A NOVEL MODE SELECTION-BASED FAST INTRA PREDICTION ALGORITHM FOR SPATIAL SHVC

Dayong Wang[1], Yu Sun[2], Weisheng Li[3], Lele Xie[4], Xin Lu[5], Frederic Dufaux[6], Ce Zhu[7]

[1]Chongqing Key Laboratory on Big Data for Bio Intelligence,[3]Chongqing Key Laboratory of Image Cognition,[4]School of Information and Communication Engineering, Chongqing University of Posts and Telecommunications.
[2]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,
[7]School of Information and Communication Engineering,
University of Electronic Science and Technology of China
wangdayong@cqupt.edu.cn, yusun@uca.edu, liws@cqupt.edu.cn, 1435439126@qq.com,
xin.lu@dmu.ac.uk, frederic.dufaux@l2s.centralesupelec.fr, eczhu@uestc.edu.cn

ABSTRACT

Due to multi-layer encoding and Inter-layer prediction, Spatial Scalable High-Efficiency Video Coding (SSHVC) has extremely high coding complexity. It is very crucial to improve its coding speed so as to promote widespread and cost-effective SSHVC applications. In this paper, we have proposed a novel Mode Selection-Based Fast Intra Prediction algorithm for SSHVC. We reveal the RD costs of Inter-layer Reference (ILR) mode and Intra mode have a significant difference, and the RD costs of these two modes follow Gaussian distribution. Based on this observation, we propose to apply the classic Gaussian Mixture Model and Expectation Maximization in machine learning to determine whether ILR is the best mode so as to skip the Intra mode. Experimental results demonstrate that the proposed algorithm can significantly improve the coding speed with negligible coding efficiency loss.

Index Terms— SHVC, ILR mode, rate distortion costs, GMM-EM

1. INTRODUCTION

VIDEO applications, such as digital TV broadcasting, video conferencing, wireless video streaming, and smartphone communications, are more and more widely used in our daily life. At the same time, more and more new terminal devices with different spatial resolutions are emerging. This requires that video streaming must be adaptive to different screen resolutions. Scalable High-Efficiency Video Coding (SHVC) is an efficient solution to this requirement. In order to adapt to different screen resolutions, Spatial SHVC (SSHVC) encodes different layers with different spatial resolutions. By selecting an appropriate layer, SSHVC can adapt to different screen resolutions.

SSHVC consists of a base layer (BL) and one and more enhancement layers (ELs). A BL only has Intra-layer prediction, while an EL further includes Inter-layer prediction. The coding process of Intra-layer prediction is the same as that of HEVC. Since the contents between BL and EL are the same except for their different resolutions, a coding unit (CU) in BL can be up-sampled to predict the co-located CU in EL. This prediction process is denoted as Inter-layer prediction, and its prediction mode is called Inter-layer Reference (ILR) mode. The SSHVC encoding process is very complex [1], which definitely restricts its wide applications, especially for wireless and real-time applications. Therefore, it is crucial to accelerate its encoding speed.

For this purpose, in this paper, we propose a novel fast intra prediction algorithm based on Mode Selection for SSHVC.

2. RELATED WORK

In SHVC [1], each CU contains both ILR mode and Intra mode which includes 35 DMs. By checking both modes to select the best one, SHVC can obtain the best coding efficiency. However, doing so can result in a very complicated coding process. In order to improve coding speed, a number of algorithms have been developed, which are reviewed and discussed below.

Generally speaking, the current CU and its relevant CUs are very similar, so relevant CUs can be used to predict candidate coding modes. Tohidypour et al. [2–4] use relevant CUs to predict likelihood modes and skip unlikely modes in EL. Wang et al. [5], [6] and [7] first check ILR mode and merge mode, and then compare the difference of their RD costs to early terminate mode selection. The above algorithms are proposed based on correlations only. Lu et al. [8] and [9] jointly use texture complexity and spatio-temporal correlation to predict candidate modes.

If a mode is predicted very well, its residual coefficients should be very small and follow the Gaussian distribution [10] or the Laplacian distribution [11]. The research in [7] also first checks ILR mode and then calculates its part-zero block based on the distribution of its residual coefficients to early terminate mode selection. Wang et al. [12] first check ILR mode and then decide whether its residual coefficients follow Gaussian distribution so as to early skip Intra mode. Pan et al. [13] combine depth correlation and all-zero block to early terminate mode selection. Obviously, a mode's probability of being selected as the best mode is strongly related to mode selection. Wang et al. [14] combine probabilities of ILR mode with its residual coefficients to early skip Intra mode.

Although all the above algorithms can improve coding speed, the underlying mechanisms of mode selection have not yet been investigated, which hinders the further improvement of SSHVC coding speed. Therefore, we develop a novel mode-distribution-based fast Intra prediction algorithm for SSHVC.

3. JUSTIFICATION OF THE PROPOSED MODE SELECTION ALGORITHM

In order to improve the coding speed, we have conducted extensive experiments to investigate Intra coding in SSHVC. In order to ensure the generality of features and rules, different types of sequences, including motion and texture from simple to complex, are selected. More specifically, we use Blue-sky, Ducks, Park_Joy, Pedestrian, Tractor, Town and Station2 in our experiments. According to common SHM test conditions (CSTC) [15], QPs set in BL and In EL are set as (22, 26, 30, 34) and (24, 28, 32, 36) in conducting experiments, respectively.

We can obtain the RD costs of ILR mode and Intra mode, which are listed in Table 1.

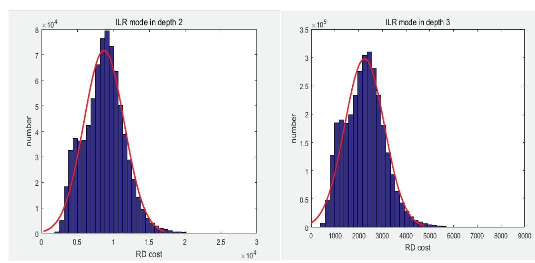
From Table 1, we can observe that the average values of the RD costs of ILR mode at depths from 0 to 2 are significantly larger than those of Intra mode. In contrast, for depth 3, the average value of the RD costs of ILR mode is smaller than that of Intra mode, but not very significantly.

In addition to the RD cost relationships of ILR mode and Intra mode, we further investigate their RD cost distribution. We have conducted extensive experiments on the RD cost distribution of ILR mode and Intra mode. In Fig. 1, the horizontal axis represents RD cost, and the vertical axis represents the histograms, i.e., the corresponding number of CUs in each bin. Fig. 1 (a) and Fig. 1 (b) show the RD cost distribution of

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	49801	11074	12172	3198	3039	1523	778	946
Ducks	140109	49873	34478	13686	8701	5732	2247	2444
Park_Joy	166607	21382	41191	5616	10308	3066	1888	2621
Pedestrian	37458	26252	9333	7063	2351	2189	609	763
Tractor	57344	22816	14445	5734	3719	2433	988	1189
Town	108752	63737	27495	16467	6875	4908	1721	1981
Station2	44750	12625	11201	4611	2839	2300	742	901
Average	86403	29680	21474	8054	5405	3164	1282	1549

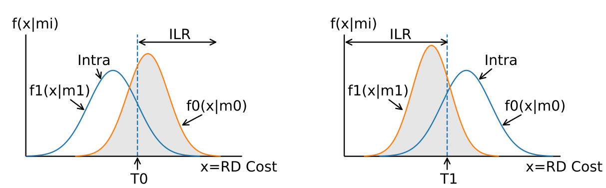
ILR mode for sequence "Blue_sky" in depth 2 and depth 3, respectively. From Fig. 1, we can observe that both the RD costs of ILR in depths 2 and 3 follow the Gaussian distribution. Our extensive experimental results have demonstrated that the RD cost distribution of both ILR mode and Intra mode in all sequences follows Gaussian distribution.



(a) Depth 2 (b) Depth 3

Fig. 1. The RD cost distribution of ILR mode for sequence "Blue_sky"

The RD cost distributions for both the ILR mode and the Intra mode at coding depths [0, 2] and coding depth of 3 are shown in Fig. 2 (a) and Fig. 2 (b), respectively.



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

Fig. 2. RD cost distributions of ILR and Intra modes at different depth levels.

4. GMM-EM-BASED MODE SELECTION

It is observed from Fig. 2 that, despite the different average RD costs, the distribution for the ILR mode overlaps with that for the Intra mode. Therefore, it is impossible to directly decide which is the better choice. GMM-EM is very suitable for classifying the data that follow the Gaussian distribution but have significantly different average values.

GMM-EM is a widely used machine learning algorithm for clustering. It uses mixed Gaussian distribution as the parametric model and utilizes the maximum Expectation (EM) algorithm for training. Since the average value of RD costs of both ILR mode and Intra mode follow Gaussian distribution, we add them into a mixed Gaussian distribution model. Based on the model, the probability of sample i belonging to each part can be derived based on the currently available mixture parameters. Then, the mixture parameters are refined. Repeating the process until converge, then the probabilities of all samples belonging to each part can be obtained.

Since ILR mode is more likely to be selected as the best mode [7] [12], we encode the current CU using ILR mode first and then use GMM-EM to determine whether ILR mode is the best mode based on its RD cost. In the affirmative, we can early terminate mode selection; otherwise, we need to further check Intra mode.

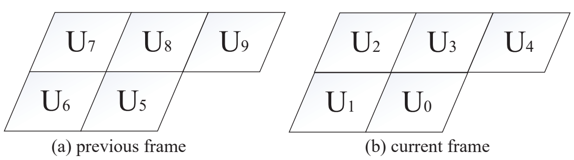


Fig. 3. The current CU and its relevant CUs

As shown in Fig. 3, U_0 represents the current CU, U_1, U_2, U_3 and U_4 are the neighbouring CUs of the current CU, 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. Since there are only two parts, i.e., ILR mode and Intra mode, let part one represent ILR mode, and part two indicates Intra mode. The corresponding Gaussian Mixture Model can be written as:

$$p\left(rd|\pi,\mu,\Sigma\right)=\pi_1 N\left(rd|\mu_1,\Sigma_1\right)+\pi_2 N\left(rd|\mu_2,\Sigma_2\right), \quad (1)$$

where rd is the RD cost of a CU, π_1 is the probability of relevant CUs using ILR mode, μ_1 and Σ_1 are respectively the expected value and the variance of the RD costs of these CUs selecting ILR mode as the best mode; and π_2 is the probability of relevant CUs using Intra mode, μ_2 and Σ_2 are respectively the expected value and the variance of RD costs of these CUs selecting Intra mode as the best mode.

In order to obtain these six parameters, based on Eq. (1), the corresponding maximum likelihood estimation function is:

$$f=\prod_{i=1}^{M} p\left(rd|\pi,\mu,\Sigma\right)=\prod_{i=1}^{M}\left(\pi_1 N\left(rd|\mu_1,\Sigma_1\right)+\pi_2 N\left(rd|\mu_2,\Sigma_2\right)\right), \quad (2)$$

where M is the number of the current CU and its relative CUs, and it is equal to 10.

The logarithm of the maximum likelihood function is:

$$\log\left(f\right)=\sum_{i=1}^{M}\log\left(\pi_1 N\left(rd_i|\mu_1,\Sigma_1\right)+\pi_2 N\left(rd_i|\mu_2,\Sigma_2\right)\right), \quad (3)$$

where π_k, μ_k and Σ_k (k=1 or 2) can be calculated by:

$$\frac{\partial\log\left(f\right)}{\partial\pi_k}=0, \frac{\partial\log\left(f\right)}{\partial\mu_k}=0, \frac{\partial\log\left(f\right)}{\partial\Sigma_k}=0. \quad (4)$$

We can derive:

$$\mu_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma\left(i,k\right)rd_i, \Sigma_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma\left(i,k\right)\left(rd_i-\mu_k\right)\left(rd_i-\mu_k\right), \quad (5)$$

where $N_k=\sum_{i=1}^{N}\gamma\left(i,k\right)$, then we have:

$$\pi_k=\frac{N_k}{N}. \quad (6)$$

$\gamma\left(i,k\right)$ is the probability that the i-th CU, denoted as U_i, belongs to the k-th part, and it can be obtained by:

$$\gamma\left(i,k\right)=\frac{\pi_k N\left(x_i|\mu_k,\Sigma_k\right)}{\pi_1 N\left(x_i|\mu_1,\Sigma_1\right)+\pi_2 N\left(x_i|\mu_2,\Sigma_2\right)}. \quad (7)$$

In order to start iterations, we must obtain their initial values first. For relative CUs selecting ILR mode as the best mode, we calculate the expected value and variance of their RD costs as initial μ_1 and Σ_1, and set the ratio between the number of these CUs and that of all the relevant CUs as initial π_1. Using the same way, we can obtain initial μ_2, Σ_2 and π_2. After these six values are obtained, we use Eq. (7) to derive initial $\gamma\left(i,k\right)$. Repeat iterations (5), (6) and (7) until $\gamma\left(i,k\right)$ converges.

As mentioned above, the current CU is U_0 and ILR mode is part one, hence $\gamma\left(0,1\right)$ denotes the probability of the current CU selecting ILR mode as the best one. We denote the i-th iteration of $\gamma\left(0,1\right)$ as $\gamma_i\left(0,1\right)$. In order to avoid unnecessary iterations, if the absolute difference between $\gamma_{i-1}\left(0,1\right)$ and $\gamma_i\left(0,1\right)$ is small enough, we can terminate the iteration. We empirically select 0.01 as the threshold, then we obtain the early termination condition below:

$$\left|\gamma_i\left(0,k\right)-\gamma_{i-1}\left(0,k\right)\right|\leq 0.01. \quad (8)$$

If condition (8) is met, we can terminate the iteration and obtain the probability of the current CU selecting ILR mode as the best mode. Since this probability is obtained based on RD cost, we define it as the RD-based ILR probability.

Since relevant CUs are usually very similar, we can use them to predict the probability of the current CU using ILR mode. Obviously, if more relevant CUs use ILR mode, the current CU is more likely to use this mode, and vice versa. In other words, the probability of the current CU using ILR mode is proportional to the number of relevant CUs using ILR mode. As shown in Fig. 3, there are 9 relevant CUs, so we simply set the possibility of the current CU using ILR mode as $\frac{k}{9}$, where k is the number of relevant CUs using ILR mode. Since this probability is obtained based on the number of relevant CUs using ILR mode, we define it as the number-based ILR probability.

Since both the RD-based ILR probability and the number-based ILR probability strongly relate to ILR mode selection, we combine them to further predict the probability of the current CU using ILR mode. Let p(A) and p(B) denote the RD-based ILR probability and the number-based ILR probability, respectively. Since both of them are independent, we further derive the probability of the current CU selecting ILR mode as the best mode, p_r, by:

$$p_r = p\left(A + B\right) = p(A) + p(B) - p(A)p(B). \qquad (9)$$

Theoretically, if $p_r \geq 0.95$, the current CU is very possible to use ILR mode. Therefore, we select 0.95 as the threshold value for p_r.

5. EXPERIMENTAL RESULTS

In order to assess the performance of the proposed Mode Selection-based fast Intra prediction algorithm for SSHVC, we use the reference software (SHM 11.0) and test the proposed algorithm on a server with Intel (R) 2.0 GHz CPU and 30 GB memory. The performances of algorithms are evaluated by coding efficiency and coding speed. Coding efficiency is measured by bitrate and visual quality together, which is indicated by BDBR [16]. It refers to the bitrate difference at an equal PSNR compared with the reference software in EL. Coding speed improvement is denoted by TS, which is the percentage of encoding run-time savings only in EL.

In order to demonstrate the performance of the proposed algorithm, we conduct a performance evaluation of our algorithm with FIICA [9] and PBFIP [14]. To the best of our knowledge, these two algorithms are the best and most recent relative algorithms for SSHVC. For fair comparisons, all algorithms are tested on the same computing platform. According to the CSTC [15], the QPs in the BL are set to (22, 26, 30, 34), and the corresponding QPs in the ELs are set to (22, 26, 30, 34) and (24, 28, 32, 36), respectively. We denote QP sets of (22, 26, 30, 34) and (24, 28, 32, 36) for EL as case 1 and case 2, respectively.

In Table 2 (case 1), the average BDBRs of the proposed algorithm, FIICA and PBFIP are 0.36%, 0.38 and 0.58%, respectively. While the average TS of the proposed algorithm, FIICA and PBFIP are 53.65%, 42.22% and 41.27% correspondingly. In Table 3 (case 2), the average BDBRs of the proposed algorithm, FIICA and PBFIP are 0.31%, 0.40% and 0.43%, respectively. While the average TS of the proposed algorithm, PAPS, EETBS and FIICA are 59.39%, 43.25% and 47.81% correspondingly. Compared with the other two algorithms, we can observe that the average BDBRs of the proposed algorithm are smaller, meanwhile, the average TSs of the proposed algorithm are significantly faster in both cases.

Table 2. Overall performance comparisons with case 1

Sequence	Proposed (%)		FIICA(%) [9]		PBFIP(%) [14]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.04	58.21	0.41	36.37	0.08	44.61
PeopleOnStreet	0.23	55.44	0.10	39.43	0.03	40.30
Kimono	-0.23	65.42	-0.13	60.27	-0.22	53.30
ParkScene	-0.17	62.51	0.22	36.49	-0.16	32.32
Cactus	0.59	53.23	0.89	37.92	0.91	39.41
BasketballDrive	0.96	46.76	0.71	41.48	2.09	45.61
BQTerrace	1.18	33.95	0.50	43.56	1.30	33.30
Average	0.36	53.65	0.38	42.22	0.58	41.27

Table 3. Overall performance comparisons with case 2

Sequence	Proposed (%)		FIICA(%) [9]		PBFIP(%) [14]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.18	64.02	-0.30	37.89	0.01	45.82
PeopleOnStreet	-0.29	62.66	-0.23	40.15	-0.23	42.65
Kimono	-0.30	66.29	0.19	60.18	-0.21	59.21
ParkScene	-0.15	65.04	0.11	38.13	0.11	41.81
Cactus	0.41	60.33	0.70	39.29	0.52	48.09
BasketballDrive	0.96	54.42	1.72	42.74	1.72	50.45
BQTerrace	1.69	42.95	0.61	44.37	1.09	46.62
Average	0.31	59.39	0.40	43.25	0.43	47.81

6. CONCLUSION

In this paper, we fully investigate the special features of SSHVC and propose a new Mode Selection-based Intra prediction algorithm for SSHVC. Experimental results show that the proposed algorithm can improve the coding speed significantly with negligible coding efficiency losses. Deep learning has been a hot research topic recently, we will plan to explore it to further improve the coding speed in our future work.

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 Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2016.

[2] H. R. Tohidypour M. T. Pourazad and P. Nasiopoulos, “Content adaptive complexity reduction scheme for quality/fidelity scalable HEVC,” 2013.

[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,” 6th Balkan Conf. Informat. BCI, Thessaloniki, Greece, pp. 61–65, 2013.

[4] H. R. Tohidypour M. T. Pourazad and P. Nasiopoulos, “An encoder complexity reduction scheme for Quality/Fidelity scalable HEVC,” *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 664–674, Sep. 2016.

[5] D. Wang Y. Sun W. Li C. Zhu F. Dufaux, “Fast Inter Mode Predictions for SHVC,” IEEE Int. Conf. Multimedia Expo (ICME), Shanghai, China, Jul. 2019, pp.1696-1701.

[6] 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, Apr. 2020.

[7] 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, Dec. 2020.

[8] X. Lu C. Yu Y. F Gu and G. Martin, “A fast Intra coding algorithm for spatial scalability in SHVC,” IEEE Int. Conf. Image Processing (ICIP), Athens, Greece, Oct. 2018, pp. 1792-1796.

[9] X. Lu C.Yu and G.R.Martin, “Fast intra-and intercoding algorithms for the spatially scalable extension of H.265/HEVC,” *Multimedia Tools and Applications*, vol. 79, no. 35, pp. 26447–26465, Sept. 2020.

[10] 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, Jun. 1999.

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

[12] 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, Apr. 2019.

[13] 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, Jun. 2014.

[14] D. Wang Y. Sun J. Liu F. Dufaux Xi. Lu and B. Hang, “Probability-Based Fast Intra Prediction Algorithm for Spatial SHVC,” *IEEE Transactions on Broadcasting*, vol. 68, no. 1, pp. 83–96, 2022.

[15] “Common SHM Test Conditions and Software Reference Configurations,” document JCTVC-Q1009, ITU-T SG 16 WP 3 ISO/IEC JTC1/SC29/WG., Mar. 2014.

[16] G. Bjontegaard, “Calculation of Average PSNR Differences between RD-curves,” in *Proc. 13th VCEG-M33 Meet., Austin, TX, USA*, Apr. 2001, pp. 1–4.