

# Analysis of the spatial layer discrete cosine transform coefficient distribution and its application to rate model for H.264/SVC encoder

Szu-Wei Lee

Graphic Software Development and Validation Team, Visual and Parallel Computing Group, Intel, Santa Clara, CA, USA  
E-mail: swli0311@yahoo.com.tw

Published in *The Journal of Engineering*; Received on 2nd November 2013; Accepted on 10th February 2014

**Abstract:** Knowledge of the discrete cosine transform coefficient distribution (DCT-DIST) is important for the encoder design. For example, rate control relies on this knowledge to estimate a possible bit rate and then decide proper coding parameters before the actual encoding task is performed. Therefore the rate control performance is fairly dependent on how accurately the DCT-DIST is modelled. The spatial enhancement layer (SL) DCT-DIST for H.264 scalable video coding (SVC) is studied in this Letter. SL DCT-DIST knowledge is furthermore used to derive a novel rate model. Our results can help design a proper rate control module for the H.264/SVC encoder.

## 1 Introduction

Recently, the scaling video coding (SVC) extension of H.264/AVC [1] has been standardised, which allows a video sequence to be coded into one base layer (BL) and several enhancement layers so as to support temporal, spatial and quality scalability. In this Letter, the H.264/SVC spatial scalable encoder is taken into consideration, and the distribution of the spatial enhancement layer (SL) discrete cosine transform (DCT) coefficients will be studied. Moreover, knowledge of the SL DCT coefficient distribution (DCT-DIST) will be applied to develop a novel rate model to estimate the SL bit rate. Our study of the SL DCT-DIST and the proposed rate model could help design a proper rate control algorithm for the H.264/SVC spatial scalable encoder.

The DCT-DIST has been studied over the past few decades. Gaussian or Laplacian distributions were popular to estimate the DCT-DIST in practice. However, it is observed that the DCT-DIST in image and video applications could differ from Gaussian or Laplacian distributions in most cases. Kamaci *et al.* [2], therefore, proposed that Cauchy distribution is more accurate to model the DCT-DIST for H.264/AVC. Although the DCT-DIST can be well estimated by Cauchy distribution, the result may not be able to apply to H.264/SVC when coding the SL DCT coefficients since the video signal could be predicted by the reconstructed signal of the reference layer, which can be the previous coded SL or the BL. To be more specific, the SL DCT coefficients contain the quantisation noise because of the quantisation effect in the reference layer, and the distribution of the quantisation noise in the reference layer could affect the SL DCT-DIST. Recently, Liu *et al.* [3] proposed a rate control algorithm for H.264/SVC. The SL DCT-DIST is assumed to be Cauchy distributed. Although their algorithm could provide good rate control results, the key assumption that the SL DCT-DIST is Cauchy distributed is not discussed in detail while it was verified experimentally only. In other words, there is no clue as to why the SL DCT-DIST is Cauchy distributed in the previous work.

The SL DCT-DIST for H.264/SVC will be investigated in this Letter. Furthermore, one application of SL DCT-DIST knowledge to bit rate modelling will be presented. There are three main contributions in our work. First, we provide the mathematical derivation to show why the SL DCT-DIST is Cauchy distributed in detail. Second, a novel SL rate model is proposed. The proposed SL rate model is more accurate and simpler as compared with the previous rate model [3]. Finally, several experiments will be conducted to verify the accuracy of the proposed SL rate model through various sizes of sequences while high-definition sequences were

not considered in [3]. We conclude that the proposed SL rate model can provide fairly accurate estimation results for various sizes of bit streams.

## 2 Overview of H.264 spatial scalable encoder

Without loss of generality, we consider the case that two layers of bit streams (i.e. one BL and one SL) are generated by the H.264/SVC encoder. Under such a scenario, the BL must be used as the reference layer when coding the SL bit stream.

To improve the coding efficiency, the H.264/SVC encoder allows either the H.264/AVC conventional coding scheme or the so-called inter-layer prediction coding scheme to code macro blocks (MBs) in the SL. When the H.264/AVC conventional coding scheme is selected, all H.264/AVC existing coding methods can be used to generate the prediction signal, and the current coding MB can be coded as either the intra- or the inter-mode. On the other hand, the reconstructed co-located MB in the BL is used as the prediction signal when the inter-layer prediction coding scheme is selected. MB coding information, such as MB partition modes, motion vectors, reference frame indices and so on, of the current coding MB is derived from the co-located MB in the BL, and the current coding MB is marked as a base mode MB. Notably, when decoding a base mode MB, the motion compensation process only needs to be performed in the target decoding layer because MB coding information can be directly derived from the corresponding reference layer. This is so-called single-loop decoding. More information can be found in [1].

Dataflow of the H.264/SVC spatial scalable encoder using the inter-layer prediction coding scheme is shown in Fig. 1. Table 1 also summarises the symbols employed in this Letter. Let  $X$  be the source video signal. Its down-scaled video signal ( $X_{BL}$ ), which contains the low-frequency component of  $X$ , is first fed

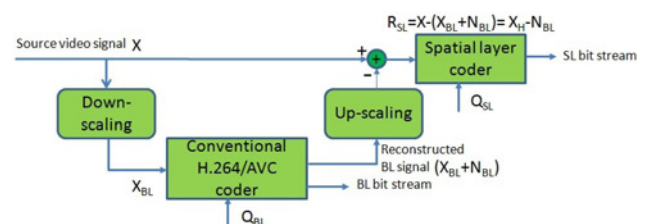


Fig. 1 Dataflow of H.264/SVC spatial scalable encoder using the inter-layer prediction coding scheme

**Table 1** Summary of symbols

Symbols	Definitions
DCT	discrete cosine transform
DCT-DIST	DCT coefficient distribution
BL	base layer
SL	spatial enhancement layer
MB	macro block
$X$	source video signal
$X_{BL}$	low-frequency component of $X$
$X_H$	high-frequency component of $X$
$Q_{BL}$	BL quantisation step size
$Q_{SL}$	SL quantisation step size
$N_{BL}$	BL quantisation noise
$R_{SL}$	SL residual signal
PDF	probability density function
$f_{N_{BL}}(\cdot)$	PDF for the distribution of $N_{BL}$
$f_{SL}(\cdot)$	PDF for the SL DCT-DIST
$\mu_N$	Cauchy scale parameter for the distribution of $N_{BL}$
$\mu_{X_H}$	Cauchy scale parameter for the DCT-DIST of $X_H$
$\mu_{SL}$	Cauchy scale parameter for the SL DCT-DIST
$B$	bit number
$\rho$	percentage of zero quantised DCT coefficients
$\theta$	$\rho$ -domain rate model parameter
$B_{SL}$	SL bit number
$\rho_{SL}$	percentage of zero quantised SL DCT coefficients
$\theta_{SL}$	SL $\rho$ -domain rate model parameter
$\eta_{SL}, \gamma, \delta$	proposed SL rate model parameters

into the conventional H.264/AVC coder with the BL quantisation step size ( $Q_{BL}$ ) to generate the BL bit stream. Then, the reconstructed BL signal, which includes not only  $X_{BL}$  but also the quantisation noise ( $N_{BL}$ ) because of the quantisation effect by  $Q_{BL}$ , is up-scaled and used as the prediction signal of the source video signal,  $X$ . Finally, the SL residual signal ( $R_{SL}$ ), which is the difference between  $X$  and the reconstructed BL signal, is coded by the SL coder with the SL quantisation step size ( $Q_{SL}$ ) to produce the SL bit stream. Notably,  $R_{SL}$  contains not only the high-frequency component of  $X$  (denoted by  $X_H$ ) but also  $N_{BL}$ .

### 3 SL DCT coefficient distribution

The H.264/SVC spatial scalable encoder using the inter-layer prediction coding scheme and the DCT-DIST of  $R_{SL}$  (i.e. SL DCT-DIST) are taken into consideration in this section. Since  $R_{SL}$  contains both  $X_H$  and  $N_{BL}$ , their distributions will be considered separately. We first borrow the DCT-DIST result in [2] to derive the distribution of  $N_{BL}$ . Then, the SL DCT-DIST is derived accordingly.

**Lemma 1:** Distribution of the BL quantisation noise is Cauchy distributed.

*Proof:* The probability density function (PDF) of the quantisation noise can be expressed as a function of the quantisation step size and the characteristic function of the source coding signal [4]. Consider the conventional H.264/AVC coder with the BL quantisation step size,  $Q_{BL}$ . Since the quantisation is performed in the DCT domain, the PDF for the distribution of  $N_{BL}$  (i.e.  $f_{N_{BL}}(q)$ ) can be written as follows

$$f_{N_{BL}}(q) = \frac{1}{Q_{BL}} + \frac{1}{Q_{BL}} \sum_{k \neq 0} \Phi\left(\frac{2\pi k}{Q_{BL}}\right) e^{(-j2\pi kq/Q_{BL})} \quad (1)$$

where  $\Phi(\cdot)$  is the characteristic function of the BL DCT coefficients.

Since the BL DCT-DIST is approximately Cauchy distributed [2], its characteristic function can be expressed by

$$\Phi(\omega) = e^{-\mu|\omega|} \quad (2)$$

where  $\mu$  is the scale parameter of the Cauchy distribution function. Substituting (2) into (1) yields

$$f_{N_{BL}}(q) = \frac{1}{Q_{BL}} \frac{\alpha^2 - 1}{\alpha^2 + 1 - 2\alpha \cos((2\pi/Q_{BL})q)} \quad (3)$$

where  $\alpha = e^{(2\pi/Q_{BL})\mu}$ .

Considering the Taylor approximation of the cosine function (i.e.  $\cos(x) \simeq 1 - (1/2)x^2$ ) and substituting it into (3) yields

$$f_{N_{BL}}(q) \simeq \tau \frac{\beta}{\beta^2 + 4\pi^2 q^2} \quad (4)$$

where  $\beta = Q_{BL}(\alpha - 1)\alpha^{-0.5}$  and  $\tau = (\alpha + 1)\alpha^{-0.5}$ .

Equation (4) suggests that the distribution of the BL quantisation noise,  $N_{BL}$ , is approximately Cauchy distributed. Thus, its PDF can be written as

$$f_{N_{BL}}(q) = \frac{1}{\pi} \frac{\mu_N}{\mu_N^2 + q^2} \quad (5)$$

where  $\mu_N$  is the Cauchy scale parameter and it can be expressed as a function of  $Q_{BL}$ .  $\square$

**Lemma 2:** SL DCT-DIST is Cauchy distributed.

*Proof:* Since  $X_H$  and  $N_{BL}$  result from the high- and low-frequency components of  $X$ , it is reasonable to assume that  $X_H$  and  $N_{BL}$  are independent. Furthermore, the DCT-DIST of  $X_H$  is approximately Cauchy distributed [2], and the distribution of  $N_{BL}$  is Cauchy distributed, too, as shown in Lemma 1. We conclude that the SL DCT-DIST, which is the DCT-DIST of  $R_{SL}$ , is Cauchy distributed because  $R_{SL} = X_H - N_{BL}$  and the sum of two independent Cauchy random variables is Cauchy distributed [5]. Notably, the scale parameter for the sum of two independent Cauchy random variables is the sum of individual scale parameters for these two Cauchy random variables. Therefore the PDF of the SL DCT-DIST can be expressed as

$$f_{SL}(x) = \frac{1}{\pi} \frac{\mu_{SL}}{\mu_{SL}^2 + x^2} \quad (6)$$

where  $\mu_{SL} = \mu_{X_H} + \mu_N$ , and  $\mu_{X_H}$  is the Cauchy scale parameter for the DCT-DIST of  $X_H$ . Since  $\mu_N$  can be written as a function of  $Q_{BL}$  (i.e.  $\mu_N = f(Q_{BL})$ ),  $\mu_{SL}$  can be further rewritten as a function of  $Q_{BL}$  too. That is

$$\mu_{SL} = \mu_{X_H} + f(Q_{BL}) \quad (7)$$

Notably, our result of the Cauchy scale parameter for the SL DCT-DIST,  $\mu_{SL}$ , is consistent with the experimental result as shown in [3], where  $\mu_{SL}$  is simply treated as a linear function of  $Q_{BL}$ .  $\square$

### 4 SL bit rate model

Consider the  $\rho$ -domain rate model [6] that the coded bit number ( $B$ ) is expressed as

$$B = \theta(1 - \rho) \quad (8)$$

where  $\theta$  is the rate model parameter, and  $\rho$  is the percentage of zero quantised DCT coefficients.

Equation (8) suggests that the coded bit number is considered being proportional to the percentage of non-zero quantised DCT coefficients. Similarly, the SL bit number ( $B_{SL}$ ) can be modelled as a function of the percentage of non-zero quantised SL DCT coefficients. That is

$$B_{SL} = \theta_{SL}(1 - \rho_{SL}) \quad (9)$$

where  $\theta_{SL}$  is the SL rate model parameter, and  $\rho_{SL}$  is the percentage of zero quantised SL DCT coefficients.

The percentage of zero quantised SL DCT coefficients,  $\rho_{SL}$ , can be calculated by

$$\rho_{SL} = \int_{-0.5 Q_{SL}}^{0.5 Q_{SL}} f_{SL}(x) dx \quad (10)$$

where  $f_{SL}(\cdot)$  is the PDF of the SL DCT-DIST.

Since the SL DCT-DIST is Cauchy distributed as shown in Lemma 2, (9) can be rewritten as

$$B_{SL} = \theta_{SL} \left[ 1 - \frac{2}{\pi} \tan^{-1} \left( \frac{Q_{SL}}{2 \mu_{SL}} \right) \right] \quad (11)$$

Considering the Taylor expansion of  $\tan^{-1}(x)$  (i.e.  $\tan^{-1}(x) = \sum_{n=0}^{\infty} ((-1)^n / (2n+1)) x^{2n+1}$ ) and substituting it into (11) yields

$$B_{SL} = \theta_{SL} \left[ 1 - \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)} \left( \frac{Q_{SL}}{2 \mu_{SL}} \right)^{2n+1} \right] \quad (12)$$

Furthermore, since  $\mu_{SL}$  can be expressed as a function of  $Q_{BL}$  as shown in (7),  $\mu_{SL}$  in (12) can be replaced by its Taylor approximation. That is

$$\mu_{SL} \simeq \mu_{X_H} + c_1 Q_{BL} + c_2 Q_{BL}^2 \quad (13)$$

As a result, (12) can be expressed as

$$B_{SL} \simeq \theta_{SL} \left[ 1 - \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)} \left( \frac{0.5 Q_{SL}}{\mu_{X_H} + c_1 Q_{BL} + c_2 Q_{BL}^2} \right)^{2n+1} \right] \quad (14)$$

Finally, we simplified the above equation and used the following approximate equation as the proposed SL rate mode

$$B_{SL} \simeq \eta_{SL} Q_{SL}^{\gamma} Q_{BL}^{\delta} \quad (15)$$

where  $\eta_{SL}$ ,  $\gamma$  and  $\delta$  are the proposed SL rate model parameters.

As compared with the previous rate model proposed in [3], the SL bit number was modelled by two different equations, and the selection of a proper equation is dependent on  $Q_{BL}$  and  $Q_{SL}$ . The proposed rate model, however, is much simpler for a single equation is used to estimate the bit rate. The accuracy of the proposed SL bit rate model will be examined in the next section.

## 5 Experimental result

The proposed SL rate model was implemented with JSVM 9.19.15. In our experiment, all test bit streams contain two layers, where BL and SL are of QCIF-CIF, CIF-4CIF and 540P-1080P resolutions. As shown in (15), the proposed model contains three parameters (i.e.  $\eta_{SL}$ ,  $\gamma$  and  $\delta$ ). These three parameters have to be decided first

**Table 2** Accuracy of the proposed SL rate model

QCIF-CIF bit stream	Bit rate, bits/s	Error, %
foreman	32 K	5.09
bus	48 K	3.30
mobile	56 K	5.69
coast	72 K	5.74
tempe	88 K	6.10
CIF-4CIF bit stream	bit rate, bits/s	error, %
harbour	384 K	2.55
city	512 K	2.43
540P-1080P bit stream	bit rate, bits/s	error, %
blue sky	2048 K	2.06
sunflower	6400 K	1.69
rush hour	8120 K	1.61

so that the proposed model can be used to estimate a possible SL bit rate.  $\eta_{SL}$  is updated by the least mean square approach per rate control unit. The rate control unit is one of the encoding parameters and it may have one or several MBs. On the other hand,  $\gamma$  and  $\delta$  are fixed in our experiment and were determined by the following training process: (i) Foreman and Mobile sequences were selected to generate several training bit streams of QCIF-CIF resolution whose BL and SL were coded by different QPs (QP = 15, 25, 35 and 45). (ii) Coded SL bit numbers and QPs for these training bit streams were used to determine  $\gamma$  and  $\delta$  by the least square approach. These trained model parameters are

$$\gamma = -0.6846, \quad \delta = 0.7769 \quad (16)$$

For model verification, each test bit stream contains 15 frames. Rate control was enabled to code the BL under different bit rate constraints. The rate control unit has exactly 11, 22 and 120 MBs for QCIF-CIF, CIF-4CIF and 540P-1080P bit streams, respectively. All SL bit streams were coded with *default\_base\_mode\_flag* equal to one. This implies that the inter-layer prediction coding scheme is always selected in the encoder. The percentage error between the actual coded SL bit number and the estimated SL bit number calculated by the proposed SL rate model is listed in Table 2.

It can be seen that the proposed SL rate model can provide good estimation results of the SL bit rate. As compared with the experimental results in [3], the proposed model can provide even more accurate data for various bit streams.

## 6 Conclusion

The SL DCT-DIST and the application of SL DCT-DIST knowledge to bit rate modelling were studied in this Letter. We first showed mathematically how the SL DCT-DIST is Cauchy distributed. After that a novel SL rate model was proposed. Experimental results show that the proposed SL rate model could result in a fairly good estimation to the SL bit rate. As compared with the previous work, the proposed rate model is more accurate and simpler. Our results can help design a good bit rate control algorithm for the H.264/SVC spatial scalable encoder.

## 7 Acknowledgment

The author would like to thank the Intel for supporting this work.

## 8 References

- [1] Schwarz H., Marpe D., Wiegand T.: 'Overview of the scalable video coding extension of the H.264/AVC standard', *IEEE Trans. Circuits Syst. Video Technol.*, 2007, 17, (8), pp. 1103–1120

- [2] Kamaci N., Altunbasak Y., Mersereau R.M.: 'Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models', *IEEE Trans. Circuits Syst. Video Technol.*, 2005, **15**, (8), pp. 994–1006
- [3] Liu J., Cho Y., Guo Z., Kuo C.C.: 'Bit allocation for spatial scalability coding H.264/SVC with dependent rate-distortion analysis', *IEEE Trans. Circuits Syst. Video Technol.*, 2010, **20**, (7), pp. 967–981
- [4] Sripad A.B., Snyder D.L.: 'A necessary and sufficient condition for quantization errors to be uniform and white', *IEEE Trans. Acoust. Speech Signal Process.*, 1977, **25**, (5), pp. 442–448
- [5] Wikipedia the free encyclopedia: 'Cauchy distribution'. Available at [http://www.en.wikipedia.org/wiki/Cauchy\\_distribution](http://www.en.wikipedia.org/wiki/Cauchy_distribution)
- [6] He Z., Kim Y.K., Mitra S.K.: 'Low delay rate control for DCT video coding via  $\rho$ -domain source modeling', *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**, (8), pp. 928–940