|  |  |
| --- | --- |
| **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**  2nd Meeting: Shanghai, CN, 13–19 Oct. 2012 | Document: JCT3V-B0050 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **3D-CE5.h related: Improved temporal motion vector prediction for merge** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Li Zhang  Ying Chen 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

In current 3D-HEVC, a target reference index for temporal merging candidate is set according to the neighboring prediction unit. When the target reference index corresponds to a reference picture in the same view while the motion vector of the co-located prediction unit (PU) points to an inter-view reference picture and vice versa, temporal motion vector prediction (TMVP) candidate is considered as unavailable. To address this issue, it is proposed that one additional target reference index is used, so that TMVP candidate can be supported for the above cases. For 3D-HEVC, the proposed method provides about 0.3% average bitrate saving for the all the coded views and 0.7% bitrate saving for the non-base views.

# Introduction

In in current 3D-HTM which is based on HM6.0 [1], when the current mode is merge, the target reference index, denoted by refIdxT, could be either 0 or the reference index of the left neighbouring PU. Suppose the target reference index equal to refIdxT corresponds to a short-term (ST) reference picture as shown in Figure 1. If the motion vector in the co-located PU, refers to a ST reference picture, it is scaled to form a merge candidate of the current PU (PU0), as shown in Figure 1, wherein MV0 is scaled to MV0’ during the merge mode.

However, if the co-located PU has a motion vector (MV1) referring to an inter-view reference picture, the motion vector is not used to predict the current PU (PU1). A motion vector pointing to a view component of a different view is called disparity motion vector in this document and this view component is referred to as an inter-view reference picture. Similarly, when the target reference index equal to refIdxT corresponds to an inter-view reference picture and TMVP refers to a ST reference picture, the temporal motion vector prediction candidate is also disabled.

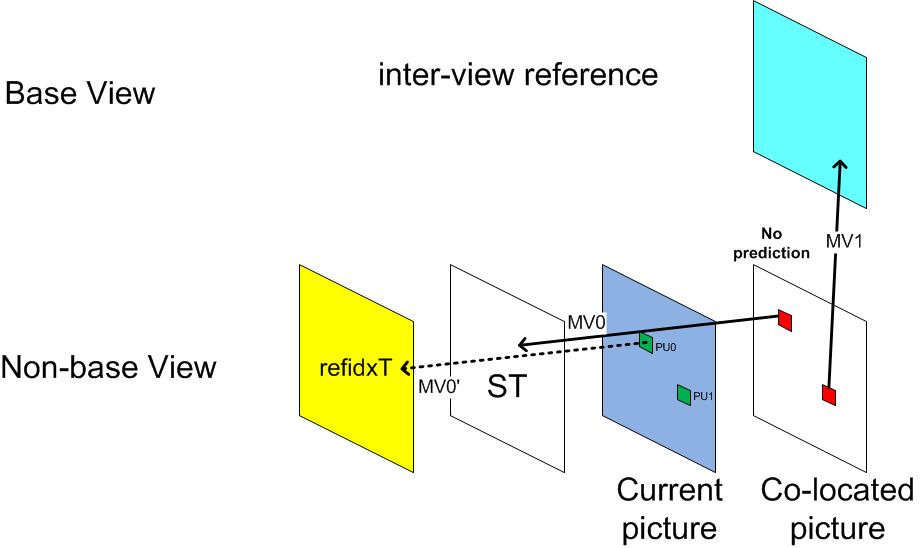


Figure 1: TMVP in MV-HEVC (merge mode).

In 3D-HEVC, there might be significant amount of co-located PUs (in the co-located picture) which have a different reference picture type than that identified by the target reference index. Therefore, disabling prediction from those motion vectors makes the merge mode less efficient.

# Proposal

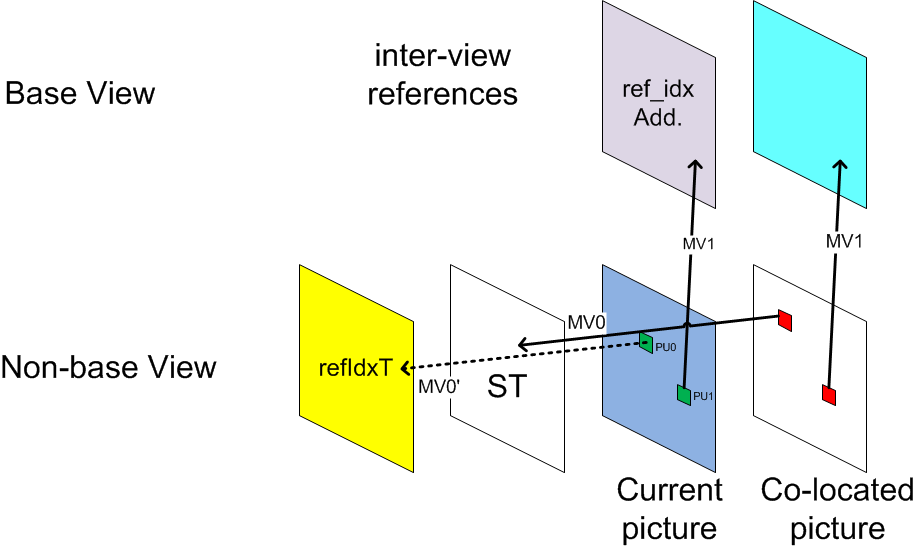


Figure 2: Proposed TMVP in 3D-HEVC (merge mode).

It is proposed that an additional target reference index is introduced when the reference picture type that refIdxT represents is different from that TMVP points to. For example, in the above example, for all inter-view reference pictures that co-located PU points to, a new additional reference index is derived, so that the motion vectors referring to an inter-view reference picture can be used to form a merge candidate and not considered as unavailable. As shown in Figure 2, MV1 of the co-located block of PU1 can be used to form a disparity motion vector candidate.

## Derivation of additional target reference indices

In the slice level, the value of the additional target reference index (refIdxLXA) and one flag to indicate the type of reference picture with index equal to 0, bZeroIdxInterViewFlagLX, for each of RefPicList0 and RefPicList1, if available, is derived.

**Derivation process for the additional target reference index for TMVP**

This process is invoked when the current slice is a P or B slice. Additional target reference indices refIdxL0A and refIdxL1A are derived which repsent different reference picture types from the reference picture with reference index equal to zero. In addtion, two flags bZeroIdxInterViewFlagL0 and bZeroIdxInterViewFlagL1 are derived to indicate whether the reference picture with index equal to 0 is an inter-view refernce picture.

Set variables refIdxL0A and refIdxL1A both to -1.

The following apply to derive refIdxL0A and refIdxL1A, bZeroIdxInterViewFlagL0 and bZeroIdxInterViewFlagL1.

for (X = 0 ; X <2; X++) {  
 bZeroIdxInterViewFlagLX = RefPicListX[ 0 ] is a short-term reference picture ? 0 : 1  
 bFound= 0  
 for( i = 1; i <= num\_ref\_idx\_lX\_active\_minus1&&! bFound; i++)  
 if ( (bZeroIdxInterViewFlagLX && RefPicListX[ i ] is an inter-view reference picture) | |  
 (!bZeroIdxInterViewFlagLX && RefPicListX[ i ] is an inter-view reference picture) ) {  
 refIdxLXA = i  
 bFound =1  
 }  
}

## Temporal motion vector prediction

During temporal motion vector predition, when the current mode is merge and one of the following conditions is satisfied, the taregt reference index needs to be modified to represent the same reference picture type as TMVP points to:

* The type of reference picture of the co-located block is different from the type of the reference picture represented by refIdxT.

Note the type here is either “short-term” or “inter-view”. Similar to the HEVC specification, the type of the reference picture of the co-located block is not the current type but is the type when it was used as a reference picture to code the co-located picture.

When the taregt reference index needs to be modified (as determined above), the following steps apply:

* If refIdxT represents a short-term reference picture
  + If bZeroIdxInterViewFlagLX is true, the taregt reference index is changed to 0
  + Otherwise if refIdxLXA is larger than 0, the taregt reference index is changed to refIdxLXA
  + Otherwise (bZeroIdxInterViewFlagLX is false and refIdxLXA is less than 0, thus not available), the TMVP candidate is set to be unavailable (as in the current 3D-HEVC)
* Otherwise (refIdxT represents an inter-view reference picture)
  + If bZeroIdxInterViewFlagLX is false, the taregt reference index is changed to 0
  + Otherwise if refIdxLXA is larger than 0, the taregt reference index is changed to refIdxLXA
  + Otherwise (bZeroIdxInterViewFlagLX is ture and refIdxLXA is less than 0, thus not available), the TMVP candidate is set to be unavailable (as in the current 3D-HEVC)

The same as in HEVC, for short-term motion vectors may be further scaled based on POC (Picture Order Count).

Note that the AMVP mode is not changed.

# Compression Performance

This section provides simulation results of the proposed method in comparison with the 3DV-HTM anchor. The proposed method is implemented on top of 3D-HTM 4.0 [2] and all the simulation tests are performed based on the common test conditions [3].

As shown in Table 1, the overall average bitrate reduction is 0.3%, 0.3%, 0.3% for decoded texture views, synthesized views, coded and synthesized views, respectively.

Table 1: Coding gain with respect to anchor for 3-view case

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Video 1 | Video 2 | Video only | Synthesized only | Coded & synthesized | Enc time | Dec time | Ren time |
| Balloons | -0.8% | -0.8% | -0.3% | -0.3% | -0.3% | 92.1% | 91.2% | 94.1% |
| Kendo | -0.7% | -0.5% | -0.3% | -0.3% | -0.3% | 101.4% | 102.3% | 99.0% |
| Newspapercc | -0.4% | -0.4% | -0.2% | -0.1% | -0.1% | 94.3% | 95.2% | 94.7% |
| GhostTownFly | -1.0% | -0.9% | -0.3% | -0.3% | -0.3% | 98.9% | 96.2% | 100.2% |
| PoznanHall2 | -1.0% | -0.7% | -0.4% | -0.4% | -0.4% | 96.7% | 93.9% | 94.2% |
| PoznanStreet | -1.0% | -1.0% | -0.3% | -0.3% | -0.3% | 97.2% | 96.8% | 100.6% |
| UndoDancer | -0.3% | -0.6% | -0.2% | -0.2% | -0.2% | 95.1% | 97.7% | 100.9% |
| 1024x768 | -0.6% | -0.6% | -0.3% | -0.2% | -0.2% | 95.9% | 96.1% | 95.9% |
| 1920x1088 | -0.8% | -0.8% | -0.3% | -0.3% | -0.3% | 97.0% | 96.1% | 98.9% |
| **average** | -0.7% | -0.7% | -0.3% | -0.3% | -0.3% | 96.5% | 96.1% | 97.6% |

# References

1. B. Bross, W.J. Han, J. Ohm, G. Sullivan, T. Wiegand, “High efficiency video coding (HEVC) text specification draft 8, ” JCTVC-J1003, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-J1003\_d2.doc, 10th Meeting: Stockholm,Sweden, 11-20 July, 2012.
2. 3DV-HTM version 4.0: <https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-4.0/>.
3. 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.

# 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).**