig. 7. The PSNR and compression ratio with 8 groups.

5. Conclusion
In this paper, we proposed an error resilient and error concealment method for 3D wavelet video coding over wireless transmission. At the encoder side, dispersive grouping which is a new encoding method composing partitioning and duplicating the lowest sub-band coefficients in a bit stream format was proposed. At the decoder side, two steps error concealment method for dispersive grouping was proposed. It was found that the proposed algorithm allows covering packet losses in wireless channel and improves video quality based on 3D wavelet video coding.

ACKNOWLEDGMENT
This study was partly supported by Hokkaido University Global COE Programs by Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES