

A GENERIC CONTEXT MODEL FOR UNIFORM-RECONSTRUCTION BASED SNR-SCALABLE REPRESENTATIONS OF RESIDUAL TEXTURE SIGNALS

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

SNR scalability involves 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 SNR layer on top of the backward compatible base layer. For SNR 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 SNR 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 show bit rate savings of around 5% relative to the total bit rate and averaged over a representative set of video sequences in a test scenario including up to three SNR layers.

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, SNR scalability has to enhance a base layer that is quantized by a uniform-reconstruction quantizer. Starting from this, a straightforward way to generate an SNR 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 SNR layer signal. For additional SNR layers, the base layer and all prior generated SNR 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 SNR layers is achieved. This is depicted in Figure 1 for one base layer and three SNR layers. However, it is not a priori clear if this scheme of SNR scalability is a good choice in terms of rate-distortion (R-D) performance.

As will be pointed out in this present study, R-D performance of SNR 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 SNR layer.

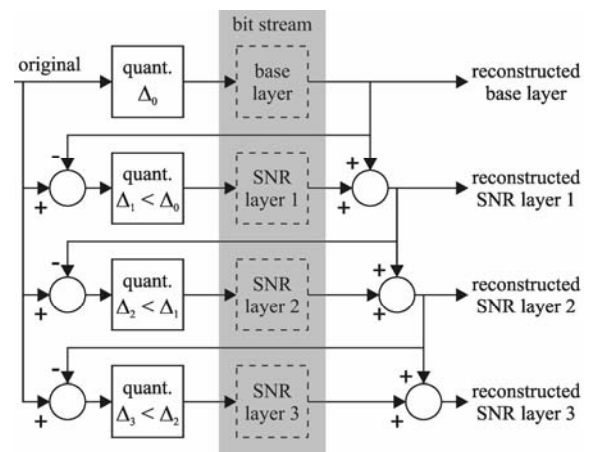


Figure 1 – Block diagram for SNR scalability with re-quantization.

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 SNR layers. When analyzing the effective decision thresholds for different choices of dead-zone parameters as applied to the quantization of SNR layers, it turns out that the resulting level information related to different SNR 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.

The next section gives a brief introduction to the basic principles of SNR 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 Sec. 4 describes an application of that modeling scheme to SVC. In Sec. 5, we present some experimental results demonstrating the effectiveness of the approach.

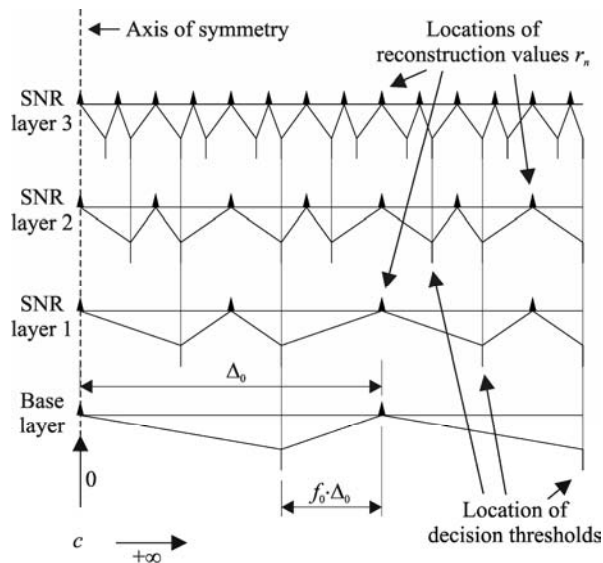


Figure 2 – Decision intervals and reconstruction values resulting from the DZ-UTQ quantization for $f_n = 1/3$ and with halved step-size from one layer to the next.

2. BACKGROUND AND PROBLEM STATEMENT

In the current SVC design, quantization of residual texture information involves a DZ-UTQ approach which is defined by the two recurrence formulas (1) and (2). The index n denotes the SNR scalable layer, c_n denotes the so-called refinement level indices of layer n , f_n denotes the so-called dead-zone parameter of layer n , Δ_n denotes the quantization step-size of layer n , r_{n-1} denotes the reconstruction value of the subordinate layer $n-1$, and c denotes the original residual texture value to be quantized.

$$c_n = \text{sgn}(c - r_{n-1}) \cdot \left[\frac{|c - r_{n-1}|}{\Delta_n} + f_n \right]. \quad (1)$$

$$r_n = r_{n-1} + c_n \cdot \Delta_n. \quad (2)$$

The base layer refinement level indices c_0 are generated by applying eq. (1) with $n = 0$ and $r_{-1} = 0$. To calculate the base layer reconstruction values r_0 from the refinement level indices c_0 , eq. (2) is applied, again with $n = 0$ and $r_{-1} = 0$. All layers different from the base layer can be generated by recursively applying eq. (1) and (2) with $n > 0$.

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 and usually takes values from 0 to 0.5. Although the step-sizes for generating different SNR 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 SNR 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.

Figure 2 depicts the quantization scheme described by eqs. (1) and (2) 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 SNR layer on top of the graph. Vertical lines in Figure 2 illustrate the decision thresholds corresponding to the quantization rule of eq. (1). The quantization scheme has mirror symmetry with the axis of symmetry at c equal to 0 and the right-hand side base layer interval (and the way of subdividing it in all SNR layers) is periodically repeated to the right. Note that due to the mirror symmetry, only the right half of the dead-zone interval of each SNR layer is depicted in Figure 2. Therefore, each decision interval that belongs to reconstruction values $r_n = 0$ is double as large as depicted in Figure 2.

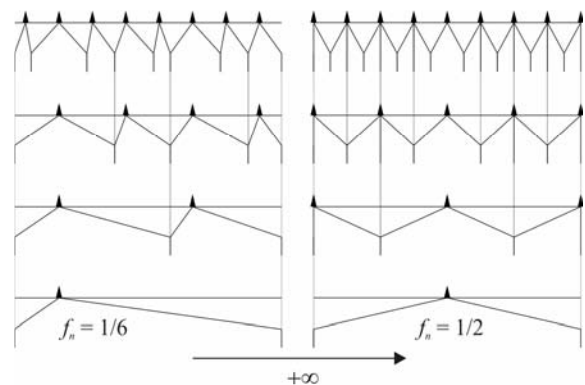


Figure 3 – Characteristic schemes for $f_n = 1/6$ on the left and for $f_n = 1/2$ on the right (showing the base layer at the bottom and 3 successive SNR layers on top of that).

A closer inspection of the location of decision thresholds involved in the quantization of two consecutive SNR layers reveals that there are two types of decision intervals. As can be seen from Figure 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 SNR 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 SNR 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 $f_n = 1/3$, Figure 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 $f_0 \dots f_n$ 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. 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 SNR 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 SNR layers is denoted as *characteristic scheme* for the remainder of this paper.

Figure 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 Figure 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 SNR layer to SNR 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) and (2) to model the probability of c_n reproducing a certain value as accurate as possible. Note that in our approach, we are primarily interested in modeling the probability of refinement levels c_n that are related to non dead-zone intervals. To this end, let us assume that the probability density function of the residual texture signal is given by $p(x)$. Let us further assume that a certain decision interval $[u, v]$ of layer n is given, which is embedded in the corresponding decision interval $[u, w]$ of layer $n-1$, i.e., it is assumed that $0 \leq u < v < w$ holds. Then we define

$$p_r(u, v, w) \stackrel{\text{def}}{=} \frac{\int_u^v p(x) dx}{\int_u^w p(x) dx},$$

where $p_r(u, v, w)$ specifies the *relative probability* of a residual texture signal value in interval $[u, w]$ being refined to sub-interval $[u, v]$.

In order to simplify our modeling approach, we make the following assumption about the relative probability. For any three values x, y, z with $0 \leq x < y < z$ and any $d \geq 0$, the condition

$$p_r(x, y, z) = p_r(x+d, y+d, z+d) \quad (3)$$

is assumed to be fulfilled. This condition ensures that the relative probability is translation invariant and, hence, it implies that it is sufficient to restrict the modeling to the characteristic scheme of one arbitrary interval. This means that only *one* single context modeling scheme is needed, regardless of the location of the characteristic scheme.

Translation invariance of relative probability according to eq. (3) is satisfied at least for uniform and Laplacian distributions. But apart from condition (3), we do not want to make any assumption about the specific nature of the probability distribution of the residual texture signal.

Under these conditions, our context modeling scheme is derived from 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 all these 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 take, 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 SNR layer $n+1$.

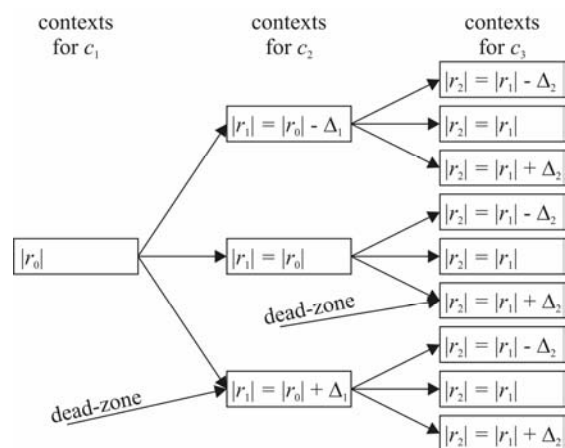


Figure 4 – Context tree for arbitrary f_n configurations.

This concept is depicted in Figure 4 with three branches per node (or sub-intervals per refinement level index). Each box represents one node and the arrows represent the branches. The number of three branches per node is not just an arbitrary

trary choice, it rather follows from the two conditions $\Delta_n = \Delta_{n-1}/2$ and $0 \leq f_n \leq 0.5$. Because of this, a tree as depicted in Figure 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 Figure 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 SNR 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 Figure 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 SNR layer are bisected exactly in the middle (cp. Figure 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 SNR 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 $p(c_2 = -1)$ is equal to 0 for the left interval, whereas the corresponding $p(c_2 = -1)$ for the right interval is usually non-vanishing. This 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. INTEGRATION INTO SVC

The generic approach has been implemented in the SVC reference software as described in [5]. The implementation employs two context models to encode refinement level indices of a certain SNR layer, where each context model conceptually relates to a distinct probability model. The first model is called 'significance context' and it is used to encode refinement indices of the dead-zone interval as well as all symbols c_0 of the base layer. The second modeling part is called 'refinement context' and it is employed to encode all refinement indices from non dead-zone intervals. A refinement context asymptotically achieves the entropy rate out of the c_n encoded with it. Conversely, significance contexts were designed to only encode c_n with a symmetrical probability distribution. This is the case for refinement level indices of dead-zone intervals, but may lead to a suboptimal modeling if used for all other intervals.

Due to complexity reasons, the SNR scalable part of SVC only uses one single significance context per SNR layer. For our generic context modeling approach, we use one significance context per layer for the dead-zone interval refinement indices and a separate refinement context for each other interval, i.e. for each tree node within the characteristic

scheme according to the given tree structure. A detailed description of this implementation can be found in [6].

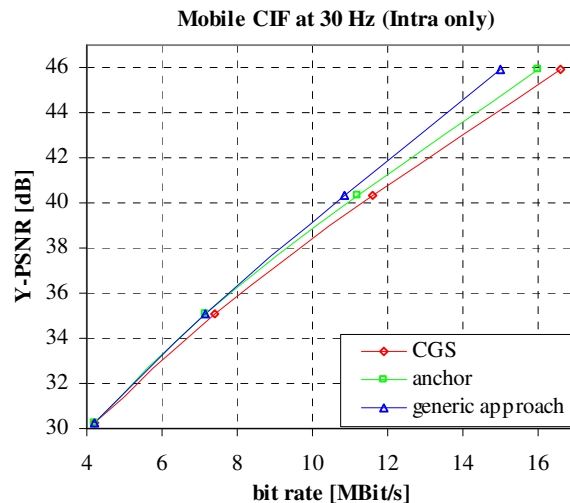


Figure 5 – R-D plot for the sequence 'Mobile', 3 SNR layers, $f_n = 1/3$

5. EXPERIMENTAL RESULTS

Three different configurations have been evaluated within the SVC-based implementation as described in the preceding section. The first configuration employs one significance context per layer to encode all refinement indices. This configuration is comparable to the typical encoding scheme for residual texture information as used in the quality scalable part of SVC, at least with respect to the context modeling part. It is therefore referred to as CGS (coarse grain scalability) in this paper. The second configuration uses one significance context for refinement indices of the dead-zone interval and one refinement context for all other refinement indices. This configuration is identical to the refinement index coding part of so-called 'progressive refinement slices' that have been developed during the standardization of SVC (also known as fine grain scalability – FGS). Since it represents a good anchor to evaluate the efficiency of the context-tree structure, it is referred to as 'anchor' in this paper. Finally, the third configuration is related to the presented generic modeling approach of Section 3 (so-called 'generic approach').

In Figure 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 SNR layer. As expected, the data point of the first SNR 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 SNR layer, our proposed approach is identical to the anchor configuration. However, for higher rate points corresponding to the second and third SNR layer, quite considerable R-D gains can be observed for our proposed approach compared to CGS as well as the anchor configuration.

	SNR layer no.	$f_n = 1/2$	$f_n = 1/3$	$f_n = 1/6$
generic approach vs. CGS	1	4.5 %	4.3 %	3.7 %
	2	14.1 %	9.7 %	6.2 %
	3	17.5 %	13.1 %	8.2 %
generic approach vs. anchor	2	7.4 %	6.3 %	0.1 %
	3	12.8 %	9.7 %	2.5 %

Table 1 – Average bit rate savings (SVC test set, CIF, Intra only).

	SNR layer no.	$f_n = 1/2$	$f_n = 1/3$	$f_n = 1/6$
generic approach vs. CGS	1	7.16 %	2.90 %	2.41 %
	2	7.10 %	4.26 %	3.17 %
	3	9.73 %	5.56 %	3.52 %
generic approach vs. anchor	2	2.24 %	2.52 %	0.52 %
	3	4.43 %	3.69 %	0.77 %

Table 2 - Average bit rate savings (SVC test set, CIF, GOP 16).

Tables 1 and 2 show averaged bit rate savings that have been achieved by our proposed modeling strategy relative to CGS and to the anchor version. The results were obtained for the eight test sequences of the SVC test set [4] in CIF resolution at 15 frames per second 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 the 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 in 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 spend for residual texture information in inter coding is usually considerably smaller than in the case of pure intra coding.

6. CONCLUSION

We have presented a generic context modeling scheme for URQ-based SNR-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.

REFERENCES

- [1] *ITU-T and ISO/IEC JTC 1*, "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG4-AVC), Version 1: May 2003, Version 2: Jan. 2004, Version 3: Sep. 2004, Version 4: July 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. J. Sullivan, "On embedded scalar quantization," in Proceedings of IEEE ICASSP, vol. IV, pp. 605-608, May 2004.
- [4] M. Wien, H. Schwarz, "AhG on Testing Conditions for Coding Efficiency and JSVM Performance evaluation," Joint Video Team Doc. JVT-P009, Poznan, Poland, July 2005.
- [5] J. Reichel, H. Schwarz, M. Wien, "Joint Scalable Video Model JSVM-8," Joint Video Team Doc. JVT-U202, Hangzhou, China, October 2006.
- [6] H. Kirchhoffer, D. Marpe, T. Wiegand, "CE3: Improved CABAC for PR slices," Joint Video Team Doc. JVT-U082, Hangzhou, China, October 2006.