Distributed video coding with decoder-driven skip

Jürgen Slowack, Jozef Škorupa, Stefaan Mys, Peter Lambert, Rik Van de Walle
Ghent University – IBBT
Multimedia Lab, ELIS Dept.
Gaston Crommenlaan 8 bus 201
B-9050 Ledeberg-Ghent, Belgium
{jurgen.slowack, jozef.skorupa, stefaan.mys, peter.lambert, rik.vandewalle}@ugent.be

Christos Grecos
School of Computing
University of the West of Scotland
Paisley, PA1 2BE, Scotland
Christos.Grecos@uws.ac.uk

ABSTRACT

In Distributed Video Coding (DVC), compression is achieved by exploiting correlation between frames at the decoder, instead of at the encoder. More specifically, the decoder uses already decoded frames to generate side information Y for each Wyner-Ziv frame X, and corrects errors in Y using error correcting information received from the encoder. Typically, this error correcting information is requested using a feedback mechanism in which the decoder requests portions of bits from the encoder until the frame is reliably decoded. In this paper we propose new techniques for deciding when a frame is reliably decoded, based on estimated decoded quality. We define criteria both for the coefficient level and the bitplane level that will allow us to skip coefficient bands and bitplanes. If bitplane decoding is nonetheless needed, the bitplane-level criterion allows early termination of the decoding process, by focusing only on a subset of the bits instead of all. Results indicate significant improvements especially for sequences that can be well predicted at the decoder side.

Categories and Subject Descriptors

I.4.2 [**Image Processing and Computer Vision**]: Compression (Coding)

General Terms

Algorithms

Keywords

Distributed Video Coding, Wyner-Ziv coding

1. INTRODUCTION

In typical video compression systems, it is the encoder that exploits the correlation between frames by motion estimation, a task making the encoder significantly more complex than the decoder. This asymmetry is well-suited for applications where video is coded only once but decoded many times, or in streaming scenarios where the encoder has much more computational resources available than the decoder. Recently, on the other hand, Distributed Video Coding (DVC) systems have been developed. In DVC, the correlation between frames is exploited by the decoder instead of by the encoder. Such systems – characterized by lightweight encoding and complex decoding – are better suited for applications such as networked camcorders, wireless video cameras, and visual sensor networks [7].

The DVC codec in this paper is based on the widely-used transform-domain Stanford architecture [2]. The frame sequence is split up into key frames and Wyner-Ziv (WZ) frames. Key frames are coded without using other frames as reference, i.e., by applying H.264/AVC intra coding. For each WZ frame, the decoder generates a prediction Y (called side information) by following the techniques used in the DISCOVER codec [3]. Since Y is merely an approximation of the original frame X available at the encoder, errors in Y are corrected using channel coding techniques. To this extent, X is transformed using a block-based DCT, and coefficients are grouped into coefficient bands C_i, where C_0 represents the set of DC coefficients in X, for example. Next, bits at corresponding positions in C_i are grouped into bitplanes B_i^j, where B_1^0 contains the most significant bits of the first transformation coefficient, for example. These bitplanes are fed as input to a turbo coder, which calculates parity bits and sends them in portions to the turbo decoder, upon request. The correlation noise between X and Y – used by the turbo decoder and the reconstruction process – is estimated online [9].

Different criterions are used for deciding at the decoder-side when to stop requesting for more parity bits. We compared several criterions in our previous work – including the one used in the DISCOVER codec [5] – and showed that the sign-difference ratio has best performance [8].

In this paper we propose a new technique by introducing a coefficient-level criterion as well as a selective bitplane-level criterion. One of the effects of this technique is that more insignificant coefficient bands and bitplanes are skipped. Firstly, we provide details on how the coefficient bands and bitplanes are used as input to the turbo coder (Sect. 2). Afterwards, we indicate how we can use the information from the turbo decoding process to develop techniques for

Mobimedia'09, September 7-9, 2009, London, UK.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7343
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7343

decoder-side driven skipping of coefficient bands (Sect. 3.1) and skipping/stopping bitplane decoding (Sect. 3.2). The results (Sect. 4) indicate that our system performs well compared to our previous work and compared to the DISCOVER codec.

2. CALCULATING PROBABILITIES FOR EACH BIT

First we will provide some context. Denote X as a random variable representing the value of a transformation coefficient in the original frame, and the random variable Y to indicate its corresponding estimate generated at the decoder. For efficient channel decoding, the decoder estimates the reliability of the virtual channel $f_{X|Y}$. For this we will use a Laplace distribution (as in most related work) of which the variance is calculated online, for each coefficient [9]. This distribution is then used to generate probabilities p_i for each quantization bin q_i, indicating the likeliness that X lies in the bin q_i:

$$p_i = \int_{q_i^L}^{q_i^H} f_{X|Y}(x|y)dx, \tag{1}$$

where q_i^L is the lower border of the bin, and q_i^H the upper border. Next, probabilities for the bits in each bitplane are generated, for example, the probability that a particular bit equals one (given the side info) is given by[1]:

$$b = \frac{\sum_{q_i \in \psi_1} p_i}{\sum_i p_i}, \tag{2}$$

where ψ_1 is the set of quantization bins that have bit one in the particular bitplane under consideration.

The turbo decoder operates on so-called log ratios:

$$L = \ln\left(\frac{b}{1-b}\right), \tag{3}$$

and updates these log ratios in an iterative fashion. Turbo decoding stops when a certain criterion is satisfied, or if the maximum number of iterations is achieved (in that case more parity bits are requested).

3. DECODER-SIDE SKIP

The key idea is that we want the turbo decoder to spend just enough parity bits so that the decoded frame is accurate enough. In contrast to the usual case where a decoder loop is used at the encoder (e.g. H.264/AVC), in DVC it is impossible to measure the quality of the output because the original frame is only available at the encoder while the decoded frame is only available at the decoder. However, it is possible to *estimate* the quality of the decoded frame. This idea plays a key role in our technique, and it will allow us to skip or stop decoding of coefficient bands and bitplanes if the quality of the decoded output is assumed to be high enough.

To estimate the quality of the output X' at a certain stage, we construct X' as follows. By rewriting Eq. 3, the update of L to L' can be formulated as an update of the bit probabilities to b':

$$b' = \frac{e^{L'}}{1+e^{L'}}. \tag{4}$$

In turn, we can interpret this as a proportional update of the bin probabilities:

$$p_i' = \begin{cases} p_i \cdot \frac{b'}{b} & , q_i \in \psi_1 \\ p_i \cdot \frac{1-b'}{1-b} & , \text{else} \end{cases} \tag{5}$$

Taking this even further, we can interpret this as a proportional update of the estimated virtual channel to:

$$f'_{X|Y}(x|y) = \frac{p_i'}{p_i} \cdot f_{X|Y}(x|y), x \in q_i. \tag{6}$$

The latter is illustrated in an example in Fig. 1.

As such, X' is constructed as the centroid of $f'_{X|Y}(x|y)$, or:

$$\begin{aligned} X' &= \frac{\sum_i \int_{q_i^L}^{q_i^H} x \cdot f'_{X|Y}(x|y)dx}{\sum_i p_i'}, \\ &= \frac{\sum_i \frac{p_i'}{p_i} \cdot \int_{q_i^L}^{q_i^H} x \cdot f_{X|Y}(x|y)dx}{\sum_i p_i'}. \end{aligned} \tag{7}$$

This formula uses information from all bins to enable generating a decoded version X' for X at any stage desired during the decoding process. This is in contrast to current techniques where reconstruction is performed after channel decoding has completed, assuming that only one quantization bin is correct [6]. By generating the reconstruction during the turbo decoding process, we can decide if more parity bits are needed from the encoder, based on the estimated quality of X'.

3.1 Skipping coefficient band decoding

Coefficient-level wise, the goal is to skip/stop decoding if all coefficients in the coefficient band are assumed to be accurate enough, i.e. if the probability is high that the error $|E| = |X - X'|$ is sufficiently small (when using absolute difference as an error metric). This is expressed as the criterion: $Prob(|E| < W) > T_c$, where T_c is the coefficient-level confidence threshold, and W is a threshold defining the desired accuracy of the decoded coefficients. In this paper we set W equal to half the WZ quantization bin width.

The probability is calculated as:
$Prob(|E| < W)$

$$\begin{aligned} &= \frac{\sum_i \int_{\max(q_i^L, X'-W)}^{\min(q_i^H, X'+W)} f'_{X|Y}(x|y)dx}{\sum_i p_i'}, \\ &= \frac{\sum_i \frac{p_i'}{p_i} \cdot \int_{\max(q_i^L, X'-W)}^{\min(q_i^H, X'+W)} f_{X|Y}(x|y)dx}{\sum_i p_i'}. \end{aligned} \tag{8}$$

Denote the set of all coefficients in this coefficient band as Ω. In this set, a subset Ω_C is already accurate while the remaining part $\Omega_{\bar{C}}$ is not accurate. Hence, the entire coefficient band can be skipped if $\Omega_{\bar{C}} = \phi$. Otherwise, we decode the first bitplane as described in the following section. After bitplane decoding, the updated L-values are used to improve the channel model and to calculate X' (Eq. 7). Next, the coefficient-level criterion (Eq. 8) is re-evaluated to update the sets Ω_C and $\Omega_{\bar{C}}$. If the current reconstructed values are

---

[1] This and following formulas are normalized to take into account the loss of the tails of the Laplace distribution used for modeling the virtual noise.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7343
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7343

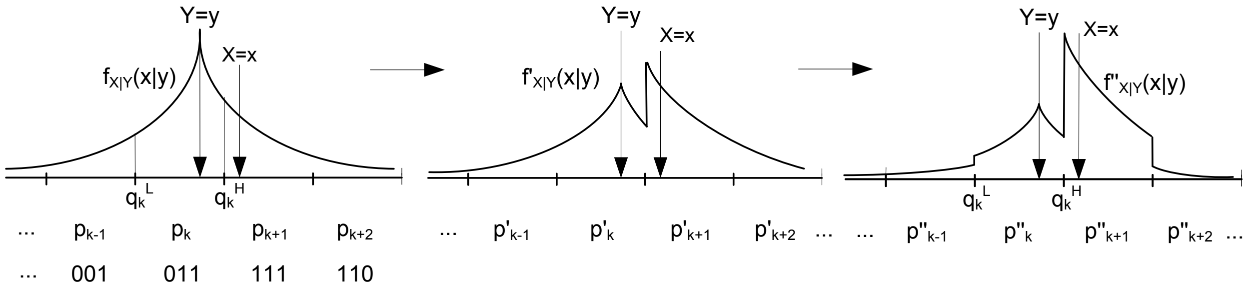

Figure 1: Example of a particular side info coefficient y (at the decoder), original coefficient x (at the encoder), and decoder-side estimated distribution $f_{X|Y}$. The figure on the left illustrates a possible labeling strategy for the quantization bins. After some iterations during the decoding of the first bitplane, the virtual channel model is adjusted, as illustrated by the center figure. The figure on the right illustrates a possible result after decoding several bitplanes.

accurate enough ($\Omega_{\bar{C}} = \phi$), coefficient band decoding is terminated. Otherwise, the following bitplane is decoded and so on.

3.2 Skipping/stopping bitplane decoding

In a DVC system, the residual between X and Y can not be coded since they are not both available at encoder or decoder. As a result, there exist cases where X and Y are close to each other (i.e., $|X - Y| < W$), even though they are in different quantization bins. Hence, the elements in Ω_C may still contain bit errors. However, it is not our goal to improve the accuracy of these elements even further so these errors should be ignored as much as possible.

Therefore, we introduce a bitplane-level criterion that is only tested on a limited set of bits, instead of testing it on all bits in the bitplane, as in current techniques. This bitplane-level criterion is given by:

$$b' < 1 - T_b \text{ or } b' > T_b, \tag{9}$$

where T_b is the (bitplane-level) confidence threshold.

Before starting to decode, the decoder determines if the bitplane can be skipped by evaluating the bitplane-level criterion only for the elements in $\Omega_{\bar{C}}$. If the criterion is valid for each element in $\Omega_{\bar{C}}$, no parity bits are requested from the encoder and bitplane decoding is skipped. Otherwise, turbo decoding is performed. After each iteration of the turbo decoder, the bitplane-level criterion is re-evaluated to determine if turbo decoding should be stopped. This time we also include the elements in Ω_C for which the bitplane-level criterion was satisfied initially (i.e. before turbo decoding). This is done for coping with inaccuracies in the virtual channel model. More specifically, if the bitplane-level criterion was initially valid then it is most likely that it will remain valid after several turbo decoding iterations, and if so, there is no difference in whether the elements were included or not. However, if the criterion does not remain valid, then this points out that the virtual channel model is not accurate and so in that case we prefer to correct this mistake.

If the maximum number of turbo decoding iterations is achieved without the bitplane-level criterion being valid for all relevant bits, more parity bits are requested from the encoder through the feedback channel and turbo decoding is restarted. Otherwise, turbo decoding is assumed to be successful and the current L-values are used by the coefficient-level solution to generate a decoded version for each coefficient (Eq. 7). Next, based on the current estimated quality, a decision is made to decode more bitplanes or not, as described in the previous section.

4. RESULTS

Tests have been performed on two different sequences ("Mother and Daughter" and "Foreman") at CIF resolution. For each sequence, 297 frames have been coded, i.e. 75 GOP's (three WZ frames per GOP) and one closing frame, at a frame rate of 30 frames per second. Only the luma component is coded (to allow comparing with DISCOVER). The quantization of intra frames and WZ frames is chosen so that both have similar quality.

Two variants of our system are considered. The first implements the sign-difference ratio as a turbo decoder stopping criterion [8], and conventional reconstruction of the transformation coefficients after turbo decoding [6]. The second variant implements the techniques proposed in this paper, with the thresholds T_c and T_b both set to 0.9 (experimentally obtained).

As a reference, we also compare our system to the DISCOVER codec, for which binaries are available online [1]. Tests have been performed in a similar setting, using quantization patterns 1, 3, 6, and 7.

The results in Fig. 2 and Fig. 3 indicate that our new technique outperforms the technique using the sign difference ratio. The gain is the largest for Mother and Daughter, where an average bit rate improvement of 19% is observed, compared to 4% for Foreman (Bjøntegaard delta rate [4]). This difference is explained by the fact that there is simply more to skip if the sequence can be more accurately predicted by the decoder. In addition, inaccuracies in the modeling of the virtual channel sometimes result in coefficient bands and bitplanes that are skipped while they should not be skipped and vice versa, which decreases the total gain achieved for sequences with more complex motion characteristics. Another observation is that the gain increases as quantization becomes coarser (i.e. for decreasing bit rates). This is explained by the fact that accuracy has been defined relative to the WZ quantization bin width.

We outperform the DISCOVER codec with rate gains of 7% for Foreman and 26% for Mother and Daughter (Bjøn-

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7343
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7343

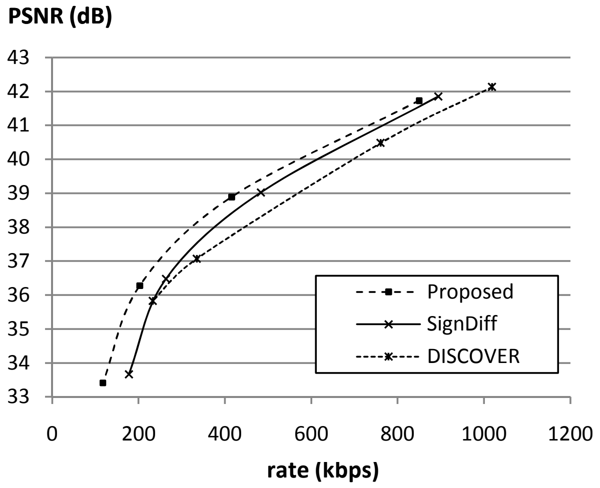

Figure 2: Results for Mother and Daughter, for the proposed technique, the system using the sign-difference ratio [8], and the DISCOVER codec [1].

tegaard delta). These gains can be explained by the use of a better correlation noise model [9] as well as the new techniques proposed in this paper.

5. CONCLUSIONS

In this paper we proposed a technique to decide at the decoder side when to skip information or stop decoding, based on the estimated quality of the current solution. The strength of our technique depends on the quality of the side information generation and the modeling of the virtual channel. Hence, by improving these techniques, even more gain should be achieved.

6. ACKNOWLEDGMENTS

The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT-Flanders), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

7. REFERENCES

[1] http://www.discoverdvc.org.

[2] A. Aaron, S. Rane, E. Setton, and B. Girod. Transform-domain Wyner-Ziv codec for video. In *Proc. SPIE Visual Communications and Image Processing*, volume 5308, pages 520–528, January 2004.

[3] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret. The DISCOVER codec: Architecture, techniques and evaluation. In *Picture Coding Symposium*, November 2007.

[4] G. Bjøntegaard. Calculation of average PSNR differences between RD-curves. Technical report, VCEG, April 2002. Contribution VCEG-M33.

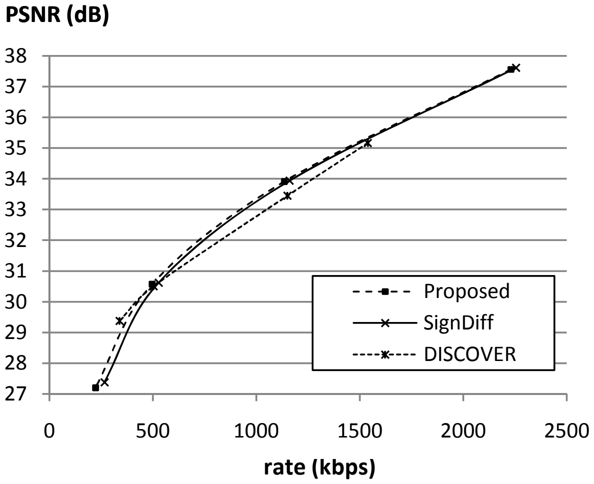

Figure 3: Results for Foreman, for the proposed technique, the system using the sign-difference ratio [8], and the DISCOVER codec [1].

[5] D. Kubasov, K. Lajnef, and C. Guillemot. A hybrid encoder/decoder rate control for a Wyner-Ziv video codec with a feedback channel. In *IEEE MultiMedia Signal Processing Workshop*, pages 251–254, October 2007.

[6] D. Kubasov, J. Nayak, and C. Guillemot. Optimal reconstruction in Wyner-Ziv video coding with multiple side information. In *IEEE MultiMedia Signal Processing Workshop*, pages 183–186, October 2007.

[7] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp. Distibuted video coding: Selecting the most promising application scenarios. *Signal Processing: Image Communication*, pages 339–352, 2008.

[8] J. Škorupa, J. Slowack, S. Mys, P. Lambert, C. Grecos, and R. Van de Walle. Stopping criterions for turbo coding in a Wyner-Ziv video codec. In *Proc. Picture Coding Symposium (PCS)*, May 2009.

[9] J. Škorupa, J. Slowack, S. Mys, P. Lambert, and R. Van de Walle. Accurate correlation modeling for transform-domain Wyner-Ziv video coding. In *Proc. Pacific-Rim Conference on Multimedia (PCM)*, pages 1–10, December 2008.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7343
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7343