GAUSSIAN DISTRIBUTION-BASED MODE SELECTION FOR INTRA PREDICTION OF SPATIAL SHVC

Dayong Wang[1], Xin Wang[2], Yu Sun[3], Weisheng Li[4], Xin Lu[5], Frederic Dufaux[6]

[1]Chongqing Key Laboratory on Big Data for Bio Intelligence, [2]School of communication and Information Engineering, [4]Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications.
[3]Department of Computer Science, University of Central Arkansas Computing,
[5]Faculty of Computing, Engineering and Media (CEM), De Montfort University,
[6]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes
wangdayong@cqupt.edu.cn, 1262126674@qq.com, yusun@uca.edu,liws@cqupt.edu.cn, xin.lu@dmu.ac.uk, frederic.dufaux@l2s.centralesupelec.fr

ABSTRACT

Due to the diversity of terminal devices, Spatial Scalable High Efficiency Video Coding (SSHVC) is an efficient solution to meet this requirement. However, its coding process is very complex, which seriously prevents its wide applications. Therefore, it is very crucial to reduce coding complexity and improve coding speed. In this paper, we propose a Gaussian Distribution-based Mode Selection for Intra Prediction of SSHVC. We show that the rate distortion costs of Inter-layer Reference (ILR) mode and Intra mode are significantly different, and both follow a Gaussian distribution. Based on this discovery, we propose to use a Bayes decision rule to determine whether ILR is the best mode so as to skip Intra mode. Experimental results demonstrate that the proposed algorithm can significantly improve coding speed with negligible coding efficiency losses.

Index Terms— SHVC, ILR mode, Bayes decision rule, Gaussian distribution

1. INTRODUCTION

Video applications are becoming increasingly popular in recent years. These applications often involve different terminal devices with various display resolutions. Therefore, video streaming must be compatible with different screen resolutions. Spatial Scalable High Efficiency Video Coding (SSHVC) is an efficient solution to meet the requirement [1]. Video sequences with the same content but different spatial resolutions are encoded at a base layer (BL) and one or more enhancement layers (ELs) in SSHVC. SSHVC permits to decode the appropriate layers to adapt to a specific display resolution.

In SSHVC, video sequences at different layers have various resolutions, but the underlying video content presented are the same. In order to improve the coding efficiency, SSHVC allows the coding unit (CU) in EL to be predicted from the upsampled co-located CU in the BL. The mechanism is referred to as inter-layer prediction, and the coding mode involved is denoted as Inter-layer reference (ILR) mode. For the base layer, the input video is encoded by a HEVC encoder. Apart from the regular coding mode, additional ILR modes are required to be evaluated in the EL. Since the coding process of HEVC is already very complex, the encoding process of SSHVC is even more complex, which seriously limits its wide usage, especially for wireless and real-time applications. Consequently, there is a urgent need to develop fast SSHVC encoders to speed up the coding process.

A novel fast intra prediction algorithm is therefore proposed for SSHEVC in this paper. It is observed that the rate-distortion (RD) costs of the ILR mode and those of the Intra mode are significantly different, but both follow a Gaussian distribution. Upon these observations, a Bayes decision rule is proposed to early determine whether the ILR mode is optimal, thus skipping the evaluation of unnecessary Intra modes.

2. RELATED WORK

In order to accelerate the processing speed, commonly-used techniques include skipping of unlikely modes and early termination of evaluations. Some representative methods are reviewed and discussed as follows.

In general, the current CU and its spatial neighbours in the same EL and it co-locate CU in the BL are highly correlated. Therefore, the modes of relevant CUs are often used

This work was supported in part by the Natural Science Foundation of Chongqing under Grant cstc2020jcyj-msxmX0766; in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K202100604; in part by the Jiangxi Provincial Natural Science Foundation under Grant 20202BABL202006.

to predict the mode candidates for the current CU. To reduce the computational complexity, Tohidypour et al. [2] use the relevant CUs to predict possible coding mode and skip unlikely modes for CUs in EL. The algorithm proposed in [3] combines depth and the co-located CU's mode in BL to predict likely modes, and then jointly use Inter-layer and spatial correlations to eliminate unlikely modes. The algorithms in [2] and [3] are developed based on the Inter-layer and spatial correlations only. In fact, the texture complexity can also be used to assist the mode decision process. Lu et al. [4] and [5] jointly use the temporal-spatial correlation and the texture complexity to predict the mode candidates and exclude unlikely modes.

In order to reduce encoding time, RD costs and residual coefficients are also widely used to early terminate coding process. Given the fact that the RD costs of neighbouring CUs are similar, Tohidypour et al. [6] have consequently determined whether the current mode is optimal according to RD cost hence avoid the unnecessary evaluations. In our previous work [7], [8], the difference of RD costs for both ILR mode and the merge mode was compared to early terminate the mode selection process. RD costs formed the basis for the above algorithms.

Generally, most of the residual coefficients for the best mode are small and the distribution often obey a certain pattern, e.g. Gaussian distribution [9] or Laplacian distribution [10]. Wang et al. [11] make decisions on whether to skip the Intra mode or not by checking the distribution of the residual coefficients of ILR mode. Since the probabilities of modes are strongly relative to mode selection, Wang et al. [12] combine the probability of ILR mode with its residual coefficients to early skip Intra mode. The works in [13] [14] adopt zero blocks to early terminate mode selection.

Although the aforementioned algorithms can reduce the encoding time to some extent, the underlying mechanisms of mode selection have not yet been investigated. This limits the reduction in computational complexity of the SSHVC encoder. In this paper, a further investigation on mode selection between ILR mode and Intra mode is presented. Consequently, a fast mode selection algorithm for SSHVC is developed.

3. THE FUNDAMENTALS OF MODE SELECTION BETWEEN ILR MODE AND INTRA MODE

In order to investigate the fundamentals of mode selection between ILR mode and Intra mode, extensive experiments have been carried out. Video sequences including Blue_sky, Ducks, Pedestrian, Tractor, Town and Station2 are processed in our experiments. These sequences are representative since they cover different motion and texture from simple to complex. According to common SHM test conditions (CSTC) [15], we use QP with 22 in BL and 24 in EL in conducting experiments. Extensive experimental results show other QPs produce similar performance. In SHVC, Coding Tree Unit (CTU) can be encoded at four coding depth levels 0, 1, 2 and 3, corresponding to CU sizes of 64×64, 32×32, 16×16, and 8×8. The average values of RD Costs of both ILR mode and Intra mode at four coding depth levels are listed in Table 1.

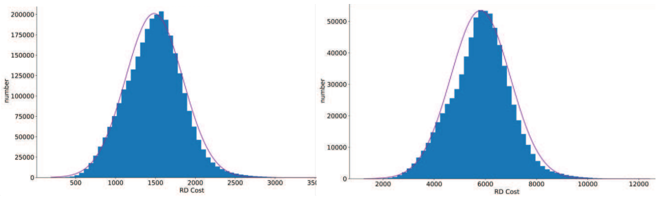
Table 1. The RD Costs of ILR mode and Intra mode in All Depths

Sequence	Depth 0		Depth 1		Depth 2		Depth 3	
	ILR	Intra	ILR	Intra	ILR	Intra	ILR	Intra
Blue_sky	23285	9296	5806	2243	1504	581	398	430
Ducks	91351	24112	22537	11784	5662	4848	1438	1512
Pedestrian	26302	10307	6534	5057	1646	1497	416	483
Tractor	35746	15127	8992	3386	2309	1396	606	615
Town	76524	57183	19207	14526	4850	3778	1138	1221
Station2	29918	10248	7498	3334	1898	1366	487	543
Average	47187	21045	11762	6722	2978	2244	747	801

As shown in Table 1, the average RD Cost of ILR mode is greater than that of Intra mode for depths 0, 1, and 2. It is the opposite for depth 3, but the difference in average RD costs is insignificant. The way to generate the predicted CU in EL explains this observation. To be specific, ILR mode uses the collocated CU in the BL to predict the current CU in the EL, while Intra mode refers to the reference pixels in the same layer. For CUs containing complex textures at depth levels [0,2], Intra mode is often unable to provide an accurate prediction, as the predicted pixels and the reference pixels are a great distance apart due to the large CU sizes. In this case, ILR mode which uses the co-located pixels in BL as reference usually performs better than Intra mode. On the contrary, when the current CU contains simple textures, Intra mode often makes a better prediction than ILR mode, as neighbouring CUs tend to contain similar textures. For CUs at depth level 3, the predicted pixels are very close to the reference pixels, as the size of the CUs is small. Consequently. the predicted pixels are similar to the reference pixels, even if the current CU contains rich details. Intra mode outperforms ILR mode in such a case. When the texture is simple, both Intra mode and ILR mode perform well in prediction. There is a margin difference only between the RD costs incurred by Intra mode and ILR mode.

In addition, the RD cost distributions of both ILR mode and Intra mode are further investigated. The RD cost distribution of ILR mode for the 'Ducks' sequence at depth levels 2 and 3 are shown in Fig.1. (a) and Fig.1 (b), respectively. In Fig.1, the x-axis is the RD cost, and the y-axis represents the histograms, i.e., the number of CUs in each bin.

It is observed from Fig.1 that both the RD costs of ILR mode at depth levels 2 and 3 follow a Gaussian distribution. Kolmogorov-Smirnov (KS) test, as a non-parametric distribution estimation method, is robust and widely used in distribu-



(a) Depth 2 (b) Depth 3

Fig. 1. The RD cost distribution of ILR mode for sequence “Ducks”.

tion estimation. It is verified by the KS test that, the RD costs of ILR mode at depth levels 2 and 3 both follow a Gaussian distribution. Similarly, this observation holds for the other video sequences.

It can be concluded that the average RD costs for both ILR mode and Intra mode are distinct, but they both follow a Gaussian distribution for all depth levels.

4. THE PROPOSED MODE SELECTION-BASED FAST INTRA PREDICTION ALGORITHM

Compared with the other coding modes, ILR mode is more likely to be selected as the best mode [11] [13], we therefore firstly evaluate this mode’s RD cost to determine whether it is the best mode, thus skipping Intra mode as early as possible. Suppose m_0 and m_1 refer to ILR mode and Intra mode respectively, x is the RD cost of the current CU, the probability of ILR mode being selected as the best mode can be derived by [16]

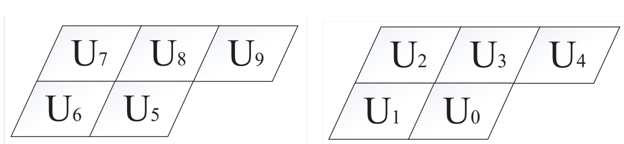
$$p\left(m_0|x\right) = \frac{p\left(x|m_0\right)p\left(m_0\right)}{p\left(x\right)}, \tag{1}$$

where $p(x|m_0)$is the conditional probability of x when ILR mode is chosen for the current CU, $p(m_0)$ is the probability of ILR mode being chosen, $p(x)$ is the probability of the RD cost being equal to x.

As there are two modes only, i.e. ILR mode and Intra mode, $p(x)$ is calculated by

$$p\left(x\right) = p\left(x|m_0\right)p\left(m_0\right) + p\left(x|m_1\right)p\left(m_1\right). \tag{2}$$

As the current CU is usually very similar to its relevant CUs, the latters can therefore be used to predict the current CU.

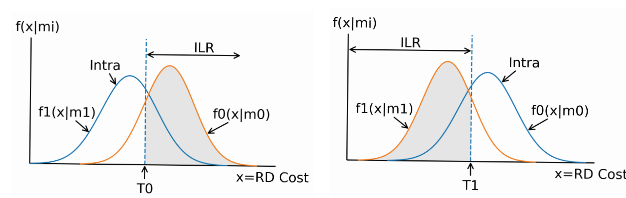


(a) Previous frame (b) Current frame

Fig. 2. The current CU and its relevant CUs.

As shown in Fig.2, U_0 represents the current CU, U_1, U_2, U_3 and U_4 are its neighboring CUs, U_5, U_6, U_7, U_8 and U_9 are the co-located CUs of U_0, U_1, U_2, U_3 and U_4 in the previous frame. Let k be the number of relevant CUs using ILR mode, since there are nine relevant CUs, from U_0 to U_9, $p(m_0)$ and $p(m_1)$ are $\frac{k}{9}$ and 1-$\frac{k}{9}$, respectively.

As mentioned above, the RD costs of both ILR and Intra modes follow a Gaussian distribution. For CUs at depth levels [0,2], the average RD costs of ILR mode is greater than that of Intra mode, the corresponding RD cost distributions are shown in Fig. 3(a). Conversely, for CUs at depth level 3, the average RD costs of ILR mode is less than that of Intra mode, the corresponding RD cost distributions are shown in Fig. 3(b).



(a) depths from 0 and 2 (b) depths 3

Fig. 3. RD cost distributions of ILR and Intra modes at different depth levels.

In Fig.3, the x-axis represents the RD cost, the y-axis represents the probability density function. The blue curve is the probability density function of RD cost when using Intra mode, the yellow curve is the probability density function of RD cost when using ILR mode. T_0 and T_1 are the thresholds used in the mode decision process in Fig. 3(a) and Fig. 3(b), respectively. In Fig. 3(a), if $x \geq T_0$, ILR mode is selected as the best one. While in Fig. 3(b), if $x \leq T_1$, ILR mode is selected as the best one. Therefore, the early termination condition for ILR mode can be written as

$$\left\{ \begin{array}{l} x \geq T_0 \\ x \leq T_1 \end{array} \right. \tag{3}$$

Consequently, if one of the conditions is satisfied, Intra mode can be directly skipped. The key here is to determine suitable T_0 and T_1. Suppose $f_0\left(x|w_0\right)$ is the conditional probability density function of x when ILR mode is chosen, the corresponding probability of T_0 can be calculated by

$$p\left(T_0|m_0\right) = \int_{T_0}^{\infty} f_0\left(x|w_0\right)dx. \tag{4}$$

Eq. (4) can be rewritten as

$$p\left(T_0|m_0\right) = 1 - \int_{-\infty}^{T_0} f_0\left(x|m_0\right)dx. \tag{5}$$

Suppose $f_1\left(x|m_1\right)$ is the conditional probability density function of x when Intra mode is chosen, we can calculate the corresponding probability of T_0 as shown below

$$p\left(T_0|m_1\right)=1-\int_{-\infty}^{T_0} f_1\left(x|m_1\right)dx. \tag{6}$$

Suppose α is the probability of using ILR mode, according to Eq. (1), it can be derived as

$$\alpha=\frac{p\left(m_0\right)\left(1-\int_{-\infty}^{T_0} f_0\left(x|m_0\right)dx\right)}{p\left(m_0\right)\left(1-\int_{-\infty}^{T_0} f_0\left(x|m_0\right)dx\right)+p\left(m_1\right)\left(1-\int_{-\infty}^{T_0} f_1\left(x|m_1\right)dx\right)}. \tag{7}$$

Obviously, if $\alpha \geq 0.95$, the probability of using ILR mode is very high, so ILR is very likely to be chosen as the best mode. Therefore, α is set to 0.95. The current CU often contains highly-similar video content with the relevant CUs. When checking ILR mode, the expectation μ_0 and variance σ_0 of the RD costs of the relevant CUs using ILR mode are used to estimate the corresponding values of the current CU. Similarly, when checking Intra mode for the current CU, the expectation μ_1 and variance σ_1 of the RD costs generated from the relevant CUs using Intra mode are used for estimations. By checking the cumulative standard normal distribution table, we can obtain upper bound of the integral of the cumulative standard normal distribution z, then T_0 can be calculated as

$$T_0=\mu_0-z\cdot\sigma_0. \tag{8}$$

In a similar way, we can obtain a threshold T_1 for depth level 3.

5. EXPERIMENTAL RESULTS

To test the performance of the proposed fast Intra prediction algorithm for SSHVC, we implement our algorithm on the reference software (SHM 11.0) with AI configuration and carry out testing on a server with Intel (R) 2.0 GHz CPU and 30 GB memory. Commonly-used performance measures are coding efficiency and coding speed. Coding efficiency in EL is measured by Bjontegaard Delta Bit Rate (BDBR) [17], which represents the bitrate difference at an equal PSNR compared with the reference software. Coding speed is denoted by TS, which is the percentage of encoding run-time savings only for EL.

In order to demonstrate the performance of the proposed algorithm, we compare the performance of our algorithm with FIICA [5] and PBFIP (%) [12]. According to common SHM test conditions (CSTC) [15], only one QP set is used for BL: (22, 26, 30, 34) and two QP sets for EL: (22, 26, 30, 34) and (24, 28, 32, 36). We denote QP sets of (22, 26, 30, 34) and (24, 28, 32, 36) for EL as case 1 and case 2. The corresponding performance comparisons are listed in Table 2 (case 1) and Table 3 (case 2), respectively.

In Table 2 (case 1), the average BDBRs of the proposed algorithm, FIICA and PBFIP are 0.26%, 0.38% and 0.58%, respectively, while the average TS of the proposed algorithm, FIICA and PBFIP are 55.85%, 42.22% and 41.27% correspondingly. In Table 3 (case 2), the average BDBRs of the proposed algorithm, FIICA and PBFIP are 0.24%, 0.40% and 0.43%, respectively, while the average TS of the proposed algorithm, FIICA and PBFIP are 58.91%, 43.25% and 47.81% correspondingly. In both case 1 and case 2, the proposed algorithm achieve the lowest BDBR increment and significantly better encoding time reduction than the other two algorithms.

Table 2. Overall performance comparisons with case 1

Sequence	Proposed (%)		FIICA(%) [5]		PBFIP(%) [12]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	0.20	60.5	0.41	36.37	0.08	44.61
PeopleOnStreet	-0.20	58.58	0.10	39.43	0.03	40.30
Kimono	-0.20	61.80	-0.13	60.27	-0.22	53.30
ParkScene	-0.10	59.97	0.22	36.49	-0.16	32.32
Cactus	0.50	55.57	0.89	37.92	0.91	39.41
BasketballDrive	0.80	48.85	0.71	41.48	2.09	45.61
BQTerrace	0.80	45.66	0.50	43.56	1.30	33.30
Average	0.26	55.85	0.38	42.22	0.58	41.27

Table 3. Overall performance comparisons with case 2

Sequence	Proposed (%)		FIICA(%) [5]		PBFIP(%) [12]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.10	64.19	-0.30	37.89	0.01	45.82
PeopleOnStreet	-0.30	63.70	-0.23	40.15	-0.23	42.65
Kimono	-0.30	63.69	0.19	60.18	-0.21	59.21
ParkScene	-0.10	62.10	0.11	38.13	0.11	41.81
Cactus	0.60	57.99	0.70	39.29	0.52	48.09
BasketballDrive	1.00	51.04	1.72	42.74	1.72	50.45
BQTerrace	0.90	49.68	0.61	44.37	1.09	46.62
Average	0.24	58.91	0.40	43.25	0.43	47.81

6. CONCLUSION

In this paper, we first investigate the fundamentals of mode selection, and then propose to use a Bayes decision rule to speed up the coding process of Intra prediction algorithm for SSHVC. Experimental results show that the proposed algorithm can reduce the encoding time effectively with negligible coding efficiency losses.

7. REFERENCES

[1] J. M. Boyce Y. Yan J. Chen and A. K. Ramasubramonian, "Overview of shvc: Scalable extensions of the

high efficiency video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, pp. 20–34, January 2016.

[2] H. R. Tohidypour M. T. Pourazad and P. Nasiopoulos, "Content adaptive complexity reduction scheme for quality/fidelity scalable hevc," in *Int. Conf. Acoust., Speech Signal Process*. IEEE, 2013, pp. 1744–1748.

[3] H. R. Tohidypour H. Bashashati M. T. Pourazad and P. Nasiopoulos, "Fast mode assignment for quality scalable extension of the high efficiency video coding (hevc) standard: A bayesian approach," in *6th Balkan Conf. Informat*, 2013, pp. 61–65.

[4] X. Lu C. Yu Y. F Gu and G. Martin, "A fast intra coding algorithm for spatial scalability in shvc," in *Int. Conf. Image Processing (ICIP)*. IEEE, 2018, pp. 1792–1796.

[5] X. Lu C.Yu and G.R.Martin, "Fast intra-and inter-coding algorithms for the spatially scalable extension of h.265/hevc," *Multimedia Tools and Applications*, vol. 79, pp. 26447–26465, July 2020.

[6] H. R. Tohidypour M. T. Pourazad and P. Nasiopoulos, "An encoder complexity reduction scheme for quality/fidelity scalable hevc," *IEEE Trans. Broadcast*, vol. 62, pp. 664–674, September 2016.

[7] D. Wang Y. Sun W. Li C. Zhu F. Dufaux, "Fast inter mode predictions for shvc," in *Int. Conf. Multimedia Expo (ICME)*. IEEE, 2019, pp. 1696–1701.

[8] D. Wang Y. Sun W. Li C. Zhu F. Dufaux, "Fast depth and inter mode prediction for quality scalable high efficiency video coding," *IEEE Trans. Multimedia*, vol. 22, pp. 833–845, April 2020.

[9] I.-M. Pao and M.-T. Sun, "Modeling dct coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol*, vol. 9, pp. 608–616, June 1999.

[10] N. Hu and E. H. Yang, "Fast motion estimation based on confidence interval," *IEEE Trans. Circuits Syst. Video Technol*, vol. 24, pp. 1310–1322, August 2014.

[11] D. Wang C. Zhu Y. Sun F. Dufaux and Y. Huang, "Efficient multi-strategy intra prediction for quality scalable high efficiency video coding," *IEEE Trans. Image Process*, vol. 28, pp. 2063–2074, April 2019.

[12] D. Wang Y. Sun J. Liu F. Dufaux Xi. Lu and B. Hang, "Probability-based fast intra prediction algorithm for spatial shvc," *IEEE Trans. Broadcast*, DOI: 10.1109/TBC.2021.3126277.

[13] D. Wang Y. Sun C. Zhu W. Li F. Dufaux and J. Luo, "Fast depth and mode decision in intra prediction for quality shvc," *IEEE Trans. Image Process*, vol. 29, pp. 6136–6150, December 2020.

[14] Z. Pan S. Kwong M. T. Sun and J. Lei, "Early merge mode decision based on motion estimation and hierarchical depth correlation for hevc," *IEEE Trans. Broadcast*, vol. 60, pp. 405–412, June 2014.

[15] "Common shm test conditions and software reference configurations," in *document JCTVC-Q1009 ITU-T SG 16 WP 3 ISO/IEC JTC1/SC29/WG*.

[16] S. Cho and M. Kim, "Fast cu splitting and pruning for suboptimal cu partitioning in hevc intra coding," *IEEE Trans. Circuits Syst. Video Technol*, vol. 23, pp. 1555–1564, September 2013.

[17] G. Bjontegaard, "Calculation of average psnr difference between rd-curves," in *Proc. 13th VCEG-M33 Meet*. IEEE, 2001, pp. 1–4.