Improving SHVC Performance with a Joint Layer Coding Mode

Xiem HoangVan[1], João Ascenso[2], Fernando Pereira[3]
[1]*VNU – University of Engineering and Technology, Hanoi, Vietnam*
[1,2,3]*Instituto Superior Técnico – Instituto de Telecomunicações, Lisbon, Portugal*
hoang.xiem@lx.it.pt, joao.ascenso@lx.it.pt, fp@lx.it.pt

***Abstract*— The growing need for a powerful scalable video coding engine targeting the heterogeneous landscape of network, devices, and consumption environments has led to the development of the Scalable High Efficiency Video Coding (SHVC) standard, an extension of the High Efficiency Video Coding (HEVC) standard. To improve the SHVC compression efficiency, this paper proposes a novel joint layer coding mode to be integrated in the SHVC codec. In the proposed coding mode, the base layer (BL) and enhancement layer (EL) decoded information are linearly combined at the pixel level to create an additional coding mode. To fuse the BL and EL driven predictions, a weighting term is defined to indicate the contributions of each of them for the final joint layer prediction. To reach high adaptability, these weights are computed at pixel level in the prediction unit. Moreover, to achieve the highest compression efficiency, the proposed joint layer coding mode is adaptively selected using a rate distortion optimization (RDO) mechanism. Experiments conducted for a rich set of test conditions have shown that significant compression efficiency gains can be achieved with the proposed joint layer coding mode, notably up to 4.3 % in BD-Rate savings regarding the standard SHVC quality scalable codec.**

Index Terms – *HEVC, SHVC, best prediction, joint layer mode*

I. INTRODUCTION

Multimedia applications have been playing a major role in the current society with video coding technologies largely driving the development of new services and applications with increasing quality of experience. Applications such as multimedia messaging, video conferencing, video surveillance, digital television, and mobile and Internet streaming typically deploy a powerful video compression engine, following the so-called predictive coding paradigm, largely adopted by the available video coding standards. The state-of-the-art on predictive video coding is nowadays represented by the recent High Efficiency Video Coding standard [1], which targets all resolutions, now up to ultra-high definition (UHD) video content. As usual, this standard provides about 50% bitrate reduction for the same perceptual quality comparing to the previous standards, notably the largely market deployed H.264/AVC (Advanced Video Coding) standard [2].

While the H.264/AVC and HEVC standards are highly efficient, they are not able to provide the adaptation capabilities necessary for the growing heterogeneity of networks, devices, consumption environments and user preferences. Moreover, several connection types may have different capabilities and characteristics along time, thus asking for a dynamic adaptation of the transmitted video streams in order the best user experience is provided. To address these adaptability requirements, a scalable video coding extension of the H.264/AVC standard, named Scalable Video Coding (SVC) standard, was developed around 2007 [3]. The SVC standard adopts a hierarchical layered design where subsets of the video bitstream can be decoded with different qualities or/and spatial and temporal resolutions commensurate to the consumed bitrate. However, the posterior development of the HEVC standard, with its 50% perceptual compression gains compared to H.264/AVC, and the increasing market relevance of heterogeneous and dynamic transmission environments, have boosted the development of a new scalable video coding standard, with a compression performance far beyond the SVC standard. In the new HEVC scalable extension (SHVC) [4], again a layered coding approach has been adopted as in the prior SVC standard [3]. However, SHVC does not follow the same conceptual approach adopted in the SVC standard [3] where new macroblock-level signaling capabilities are defined to indicate whether the EL macroblock is predicted from the BL or the EL layers. Instead, in SHVC, the BL reconstructed picture is taken as an inter-layer reference (ILR) picture to be included in the EL prediction buffer, eventually after some inter-layer processing. In this context, the standard HEVC reference index signaling capabilities are enough to identify whether the EL block level prediction comes from the BL or the EL. The main advantage is that this different scalable coding approach requires changes only in the HEVC high level syntax (HLS) and no changes in terms of the HEVC block level coding process, thus increasing the compatibility between HEVC and SHVC and easing its implementation and deployment.

To further improve the SHVC compression efficiency, this paper proposes a novel joint layer coding mode which should be taken as an additional coding mode to be added to the already available SHVC EL coding modes. In the proposed joint layer coding mode, the best conventional EL prediction and the BL reconstructed information are linearly combined at pixel level to create a so-called *joint layer prediction. This type of prediction mode does not exist in any scalable video coding standard*. In this context, the weights to fuse the BL reconstruction and conventional EL best prediction play a critical role for the final joint layer prediction quality. Therefore, this paper proposes an adaptive mechanism to compute the pixel level weights that considers not only the decoded information associated to the current pixel but also its neighboring pixels. Finally, the joint layer prediction and the conventional best EL prediction are adaptively selected using a RDO mechanism. To create a similar prediction at the decoder, a binary flag is added to the bitstream to indicate the selected coding mode. Experimental results shown that a significant SHVC quality scalability compression performance gain can be achieved with the proposed joint coding mode, notably up to 4.3 % BD-Rate reduction regarding the SHVC standard.

After this Introduction, this paper is organized as follows: Section II briefly summarizes the relevant background work, notably the standard SHVC EL coding modes. Next, Section III describes the proposed joint layer coding mode to be integrated in the quality SHVC codec. After, Section IV evaluates the proposed solution regarding the relevant benchmarks in terms of compression efficiency. Finally, Section V presents the main conclusions and ideas for future work.

II. BACKGROUND WORK

As mentioned above, the background codec for the proposed novel prediction tool is the SHVC standard. Therefore, this section will briefly review the prediction tools adopted in the SHVC standard and the related work on improving the SHVC compression efficiency.

SHVC prediction brief review: Similar to HEVC, the prediction unit (PU) in SHVC is defined to efficiently code a coding unit (CU) using either Inter prediction or Intra prediction coding modes. A CU can be split into one of eight partitions: PART_2N×2N, PART_N×N, PART_2N × N, PART_N × 2N, PART_2N × nU, PART_2N × nD, PART_nL×2N, and PART_nR×2N as illustrated in Figure 1.

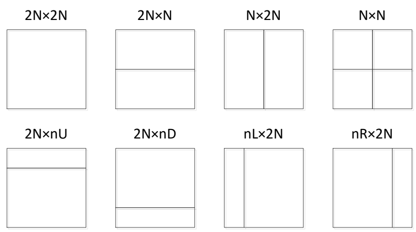

Figure 1. Possible CU partitions in SHVC.

While these eight partition modes can be used for Inter-coded CUs, only two square partitions, PART_2N × 2N and PART_N × N, are defined for Intra-coded CUs. As usual, Intra prediction exploits only the decoded spatially adjacent information to create the PU prediction. SHVC Intra prediction considers up to 35 different prediction modes, notably 33 directional, 1 DC and 1 planar mode. Depending on the block size and Intra prediction direction, the array of neighboring samples may be filtered before creating the prediction.

In SHVC, all Inter coding CU partitions are associated to motion parameters obtained using two alternative coding modes: traditional motion estimation (ME) and the new Merge mode. The main difference between these two prediction modes is that while the ME prediction requires to perform motion estimation and codes the motion vector difference between the best motion vector and the predicted motion information, the Merge mode just indicates which Merge motion candidate (spatial, temporal and inter-layer candidates) is selected, this means which of these neighboring block provides the best motion data, through a motion vector competition. For each PU, the encoder selects one of these alternative coding modes and signals the selected mode in the bitstream with a flag which is followed by the remaining auxiliary data associated to each mode, e.g. differential motion data. Moreover, as the HLS approach is adopted in SHVC, the BL reconstructed picture is taken as an inter-layer reference picture to be included in the EL decoded picture buffer. Therefore, just a reference index is used to signal whether the EL PU comes from BL or EL decoded information.

Related work: Since SHVC adopted the HLS approach mentioned above, recent research to improve its compression efficiency mainly focused on the inter-layer processing. The main target is to provide better predictions for the EL CUs based on the available BL decoded data. In [5], a method called *generalized inter-layer residual prediction* is proposed to improve the prediction accuracy by combining the obtained EL prediction with a new residual derived from both the BL and EL. Later, a *combined temporal and inter-layer prediction* solution is proposed in [6] which is able to create more efficient predictions for the current EL block. In [7, 8], a different approach is followed as adaptive filters are proposed to be directly applied to the BL reconstructed samples. In [9], a method is proposed to enhance the ILR picture quality by combining the high frequency information from the EL reference picture with the low frequency information from the BL reconstructed picture. Similarly, differential coding methods are proposed in [10] to improve the SHVC performance by adding the high frequency component present in the previous decoded EL pictures to the BL reference picture. Recently, the importance of the Merge mode in the SHVC performance is considered [11] to propose an *improved EL Merge mode* solution which adaptively refines the Merge motion information and linearly combines the BL reconstructed picture with a refined EL Merge prediction to achieve a better Merge prediction quality, thus improving the overall SHVC RD performance.

III. Proposed Joint Layer Coding Mode

This section describes the proposed additional joint layer coding mode targeting to improve the overall SHVC RD performance. First, the overall architecture of the proposed EL coding mode selection process is presented. After, the algorithms associated to each novel module are presented in detail.

A. Proposed EL coding mode selection architecture

Figure 2 shows the proposed EL coding mode selection architecture, highlighting the integration of a novel joint layer coding mode based on the fusion of the conventional best EL prediction, P_{Best}^{EL}, and the BL reconstruction, $\hat{X}_t^{BL}$. Naturally, when the best EL prediction is the BL reconstruction, the proposed joint layer coding mode is not checked as no compression may be achieved. The novel tools to be later presented in detail are highlighted in yellow.

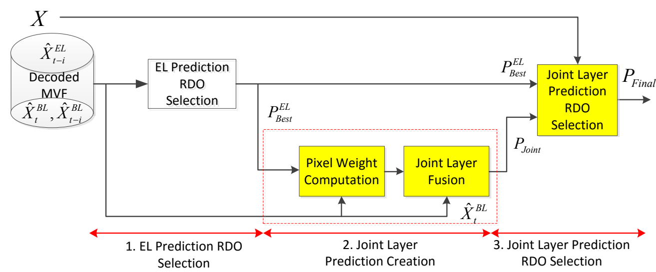

Figure 2. Proposed EL coding mode selection (*novel techniques are highlighted in yellow*).

The basic target of the proposed joint layer prediction creation process is to obtain better CU predictions (this means better PUs) than with the conventional EL coding modes, notably the standard set of EL Inter and EL Intra coding modes. To achieve this target, first the usual RDO based EL coding mode selection is performed to find the best EL prediction among the EL Intra, Inter ME and Inter Merge coding modes. It is here proposed that this best EL prediction may be combined with the BL reconstruction to create a so-called *joint layer prediction*, P_{Joint}. Finally, a joint layer prediction RDO selection is performed to select the final EL prediction, P_{Final}, between P_{Best}^{EL} and P_{Joint}. In this regard, a binary flag is requested to signal the final selected EL prediction (or coding mode) to guarantee that a similar EL prediction can be created at the decoder. The novel techniques associated to the novel joint layer coding mode are detailed in the next sub-sections.

B. Joint layer prediction creation

As shown in Figure 2, the joint layer prediction process includes two main steps which are described in the reverse order for better reading.

Joint layer fusion: The proposed joint layer coding mode creates a new prediction exploiting both the BL and EL decoded information. To achieve the highest joint layer prediction quality, the conventional EL best prediction, P_{Best}^{EL}, is appropriately fused with the BL reconstruction, $\hat{X}_t^{BL}$. In this case, the joint layer prediction, P_{joint}, created for each pixel (x, y) can be formulated as a linear combination between P_{Best}^{EL} and $\hat{X}_t^{BL}$ as:

$$P_{Joint}(x,y,w(x,y)) = w(x,y) \times P_{Best}^{EL}(x,y) + (1 - w(x,y)) \times \hat{X}_t^{BL}(x,y) \quad (1)$$

In this fusion, the weighting term, $w(x, y)$, associated to each pixel (x, y), determines the contributions of $P_{Best}^{EL}(x, y)$ and $\hat{X}_t^{BL}(x, y)$ to the joint layer prediction, thus playing a critical role in the final joint prediction quality. Therefore, $w(x,y)$ has to be carefully designed as explained in the following. After computing the regularized weights as will be defined in (7), the joint layer prediction for each pixel is created by using the fusion process defined in (1).

Pixel weight computation: Naturally, the prediction quality for each pixel of the conventional EL best prediction versus the BL reconstruction can be easily assessed at the encoder if the availability of the originals is exploited to compute the weighting term, $w(x,y)$. In this case, the prediction quality of the conventional EL best prediction and BL reconstruction can simply be taken as inversely proportional to the square difference/error, SD_{BL}, between the original information, X_t, and $\hat{X}_t^{BL}$ and the square difference/error, SD_{EL}, between the X_t and P_{Best}^{EL}. These SD values can be used to compute oracle weight values which accurately express the ideal prediction goodness of P_{Best}^{EL} and $\hat{X}_t^{BL}$or the joint layer prediction.

However, to use these oracle weights, the encoder would have to code and send the weights associated to each pixel to the decoder to guarantee that a similar joint layer prediction could be regenerated at the decoder. Naturally, this is not an efficient solution due to the associated bitrate overhead. To overcome this problem, this paper proposes a pixel level weight computation solution exploiting only the available decoded information to compute the weighting values; thus, no bitrate overhead is required and the weight computation can be synchronously performed at both encoder and decoder.

In the proposed weight computation, the quality (or prediction goodness) of the conventional best EL prediction is indirectly assessed through the square difference, SD_{EL}^*, between the $\hat{X}_t^{BL}$ and the best BL prediction, P_{Best}^{BL}, while the BL reconstruction quality is assessed through the difference, SD_{BL}^*, between P_{Best}^{EL} and P_{Best}^{BL}. If these *SDs* are small, the quality of the associated predictions should be high and vice-versa. P_{Best}^{BL} corresponds to the selected prediction mode when coding the collocated BL block. The relationship between these *SDs* and the oracles, SD_{BL},SD_{EL} defined above is illustrated in Figure 3.

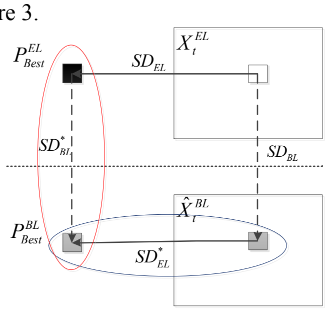

Figure 3. Illustration of the square difference calculations.

However, since no original information may be used, it may happen that the computed weights do not always accurately express the ideal contribution of $P_{Best}^{EL}(x,y)$ and $\hat{X}_t^{BL}(x,y)$ to the joint layer prediction, $P_{Joint}(x,y)$. Moreover, it is also observed that the difference between the joint layer prediction of a specific pixel and its close neighborhood pixels is usually small which motivates the use of some regularization approach. Therefore, to regularize the initially computed weight values it is proposed to exploit neighboring data, notably the neighboring weights and neighboring joint layer predictions. In summary, the proposed weight computation proceeds with the following steps:

Step 1: Pixel weight initialization - First, the square differences for each pixel, SD_{BL}^* and SD_{EL}^* are computed as:

$$SD_{BL}^*(x,y) = \left(P_{Best}^{EL}(x,y) - P_{Best}^{BL}(x,y)\right)^2 \tag{2}$$

$$SD_{EL}^*(x,y) = \left(\hat{X}_t^{BL}(x,y) - P_{Best}^{BL}(x,y)\right)^2 \tag{3}$$

Then, the initial weigh for each pixel, $w_{ini}(x,y)$, is determined as:

$$w_{ini}(x,y) = \frac{SD_{BL}^*(x,y)+1}{SD_{BL}^*(x,y)+SD_{EL}^*(x,y)+2} \tag{4}$$

Here, 1 is added to both SD_{BL}^* and SD_{EL}^* to avoid dividing by zero when both *SDs* equal to zero.

Step 2: Pixel weight regularization – Figure 4 illustrates the spatial neighborhood considered for weight regularization where the yellow node signals the current pixel, the grey nodes the pixels for which the joint layer prediction value is already available and the white nodes the pixels for which no joint layer prediction is yet available.

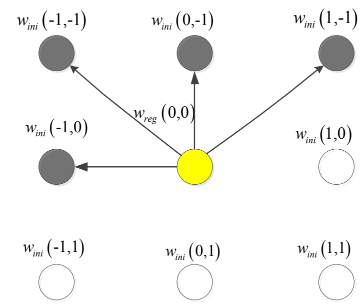

Figure 4. Spatial neighborhood considered for weights regularization.

The pixel weights regularization process proceeds as follows:

- **Weight candidates definition** – To regularize the initial weight value, $w_{ini}(x,y)$, using the neighboring information, a weight candidate list, W_List, for each pixel (x,y) is defined as:

  $$W\_List = \{w_{ini}(i,j)\}\,;\,(i,j) = \{-1,0,1\}$$

  Here $w_{ini}(0,0)$ corresponds to the initial weight of the current pixel (x,y) and $w_{ini}(i,j)$ with $(i,j) \neq (0,0)$ correspond to the initial weights of the surrounding pixels as illustrated in Figure 4.

- **Joint layer prediction candidates** - Next, the joint layer prediction associated to each weight candidate, $w_{ini}(i,j)$, in the selected window, $P(x,y,w_{ini}(i,j))$, is computed as:

  $$P(x,y,w_{ini}(i,j)) = w_{ini}(i,j)\times P_{Best}^{EL}(x,y) + (1-w_{ini}(i,j))\times \hat{X}_t^{BL}(x,y) \tag{5}$$

- **Spatial coherence measurement definition** – As it is observed that an accurate weight in the joint layer prediction is not very different from its neighooring weights, it is proposed here to regularize the weights by adopting a *spatial coherence metric,* expressing the difference between the current joint layer prediction and the one obtained by its neighbors. Moreover, as shown in Figure 4, although there are 8 neighboring pixels, only 4 neighboring pixels, the lelf, top-left, top, and top-right (grey nodes) provide 'reliable' joint layer prediction since their weight have already been regularized. Therefore, the spatial coherence metric is defined as the sum of square differences, $SSD(x,y,w_{ini}(i,j))$ between $P(x,y,w_{ini}(i,j))$ and its four 'reliable' neighboring pixels, $P(m,n,w(m,n))$ with $(m,n) \in P\_List$ and P_List is:

  $$P\_List = \{(-1,0);(-1,-1);(0,-1);(1,-1)\}$$

  In this case, the spatial coherence metric is computed as:

  $$SSD(x,y,w_{ini}(i,j)) = \sum_{(m,n)\in P\_List}\left(P(x,y,w_{ini}(i,j)) - P(m,n,w(m,n))\right)^2 \tag{6}$$

- **Regularized weight creation** – Finally, the regularized weight for each pixel, $w_{reg}(x,y)$, is obtained by selecting the weight candidate that minimizes $SSD(x,y,w_{ini}(i,j))$ as follows:

  $$w_{reg}(x,y) = \underset{w_{ini}(i)\in W\_List}{\operatorname{argmin}}\ SSD(x,y,w_{ini}(i,j)) \tag{7}$$

To conclude the fusion process, the regularized weights from (7) are applied to (1) to create the joint layer prediction for each pixel.

C. Joint layer prediction RDO selection

Naturally, the joint layer prediction RDO selection first accounts for the rate and distortion of the conventional EL best prediction and of the joint layer coding mode to select the best one. The overall rate associated to the novel joint layer coding mode must include the rate associated to signal the conventional EL best prediction, notably the selected coding mode flags and prediction information. Moreover, as the joint layer coding mode is adaptively selected using a second RDO mechanism, an additional binary flag is required to signal the final selected coding mode to guarantee that the same EL prediction is created at the decoder.

In comparison to the improved Merge mode solution proposed in [11] that also exploits both the BL and EL decoded data to improve the SHVC compression efficiency, the proposed joint layer coding mode comes with several major differences as i) it provides an additional prediction mode; ii) it introduces a different weight regularization approach; and iii) more importantly, allows to combine the BL and EL decoded information not only for the Inter Merge mode but also for the Inter ME and Intra modes.

IV. PERFORMANCE EVALUATION

This section presents the performance gains of the proposed joint layer coding mode after integration in the SHVC codec. First, the test conditions are presented; after, the RD performance is analyzed.

A. Test conditions

To assess the compression efficiency of the SHVC standard with the proposed joint layer coding mode, five video sequences with different motion and texture characteristics were used as shown in Table I. To consider several coding scenarios, three different group of pictures (GOP) sizes with four BL prediction structures were tested, notably: i) IBP for GOP 2; ii) IPP for GOP 4 Low Delay–P (LD-P); iii) IBB for GOP 4 Low Delay–B (LD-B); and iv) IBB for GOP 8 Random Access (RA). The selected benchmarks are the SHVC standard with the conventional coding modes [4] and the improved SHVC solution proposed in [11]. The SHVC reference software version 6.0 has been used for the experiments.

TABLE I. SUMMARY OF TEST CONDITIONS

Sequences	**Spatial resolution**	**Temporal resolution**	**Number of test frames**
RaceHorses	416×240	30 Hz	297
BlowingBubbles		50 Hz	497
BasketballPass		50 Hz	497
PartyScene	832×480	50 Hz	497
BQMall		60 Hz	600
GOP size	2, 4 (LD-P, LD-B), 8 (RA)		
Quantization parameters	$QP_{BL} = 34$ $QP_{EL} = \{32; 30; 28; 26\}$		

B. RD performance and discussion

As common in the SHVC RD performance assessment [5-11], the BD-Rate [12] metrics are computed for both the BL and EL bitrates as shown in Table II. Moreover, the BD-Rate savings regard both the SHVC standard [4] and the improved SHVC solution [11] (labelled *as* Ref [11]).

From the experimental results, some conclusions can be derived:

- **Proposed SHVC extension versus SHVC standard:**
  - The proposed SHVC extension with the novel joint layer coding mode always outperforms SHVC with the standard prediction modes for all test sequences and all GOP sizes; notably up to 4.3% BD-Rate reduction are obtained for GOP 4 (LD-P) and for the *BQMall* video sequence.
  - The higher gains are obtained for the smaller GOP sizes. This results from the fact that, for lower GOP sizes, the temporal distance between the current and the reference pictures is smaller; thus, the weight computed with the proposed solution is more accurate and the created joint layer prediction has higher quality.
- **Proposed SHVC extension versus SHVC with improved Merge mode [11]:**
  - The proposed SHVC extension always outperforms the SHVC solution with the improved Merge mode [11] for all GOP sizes; notably about 2.24% average BD-Rate reduction is obtained for the GOP 4(LD-P) test condition.
  - The proposed SHVC extension brings significant compression efficiency gains not only for low motion sequences like *BlowingBubbles* and *BQMall* but also for high motion sequences like *RaceHorses* and *BasketballPass*. This is a remarkable result, notably when compared to the improved Merge solution of [11]. The reason for this coding achievement is that the proposed joint layer coding mode is taken as an additional coding mode and adaptively selected with a RDO mechanism; thus, when the joint layer coding mode is inefficient, the conventional EL best prediction is selected.
  - Moreover, the proposed SHVC extension also requires lower processing complexity than the improved Merge mode solution in [11] especially due to the absence of the high complexity motion refinement process to obtain better Merge predictions.

V. CONCLUSIONS

This paper proposes a novel joint layer coding mode for the SHVC standard jointly exploiting the EL and BL decoded information. The novel coding mode efficiently combines, at pixel level, the EL best prediction and the BL reconstruction by performing an adaptive fusion with regularized weights. Experimental results show that SHVC with the proposed joint layer coding mode outperforms the standard SHVC solution, notably up to 4.3 % BD-Rate saving. As the weight accuracy plays a critical role in creating the joint layer prediction, future work will consider improving the accuracy of fusion weights.

TABLE II. BD-RATE SAVINGS REGARDING SHVC (BOTH BL AND EL BITRATE)

Sequences	**GOP 2**		**GOP 4 (LD-P)**		**GOP 4 (LD-B)**		**GOP 8 (RA)**	
	Ref [11]	**Proposed**	**Ref [11]**	**Proposed**	**Ref [11]**	**Proposed**	**Ref [11]**	**Proposed**
RaceHorses	-3.27	-4.04	-0.32	-3.44	-0.64	-2.69	-0.21	-1.79
BlowingBubbles	-3.49	-4.06	-2.14	-3.76	-1.90	-2.78	-2.36	-3.64
BasketballPass	-3.01	-3.86	-0.33	-3.22	-0.45	-2.06	-0.08	-1.85
PartyScene	-2.62	-3.12	-1.85	-3.26	N	N	-1.51	-2.57
BQMall	-2.93	-3.88	-2.12	**-4.30**	N	N	-1.19	-3.11
Average BD-Rate to SHVC	**-3.06**	**-3.79**	**-1.35**	**-3.60**	**-1.00**	**-2.51**	**-1.07**	**-2.59**
Average BD-Rate to [11]		***-0.73***		***-2.24***		***-1.51***		***-1.52***

REFERENCES

[1] G. J. Sullivan, J. -R. Ohm, W. –J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.,* pp. 1649-1668, vol. 22, no. 12, Dec. 2012.

[2] T. Wiegand G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.* , pp. 560-576, vol. 13, no. 7, Jul. 2003.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.,* pp. 1103-1120, vol. 17, no. 9, Sep. 2007.

[4] G. J. Sullivan, J. M. Boyce, Y. Chen, J. –R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of High Efficiency Video Coding (HEVC)," *IEEE Journal Selected Topics in Signal Processing,* pp. 1001-1016, vol. 7, no. 6, Dec. 2013.

[5] X. Li, J. Chen, K. Rapaka, and M. Karczewicz, "Generalized inter-layer residual prediction for scalable extension of HEVC," in *IEEE International Conference on Image Processing*, pp. 1559-1562, Melbourne, VIC, Australia, Sept. 2013.

[6] P. Lai, S. Liu, and S. Lei, "Combined temporal and inter-layer prediction for scalable video coding using HEVC," in *Picture Coding Symposium*, pp. 117-120, San Jose, CA, USA, Dec. 2013.

[7] M. Guo, S. Liu, and S. Lei, "Inter-layer adaptive filtering for scalable extension of HEVC," in *Picture Coding Symposium*, pp. 165-168, San Jose, CA, USA, Dec. 2013.

[8] P. Lai, S. Liu, and S. Lei, "Low latency directional filtering for inter-layer prediction in scalable video coding using HEVC," in *Picture Coding Symposium*, pp. 269-272, San Jose, CA, USA, Dec. 2013.

[9] T. Laude, X. Xiu, J. Dong, Y. He, Y. He, and J. Ostermann, "Scalable extension of HEVC using enhanced inter-layer prediction," in *IEEE International Conference on Image Processing*, pp. 3739-3743, Paris, France, Oct. 2014.

[10] A. Aminlou, J. Lainema, K. Ugur, M. M. Hannuksela, and M. Gabbouj, "Differential coding using enhanced inter-layer reference picture for scalable extension of H.265/HEVC video codec," *IEEE Trans. Circuits Syst. Video Technol.,* pp. 1945-1956, vol. 24, no. 11, Nov. 2014.

[11] X. HoangVan, J. Ascenso, and F. Pereira, "Improving enhancement layer merge mode for HEVC scalable extension," in *Picture Coding Symposium*, pp. 15-19, Cairns, QLD, Australia, Jun. 2015.

[12] G. BjØntegaard, "Calculation of average PSNR differences between RD curves," *Doc. VCEG-M33, 13th ITU-T VCEG Meeting*, Austin, TX, USA, Apr. 2001.