

Adaptive Spatial and Transform Domain FGS Coding

Li Zhang¹, Xiangyang Ji¹, Wen Gao^{2,3}, Debin Zhao¹

¹Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, 100080, China

²Graduate School of the Chinese Academy of Sciences, Beijing, 100039, China

³Peking University, Beijing, 100871, China
{zhanglili, xyji, wgao, dbzhao}@jdl.ac.cn

ABSTRACT

In inter-picture coding, block-based frequency transform is usually carried out on the predicted errors for each inter-block to remove the spatial correlation among them. However, it can not always do well since the predicted errors in some inter-blocks have marginal or diagonal correlation. A good solution is to omit transform operations for the predicted errors of those inter-blocks with low correlation before quantization operation. The same phenomenon also can be observed in Fine Grain Scalability (FGS) layer coding. In this paper, an adaptive prediction error coding method in spatial and frequency domain with lower complexity is considered for FGS layer coding. Transform operation is only needed when there are non-zero reconstructed coefficients in spatially co-located block in base layer. The experimental results show that compared with FGS coding in JSVM, higher coding efficiency can be achieved with lower computational complexity at decoder since inverse transform is no longer needed for those predicted errors coded in spatial domain at encoder.

Index Terms—FGS, adaptive spatial and frequency domain

1. INTRODUCTION

Scalability started to be a research and application issue in video coding quite a long time ago. Nowadays, it becomes a more and more challenging issue due to the wide spreading of wireless and heterogeneous networks. Numerous scientific papers and international standardization activities prove the importance of scalability. Currently, ISO and ITU are working on an international standard of scalable video coding [1] that will be an extension of the H.264/MPEG-4 AVC video compression standard [2]. This scalability technology comprises sophisticated tools for spatial, temporal, and quality scalability. A Fine Grain SNR (Quality) Scalability (FGS)-based sequence consists of two kinds of streams: a non-scalable base layer which provides basic quality and enhancement layers that can add incremental quality refinements proportional to the number of bits received.

In FGS encoding, all enhancement layers in a predicted frame can be coded from the reconstructed version of the highest-quality frame in the reference frame. The temporal redundancy can then be reduced by temporal prediction. The prediction error signal is encoded and transmitted afterwards. In order to reduce the remaining spatial redundancy of the prediction error signal, block-wise transform coding is applied accordingly. The block size of the transform allows for 4×4 or 8×8 in current reference software [3]. The transformed coefficients are quantized with the step size controlled by the quantization parameter QP and the quantization step corresponding to QP which is repeatedly decreasing for successive enhancement layers. The quantized transform coefficients are ZigZag scanned and then entropy-coded. Transform plays an important role for high compression if the prediction error samples are highly correlated. For those well predicted samples, the transform is possibly inefficient. Therefore, it is proposed in this paper to code these prediction errors in the spatial domain. Each prediction error block in FGS layers is coded either in spatial or in frequency domain according to the corresponding Coded Block Pattern (CBP) in base layer. If CBP of the luma component in the same position in base layer is zero (there are no non-zero coefficients) and it is not an intra block, the corresponding macroblock in enhancement layer will be coded in the spatial domain. Otherwise, frequency transform is required. In [4], it is proposed to skip the transform for non-scalable coding. Different from our scheme, the mode selection in [4] is based on Rate-Distortion (RD) cost and encoded with a flag.

The remainder of this paper is structured as follows. In the next section, we investigate the correlation of the prediction errors for different cases in enhancement layers. Thereby the criterion for determining spatial or frequency domain coding is deduced. Section 3 describes the quantizer design for the spatial domain. The entropy coding, in Section 4 and the complexity reduction of our proposed scheme, in Section 5, are explained in detail. Experimental results and the conclusions are given in Sections 6 and 7, respectively.

2. CHARACTERISTICS OF RESIDUALS IN FGS LAYERS

As aforementioned, transform operation usually works well for the inter-predicted coefficients with high spatial correlation. Otherwise, it may play a negative role. In general, if there are non-zero reconstructed coefficients in a block of the base layer, which usually is the coarse quantization version of the original picture, the prediction errors of this block still have high spatial correlation. Hence, transform operation should be used to further reduce its spatial correlation in enhancement layers. Consequently, the value of CBP in the base layer can be used to determine whether the predicted errors need transform operation or not.

In order to check whether our conclusion is correct, an experiment is conducted as shown in Figure 1, where the coding of two prediction error blocks with size of 4×4 is shown. The first column of 4×4 block represents the typical distribution of prediction errors in an enhancement layer where in base layer there are some non-zero coefficients. The block in the second column is obtained from a block with CBP equal to zero in spatially co-located block in base layer, that is, all coefficients in the block of base layer are equal to zero. The quantization parameter QP for base layer is 32 and one FGS layer is appended on top of base layer and D_c is the corresponding calculated cost.

17	7	-1	-3	Prediction Errors	1	1	-5	14
3	2	-2	-2		2	1	-2	6
1	-3	-2	0		2	1	10	8
-3	-2	-3	-3		7	-16	14	15

0	0	0	0	After Quantization	1	-1	1	0
1	0	0	0		0	0	0	-1
1	0	0	0		0	0	0	0
0	0	0	0		0	0	0	0

9	7	2	-1	After IDCT	0	3	-3	13
5	3	1	0		1	1	0	12
-4	-3	-1	1		4	-4	5	9
-9	-6	-1	1		5	-7	7	8

$$D_c = 181 + \lambda \times 16 = 489$$

(a)

$$D_c = 289 + \lambda \times 18 = 636$$

(b)

Figure 1. Examples for the coding of prediction error blocks with base layer CBP equal to zero (a) or not (b).

For the two example blocks, the Lagrangian costs of the transform coder are determined by

$$D_c = SSD + \lambda \times R \quad (1)$$

Here, SSD is the sum of squared differences between prediction errors and reconstructed residuals, R is the rate of the transform coder required to encode the quantized coefficients using CABAC and λ is the Lagrange multiplier as commonly used for the coder control of FGS coding. The costs of the transform coder D_c are shown for the two example blocks in Figure 1 respectively. It can be observed that for the block with base layer CBP equal to zero, the costs of the block are much higher than that of the left block with CBP equal to zero in base layer. This proves that the transform coder is only adjusted for blocks with vertically or horizontally correlated samples, also it indicates that the prediction errors are weakly low correlated if base layer CBP equals to zero in the same spatial position. For these marginally correlated coefficients, we shall code these blocks in spatial domain in order to improve the coding efficiency. The framework of this adaptive frequency and spatial domain FGS coding is depicted in Figure 2. The red part is what we propose in this paper.

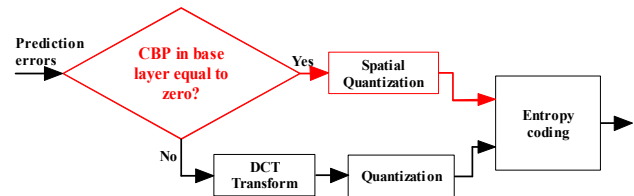


Figure 2. Block diagram of proposed scheme for prediction errors in FGS layers.

3. SPATIAL QUANTIZATION DESIGN

Due to the different residual distribution in the frequency and spatial domain, a special scalar quantizer is needed to be designed for the spatial domain. Since the distribution of the prediction error is close to a Laplacian distribution, as shown in Figure 3, an effective scalar dead-zone plus uniform threshold (DZ+UTQ) quantizer [4] is used in the case of mean squared quantization error optimization. The dead-zone plus uniform threshold quantizer is defined as follows.

$$Q(x) = \text{sign}(x) \times \max \left[0, \frac{|x|}{s} - \frac{z}{2} + 1 \right] \quad (2)$$

where the notation $\lfloor \cdot \rfloor$ denotes the maximum integer less than or equal to the argument, while the reconstruction rule for DZ+UTQ is represented by the following equation

$$Q^{-1}(x) = \text{sign}(x) \times \left[\left(|x| + \frac{z}{2} - 1 \right) s + \left(1 - \frac{z}{2} \right) s \right] \quad (3)$$

Here s and z represent the step size for all steps other than the dead-zone region and the ratio of dead-zone size to the size of the other step s respectively. In equations (2) and (3),

the variable x indicates the input to quantizer or inverse quantizer. The visualized presentation of DZ+UTQ is depicted in Figure 4. For our experiment, we use the uniform reconstruction quantizer (URQ) as defined in [5].

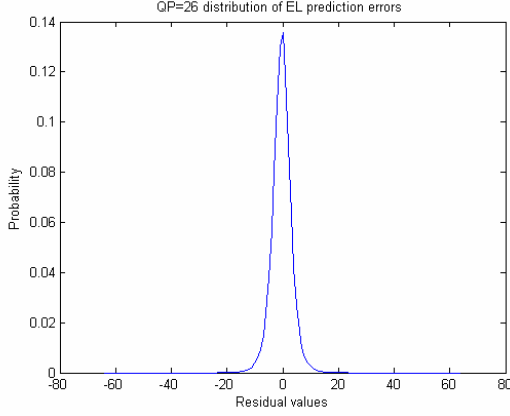


Figure 3. Distribution of prediction errors for the quantization parameter $QP=26$ in Enhancement layer.

In FGS coding, rate-distortion-based quantization is employed, and the typical cost function is defined as

$$Cost = D + \lambda \times R \quad (4)$$

Here, the Rate (R) and distortion (D) can be represented by equations (5) and (6), respectively.

$$R = - \sum_{k=0}^n p_k (\log_2 p_k) \quad (5)$$

$$D = E\{d(X-Q(X))\} \\ = 2 \times \left(\sum_{k=1}^n \int_{(z+2k-2)s/2}^{(z+2k)s/2} d(x-k)f(x)dx + \int_0^{zs/2} d(x-0)f(x)dx \right) \quad (6)$$

where p_k is given by

$$p_k = \begin{cases} \int_{-s/2}^{s/2} f(x) dx & k = 0 \\ \int_{(z+2|k|-2)s/2}^{(z+2|k|)s/2} f(x) dx & k \neq 0 \end{cases} \quad (7)$$

Therefore, for a given Laplacian distribution with variance σ , the aim of quantization design is to minimize the rate-distortion cost. According to equations (2)-(7), we can get the following relationships among the Laplacian variance σ , quantization step s , dead-zone ratio z and λ . For each quantization parameter QP , the corresponding parameters are determined by formulas (8) and (9). Table I shows some parameters s , z for commonly used QPs .

$$\frac{\lambda \times \sqrt{2}}{\sigma \times \ln 2} = s \times \frac{\exp(\frac{\sqrt{2}}{\sigma} s) + 1}{\exp(\frac{\sqrt{2}}{\sigma} s) - 1} - \sqrt{2} \sigma \quad (8)$$

$$(1 - \frac{z}{2})s = \frac{\sigma}{\sqrt{2}} \left\{ 1 - \frac{\frac{\sqrt{2}}{\sigma} s \cdot \exp(-\frac{\sqrt{2}}{\sigma} s)}{1 - \exp(-\frac{\sqrt{2}}{\sigma} s)} \right\} \quad (9)$$

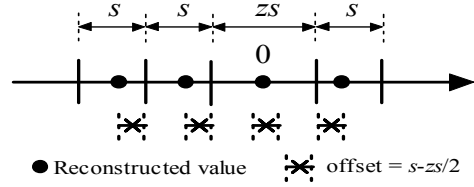


Figure 4. DZ + UTQ Quantizer for the spatial domain

TABLE I. PARAMETERS s , z OF THE SPATIAL QUANTIZER

QP	s	z
23	10.32	1.43
26	14.83	1.47
29	21.46	1.53
32	31.18	1.57

4. ENTROPY CODING

For entropy coding of the quantized residuals in the spatial domain, Context-based Adaptive Binary Arithmetic Coding (CABAC) for the quantized coefficients in the frequency domain can be easily extended to coefficients coding in spatial domain by similar method. However, note that the context models in the spatial domain are separated from those in the frequency domain and the initialization of the models is derived from the statistics of the quantized samples. This is because of the different residual distributions in frequency and spatial domain.

TABLE II. RATIO OF SKIPPED IDCT IN FGS LAYER

QP	BLR	ELR
Bus(qcif@15HZ)	75%	26%
Foreman(qcif@15HZ)	85%	19%
Mobile(qcif@15HZ)	89%	26%
Football(qcif@15HZ)	48%	21%

5. ANALYSIS OF COMPUTATIONAL COMPLEXITY IN THE DECODER

This adaptive spatial and transform domain FGS coding scheme can reduce the computational complexity on the decoder side since those macroblocks coded in spatial domain don't need inverse transform. Thus, the computational complexity reduction is analyzed in this section. Table II shows the percentage of spatial domain coded macroblocks in different cases. All these sequences are standard testing sequences. BLR represents the percentage for macroblocks with CBP equal to 0 and ELR

indicates the percentage for macroblocks with CBP equal to 0 at base layer but not equal to 0 at enhancement layer.

6. EXPERIMENTAL RESULTS

In this paper, we propose an adaptive frequency and spatial FGS coding scheme. To evaluate the performance of this proposed method, experiments were performed on various common test sequences. The initial QP for base layer is set to be 40, and for enhancement layer QP is decreased by step of 6. One FGS enhancement layer is appended on top of the base layer for coding. Only the first key frame is coded as intra-picture, and the remaining key frames are coded as P or B pictures with close loop at the highest rate point. All of the test sequences were coded in hierarchy B with GOP size of 16 and only one enhancement layer was applied. For the entropy part, CABAC is applied. In order to fairly and clearly verify the advantages of our proposed FGS scenario, we first encode the original sequences and then extract and decode the encoded sequences at five different bit rates including their corresponding base layers' bit rate. The SVC reference software is [3] and the corresponding bit rate savings compared to the FGS coding of SVC standard are shown in table III. Besides, the better performance on two sequences 'Mobile' and 'Paris' in QCIF format are illustrated in Figure 5. It can be observed that the bit rate savings can be up to 4%. This coding gain comes from the adaptive scheme which avoids DCT transform for marginally correlated prediction errors.

TABLE III. BIT RATE REDUCTION COMPARED TO FGS CODING IN SVC

Sequence name	Bit rate Reduction
Mobile	2.23%
Paris	2.60%
Foreman	4.03%
Mother_daughter	3.02%

7. CONCLUSIONS

In this paper, an efficient frequency and spatial domain FGS coding scheme is introduced. For each block of the luminance signal in FGS layers, either frequency domain or spatial domain is applied whereas the selection is based on the value of CBP in spatially co-located macroblock in base layer. Experimental results have proven that higher coding efficiency can be achieved at low bit rates and it also reduces the computational complexity since inverse transform is no longer needed when reconstructing the predicted errors, which are encoded in spatial domain. What's more, better quantization and entropy coding and how to extend the macroblock-based spatial and frequency domain coding decision to block base shall be investigated and further improvement of coding gain is expected.

8. REFERENCES

- [1] ITU-T and ISO/IEC JTC1, Available: http://ftp3.itu.ch/av-arch/jvt-site/2006_07_Klagenfurt/JVT-T201.zip.
- [2] ITU-T and ISO/IEC JTC1, "Advanced Video Coding for Generic Audiovisual Services," ITU-T Recommendation H.264 – ISO/IEC 14496-10 AVC, 2003.
- [3] ITU-T and ISO/IEC JTC1, "JSVM 7 Software," JVT-T203, July 2006.
- [4] Matthias Narroschke, Hans Georg Musmann, Available: http://ftp3.itu.ch/av-arch/video-site/0601_Ban/VCEG-AB06.doc
- [5] G.J.Sullivan, S.Sun, "On Dead-Zone Plus Uniform Threshold Scalar Quantization," Proceedings of VCIP, Beijing, China, 2005.

ACKNOWLEDGEMENT

This work is supported by Special Foundation of President of The Chinese Academy of Sciences (No. 20064020).

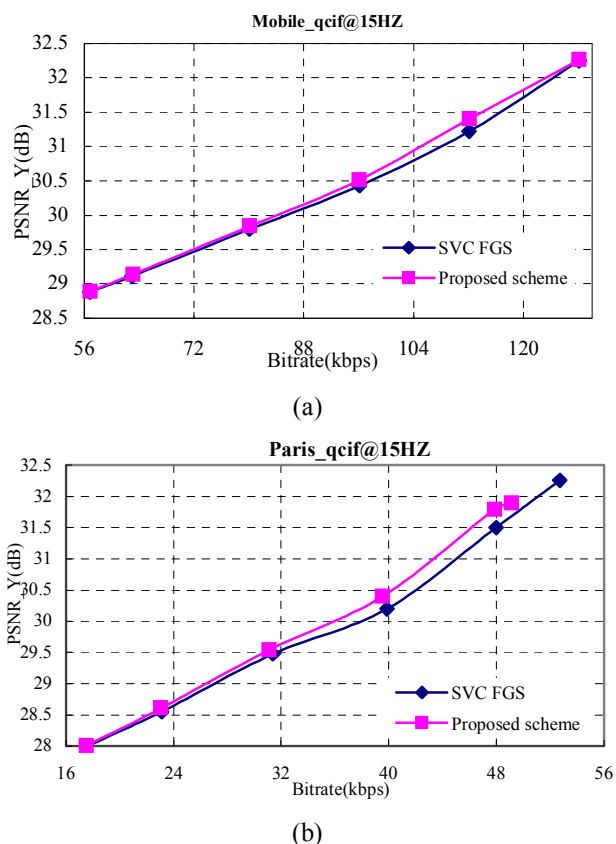


Figure 5. Comparison of rate distortion curves between original FGS and proposed adaptive FGS coding schemes for two test sequences in QCIF@15HZ. (a) Mobile. (b) Paris.