OPTIMAL JOINT SOURCE-CHANNEL CODING USING UNEQUAL ERROR PROTECTION FOR THE SCALABLE EXTENSION OF H.264/MPEG-4 AVC

M. Stoufs, A. Munteanu, P. Schelkens and J. Cornelis

Vrije Universiteit Brussel
Department of Electronics and Informatics (ETRO) – Interdisciplinary Institute for Broadband Technology (IBBT)
Pleinlaan 2, 1050 Brussels, Belgium

ABSTRACT

This paper proposes an optimized joint source-channel coding methodology with unequal error protection for the transmission of video encoded with the recently developed scalable extension of H.264/MPEG-4 AVC. The proposed methodology uses a simplified Viterbi-based search method which significantly outperforms the classical exhaustive search method in terms of computational complexity, leading to a practically applicable solution at the expense of a minimal loss of optimality. Experimental results show the effectiveness of our protection methodology and illustrate its capability to provide graceful degradation in the presence of channel mismatches.

***Index Terms*—** Scalable video coding, error resilience, joint source-channel coding, unequal error protection

1. INTRODUCTION

Scalable video coding is of particular importance when transmitting digital video over heterogeneous, error-prone networks to a large variety of devices, since it allows on-the-fly adaptation of the original encoded bit-stream to meet the requirements set by the end-user's connection and terminal characteristics. Recently, the standardization of SVC (Scalable Video Coding), a scalable extension to the state-of-the-art H.264/MPEG-4 AVC video coding standard has reached its final stages [1]. This highly effective scalable video compression technique provides quality, resolution and frame-rate scalability while its rate-distortion performance is on par with single layer H.264.

However, when the SVC-streams are transmitted over error-prone channels, the incurred errors typically result in a loss of synchronization of the entropy decoder which eventually leads to irrecoverable decoder failure. Error control mechanisms are therefore of vital importance when the SVC-streams are sent over error-prone channels. In this context, error protection in packet-based networks is typically achieved with forward error correction codes. Although according to [2] source and channel coding can be performed separately with optimality, this is only valid under the assumption of asymptotically long block lengths of transmitted data and unlimited complexity and delay. In practical applications, joint source-channel coding (JSCC), allowing for the optimal allocation of the available bits between the source and channel codes, has been shown to deliver better results. In the context of SVC, to the best of our knowledge, only one JSCC-algorithm has been recently presented in [3]. In [3], a rate allocation approach for the transmission of SVC bit-streams in variable-length channel packets over MIMO systems is proposed.

In contrast to the approach of [3], where a computationally prohibitive exhaustive search for the optimum solution is performed, in this paper, we develop a novel optimal JSCC-methodology based on Lagrangian-optimization for the transmission of SVC-streams over error-prone channels in fixed-length channel packets, since this provides the advantage of an easier cross-layer design [4]. We also derive a Viterbi-based search algorithm which significantly reduces the computational complexity of the rate allocation and minimizes the end-to-end distortion. Channel coding is performed with state-of-the-art low-density-parity-check (LDPC) codes [5] and transmission over binary erasure channels (BECs) is considered. The proposed JSCC-approach assumes the presence of an interleaver in the transmission scheme, which translates the packet-loss model into a BEC model. Finally, as the source is encoded by a scalable codec producing layers with different levels of importance, our JSCC-design incorporates unequal error protection (UEP) of the source packets such that error resilience with graceful degradation is achieved up to a certain level of transmission errors.

The paper is structured as follows. Section 2 briefly describes the recently developed SVC-standard. In section 3 and 4, a methodology to solve the JSCC-problem is proposed. Section 5 introduces a novel Viterbi-base search algorithm as a practical alternative to the computationally prohibitive exhaustive search method. Next, in section 6 the proposed JSCC solution is experimentally evaluated. Finally, in section 7, the conclusions are drawn.

2. SCALABLE VIDEO CODING (SVC) CODER

SVC generates a fully-scalable bit-stream consisting of a base layer, representing the version of the input video with the lowest supported resolution and quality and a number of enhancement layers, representing versions of the input sequence with a higher resolution and/or quality. To encode each (coarse-grain) quality/resolution layer, the frames are first temporally decomposed using hierarchical bi-directional motion compensated prediction (hierarchical B-frames). By omitting the lower levels of the hierarchy, temporal scalability can be supported.

Each spatial resolution layer is composed of a base layer and some quality enhancement layers that may be either coarse-grain scalable, medium-grain scalable or fine-grain scalable (FGS) [1]. FGS is supported by encoding successive refinements of the transform coefficients, starting with the minimum quality provided by AVC-compatible intra/residual coding. This is done by repeatedly decreasing the quantization parameter QP by 6, thereby approximately dividing the quantization step size by two and applying a modified entropy coding similar to bitplane coding. The setup of an SVC-encoder producing a bit-stream with one resolution and two FGS-layers is given in Figure 1.

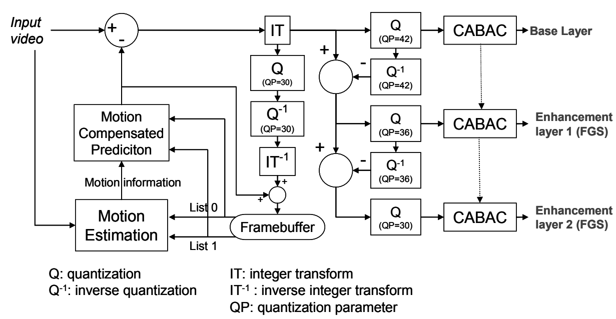

Figure 1: SVC encoder generating one base layer and two enhancement layers.

3. JSCC PROBLEM FORMULATION

We consider an original video sequence consisting of F frames. This sequence is encoded with a single resolution layer and a number of FGS-layers. We focus on the transmission of M_l packets of fixed-length N over a BEC with erasure parameter ε and total capacity of the channel R_{tot}. We also define a number of protection levels d and impose UEP by forcing the source rate in the M_l packets to be non-decreasing. As in [6] it is assumed that the total expected distortion can be expressed as a sum of the individual frame distortions. In this context, the problem is to find, for every encoded frame l, the optimum source and channel rates, $R_{s,l}$ and $R_{c,l}$ respectively, as well as the corresponding number of packets M_l to be transmitted. Additionally, for every frame, the optimal rate distribution between the source and channel coders at the level of every packet needs to be determined. Denote by $R_{s_i,l}, R_{c_i,l}$ the source and channel rates respectively used in packet i of the enhancement information of frame l. The JSCC-problem can then be defined as the minimization of the distortion

$$\left(\overline{D}_{tot} = \sum_{l=1}^{F} \overline{D_l}(R_{s,l}, R_{c,l})\right), \qquad (1)$$

under the constraint that the target rate is met:

$$R = \sum_{l=1}^{F}(R_{s,l} + R_{c,l}) = \sum_{l=1}^{F}(R_{s,l} + R_{c,l})$$
$$\Leftrightarrow R = \sum_{l=1}^{F}\left(\sum_{i=0}^{M_l} R_{s_i,l} + \sum_{i=0}^{M_l} R_{c_i,l}\right) \leq R_{tot} \qquad (2)$$

As the SVC reference decoder typically fails if the base layer is missing and no error concealment is performed, the base layer of each frame is by default protected with the highest protection level. The available channel rate must therefore be larger than the sum of protected base layer rates $R_{0,l}$ of each frame l: $R_{tot} > R_0 = \sum_{l=1}^{F} R_{0,l}$. The remaining available channel rate $R'_{tot} = R_{tot} - R_0$ must then be allocated by the JSCC-algorithm between the enhancement information of all frames and between the source and channel coders such that the overall distortion is minimized.

4. PROPOSED SOLUTION

4.1. Recursive Distortion Formula

Let $r_{l,i}$ define the code rate of channel packet $i, 0 \leq i \leq M_l$ of frame l, i.e. $r_{l,i} \triangleq K_{l,i}/N$, with $K_{l,i}$ the number of source bits and N the total number of bits in the channel packet. Denote by $p_f(r_{l,i},\varepsilon)$ the probability of losing channel packet i of frame l when transmitted over a BEC with parameter ε. The average expected distortion of frame l when transmitting M_l packets is then given by:

$$\overline{D_l}(R_{s,l}, R_{c,l}) = \sum_{m=0}^{M_l}\left(\prod_{i=0}^{m}(1 - p_f(r_{l,i},\varepsilon))\right) \cdot p_f(r_{l,m+1},\varepsilon) \cdot D_l\left(\sum_{i=0}^{m} R_{s_i,l}\right)$$

where $D_l(\sum_{i=0}^{m} R_{s_i,l})$ is the source distortion given that all packets up to packet m were received, and where $r_{l,0} = 0$, $p_f(r_{l,0},\varepsilon) = 0$, $p_f(r_{l,M_l+1},\varepsilon) = 1$ and $R_{s_0,l} = R_{c_0,l} = 0$.

In our transmission scenario, N is fixed. $D_l(\sum_{i=0}^{m} R_{s_i,l})$ and $\overline{D_l}(R_{s,l}, R_{c,l})$ can therefore equivalently be formulated as a function of the code rates $r_{l,i}$ used in packet i of frame l as: $D_l(\sum_{i=0}^{m} r_{l,i})$ and $\overline{D_l}(r_{l,0}, r_{l,1}, ..., r_{l,M_l})$ respectively.

Denote: $\alpha_{l,m} = \prod_{i=0}^{m}(1 - p_f(r_{l,i},\varepsilon))$ and $\widetilde{r_{l,k}} = \sum_{i=0}^{k} r_{l,m}$.

We can then write for $m, 0 \leq m \leq M_l$:

$$\alpha_{l,m+1} = \alpha_{l,m}(1 - p_f(r_{l,m+1},\varepsilon))$$
$$\Leftrightarrow \alpha_{l,m} p_f(r_{l,m+1},\varepsilon) = (\alpha_{l,m} - \alpha_{l,m+1}) \qquad (4)$$

Call the set of code rates $\left(r_{l,0}, r_{l,1}, ..., r_{l,M_l}\right)$ assigned to the M_l packets of frame l the path Π_{M_l}. A *recursive* formula for the average expected distortion $\overline{D_l}$ of a reconstructed frame when taking path Π_{M_l} can then be found as:

$$\overline{D_l}(\Pi_{M_l}) = \overline{D_l}(\Pi_{M_l-1}, r_{l,M_l}) =$$
$$\overline{D_l}(\Pi_{M_l-1}) + \alpha_{l,M_l-1}(1 - p_f(r_{l,M_l},\varepsilon))\left(D_l(\widetilde{r_{l,M_l}}) - D_l(\widetilde{r_{l,M_l-1}})\right) \qquad (5)$$

Equivalently, for a number of packets $k, 1 \le k \le M_l$ sent, one can recursively compute the average expected distortion as:

$$\overline{D_l}(\Pi_k) = \overline{D_l}(\Pi_{k-1}, r_{l,k}) =$$
$$\overline{D_l}(\Pi_{k-1}) + \alpha_{l,k-1}(1 - p_f(r_{l,k}, \varepsilon))\left(D_l(\widetilde{r_{l,k}}) - D_l(\widetilde{r_{l,k-1}})\right). \qquad (6)$$

The source rate-distortion (RD-)points of the SVC-stream can be found at each FGS-layer as in [6, 7], where the distortion of each quality level q of each frame i with dependency identifier d and temporal level t is calculated:

$$D(d,i,q) = (\log(D_{ind}(q+1)) - \log(D_{ind}(q))) +$$
$$(\log(D_{dep}(t+1)) - \log(D_{dep}(t)))$$

The first and second term express the distortion reductions resulting from decoding quality layer q and from decoding temporal layer t respectively. The SVC RD-points lie on a convex hull with monotonically decreasing slopes [6].

Using the above recursive formula two lemmas were proven in [8]. These lemmas show that the average expected distortion $\overline{D_l}(\Pi_{M_l})$ of a frame is always convex with monotonically decreasing slopes, no matter what path is taken with bounds on the allowable code rates when transmitting an increasing number of fixed-length packets given as: $r_{l,k} > \log_4 e/e \simeq 0.2654$ for all $k, 1 \le k \le M_l$.

5. NEAR-OPTIMAL JSCC-METHOD

A JSCC rate-allocation mechanism can now be derived. As can be seen from Figure 2, one cannot assume the existence of a single path which delivers the best protection for any number of transmitted packets of frame l.

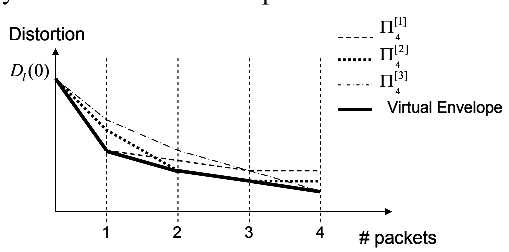

Figure 2: Example of three convex protection paths $\Pi_4^{[1]}, \Pi_4^{[2]}, \Pi_4^{[3]}$ and their convex virtual envelope hull.

We therefore propose to solve the global optimization problem by retaining from all constructed convex hulls, the paths which result in a minimal distortion at each subsequent packet (see Figure 2). With these virtual envelopes we can perform an optimal rate-allocation in between the frames by using a Lagrangian-base optimization method applied on the virtual envelopes computed for each frame l.

Because of the explosive growth of the search space when using an exhaustive search method to evaluate all the possible protection paths, we propose a simplified Viterbi-search method as follows. For each packet k, the average expected distortions of all paths $\Pi_k = \left(\Pi_{k-1}, r_{l,k}\right)$ are evaluated by using our recursive formula (6). From these paths, the path Π_k that delivers the smallest average expected distortion at each protection level $r_{l,k}$ of packet k is kept as a candidate for the calculations of the subsequent paths $\Pi_{k+1} = \left(\Pi_k, r_{l,k+1}\right)$ at packet $k+1$. Additionally, from all evaluated paths Π_k, the single path that results in the minimal expected distortion, defines the k^{th} point of the virtual envelope. This is done recursively until the source rate $R_{s,l}$ of frame l is exhausted and the virtual envelope of frame l is constructed. UEP is incorporated by imposing: $r_{l,k} \le r_{l,k+1}$, $\forall k, 0 \le k \le M_l$.

The total complexity in terms of the total number of paths to compute when transmitting M_l packets is then reduced to an order of $O(d^2.M_l)$. It was shown in [8] that this method results in very close-to-optimal results compared to an exhaustive search.

6. EXPERIMENTAL RESULTS

For our experiments we use the Joint Scalable Video Model reference software, JSVM v7.10, in scalable coding mode. We consider the transmission of the 30Hz SVC-encoded classical CIF-resolution “Bus” (150 frames) and “Footbal” (260 frames) test sequences in packets of exactly 256 bytes over BECs with 10% of bit erasures and several target bit-rates. The source coding (with a single resolution layer and three FGS quality enhancement layers and base layer QP set to 40, GOP size equal to 8, intra period equal to 16 and search range equal to 32) and the determination of the source RD-points using the quality level assigner [7] is performed once for each video sequence and stored. To measure the performance, we first compute the mean square error (MSE) over the whole sequence in Y, U and V for each transmission. For each transmission the PSNR in Y, U and V is then obtained as (with $X = Y, U, V$): $PSNR_X = 10\log_{10}\left(255^2 / MSE_X\right)$. These PSNR-values are finally averaged over all transmissions.

We repeated the transmissions 100 times. We do not consider the (small) amount of rate used by the first supplemental enhancement information message, the sequence parameter sets and the picture parameter sets in our algorithm and assume that this information reaches the decoder intact. For the Bus and Football sequences this information is about 400 bytes.

For the channel coding, we employ punctured regular (3,6)-LDPC codes. The performance of these codes was measured off-line and is given in Table 1. Iterative LDPC-decoding is allowed up to 100 iterations.

In Table 2, the average PSNR-values obtained for the transmission of protected SVC-streams over channels with 10% binary erasures when using a single code (equal error protection (EEP)) and using five channel codes (UEP) in our JSCC-algorithm are given. The single code used for EEP is perfectly matched with the BEC with 10% bit erasures such that no packet losses occur with this protection level. The strongest code used for UEP is slightly stronger than the single EEP code. As can be seen from the results in Table 2, choosing UEP or EEP has almost no influence on the average performance in terms of quality (the standard deviation on the average PSNR in Y, U and V when using UEP is always smaller than 0.03dB).

	N (bytes)	K (bytes)	P (bytes)	probability of packet loss
1 code (EEP)	392	196	136	0
5 codes (UEP)	388	194	132	0
	396	198	140	4.84E-05
	400	200	144	0.00121221
	404	202	148	0.0363238
	408	204	152	0.149072

Table 1: Average packet loss probability for the employed punctured regular (3,6) LDPC-codes when transmitted over a BEC with $\varepsilon = 10\%$ erasures. N: total number of bytes in the codeword. K: number of source bytes in the codeword. P: number of redundant bytes punctured from the codeword.

BUS 150 frames										
	1 code (EEP)					5 codes (UEP)				
Total target rate (kbps)	Total rate met (kbps)	Source rate (kbps)	Avg. PSNR Y (dB)	Avg. PSNR U (dB)	Avg. PSNR V (dB)	Total rate met (kbps)	Source rate (kbps)	Avg. PSNR Y (dB)	Avg. PSNR U (dB)	Avg. PSNR V (dB)
1000	999.83	765.50	31.52	39.82	41.70	999.83	767.17	31.52	39.81	41.70
1500	1499.96	1148.40	33.24	40.65	42.65	1499.96	1155.74	33.27	40.66	42.66
2000	1999.67	1531.00	34.69	41.90	43.91	1999.67	1540.50	34.69	41.90	43.90
2500	2499.79	1913.90	35.67	42.08	44.07	2499.79	1927.79	35.69	42.08	44.06
FOOTBALL 260 frames										
	1 code (EEP)					5 codes (UEP)				
1000	999.82	765.49	31.64	38.06	40.23	1000.05	766.97	31.64	38.07	40.23
1500	1499.84	1148.32	33.26	38.83	40.74	1500.08	1155.18	33.28	38.83	40.75
2000	1999.87	1531.15	35.09	40.40	42.09	1999.87	1540.24	35.11	40.43	42.11
2500	2499.90	1913.99	36.01	41.04	42.59	2499.90	1927.51	36.02	41.04	42.59

Table 2: PSNR-results of the proposed JSCC-methodology using our simplified Viterbi-search with EEP and with UEP for transmission over BECs with $\varepsilon = 10\%$ bit erasures.

When channel mismatches are present, we use the frame copy error concealment mode present in JSVM7.10 [7]. This error concealment mode only supports spatial scalability. Therefore, we extended the error concealment mode such that available FGS-enhancement layers are decoded as well. Due to space limitations, in Table 3, we only show three mismatched channels at one channel rate of 2500 kbps for bus and football. From Table 3 one notes that the SVC-stream protected with UEP brings graceful degradation when mismatches on the channel occur. Also, the SVC-stream protected with UEP can maintain good average PSNR-values over a broad range of mismatched channels, while the SVC-stream protected with EEP fails completely at a certain mismatch level.

7. CONCLUSIONS

In this paper we have proposed a novel Lagrangian-based JSCC-methodology with UEP that minimizes the end-to-end distortion when transmitting SVC bit-streams over BECs in packets of fixed-length. Additionally, we introduced a simplified Viterbi-based search method that presents reduced computational complexity compared with an exhaustive search. Experimental results show that equipping the JSCC-approach with UEP offers graceful quality degradation when the effective channel losses are mismatched to the expected channel losses.

	effective BEC (%)	Total rate met (kbps)	Source rate (kbps)	Y (dB)	U (dB)	V (dB)
	BUS 150 frames					
1code	11	2500	1914	35.46	42.06	44.03
1code	12	2500	1914	22.13	39.21	40.24
1code	12.5	2500	1914	15.43	35.36	35.75
5 codes	11	2500	1928	34.32	41.26	43.17
5 codes	12	2500	1928	29.48	38.52	40.11
5 codes	12.5	2500	1928	25.11	37.84	39.17
	FOOTBALL 260 FRAMES					
1code	11	2500	1914	35.56	40.80	42.49
1code	12	2500	1914	24.66	32.96	38.11
1code	12.5	2500	1914	19.16	27.59	33.85
5 codes	11	2500	1928	34.87	40.15	41.91
5 codes	12	2500	1928	30.27	36.68	39.17
5 codes	12.5	2500	1928	27.05	34.55	38.02

Table 3: Average performance of using EEP and UEP when the expected channel erasure parameter is $\varepsilon = 10\%$ while the effective channel parameter is $\varepsilon' = 11$, 12 and 12.5%.

8. ACKNOWLEDGMENTS

This research was funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (PhD grant M. Stoufs), the Fund for Scientific Research—Flanders (post-doctoral fellowships A. Munteanu and P. Schelkens) and by the European FP6 Project "Scalable, Ultra-fast and Interoperable Interactive Television (SUIT)".

9. REFERENCES

[1] T. Wiegand, G. J. Sullivan, J. Reichel, H. Schwarz, and M. Wien, Joint Draft 7 of SVC Amendment (revision 2), Klagenfurt, Austria, JVT-T201, 15–21 July, 2006.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, July 1948.

[3] M. K. Jubran, M. Bansal, R. Grover, and L. P. Kondi, "Optimal Bandwidth Allocation for Scalable H.264 Video Transmission Over MIMO Systems," in *Proc.* MILCOM2006, Washington, USA, vol., pp. 1-7, 23-25 October 2006.

[4] R. Hamzaoui, V. Stankovic, and Z. Xiong, "Optimized Error Protection of Scalable Image Bit Streams," *IEEE Signal Processing Magazine*, vol. 22, pp. 91-107, November 2005.

[5] R. Gallager, *Low-Density Parity-Check Codes*, Massachusetts Institute of Technology, PhD thesis, 1963

[6] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Optimized Rate-Distortion Extraction with Quality Layers," in *Proc.* International Conference on Image Processing, Atlanta, GA, USA, vol. ICIP2006, pp. 173-176, October 2006.

[7] J. Reichel, H. Schwarz, and M. Wien, Joint Scalable Video Model JSVM-6, Draft Output Document from JVT, JVT, Geneva, Switzerland JVT-S202, April 2006.

[8] A. Munteanu, M. R. Stoufs, J. Cornelis, and P. Schelkens, "Scalable and Channel-Adaptive Unequal Error Protection of Images with LDPC Codes," *Lecture Notes in Computer Science*, vol. 4179/2006, pp. 722-733, 2006.