A CONTEXT MODELING SCHEME FOR CODING OF TEXTURE REFINEMENT INFORMATION

Heiner Kirchhoffer, Detlev Marpe, and Thomas Wiegand

Image Communication Group, Image Processing Department
Fraunhofer Institute for Telecommunications - Heinrich Hertz Institute
Einsteinufer 37, D-10587 Berlin, Germany, [kirchhof|marpe|wiegand]@hhi.fraunhofer.de

ABSTRACT

Fidelity scalability involves the refinement of residual texture information. The entropy coding of texture refinement information in the scalable video coding (SVC) extension of H.264/AVC relies on a simple statistical model that is tuned to an encoder-specific way of quantization for generating a single fidelity enhancement layer on top of the backward compatible base layer. For fidelity enhancement layers above the first layer, we demonstrate how and why the current model fails to properly reflect the statistics of texture refinement information. By analyzing the specific properties of the typical quantization process in fidelity scalable coding of SVC, we are able to derive a generic modeling approach for coding of refinement symbols, independent of the specific choice of dead-zone parameters and classification rules. Experimental results for a broad range of quantization parameters show averaged bit-rate savings of around 5% (relative to the total bit rate) by using our proposed context modeling approach for a representative set of video sequences in a test scenario including up to three fidelity enhancement layers.

Index Terms— H.264, AVC, SVC, entropy coding, fidelity scalability

1. INTRODUCTION

Uniform-reconstruction quantizers (URQ) are specified in H.264/AVC [1]. Since the SVC extension [2] of H.264/AVC is designed in a way that the base layer of a scalable bit-stream is conforming to H.264/AVC, fidelity scalability has to enhance a base layer that is quantized by a uniform-reconstruction quantizer. Starting from this, a straightforward way to generate a fidelity enhancement layer is to subtract the (coarse quantized) residual texture signal of the base layer from the original residual texture signal and to quantize this difference with a smaller step-size. Then, the reconstruction is the sum of the base layer signal and the fidelity enhancement layer signal. For additional fidelity enhancement layers, the base layer and all prior generated fidelity enhancement layers are subtracted from the original to calculate the difference to be quantized. In principle, this procedure can be repeated until the desired number of fidelity enhancement layers is achieved. This is depicted in Fig. 1 for one base layer and three fidelity enhancement layers. However, it is not a priori clear if this scheme of fidelity scalability is a good choice in terms of rate-distortion (R-D) performance.

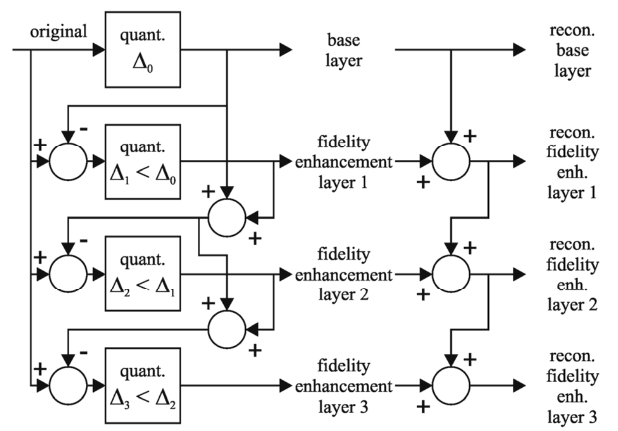

Fig. 1: Block diagram for fidelity scalability with re-quantization.

As will be pointed out in this present study, R-D performance of fidelity scalability critically depends not only on a suitable choice of encoding rules for the uniform-reconstruction quantizer but also on a carefully designed related context modeling scheme for subsequent entropy coding of quantizer levels in each fidelity enhancement layer. In the current SVC reference encoder design, the URQ encoding rule involves a so-called *dead-zone plus uniform threshold quantization* (DZ-UTQ) approach [3] for generation of fidelity enhancement layers. When analyzing the effective decision thresholds for different choices of dead-zone parameters as applied to the quantization of fidelity enhancement layers, it turns out that the resulting level information related to different fidelity enhancement layers are highly correlated with each other and with the corresponding base layer level information. Based on this observation, a generic context model is derived that exploits those statistical dependencies improving the R-D performance, as shown by experimental results.

This paper is organized as follows. The next section gives a brief introduction to the basic principles of fidelity scalable coding and points out some problems related to its realization in the context of SVC. Section 3 introduces the proposed generic context modeling scheme for coding of refinement symbols, and in Section 4, we present some experimental results demonstrating the effectiveness of the approach.

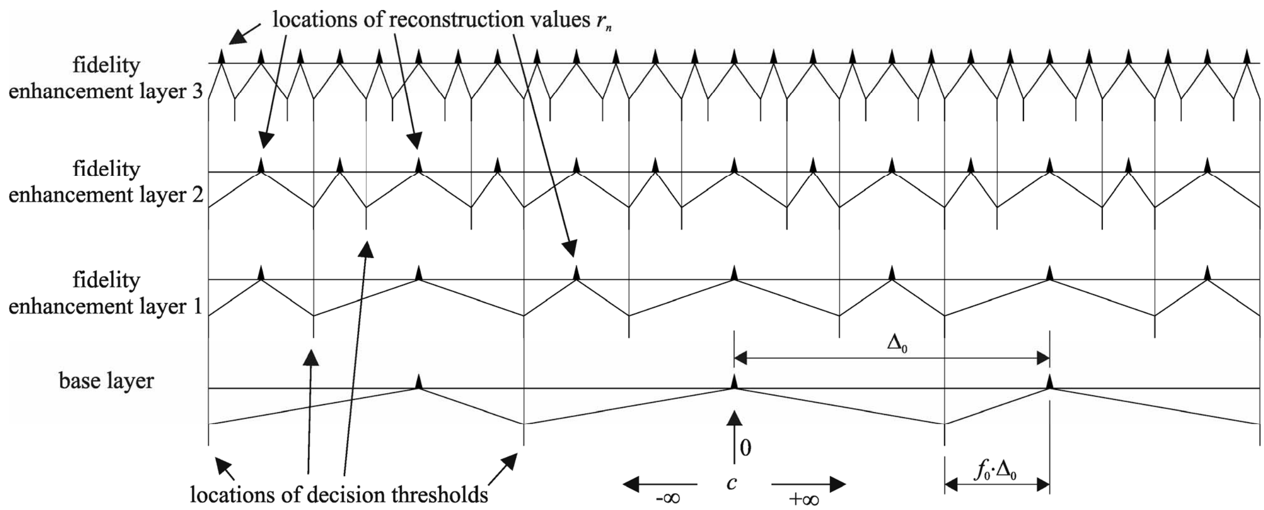

Fig. 2: Decision intervals and reconstruction values resulting from the DZ-UTQ quantization for f_n=1/3 and with halved step-size from one layer to another.

2. BACKGROUND AND PROBLEM STATEMENT

In the current SVC design, quantization of residual texture information involves a DZ-UTQ approach. First, the base layer is generated according to eq. (1). f_0 denotes the so-called dead-zone parameter with $0 \le f_0 \le 0.5$, Δ_0 denotes the quantization step-size, and c denotes an original residual texture value that is quantized to a so-called refinement level index c_0.

$$c_0 = \operatorname{sgn}(c) \cdot \left\lfloor \frac{|c|}{\Delta_0} + f_0 \right\rfloor . \tag{1}$$

To calculate a quantized (reconstruction) value r_0 from a refinement level index c_0, a simple multiplication with the step-size Δ_0 is performed according to following equation:

$$r_0 = c_0 \cdot \Delta_0 . \tag{2}$$

Quantization of fidelity enhancement layers is carried out by recursively applying eq. (3), where Δ_n denotes the step-size of fidelity enhancement layer n satisfying $\Delta_n < \Delta_{n-1}$, f_n is the dead-zone parameter of fidelity enhancement layer n, and c_n is the refinement level index of layer n:

$$c_n = \operatorname{sgn}(c - r_{n-1}) \cdot \left\lfloor \frac{|c - r_{n-1}|}{\Delta_n} + f_n \right\rfloor . \tag{3}$$

Accordingly, reconstruction of fidelity enhancement layers is performed by recursively applying eq. (4) with r_n denoting a reconstruction value of fidelity enhancement layer n.

$$r_n = r_{n-1} + c_n \cdot \Delta_n . \tag{4}$$

The width of the so-called dead-zone of each layer (that is the center interval mapped to reconstruction values r_n being equal to zero) is controlled by the dead-zone parameter f_n, which may vary from layer to layer. Although the step-sizes for generating different fidelity enhancement layers may be chosen arbitrarily (as long as $\Delta_n < \Delta_{n-1}$ is fulfilled), it makes sense to halve the step size from one layer to the next layer ($\Delta_n = \Delta_{n-1}/2$). In this way it is ensured that the reconstruction values stay equidistant (with the step-size of the current fidelity enhancement layer being the distance between two neighboring values) and thus represent a uniform-reconstruction system with the useful property of a low-complexity reconstruction rule.

Fig. 2 depicts the quantization scheme described by eqs. (1)-(4) with the choice of f_n = 1/3 and $\Delta_n = \Delta_{n-1}/2$. The locations of the reconstruction values r_n for the different layers are shown as small solid triangles on the horizontal lines, where each such line indicates a different layer starting from the base layer at the bottom, up to the third fidelity enhancement layer on top of the graph. Vertical lines in Fig. 2 illustrate the decision thresholds corresponding to the quantization rule of eq. (1) and (3). Note that the right-hand side base layer interval (and the way of subdividing it in all fidelity enhancement layers) is periodically repeated to the right, as well as the left-hand side interval subdivision is periodically repeated to the left.

A closer inspection of the location of decision thresholds involved in the quantization of two consecutive fidelity enhancement layers reveals that there are two types of decision intervals. As can be seen from Fig. 2, there is one type of decision interval whose interval width keeps constant and another type of decision interval that gets subdivided into three sub-intervals. In other words, for reconstruction values r_n belonging to the first type (and indicated by the smaller intervals), the reconstruction values r_{n+1} of the next fidelity enhancement layer will all stay the same ($r_{n+1} = r_n$). Conversely, for each reconstruction value r_n belonging to a larger interval, the corresponding reconstruction value r_{n+1} of the next fidelity enhancement layer will be one out of the three values r_n, $r_n + \Delta_{n+1}$, and $r_n - \Delta_{n+1}$. In fact, for the special case of a fixed choice of f_n = 1/3, Fig. 2 shows that two types of intervals are alternating and therefore, it would be possible to locate those intervals that keep constant (as well as the related quantized values) at the decoder side and avoid signaling any refinement level indices c_n for the corresponding quantized values. Even if we fix the ratio $\Delta_n = \Delta_{n-1}/2$, because the choice of the dead-zone parameters $f_0 \ldots f_n$ at the encoder is not known to the decoder and because arbitrary choices are possible, it is not advisable to establish a solution that is tailored to specific encoder settings.

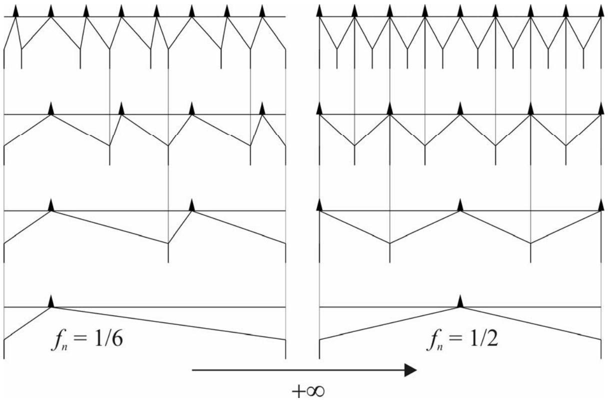

Fig. 3: Characteristic schemes for $f_n = 1/6$ on the left and for $f_n = 1/2$ on the right (showing the base layer and 3 fidelity enhancement layers).

To explore other configurations of f_n, it is sufficient to only investigate the subdivision of one base layer interval (except the center interval) in all fidelity enhancement layers. This is because the scheme has mirror symmetry with the axis of symmetry at $c = 0$ *and* because the right half of the center base layer interval is identical to that part of the right-hand side base layer interval that ranges from its reconstruction value to its right-hand side interval boundary. For this reason, the right-hand side base layer interval and its subdivision in all fidelity enhancement layers is denoted as *characteristic scheme* for the remainder of this paper.

Fig. 3 depicts two characteristic schemes, one for $f_n = 1/2$ and one for $f_n = 1/6$ (both with $\Delta_n = \Delta_{n-1}/2$). Similar to the case of $f_n = 1/3$, as shown in Fig. 2, different types of decision intervals can be distinguished depending on how each of the intervals gets subdivided from one layer to another. However, in contrast to the case $f_n = 1/3$, the number of types or classes of decision intervals differs from fidelity enhancement layer to fidelity enhancement layer as well as from one choice of f_n to another.

3. GENERIC CONTEXT MODELING APPROACH

For encoding a certain refinement level index c_n, it is desirable to use the knowledge about the behavior of the quantization scheme described by eqs. (1)-(4) to model the probability of c_n reproducing a certain value as accurate as possible. But since we do not want to make any assumption about the probability distribution of the residual texture signal, our context modeling scheme is based on the following partitioning principle:

***Partitioning Principle**: Whenever the subdivision of a certain quantization interval **A** leads to different sub-intervals compared to the subdivision of another quantization interval **B**, whether in terms of sub-interval size or relative location of corresponding reconstruction values, then, for these two types of intervals **A** and **B**, distinct probability models, i.e., contexts should be used for encoding of their corresponding refinement level indices* c_n.

This partitioning principle follows from the assumption that intervals of differing sizes usually lead to differing relative probabilities of refinement levels c_n. To stay independent of the encoder's choice of f_n, it is therefore necessary to assume that the widths of all intervals inside the characteristic scheme may differ from each other. Consequently, for each sub-interval a separate probability model has to be maintained for the encoding of its corresponding level values c_n, regardless whether some of the intervals are subdivided in the same way and thus could share one probability model.

An easy way to discriminate between the sub-intervals is the usage of a tree structure, where the base layer interval of the characteristic scheme is given as the root and where each node represents one possible sub-interval. For each value that the refinement level indices c_n of a certain (sub-)interval can assume, one branch is added to the corresponding node, where each of these branches leads to a new node representing a sub-interval at the next fidelity enhancement layer n+1.

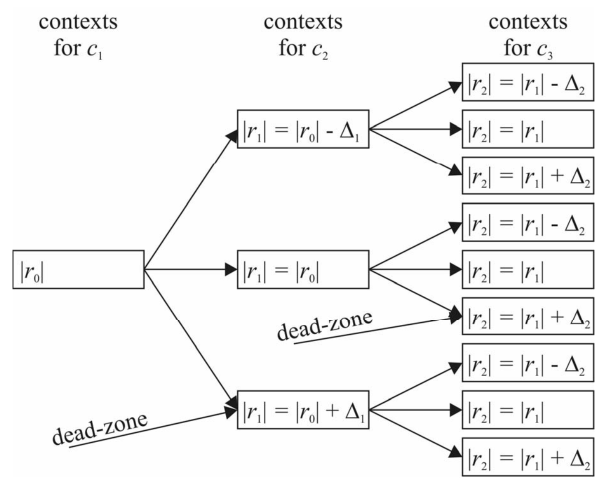

Fig. 4: Context tree for arbitrary f_n configurations.

This concept is depicted in Fig. 4 with three branches per node (or sub-intervals per refinement index). Each box represents one node and the arrows represent the branches. The number of three branches per node is not just an arbitrary choice, it rather follows from the two conditions $\Delta_n = \Delta_{n-1}/2$ and $0 \le f_n \le 0.5$. Because of this, a tree as depicted in Fig. 4 is always sufficient to implement the presented generic modeling approach, although some of the nodes may not be in use for particular choices of f_n.

As illustrated in Fig. 2, all types of decision intervals can be found within the characteristic scheme except for the dead-zone interval of each layer. Therefore, a separate probability model is used for each fidelity enhancement layer to encode refinement level indices resulting from the dead-zone interval. Conversely, the two outer sub-intervals that are generated, whenever a dead-zone interval is subdivided, can also be found in the characteristic scheme and thus can be taken into account for modeling as depicted in Fig. 4 by the arrows named 'dead-zone'.

As shown in the characteristic scheme for the special choice of $f_n = 1/2$, the left and right interval at the first fidelity enhancement layer are bisected exactly in the middle (cp. Fig. 3). Because of this, it seems to be unnecessary to maintain different contexts for these two intervals. But note that the reconstruction value of the left interval (in fidelity enhancement layer 1) lies on the interval's left boundary, whereas the reconstruction value of the right interval lies on the interval's right boundary. This means that the refinement level indices c_2 for the left interval can only take the values 0 and +1, while the refinement level indices c_2 for the right interval can only take the values 0 and -1. In other words, the relative probability $p(c_2\text{=-}1)$ is equal to 0 for the left interval, whereas the corresponding $p(c_2\text{=-}1)$ for the right interval is usually non-vanishing. That is exactly why the usage of different context models for the left and right interval, even in this particular case of having equal sized decision intervals, is required according to our partitioning principle.

4. EXPERIMENTAL RESULTS

The generic modeling approach, as described in the previous section, has been implemented into the SVC reference software. The details of this implementation can be found in [4]. The anchor which has been used in our simulations corresponds to so-called "progressive refinement slices" [2] which were developed during the standardization process of SVC and which provide a functionality known as fine grain scalability – FGS. To evaluate the benefit in terms of coding performance, this anchor version has been used together with a degenerated context tree consisting only of the root node. Thus, the anchor configuration uses only one context per fidelity enhancement layer for all refinement level indices c_n from non dead-zone intervals, whereas the proposed approach uses one context per node in the full context tree for the same kind of symbols, as depicted in Fig. 4. In addition, both configurations conceptually use one distinct context per fidelity enhancement layer for coding of refinement level indices c_n stemming from the dead-zone interval of the previous layer.

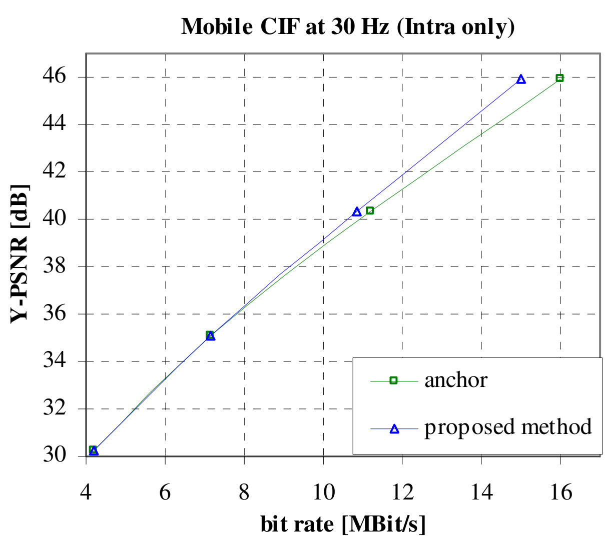

Fig. 5: R-D plot for the sequence 'Mobile', 3 fidelity enhancement layers, $f_n = 1/3$.

In Fig. 5, a sample R-D curve is depicted for each configuration. The data points at the lowest bit rate correspond to the base layer and each of the other points represents one fidelity enhancement layer. As expected, the data point of the first fidelity enhancement layer of the anchor curve is identical to the corresponding point of the curve belonging to the generic approach. This is because for the first fidelity enhancement layer, our proposed approach is identical to the anchor configuration. However, for higher rate points corresponding to the second and third fidelity enhancement layer, quite considerable R-D gains can be observed for our proposed approach.

Tables 1 and 2 show averaged bit-rate savings that have been achieved by our proposed modeling strategy relative to the anchor. The results were obtained for the SVC test set in CIF resolution with the encoder configured to use intra coding (Table 1) or a group of 15 hierarchical B-frames (GOP16), as shown in Table 2. For encoding of the base layer, the quantization parameter was set equal to 34 which results in base layer PSNR values of around 30-35 dB. Each entry in Tables 1 and 2 represents the percentage of bit-rate reduction associated with the corresponding layer only (averaged over all sequences of the chosen test set). In this way, each layer is evaluated independently, excluding the rate savings of all subordinate layers. As can be seen from Tables 1 and 2, the measured coding gains for intra coding are higher than those for inter coding (GOP16). This is due to the fact that, as a result of the efficacy of motion-compensated prediction, the amount of bits spent for residual texture information in inter coding is usually considerably smaller than in the case of pure intra coding.

fidelity enhancement layer no.	f_n **= 1/2**	f_n **= 1/3**	f_n **= 1/6**
2	7.45 %	6.31 %	0.15 %
3	12.87 %	9.78 %	2.53 %

Table 1: Average bit-rate savings for the second and third enhancement layer (SVC test set, CIF, Intra only).

fidelity enhancement layer no.	f_n **= 1/2**	f_n **= 1/3**	f_n **= 1/6**
2	2.24 %	2.52 %	0.52 %
3	4.43 %	3.69 %	0.77 %

Table 2: Average bit-rate savings for the second and third enhancement layer (SVC test set, CIF, GOP 16).

5. CONCLUSION

We have presented a generic context modeling scheme for URQ-based fidelity-scalable representations of residual texture data. With the presented approach it is possible to exploit structural dependencies due to the recursively applied quantization process. Experimental coding results for various encoder configurations have shown the anticipated increase in coding efficiency following from our analysis of DZ-UTQ quantization.

6. REFERENCES

[1] *ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC*, "Advanced Video Coding for Generic Audiovisual Services," Version 3, 2005.

[2] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, M. Wien, "Joint Draft 8 of SVC Amendment," Joint Video Team Doc. JVT-U201, Hangzhou, China, October 2006.

[3] G. Sullivan, "On embedded scalar quantization," Proceedings of IEEE ICASSP, vol. IV, pp. 605-808, May 2004.

[4] H. Kirchhoffer, D. Marpe, and T. Wiegand, "CE3: Improved CABAC for PR slices," Joint Video Team Doc. JVT-U082, Hangzhou, China, October 2006.