|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extension Development**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  3rd Meeting: Geneva, CH, 17–23 Jan. 2013 | Document: JCT3V-C0049 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **3D-CE4: Advanced residual prediction for multiview coding** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Li Zhang Ying Chen Xiang Li Marta Karczewicz  5775 Morehouse Drive San Diego, CA 92121 USA | Tel: Email: | 1-858-651-6660 [lizhang@qti.qualcomm.com](mailto:lizhang@qti.qualcomm.com)  1-858-845-6589 [cheny@qti.qualcomm.com](mailto:cheny@qti.qualcomm.com) |
| *Source:* | Qualcomm Incorporated | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

Inter-view residual prediction is enabled in the current HTM design to code the residue of dependent texture views more efficiently. This proposal is a follow-up of JCT3V-B0051 with some complexity reduction parts introduced. In this proposal, an advanced residual prediction (ARP) is proposed to further improve the coding efficiency of inter-view residual prediction. In ARP, to ensure high correlation between residues of two views, motion of the current block of picture in current view is applied to the corresponding block in a reference view picture to generate residual in the base view to be used for inter-view residual prediction. Moreover, an adaptive weighting factor is applied to the residue signal so that the prediction error is further reduced. It is shown that the proposed method achieves compression efficiency gain of 0.9% and 1.4% for coded views in terms of BD rate compared to HTM anchor and CE4 anchor, respectively.

1. **Introduction**

In the current design of 3D-HEVC, inter-view residual prediction (IVRP) is enabled in a way that the corresponding block in the base view is firstly located based on the disparity vector and then the residue of the corresponding block is used as a predictor of the residue of current coding unit (CU). The residue of the corresponding block is added to the motion compensated signal and only the resulting difference signal is transform coded and transmitted. In other words, the residue generated from motion compensation is further predicted by the residue from the base view, before it is transformed, quantized and written in the bitstream.

Although the current IVRP further improves the overall quality of prediction, it has three main drawbacks.

* First, it uses the reference view residues to directly predict those in current view regardless whether the residual block is obtained with the same motion as that in the current view. Clearly, when the two blocks are with different motion, the correlation between them is relatively low so that the prediction performance will be sub-optimal.
* Second, reconstructed residues are used in prediction, which introduces undesired quantization error in prediction and further degrades the performance.
* Third, the parsing of the flag indicating inter-view residual prediction is dependent on decoded information of the view component of the reference view.

1. **Proposed Method**

To further improve the coding efficiency, it is proposed that motion of current view is applied to the corresponding block in a reference view to have a better prediction performance. Moreover, additional weighting factor is introduced to further compensate quality difference between different views.

Figure 1 illustrates the prediction structure of ARP in multiview video coding. As shown in Figure 1, Dc represents the current block in the current view (view 1), Bc, and Dr denote the representation of Dc in the reference view (view 0) at time Tj and Dc’s temporal prediction from the same view (view 1) at time Ti. VD denotes the motion from Dc to Dr. Since Dc and Bc are actually projections of the same object in two different views, these two blocks should share the same motion information. Therefore, Bc’s temporal prediction Br in view 0 at time Ti can be located from Bc by applying the motion information of VD. The residual of Bc with motion information of VD is then multiplied by a weighting factor and used as the residual predictor for current residual.

## 1

Figure 1. Prediction structure of the proposed ARP in 3D-HEVC.



Figure 2. Relationship among current block, reference block and motion compensated block

## Decoding process of ARP

Main procedures of the proposed ARP at the decoder side can be described as follows:

1. Obtain a disparity vector as specified in the current 3D-HEVC, pointing to a target reference view. Then, in the picture of the reference view within the same access unit, the corresponding block is located by the disparity vector.
2. Re-use the motion information of the current block to derive the motion information for the reference block. Apply motion compensation for the corresponding block based the same motion vector of current block and derived reference picture in the reference view for the reference block, to derive a residue block. The relationship among current block, corresponding block and motion compensated block is shown in Figure 2. The reference picture in the reference view (V0) which has the same POC (Picture Order Count) value as the reference picture of current view (Vm) is selected as the reference picture of the corresponding block.
3. Apply the weighting factor to the residue block to get a weighted residue block and add the values of the weighted residue block to the predicted samples.

## Weighting factor

Three weighting factors are used in ARP, i.e., 0, 0.5 and 1. The one leading to minimal rate-distortion cost for the current CU is selected as the final weighting factor and the corresponding weighting index (0, 1 and 2 which correspond to weighting factor 0, 1, and 0.5, respectively) is transmitted in the bitstream at the CU level. All PU predictions in one CU share the same weighting factor. When the weighting factor is equal to 0, ARP is not used for the current CU.

## Simplification of ARP

Different from JCT3V-B0051 [1], two aspects were slightly modified to further reduce the computational complexity, in this proposal:

* The 8/4-tap luma/chroam interpolation filter used in HEVC for fractional sample interpolation process is replaced by bi-linear filter when generating the residual predictor.
* The proposed ARP is only applied to those coding units with the partitioning mode equal to PART\_2Nx2N.

1. **Compression Performance**

This section provides simulation results of the proposed ARP in comparison with the 3D-HTM anchors. The platform 3D-HTM 5.0.1 [2] is utilized and the proposed method is implemented on it. All the simulation tests are performed based on the common test conditions [3]. Two anchors are compared with ARP, i.e., the anchor of 3D-HTM5.0.1 and the anchor of CE4 with the IVRP flag always transmitted solving the parsing issue of IVRP [4].

Table 1 provides the results of ARP compared with 3D-HTM anchor. As shown in Table 1, the overall average bitrate reduction is 0.9%, 0.6% and 0.7% for texture views, synthesized views, and coded & synthesized views, respectively.

As shown in Table 2, when the parsing issue of IVRP is solved in 3D-HTM, additional coding gain of ARP could be observed, i.e., 1.4%, 1.0%, 1.1% for texture views, synthesized views, and coded & synthesized views.

Table 1: Coding gain of ARP with respect to 3D-HTM anchor for 3-view case

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Video  1 | Video  2 | Video  only | Synthesized only | Coded & synthesized | Enc  time | Dec  time | Ren  time |
| Balloons | -3.8% | -3.4% | -1.3% | -0.9% | -1.0% | 102% | 91% | 97% |
| Kendo | -3.8% | -4.0% | -1.5% | -0.8% | -1.0% | 95% | 94% | 97% |
| Newspapercc | -3.0% | -2.2% | -1.0% | -0.6% | -0.7% | 100% | 94% | 95% |
| GhostTownFly | -1.7% | -1.7% | -0.6% | -0.4% | -0.5% | 92% | 90% | 97% |
| PoznanHall2 | -0.7% | -0.4% | -0.2% | -0.2% | -0.2% | 108% | 99% | 101% |
| PoznanStreet | -3.8% | -3.4% | -1.1% | -0.8% | -0.9% | 99% | 94% | 92% |
| UndoDancer | -1.8% | -1.7% | -0.5% | -0.4% | -0.4% | 100% | 86% | 92% |
| 1024x768 | -3.5% | -3.2% | -1.3% | -0.8% | -0.9% | 99% | 93% | 97% |
| 1920x1088 | -2.0% | -1.8% | -0.6% | -0.5% | -0.5% | 99% | 92% | 95% |
| **average** | **-2.7%** | **-2.4%** | **-0.9%** | **-0.6%** | **-0.7%** | **99%** | **92%** | **96%** |

Table 2: Coding gain of ARP with respect to CE4 anchor for 3-view case

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Video  1 | Video  2 | Video  only | Synthesized only | Coded & synthesized | Enc  time | Dec  time | Ren  time |
| Balloons | -5.9% | -5.6% | -2.2% | -1.7% | -1.8% | 97% | 89% | 92% |
| Kendo | -5.7% | -6.1% | -2.3% | -1.6% | -1.7% | 102% | 98% | 101% |
| Newspapercc | -4.5% | -3.5% | -1.5% | -1.1% | -1.2% | 101% | 92% | 95% |
| GhostTownFly | -2.2% | -2.1% | -0.7% | -0.5% | -0.6% | 97% | 94% | 102% |
| PoznanHall2 | -2.0% | -1.7% | -0.8% | -0.6% | -0.6% | 106% | 98% | 104% |
| PoznanStreet | -4.4% | -4.3% | -1.3% | -1.1% | -1.1% | 109% | 104% | 102% |
| UndoDancer | -2.7% | -2.7% | -0.8% | -0.6% | -0.6% | 101% | 90% | 96% |
| 1024x768 | -5.4% | -5.1% | -2.0% | -1.5% | -1.6% | 100% | 93% | 96% |
| 1920x1088 | -2.8% | -2.7% | -0.9% | -0.7% | -0.7% | 103% | 96% | 101% |
| **average** | **-3.9%** | **-3.7%** | **-1.4%** | **-1.0%** | **-1.1%** | **102%** | **95%** | **99%** |

It is noted that if interpolation filters as in HEVC are still used to generate the residual signal, similar coding efficiency results were achieved, as shown in Table 3 and Table 4.

Table 3: Coding gain of ARP with HEVC interpolation filters compared to 3D-HTM anchor

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Video  1 | Video  2 | Video  only | Synthesized only | Coded & synthesized | Enc  time | Dec  time | Ren  time |
| Balloons | -3.6% | -3.2% | -1.2% | -0.9% | -1.0% | 111% | 94% | 97% |
| Kendo | -3.7% | -3.6% | -1.3% | -0.7% | -0.8% | 98% | 91% | 95% |
| Newspapercc | -2.6% | -1.9% | -0.8% | -0.5% | -0.6% | 104% | 95% | 97% |
| GhostTownFly | -1.5% | -1.7% | -0.4% | -0.2% | -0.3% | 94% | 92% | 92% |
| PoznanHall2 | -0.5% | -0.2% | -0.1% | 0.0% | -0.1% | 112% | 91% | 96% |
| PoznanStreet | -3.2% | -3.1% | -0.9% | -0.7% | -0.7% | 108% | 94% | 92% |
| UndoDancer | -2.5% | -2.2% | -0.6% | -0.5% | -0.5% | 99% | 87% | 91% |
| 1024x768 | -3.3% | -2.9% | -1.1% | -0.7% | -0.8% | 104% | 93% | 96% |
| 1920x1088 | -1.9% | -1.8% | -0.5% | -0.4% | -0.4% | 103% | 91% | 93% |
| **average** | **-2.5%** | **-2.2%** | **-0.8%** | **-0.5%** | **-0.6%** | **103%** | **92%** | **94%** |

Table 4: Coding gain of ARP with HEVC interpolation filters compared to CE4 anchor

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Video  1 | Video  2 | Video  only | Synthesized only | Coded & synthesized | Enc  time | Dec  time | Ren  time |
| Balloons | -5.7% | -5.4% | -2.1% | -1.6% | -1.7% | 106% | 91% | 92% |
| Kendo | -5.5% | -5.6% | -2.2% | -1.5% | -1.6% | 105% | 95% | 99% |
| Newspapercc | -4.1% | -3.3% | -1.4% | -1.0% | -1.0% | 105% | 93% | 97% |
| GhostTownFly | -2.0% | -2.1% | -0.6% | -0.3% | -0.4% | 99% | 96% | 97% |
| PoznanHall2 | -1.8% | -1.5% | -0.7% | -0.4% | -0.5% | 109% | 91% | 98% |
| PoznanStreet | -3.8% | -3.9% | -1.1% | -0.9% | -1.0% | 119% | 105% | 102% |
| UndoDancer | -3.4% | -3.2% | -0.9% | -0.7% | -0.7% | 101% | 91% | 95% |
| 1024x768 | -5.1% | -4.8% | -1.9% | -1.4% | -1.5% | 105% | 93% | 96% |
| 1920x1088 | -2.8% | -2.7% | -0.8% | -0.6% | -0.6% | 107% | 95% | 98% |
| **average** | **-3.8%** | **-3.6%** | **-1.3%** | **-0.9%** | **-1.0%** | **106%** | **94%** | **97%** |

1. **References**
2. L. Zhang, Y. Chen, X. Li, M. Karczewicz, "3D-CE5.h related: Advanced residual prediction for multiview coding," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-B0051, 2nd Meeting: Shanghai, CN, 13–19 Oct. 2012.
3. 3D-HTM version 5.0.1: <https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-5.0.1/>.
4. D. Rusanovskyy, K. Müller, A. Vetro, "Common Test Conditions of 3DV Core Experiments," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-A1100, 1st Meeting: Stockholm, SE, 16–20 July 2012.
5. L. Zhang, "Description of Core Experiment 4 (CE4): Residual Prediction," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-B1104, 2nd Meeting: Shanghai, CN, 13–19 Oct. 2012.
6. **Patent rights declaration(s)**

**Qualcomm Incorporated may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**