

EXTENSIONS OF THE HEVC TECHNOLOGY FOR EFFICIENT MULTIVIEW VIDEO CODING

Jakub Stankowski, Marek Domański, Olgierd Stankiewicz, Jacek Konieczny, Jakub Siast, Krzysztof Wegner

Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology, Poznań, Poland

ABSTRACT

The paper deals with multiview video coding using the new technology of High Efficiency Video Coding (HEVC). Implementation of multiview video coding in the framework of HEVC is described together with new specific tools proposed by the authors. Extensive experimental results are reported for compression performance comparison of MVC (ISO 14496-10), HEVC simulcast and two versions of proposed "multiview HEVC". For "multiview HEVC" the results indicate significant bitrate reduction of about 50%, as compared to the state-of-the-art MVC technology standardized as a part of AVC (MPEG-4, H.264).

Index Terms— HEVC, MVC, inter-view prediction, multiview coding

1. INTRODUCTION

Regarding 3D video, two basic multiview video formats are considered for transmission and compression:

- multiview video with depth maps (MVD),
- multiview video with no depth map.

Currently, the MPEG (ISO/IEC expert group) is working towards an international standard related to the first approach, i.e. coding of multiview video along with the respective depth maps. For the second format, i.e. multiview video with no additional depth information, there already exists a technique called Multiview Video Coding (MVC) that is incorporated into Advanced Video Coding (AVC) compression standard of ISO and ITU [1]. The MVC technique takes advantage of inter-view redundancy and provides bitrates reduced mostly by 15-25% as compared to independent compression of the views using AVC (called also as "simulcast AVC").

On the other hand, a new generic video compression technology called High Efficiency Video Coding (HEVC) is under standardization by ISO and ITU [2]. HEVC provides almost halved bitrates as compared to AVC. Therefore even independent compression of all views using HEVC (simulcast HEVC) appears to be significantly more efficient than the older but dedicated MVC technique [7]. In this paper, we are going to study this issue in details. Therefore, two main problems are addressed in this paper:

- Compression performance of "multiview HEVC" obtained by direct adoption of multiview coding tools into HEVC.
- Potential improvements of "multiview HEVC" by extensions of the prediction modes already used in HEVC.

For the sake of conciseness, in the considerations, the case of 3 views will be treated. Nevertheless the conclusions hold also for other numbers of view, with the reservation that the numbers

referred to the compression gain against MVC may be somewhat different for other numbers of views.

2. MULTIVIEW HEVC (MV-HEVC)

The classic Multiview Video Coding (MVC) [1] uses inter-view predictions in addition to the standard inter-frame prediction as used in the generic Advanced Video Coding (AVC)[1]. In MVC, the pictures from other views but the same time instant may be taken also as the reference pictures. This is done by inserting additional reference pictures into the reference picture lists. This additional inter-view prediction improves the compression efficiency as compared to independent coding of the views using the generic AVC technique. In particular, when I-frames are replaced by P- or B-frames by the use of this additional prediction, the compression improvement is particularly high.

The HEVC codec structure is quite similar to that of AVC. Therefore, the idea of multi-view coding can be easily adopted to HEVC [3]. We call the new codec Multiview-HEVC (MV-HEVC). The adoption of multiview coding tools is nearly straightforward. For MV-HEVC, the difference is, that except of I-pictures in the base view, all other pictures may be B-pictures. In Fig. 1, the boxes marked in gray denote the B-pictures in MV-HEVC that would be P-pictures in MVC.

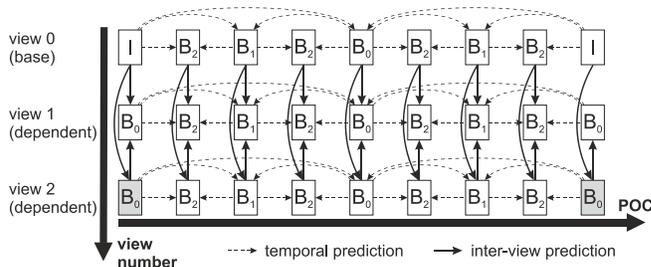


Figure 1. Temporal and inter-view prediction scheme.

3. IMPROVED PREDICTION MULTIVIEW-HEVC (IPMV-HEVC)

In the just proposed MV-HEVC, there still exists a potential for improvement of the compression performance. In the abovementioned proposal of MV-HEVC, inter-view predictions were used for texture, while now we are going to consider also modifications of motion data prediction. The MV-HEVC with such prediction tools will be called Improved Prediction Multiview HEVC (IPMV-HEVC).

In the generic HEVC, very efficient Advanced Motion Vector Prediction (AMVP) is used. The idea of AMVP is to use some predefined rules to select candidate motion vectors from the

neighboring prediction units in the same frame or from prediction units in the reference frame. These selected motion vectors and the related information like the identifier of the reference frame linked to the vector are the prediction candidates. The encoder selects one prediction candidate and transmits its identifier in the bitstream.

In this section, we describe the modification of prediction candidate derivation that exploits inter-view dependencies. The proposed modification is useful for both coding modes of AMVP: motion vector “competition” and motion data “merging”.

3.1. Vector Scaling

The motion vector scaling is used for deriving AMVP candidates. In the original HEVC design, there is temporal scaling only. A scaled vector for target (t) prediction unit is calculated from source (s) prediction unit vector. Scaling operation is performed as follows:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} \cdot \frac{t_t - t_{t,ref}}{t_s - t_{s,ref}}$$

where: x_t, y_t are the scaled vector components for target prediction unit; x_s, y_s are the vector components for source prediction unit; t_t and t_s are the values of Picture Order Count (POC) for the target and the source prediction unit, respectively; $t_{t,ref}$ and $t_{s,ref}$ are POC of the picture referenced by the target and the source unit, respectively. In the case of scaling of vectors derived from neighboring units, t_t is equal to t_s . The motion vector scaling is illustrated in Figure 2. Vectors from neighboring units (red, green, blue) are used to derive prediction candidates for the grey unit. Vectors for the current unit (grey), pointing to a reference frame (darker, POC=1) are searched for. Green neighbor vector points to the same reference picture and can be directly copied, but red and blue neighbor vectors point to different POCs and have to be scaled.

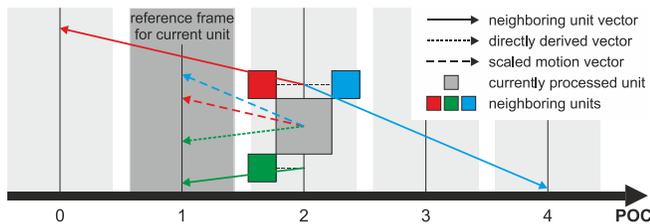


Figure 2. Illustration of motion vector scaling.

In IPMV-HEVC, for the case where reference pictures from different views are used, we propose inter-view scaling for disparity vectors. Inter-view scaling allows to use disparity vector pointing to the reference picture from a different view as a prediction candidate. The inter-view scaling is performed as follows:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} \cdot \frac{v_t - v_{t,ref}}{v_s - v_{s,ref}}$$

where: x_t, y_t are the scaled vector components for target unit; x_s, y_s are vector components for source prediction unit; v_t and v_s are the view numbers for the target and the source prediction unit, respectively; $v_{t,ref}$ and $v_{s,ref}$ are view numbers of the picture referenced by the target and the source unit, respectively.

3.2. Colocated Units

In HEVC, a prediction candidate can be derived from the colocated units in the reference frame. This is an additional candidate for

temporal motion data prediction. The derivation of a motion vector from colocated unit is illustrated in Figure 3. The colocated unit (orange) is located in the first image from the reference list. Next, the motion vector from the colocated unit motion vector is scaled for current unit and the reference frame POCs.

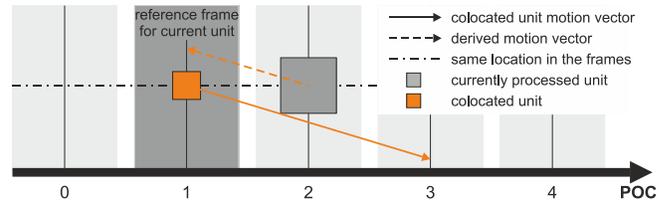


Figure 3. Derivation of a motion vector from the colocated unit.

In IPMV-HEVC, in order to increase multiview coding efficiency, the inter-view support for colocated units is proposed. There are four possibilities of using colocated units in a coder:

- derive the motion vector from the colocated unit in the temporal reference frame (the only possibility in HEVC and MV-HEVC),
- derive the disparity vector from the colocated unit in the temporal reference frame,
- derive the disparity vector from the colocated unit in the inter-view reference frame,
- derive the motion vector from the colocated unit in the inter-view reference frame.

Depending on the choice of:

- the reference frame selected for the currently processed unit,
- the frame where the colocated units is located,

the encoder chooses one of four ways to derive the colocated prediction candidate.

For example, consider the last of these four. Figure 4. shows the process of temporal motion vector derivation from the colocated unit in the inter-view reference frame. This process is used when the colocated unit is taken from inter-view reference frame but the temporal motion vector is required.

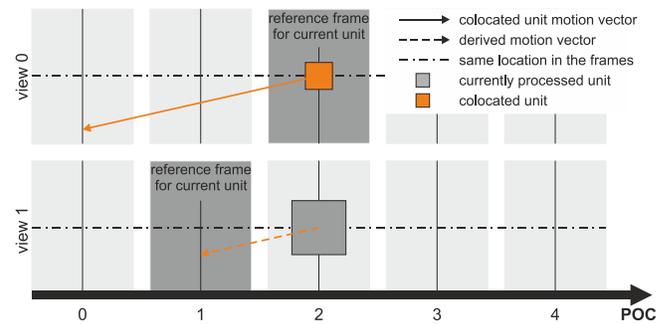


Figure 4. Derivation of a motion vector from the colocated unit in an inter-view reference frame (in IPMV-HEVC).

3.3. Nested Prediction

Nested Prediction (NP) is a new tool introduced to improve prediction of vectors of motion and disparity. The aim of this tool is to avoid situation, when there is no possibility to derive candidates from the neighboring units. The NP is useful in two following cases:

- The currently processed unit is encoded with the use of the reference image in the same view but all neighboring units have their reference frames from side views. Therefore, the motion (temporal) vector is necessary for prediction, but only

disparity (inter-view) vectors are available in neighboring units.

- The currently processed unit is encoded with the use of the reference image in the side view but all neighboring units have their reference frames from same view. Therefore, the disparity (inter-view) vector is necessary for prediction, but only motion (temporal) vectors are available in neighboring units.

Unfortunately, disparity vectors provide no information on motion and motion vectors provide no information on disparity. Therefore, in that case, the coder is unable to derive prediction candidates from neighboring units. In both described cases, the NP can provide a prediction candidate and fill in the sparse prediction candidate list.

The idea behind NP is to track the neighboring unit vector and identify the unit pointed by the neighboring unit vector. Once identified the pointed unit can be used as source of prediction candidate vectors.

Example of NP is shown in Figure 5. The disparity vector of neighboring unit (red) points to another unit (light red) in the reference frame. If the pointed unit has an appropriate motion vector (red dotted vector), the pointed unit vector can be scaled and used as a prediction candidate. In case shown on Figure 5, the NP is able to provide indirectly derived prediction candidate and increase the number of available prediction candidates.

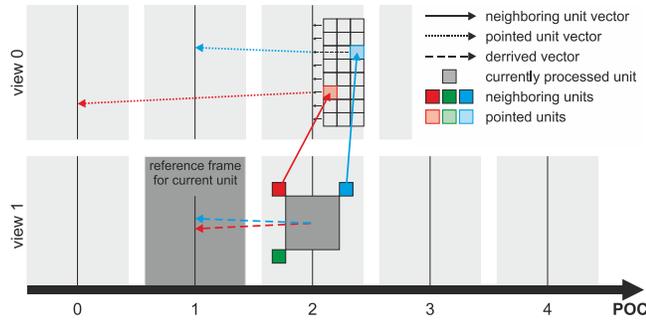


Figure 5. Example of nested prediction (NP).

4. PERFORMANCE EVALUATION AND EXPERIMENTAL RESULTS

The performance evaluation was done under the conditions described in the Call for Proposals (CfP) on 3D Video Coding (3DVC) [4] and in Common Test Conditions [5]. The CfP defines two coding scenarios: 2-view and 3-view case and gives a set of multiview sequences. The test sequences are divided into two classes: Class A – FullHD resolution (Poznan_Hall2, Poznan_Street, Undo_Dancer, GT_Fly) and Class C – XGA resolution (Kendo, Balloons, Lovebird1, Newspaper). The same coding scenarios and test sequences were used to evaluate the performance of responses to CfP on 3D Video Coding.

In both 2- and 3-view cases, the first dependent view was encoded using the base view as a reference. In the 3-view case, the second dependent view was encoded using both the base view and first dependent view as a reference. The available side views served as a reference for both anchor and non-anchor frames.

All the results are gathered for coding conditions conforming random access high efficiency setup (RA-HE) [5]. The BD-rate gain [6] has been calculated.

General comparison between MVC, simulcast HEVC and MV-HEVC are shown in Table 1. Comparison between simulcast

HEVC and MVC shown, that simulcast HEVC can provide significantly higher compression than MVC, despite not exploiting inter-view dependencies. Those results are similar to described in [7]. When compared to simulcast HEVC, the multiview HEVC (MV-HEVC) can provide about 22% and 30% bitstream reduction for 2- and 3-view case, respectively. The inter-view texture prediction applied to HEVC has similar efficiency to that in MVC.

The experimental results for IPMV-HEVC are collected in Table 2. Coding results are shown in comparison to MV-HEVC coder [3]. Any of the described modification changes base view encoding process and there is no coding gain for base view. Therefore, Table 2. shows coding efficiency gains for dependent views only.

All proposed tools provide noticeable coding performance gain without increasing the encoding and decoding time. Inter-view disparity vectors scaling itself provides coding performance gain only for 3-view case because this tool can be only used if two inter-view reference images are available on the reference list. Once the inter-view collocated is added, the highest compression gain for a single tool is achieved. This tool is useful in both 2- and 3-view case. The Nested Prediction (NP) also provides coding performance gain for both scenarios. Applying all described modifications allow increasing coding efficiency more than any single modification. The results achieved for this setup shows that both main improvements (inter-view collocated units and NP) complement each other.

Table 1. Bitrate reductions for luma.

	HEVC simulcast vs. MVC	MV-HEVC vs. HEVC simulcast	MV-HEVC vs. MVC
	2-view case		
Class A	-36.0 %	-26.8 %	-54.5 %
Class B	-47.0 %	-18.6 %	-57.3 %
Overall	-41.5 %	-22.7 %	-55.9 %
3-view case			
Class A	-30.4 %	-35.4 %	-57.3 %
Class B	-44.3 %	-25.4 %	-59.1 %
Overall	-37.4 %	-30.4 %	-58.2 %

Table 2. Bitrate reductions for variants of IPMV-HEVC vs. MV-HEVC.

	Inter-view Vector Scaling	Inter-view Collocated Units	Nested Prediction	All tools used
2-view case – dependent view				
Class A	0.00 %	-0.51 %	-0.11 %	-0.44 %
Class B	0.00 %	-0.67 %	-0.44 %	-0.94 %
Overall	0.00 %	-0.59 %	-0.28 %	-0.44 %
3-view case – 1st dependent view				
Class A	0.00 %	-0.77 %	-0.61 %	-1.08 %
Class B	0.00 %	-0.32 %	-0.38 %	-0.63 %
Overall	0.00 %	-0.55 %	-0.50 %	-0.86 %
3-view case – 2nd dependent view				
Class A	-0.49 %	-1.76 %	-0.32 %	-2.14 %
Class B	-0.33 %	-1.48 %	-0.89 %	-2.02 %
Overall	-0.41 %	-1.62 %	-0.60 %	-2.08 %

5. CONCLUSIONS

In this paper, the multiview HEVC (MV-HEVC) and improvement prediction for multiview HEVC (IPMV-HEVC) was proposed and evaluated. The MV-HEVC inter-view texture

prediction allows to achieve coding gain similar or even better than MVC coding gain. The new improvements of MV-HEVC for motion data prediction have been proposed. The IPMV-HEVC provides an additional bitstream reduction for dependent views without a significant increase in complexity.

Acknowledgement: This work was supported by the public funds as a research project.

6. REFERENCES

[1] ISO/IEC 14496-10, International Standard, “Generic coding of audio-visual objects – Part 10: Advanced Video Coding,” 6th Ed., 2010, [also:] ITU-T Rec. H.264, Edition 5.0, 2010.
 [2] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-G1103, B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, T. Wiegand (Eds.), “High Efficiency Video Coding

(HEVC) text specification Working Draft 5”, Geneva, November 2011.
 [3] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, “Multiview HEVC – experimental results” JCT-VC (MPEG/VCEG) Doc. JCTVC-G582, Geneva, November 2011.
 [4] Call for Proposals on 3D Video Coding Technology, ISO/IEC JTC1/SC29/WG11, MPEG2011/N12036, Geneva, March 2011.
 [5] F. Bossen, “Common test conditions and software reference configurations”, JCT-VC (MPEG/VCEG) Doc. JCTVC-F900, Torino, July 2011.
 [6] G. Bjøntegaard “Improvements of the BD-PSNR model,” ITU-T SG16 Q.6 Document, VCEG-A111, Berlin, July 2008.
 [7] K. Wegner, O. Stankiewicz, K. Klimaszewski, M. Domański “Comparison of multiview compression performance using MPEG-4 MVC and prospective HVC technology”, ISO/IEC JTC1/SC29/WG11 MPEG M17913, Geneva, July 2010

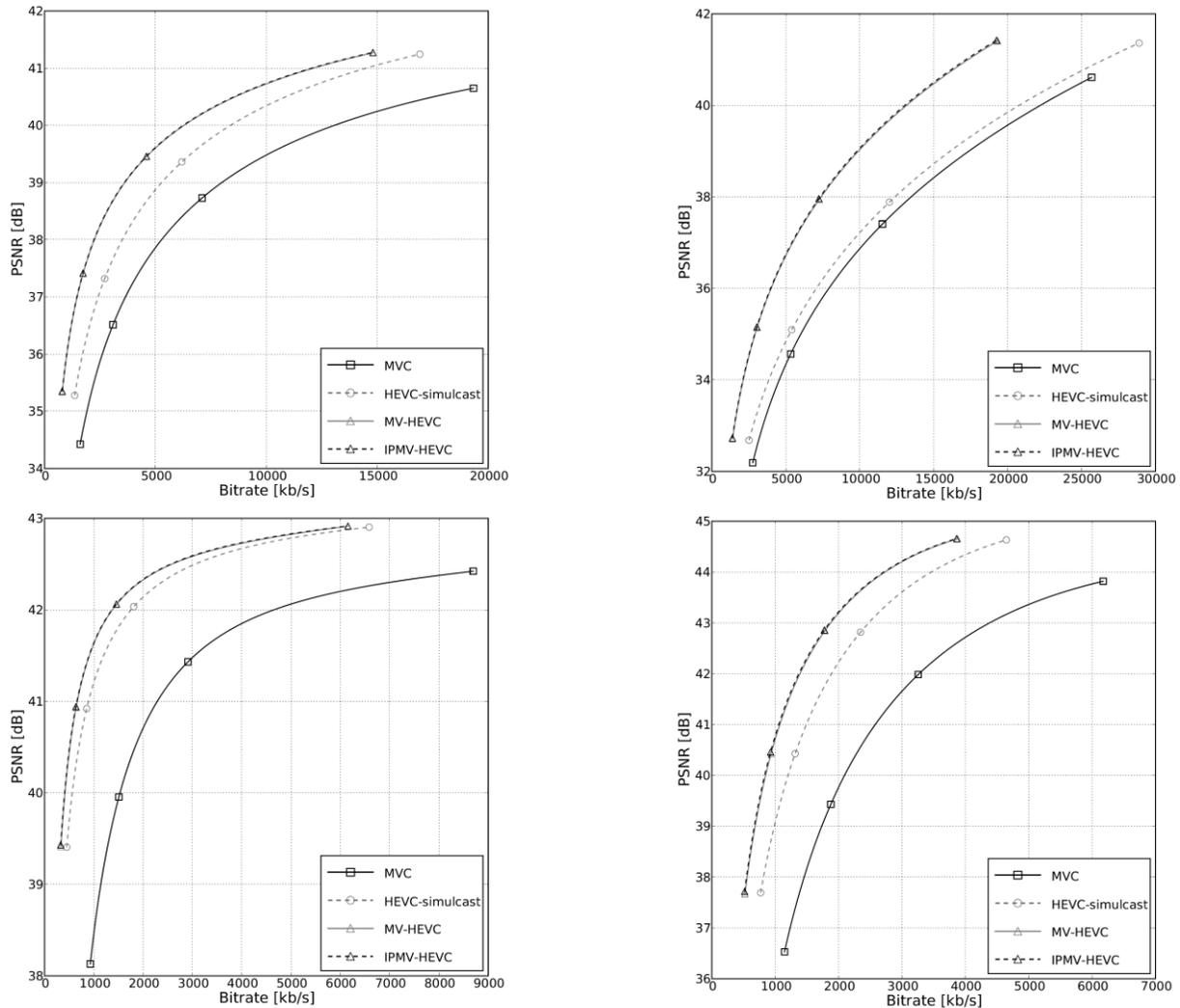


Figure 6. Experimental curves for “Poznan_Street” (left-top), “Poznan_Hall2” (right-top), “Dancer” (left-bottom), “Baloons” (right-bottom) sequences.