OPTIMUM UNEQUAL ERROR PROTECTION OF SNR-SCALABLE DPCM-CODED VIDEO

Till Halbach

Department of Telecommunications
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
E-mail: *halbach@tele.ntnu.no*

ABSTRACT

A channel code allocation scheme is developed for transmission of video over error-prone channels. The system is designed to minimize the expected distortion of the video under a quality constraint. It is based on the Viterbi algorithm and achieves a very robust bit stream, also in mismatch situations. It is further shown that the distribution of available channel codes vary significantly with respect to the channel conditions.

1. VIDEO CODING

Almost all video coding standards used today are based on the well known block-based hybrid coding scheme, here represented by the recently released H.264. A prediction of the current signal is computed and subtracted from the original. This prediction error is then transformed, quantized and entropy-encoded. H.264 processes so-called macroblocks (MBs) of a video frame. Considering only the luminance component of a YCbCr signal, a MB consists of 16×16 pixels. A comprehensive overview of H.264 can be found in [1].

To achieve high compression, various prediction processes are employed. Spatial prediction is based on samples of the current video frame. The signal's temporal prediction is made from samples of a previously en- and decoded frame. Other predictions exist as well, e.g. of motion vectors and coding modes. This work extends the standard further by SNR scalability as a highly desirable feature in e.g. erroneous transmission environments. This is achieved by incorporation of quality layers and the prediction of quantized transform coefficients of a high-quality layer by the coefficients of the next lower-lying layer [2].

Prediction and the use of variable-length coding for entropy coding make the resulting bit stream prone for propagation of transmission errors. Thus, H.264 groups MBs in spatial areas called slices which are independently decodable, i.e. there is no prediction across slice boundaries. Potential errors can, however, propagate temporally. To avoid that, this work focuses on the use of rectangular contiguous slices of which the borders have the same meaning as video frame boundaries, as well as the use of only one reference frame. Motion estimation/compensation is thus limited by slice boundaries. Temporal error propagation is further limited by inserting I slices after $N_\mathrm{f} - 1$ video frames, hereby leading to a Group of Pictures (GOP) structure of N_f frames.

An encoded slice is referred to by $\tilde{X}[f, l, y, x]$ and is identical with one source packet (SP). The subscripts f, l, y, and x denote the frame, layer, vertical and horizontal index, respectively. The slice of the original frame is $X[f, y, x]$.

2. UNEQUAL ERROR PROTECTION

Consider a quality-controlled source encoder and a set of channel codes $\mathcal{C}$. The SPs, referred to by the index i_s and with length $R_\mathrm{s}[i_\mathrm{s}]$, are split up and transmitted as payloads of length $R_\mathrm{c}[i_\mathrm{c}]$ by N_cp channel packets (CPs), each of which is identified by the index i_c. The task is then to assign each CP a channel code $\Gamma[i_\mathrm{c}] \in \mathcal{C}$ such that the overall expected distortion of the GOP to transmit is as small as possible. With other words, there is a joint source channel coding problem with respect to the optimum channel rate allocation. The channel encoder seeks to minimize the overall expected distortion,

$$\min_{\Gamma[i_\mathrm{c}] \in \mathcal{C}} \tilde{D}_\mathrm{GOP} \,, \tag{1}$$

subject to the rate constraint

$$\sum_{i=0}^{N_\mathrm{cp}-1} R_\mathrm{c}[i] = \sum_{i=0}^{N_\mathrm{sp}-1} R_\mathrm{s}[i] \,, \tag{2}$$

which is given by the number of SPs $N_\mathrm{sp} = N_\mathrm{f} N_\mathrm{l} N_\mathrm{s}$ and their lengths which are in turn quality-controlled. N_l is the number of quality layers, and $N_\mathrm{s} = N_\mathrm{y} N_\mathrm{x}$, where N_y and N_x are the numbers of vertical and horizontal slices in a frame, respectively.

CPs are of fixed length R_cp, which in this work is chosen to be 517 bytes or 4136 bits. The payload bits are followed by an 8-bit code word which specifies the channel code of the *next* transmitted CP. The specifier of the first packet is assumed known. A 16-bit cyclic redundancy check code is computed over both payload and code specifier to detect residual errors in the code stream after channel decoding. The CP is completed by appending $C[i_\mathrm{c}]$ parity bits.

The dependency between SP indices i_s and slice indices is described by $i_\mathrm{s}[f,l,x,y] = N_\mathrm{s}N_\mathrm{l}f + N_\mathrm{s}l + N_\mathrm{x}y + x$, and $f[i_\mathrm{s}] = \left\lfloor \frac{i_\mathrm{s}}{N_\mathrm{s}N_\mathrm{l}} \right\rfloor$, $l[i_\mathrm{s}] = \left\lfloor \frac{i_\mathrm{s}}{N_\mathrm{s}} \right\rfloor - f[i_\mathrm{s}]N_\mathrm{l}$, $y[i_\mathrm{s}] = \left\lfloor \frac{i_\mathrm{s}}{N_\mathrm{x}} \right\rfloor - l[i_\mathrm{s}]N_\mathrm{y} - f[i_\mathrm{s}]N_\mathrm{l}N_\mathrm{y}$, and $x[i_\mathrm{s}] = i_\mathrm{s} \bmod N_\mathrm{x}$. In words, slices are read in raster scan order into the code stream. The mapping of SP to CP indices is done with respect to the rate consumption of the channel rate allocation. By means of the accumulated source rate

$$R_\mathrm{s}^{(\mathrm{a})}[i_\mathrm{s}] = \sum_{i=0}^{i_\mathrm{s}-1} R_\mathrm{s}[i] \qquad i_\mathrm{s} = 1,\ldots,N_\mathrm{sp}-1 \tag{3}$$

(in bits), the special case being $R_\mathrm{s}^{(\mathrm{a})}[0] = 0$, and the accumulated channel payload rate

$$R_\mathrm{c}^{(\mathrm{a})}[i_\mathrm{c}] = \sum_{i=0}^{i_\mathrm{c}-1} R_\mathrm{c}[i] \qquad i_\mathrm{c} = 1,\ldots,N_\mathrm{cp}-1 \tag{4}$$

(in bits), including $R_\mathrm{c}^{(\mathrm{a})}[0] = 0$, the mapping can be defined as

$$i_\mathrm{s}[i_\mathrm{c}] = \max\Big(\big\{0, i_\mathrm{s} \mid R_\mathrm{s}^{(\mathrm{a})}[i_\mathrm{s}+1] \le R_\mathrm{c}^{(\mathrm{a})}[i_\mathrm{c}]\big\}\Big)\,. \tag{5}$$

The number of CPs which have to be transmitted for one GOP then becomes

$$N_\mathrm{cp} = \min\Big(\big\{i_\mathrm{c}+1 \mid R_\mathrm{s}^{(\mathrm{a})}[N_\mathrm{sp}] \le R_\mathrm{c}^{(\mathrm{a})}[i_\mathrm{c}]\big\}\Big)\,. \tag{6}$$

Zero padding may have to be applied to the last SP to fill up a CP. The allocation procedure has to evaluate the curves $D\big(R_\mathrm{c}^{(\mathrm{a})}[i_\mathrm{c}]\big)$ of each GOP for a set of combinations of channel code allocations, which may be a subset of the set of all possible combinations.

A stationary binary symmetric channel with channel bit error rate ϵ is considered. A set of eight channel codes is employed for encoding and error correction in the following, the code rates being $r_i = k_i/d$ with the common denominator $d = 12$ and $k_i = \{4, 5, 6, 8, 9, 10, 11, 12\}$. This results in payload lengths of 169, 212, 255, 341, 384, 427, 470, and 513 bytes. The codes consist of punctured parallel concatenated recursive convolutional codes as described in [3]. Based on these rates, the probabilities $P_\mathrm{e}\big(\Gamma[i_\mathrm{c}], \epsilon, R_\mathrm{cp}\big)$ of a CP having at least one bit error after 20 decoder iterations have been computed in extensive Monte-Carlo simulations of 10,000 blocks and can hence be tabulated for use in the channel code allocation algorithm.

The decoding strategy is that, if an error is encountered, all temporally following slices at the same spatial position, and which belong to the same or higher-quality layers, are discarded and concealed to avoid error propagation. The expected distortion after the transmission of $N_t(>1)$ CPs of a GOP then becomes

$$\tilde{D}_\mathrm{GOP}(N_\mathrm{t}) = \sum_{i=0}^{i_\mathrm{s}[1]} D_\mathrm{e}^{(\mathrm{a})}(i,0)P_\mathrm{e}(0) + \sum_{i_\mathrm{c}=1}^{N_\mathrm{t}-1} \sum_{i=i_\mathrm{s}[i_\mathrm{c}]}^{i_\mathrm{s}[i_\mathrm{c}+1]} D_\mathrm{e}^{(\mathrm{a})}(i,i_\mathrm{c})P_\mathrm{e}(i_\mathrm{c})P_\mathrm{ne}^{(\mathrm{a})}(i_\mathrm{c})\,, \tag{7}$$

where $P_\mathrm{e}(i_\mathrm{c})$ is the abbreviation for $P_\mathrm{e}(\Gamma[i_\mathrm{c}], \epsilon, R_\mathrm{cp})$.

In order to define the distortion $D_\mathrm{e}^{(\mathrm{a})}$ made by terminating decoding and carrying out a controlled concealment, three strategies are specified. If no data concealment can refer to are available, the best estimation of the lost data is to assume the segment's mean and accumulate the distortion contributions over the whole GOP,

$$D_\mathrm{c,m}^{(\mathrm{a})}(i) = \sum_{g=0}^{N_\mathrm{f}-1} d\big(\tilde{X}(g, N_\mathrm{l}-1, y, x), m_{\tilde{X}}\big)\,, \tag{8}$$

where $d(\cdot,\cdot)$ represents the 2-D MSE between two variables. Here and also in following distortion terms, i is the index of the particular SP, and f, g, l, y, and x are functions of i according to Eq. (5). In practice, $m_{\tilde{X}}$ is estimated by the value 128; this assumes an 8-bit pixel representation. If a base layer slice is lost, the concealment scheme of choice is to freeze the highest-quality content of the corresponding slice from the previous frame,

$$D_\mathrm{c,t}^{(\mathrm{a})}(i) = \sum_{g=f}^{N_\mathrm{f}-1} d\big(\tilde{X}(g, N_\mathrm{l}-1, y, x), \tilde{X}(f-1, N_\mathrm{l}-1, y, x)\big)\,. \tag{9}$$

If a lower-quality slice of the lost slice is available, all remaining frames which depend on the lost one are replaced,

$$D_\mathrm{c,l}^{(\mathrm{a})}(i) = \sum_{g=f}^{N_\mathrm{f}-1} d\big(\tilde{X}(g, l, y, x), \tilde{X}(g, l-1, y, x)\big)\,. \tag{10}$$

Also, $D_\mathrm{c,f}^{(\mathrm{a})}(i) = D_\mathrm{c,l}^{(\mathrm{a})}(i)$. Summarizing, concealment and channel distortions are given by

$$D_\mathrm{c}^{(\mathrm{a})}(i) = \begin{cases} D_\mathrm{c,m}^{(\mathrm{a})}(i) & f = 0 \wedge l = 0 \quad \text{(11a)} \\ D_\mathrm{c,f}^{(\mathrm{a})}(i) & f = 0 \wedge l > 0 \quad \text{(11b)} \\ D_\mathrm{c,t}^{(\mathrm{a})}(i) & f > 0 \wedge l = 0 \quad \text{(11c)} \\ D_\mathrm{c,l}^{(\mathrm{a})}(i) & f > 0 \wedge l > 0\ . \quad \text{(11d)} \end{cases}$$

Corresponding to Eq. (11) (with substript 'q' instead of 'c'), the definition of $D_\mathrm{q}^{(\mathrm{a})}(i)$ distinguishes among four cases, where

$$D_\mathrm{q,m}^{(\mathrm{a})}(i) = \sum_{g=0}^{N_\mathrm{f}-1} d\big(X(g, y, x), \tilde{X}(g, N_\mathrm{l}-1, y, x)\big)\,, \tag{12}$$

and $D^{(a)}_{q,f}(i) = D^{(a)}_{q,t}(i) = D^{(a)}_{q,m}(i)$. In case of layer concealment and $f > 0$, the quantization distortion becomes

$$D^{(a)}_{q,l}(i) = \sum_{g=0}^{f-1} d\big(X(g,y,x), \tilde{X}(g, N_l - 1, y, x)\big) + \sum_{g=f}^{N_f-1} d\big(X(g,y,x), \tilde{X}(g,l,y,x)\big)\,. \quad (13)$$

In case several SPs are transported with one CP, the formulation of a joint distortion is required. Let $i_{s,l} = i_s(i_c+1)$ be the index of the last included SP in the CP with index i_c. Then, the on-error distortion is affected by the first N_s and potentially all base layer slices of the following video frame. Let further $i_{s,f} = i_s(i_c)$ be the index of the first included SP in the CP, and $i_{s,s}(i) = i_s\big(f(i_{s,f}) + 1, 0, y(i), x(i)\big)$, and let $i_{s,bl} = i_s\big(f\big(i_s(i_c)\big) + 1, 0, 0, 0\big)$. The set of SP indices of these base layer slices is then written as

$$\mathcal{I}_{e,bl}(i_c) = \begin{cases} \varnothing & \text{(i)} \quad (14a) \\ \mathcal{I}_{e,bl,1}(i_c) & \text{(ii)} \quad (14b) \\ \mathcal{I}_{e,bl,2}(i_c) & \text{otherwise}\ , \quad (14c) \end{cases}$$

where $\mathcal{I}_{e,bl,1}(i_c) = \big\{i_{s,bl}, \ldots, \min\big(\{i_{s,l}, i_{s,bl}{+}N_s{-}1\}\big)\big\}$ and $\mathcal{I}_{e,bl,2}(i_c) = \big\{i_{s,bl}, \ldots, \min\big(\{i_{s,l}, i_{s,bl}{+}N_x y(i_{s,f}){+}x(i_{s,f})\}\big)\big\}$. The named conditions are (i) $i_{s,l} < i_{s,bl} \vee \big(l(i_{s,f}) = 0 \wedge y(i_{s,f}) = 0 \wedge x(i_{s,f}) = 0\big)$, and (ii) $i_{s,l} \geq i_{s,bl} \wedge l(i_{s,f}) > 0 \wedge y(i_{s,f}) = 0 \wedge x(i_{s,f}) = 0$. Now, the set of all indices of interest can be formulated as $\mathcal{I}_e(i_c) = \big\{i_{s,f}, \ldots, i_{s,f} + N_s - 1\big\} \cup \mathcal{I}_{e,bl}(i_c)$, and the on-error distortion becomes finally

$$D^{(a)}_e(i, i_c) = \begin{cases} w_{ss}(i)\big(D^{(a)}_c(i) + D^{(a)}_q(i)\big) & \text{(iii)} \quad (15a) \\ w_{ss}(i)\big(D^{(a)}_{jc}(i) + D^{(a)}_{jq,f}(i)\big) & \text{(iv)} \quad (15b) \\ w_{ss}(i)\big(D^{(a)}_{jc}(i) + D^{(a)}_{jq,l}(i)\big) & \text{(v)} \quad (15c) \\ 0 & \text{othrw.,} (15d) \end{cases}$$

where $D^{(a)}_{jq,l} = D^{(a)}_{q,l}$, as well as

$$D^{(a)}_{jc}(i) = \sum_{g=f}^{N_f-1} d\big(\tilde{X}(g,l,y,x), \tilde{X}(f, l-1, y, x)\big) \quad (16)$$

and

$$D^{(a)}_{jq,f}(i) = \sum_{g=0}^{N_f-1} d\big(X(g,y,x), \tilde{X}(g,l,y,x)\big)\,, \quad (17)$$

and with the conditions (iii) $i \in \mathcal{I}_e(i_c) \wedge i_{s,s}(i) \notin \mathcal{I}_e(i_c)$, (iv) $i \in \mathcal{I}_e(i_c) \wedge i_{s,s}(i) \in \mathcal{I}_e(i_c) \wedge f(i) = 0$, and (v) $i \in \mathcal{I}_e(i_c) \wedge i_{s,s}(i) \in \mathcal{I}_e(i_c) \wedge f(i) > 0$. The weighting factor $w_{ss}(i) = \frac{1}{N_f}\frac{N_{MB}[f,y,x]}{N_{MB}[f]}$ completes the GOP distortion average by normalizing by N_f, and also accounts for unequal slice sizes, where N_{MB} is the number of MBs of a particular slice segment or frame. The curves of typical average GOP distortions are shown in Fig. 1.

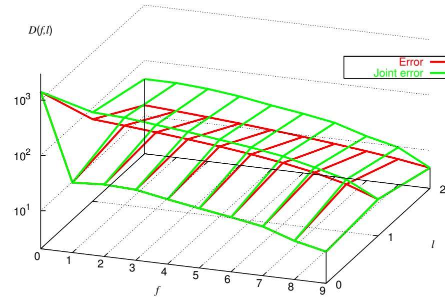

Fig. 1. Accumulation distortions of a single source segment with 10 frames and 3 layers.

$P^{(a)}_{ne}(i_c)$ in Eq. (7) is the probability for the event 'no error in previously transmitted CPs', and is thus formulated as

$$P^{(a)}_{ne}(i_c) = \prod_{k=0}^{i_c-1} P_{ne}(k, i_c) \qquad i_c > 0\,. \quad (18)$$

The probability $P_{ne}(k, i_c)$ of the event 'no error in the CP with index k' is taken into consideration in Eq. (18) only when the respective packet is a reference CP, i.e. it contains whole or parts of SPs which are referred to from the current SP (index i) of the CP with index i_c:

$$P_{ne}(k, i_c) = \begin{cases} \big(1 - P_e(k)\big) & k \in \mathcal{R}(i_c) \quad (19a) \\ 1 & \text{otherwise}\ . \quad (19b) \end{cases}$$

The set $\mathcal{R}(i_c)$ contains the indices of reference CPs of a SP, $\mathcal{R}(i_c) = \Big\{r | r \in \{0, \ldots, i_c - 1\} \ \wedge\ \big(R^{(a)}_c(r) < R^{(a)}_s(t + 1) \ \wedge\ R^{(a)}_c(r+1) > R^{(a)}_s(t)\big)\Big\}$. The indices t are in turn elements of the set of reference SPs of a SP, $t \in \mathcal{J}(i_c)$, where, with $g(s) \in \big\{0, \ldots, f(s)\big\}$ and with $m(s) \in \big\{0, \ldots, l(s)\big\}$, $\mathcal{J}(i_c) = \big\{i_s\big(g(s), m(s), y(s), x(s)\big)\big\}$. Finally, $s \in \mathcal{S}(i_c)$, where $\mathcal{S}(i_c) = \big\{i_{s,f}, \ldots, i_{s,l}\big\}$.

3. EXPERIMENTS, RESULTS, AND DISCUSSION

The optimal solution to (1) is achieved by a brute-force search through the set of all possible solutions, but this approach is limited by N_{cp}. Therefore, this work deploys the low-complexity Viterbi Algorithm (VA), which provides a potentially suboptimal solution, to determine the channel rate allocation which gives minimum expected distortion, as exposed in [4]. It is stressed that, even though only the results for the QCIF-size YCbCr sequence `Mother&Daughter` are presented below due to space limitations, they are consistent to those of other sequences. In the following, $N_f = 10$, $N_l = 3$, and $N_s = 1$. The expected distortions are averaged over all GOPs of the video.

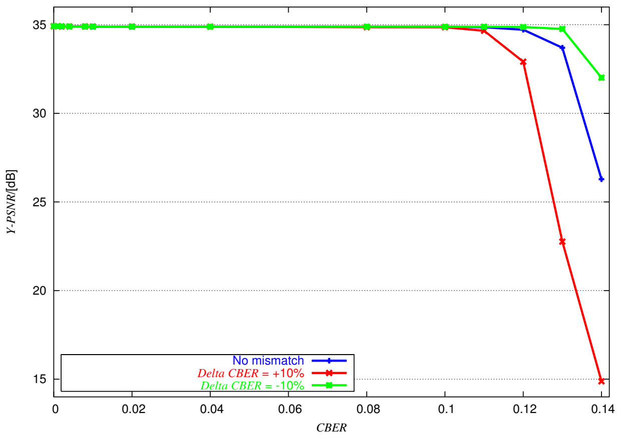

Fig. 2. Expected distortion as a function of the channel bit error rate, with and without a mismatch of $\Delta\epsilon = \pm 10\%$

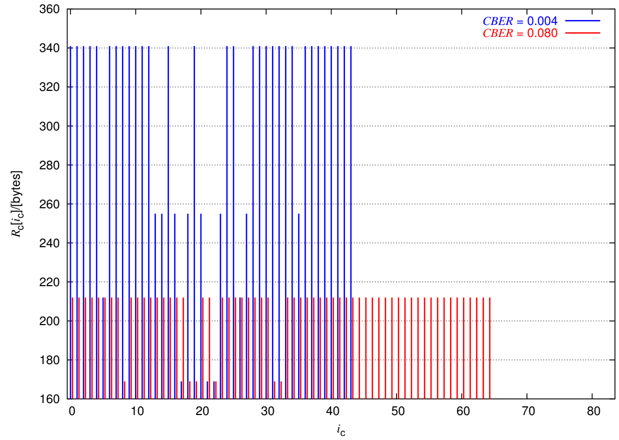

Fig. 4. Distribution of channel codes for one GOP

As seen in Fig. 2, the system aims at a constant distortion, except when the $CBER$ becomes too high to ensure an error-free transmission; then the protecting capability of the strongest channel code is exceeded, here at 0.14. The loss in mismatch situations is less than 0.2 dB, i.e. the image quality degrades gracefully.

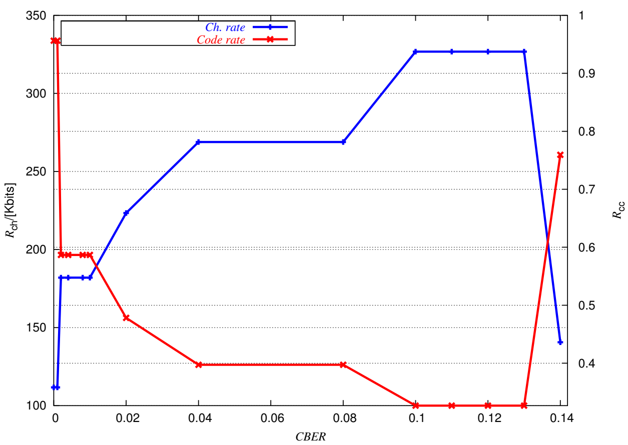

Fig. 3. Channel rate in- and average code rate decrease

All-sequence average $Y-PSNR$ and source bit rate of the three (YUV) layers are 29.23, 32.19, and 34.93 dB, and 21.7, 50.1, and 106.8 kbits/s, respectively. The computed channel rates which yield minimum distortion are plotted in Fig. 3.

As shown in Fig. 4, the code distributions are quite contrasting for the two given $CBER$s. It is stressed that, as the codec aims at the maximization of the average payload length of CPs, plain use of the strongest channel code only would not give maximum $PSNR$. Furthermore, packets, the loss of which would lead to high distortions, are protected by strong channel codes, whereas less important packets are assigned weaker codes.

4. CONCLUSIONS

A channel code allocation scheme has been developed which is optimal in a VA sense. It was shown that the proposed algorithm achieves a very robust bit stream, also in channel mismatch situations. The distribution of channel codes varies strongly according to the channel conditions. Important signals are protected by strong channel codes, and less important information is assigned weak channel codes. The algorithm aims at the same time successfully at maximizing the average code rate. The developed channel coding system is therefore highly recommendable in situations where video is transmitted over unreliable channels.

5. REFERENCES

[1] Till Halbach, "The H.264 video compression standard," in *Proc. Norwegian Signal Processing Symposium (NORSIG)*, Bergen (Norway), Oct. 2003.

[2] Till Halbach and Thomas R. Fischer, "SNR scalability by transform coefficient refinement for block-based video coding," in *Proc. SPIE's Visual Communications and Image Processing (VCIP)*, Lugano (Switzerland), July 2003.

[3] B.A. Banister, *Robust image and video transmission across heterogeneous networks with packet erasures and bit errors*, Ph.D. thesis, School of Electrical Engineering and Computer Science, Washington State University, Pullman (WA, USA), May 2002.

[4] Till Halbach and Thomas R. Fischer, "Channel rate allocation for error-robust packet transmission of SNR-scalable DPCM-coded video," in *Proc. International Symposium on Signal Processing and its Applications (ISSPA)*, Paris (France), July 2003.