



biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

3D Video Compression Based on High Efficiency Video Coding

Glenn Van Wallendael, Sebastiaan Van Leuven, Jan De Cock, Fons Bruls, Rik Van de Walle

IEEE Transactions on Consumer Electronics, Volume 58, Issue 1, pp. 137-145, February 2012.

To refer to or to cite this work, please use the citation to the published version:

G. Van Wallendael, S. Van Leuven, J. De Cock, F. Bruls, R. Van de Walle, "3D Video Compression Based on High Efficiency Video Coding", IEEE Transactions on Consumer Electronics, Volume 58, Issue 1, pp. 137-145, February 2012

3D Video Compression Based on High Efficiency Video Coding

Glenn Van Wallendael, *Student Member*, IEEE, Sebastiaan Van Leuven, *Student Member*, IEEE, Jan De Cock, *Member*, IEEE, Fons Bruls, and Rik Van de Walle, *Member*, IEEE

Abstract — *With the advent of autostereoscopic displays, questions rise on how to efficiently compress the video information needed by such displays. Additionally, for gradual market acceptance of this new technology it is valuable to have a solution offering forward compatibility with stereo 3D video as it is used nowadays. In this paper, a multiview compression scheme making use of the efficient single-view coding tools used in High Efficiency Video Coding (HEVC) is provided. Although efficient single view compression can be obtained with HEVC, a multiview adaptation of this standard under development is proposed, offering additional coding gains. On average, for the texture information, the total bitrate can be reduced by 37.2% compared to simulcast HEVC. For depth map compression, gains largely depend on the quality of the captured content. Additionally, a forward compatible solution is proposed offering the possibility for a gradual upgrade from H.264/AVC based stereoscopic 3D systems to an HEVC-based autostereoscopic environment. With the proposed system, significant rate savings compared to Multiview Video Coding (MVC) are presented¹.*

Index Terms — **Multiview video compression, HEVC, autostereoscopic 3D, H.264/AVC compatible.**

I. INTRODUCTION

To visualize 3D content on end-user displays, more video data needs to be transported than with 2D content. Therefore, an efficient compression algorithm for 3D content is vital for the adoption of 3D technology. There is a plurality of choice for 3D visualization [1], but from a compression point of view, two main techniques can be differentiated [2].

First, 3D perception can be simulated by a stereo pair of video streams. This video pair is captured in such a way that when displayed separately on each eye, a 3D effect is created. Because two views are needed to create this 3D illusion, the

encoder needs to compress twice the amount of visual information.

The second category of 3D visualization techniques makes use of a multitude of video streams to create a 3D effect. Displays using this technique are called autostereoscopic displays and generate a 3D experience without the need for additional glasses. The amount of views required by the autostereoscopic display depends on the display characteristics. In general, the number of views (N) varies between 7 and 28. Processing these N views in the whole 3D production and distribution chain would put a large load on the system turning it impractical. To reduce the number of views required by the display and to enable a common interface for all these displays, different synthesization techniques are used. These techniques convert a restricted number (M) of views to the N views required by the display. From a compression and transmission point of view, these warping techniques come in two varieties. First, additional views can be created solely with pixel information from the captured views [3]. Disparities between the M views are estimated and with this information, N views are created. As an alternative for the disparity estimation process, during the capturing or production stage, depth maps can be created. Warping is then facilitated with the combination of texture and depth information [4] [5]. Different algorithms exist, but a realistic creation of N views can already be obtained from a view synthesis of three views optionally combined with three corresponding depth maps.

In the market, compression of any of these 3D representation technologies requires forward compatibility with a 2D compression scheme. Because of its superior compression efficiency, H.264/AVC [6] is chosen in a lot of video applications as a basis to be compatible with. On top of this video compression standard, a multiview extension was defined to efficiently compress a multitude of views. This standard was called Multiview Video Coding (MVC) [7]. A subset of MVC, called Stereo Profile, enjoyed large adoption in the market of stereo 3D, e.g. Blu-ray 3D.

Following these activities, there was evidence that an even better 2D compression algorithm was needed to catch up with the ever increasing bandwidth demand of video content. Similar to H.264/AVC standardization, a collaboration of MPEG and VCEG called Joint Collaborative Team on Video Coding (JCT-VC) was created to develop this standard. The process is entering a stabilization phase in its development and the final standard will be called High Efficiency Video Coding

¹ The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), Ph.D. and post-doctoral fellow grants of the Agency for Innovation by Science and Technology (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union. Furthermore, this work was carried out using the Stevin Supercomputer Infrastructure at Ghent University.

Glenn Van Wallendael, Sebastiaan Van Leuven, Jan De Cock, and Rik Van de Walle are with the Multimedia Lab research group, Department of Electronics and Information Systems (ELIS), Ghent University - IBBT, Ghent, Belgium (e-mail: { glenn.vanwallendael; sebastiaan.vanleuven; jan.decock; rik.vandewalle}@ugent.be)

Fons Bruls is with the Video & Image Processing group, Philips Research, Eindhoven, Netherlands (e-mail: fons.bruls@philips.com)

(HEVC) [8]. Compared to H.264/AVC, this technology under development already objectively reduces bandwidth with 44% on average [9]. It is estimated that with a subjective comparison, this reduction can even be higher.

With an increased performance of single view coding, it becomes more difficult to obtain additional gains from multiview coding of the different views. In this paper, it is shown that multiview coding can still bring significant improvements even when a forward compatible solution with H.264/AVC is facilitated.

First, MVC and the techniques used in this standard to efficiently compress multiple views will be explained in Section II. Then, fundamental HEVC properties necessary to understand a multiview version of HEVC will be introduced in Section III. Section IV will discuss the proposed changes to HEVC to enable multiview encoding. An H.264/AVC forward compatible version of this multiview HEVC codec will be explained in Section V. Finally, results from both proposed HEVC-based 3D video compression schemes will be given in Section VI followed by conclusions in Section VII.

II. MULTIVIEW VIDEO CODING

MVC, the multiview extension of H.264/AVC [7], was mainly developed for efficient compression of scenes shot from different viewpoints. For a multiview compression scheme it does not matter if video streams originate from a free viewpoint setup around the scene of interest or from a 3D production camera. The essence is that neighboring views from the same scene contain a lot of correlation. Multiview compression makes use of this aspect by taking into account neighboring views during the compression process of a view.

The way MVC uses inter-view correlation is similar to how single view compression takes advantage of temporal correlation between successive frames. In a single-view block-based codec like H.264/AVC, this correlation is reduced with a motion compensation process. With the multiview extension, the same process is applied with a neighboring view which is then called disparity compensation instead of motion compensation. The process of prediction between different views is called inter-view prediction. An example multiview configuration with three views is given in Fig. 1. In this figure, horizontal arrows indicate temporal prediction within a view. Vertical arrows indicate inter-view prediction between different views.

As an extension of H.264/AVC, MVC provides forward compatibility with its single view variant. One view within MVC is always independently encoded from the other views making this view H.264/AVC compatible [10]. In Fig. 1, it can be observed that the center view is the H.264/AVC compatible video stream because only temporal referencing is used. This view is also named the base view of a multiview video stream. For the left and the right view, inter-view prediction is applied, making these views only decodable with an MVC compliant decoder. The left and right view of this example configuration can also be called enhancement views. In general, forward

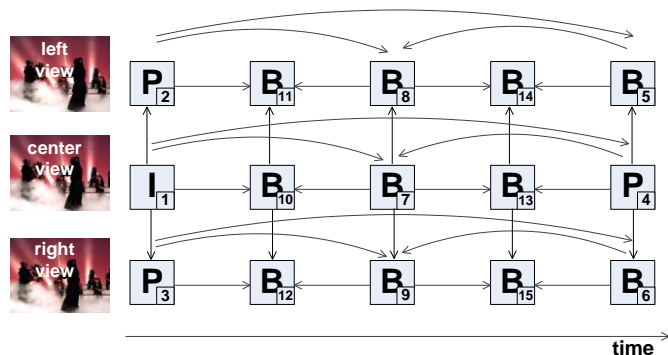


Fig. 1 Multiview compression configuration

compatibility was one of the major advantages facilitating the adoption of MVC in the market.

Not only can the H.264/AVC compatible view be decoded from the MVC video stream, it can also easily be extracted from the video stream. This property is called view scalability and can be applied to every view within the multiview video stream. In general, from an MVC video stream, a subset of views can be extracted creating a new and smaller video stream. It must be remarked that when extracting a certain view, all dependent views must remain in the video stream.

Additional to forward compatibility, MVC has a plurality of other properties which are important to maintain when trying to improve upon this compression scheme. First of all, to be realizable in practice, inter-view prediction was restricted to other views within the same access unit [11]. Inter-view prediction with frames from a different view on a different time instance was not allowed. In Fig. 1, this property can be verified by the fact that vertical arrows indicating inter-view prediction only stay within the same time instance.

A last requirement that should be maintained in the proposed HEVC multiview compression scheme is temporal random access. Single view random access can be facilitated with an Intra (I) frame. For random access purposes, H.264/AVC offers a picture type called Instantaneous Decoder Refresh (IDR). This picture indicates a full reset of the decoder before effectively decoding the I-frame contained in it. When multiple views are compressed, this restriction can be relaxed, because inter-view prediction with depending views should still be allowed. As a result, a concept of anchor pictures is created. These pictures only depend on frames from the same time instance enabling random access in all views. Such anchor pictures, indicated as P_2 and P_3 , are illustrated in the first time instance of Fig. 1.

III. HIGH EFFICIENCY VIDEO CODING

On a high level, there is a lot of correspondence between HEVC and H.264/AVC. A video stream is still divided in Network Access Layer (NAL) units. NAL units provide a network friendly representation of the video coding data and form the basic units on which scalability features can be applied. For example, with H.264/AVC, temporal scalability can be applied by discarding the NAL units of the highest

temporal layer. When modifying HEVC to a multiview codec, this NAL unit property is important to retain for view scalability purposes.

In H.264/AVC, two different NAL units were used to signal high level information to the decoder, Sequence Parameter Set (SPS) and Picture Parameter Set (PPS). In HEVC, these sets are modified and extended with two additional sets, namely Adaptation Parameter Set (APS) and Reference Parameter Set (RPS). In the APS, information that is not likely to change from one slice to the other is included. The RPS is a straightforward method replacing the Reference Picture List Modification (RPLM) Memory Management Control Operation (MMCO) messages in H.264/AVC. Furthermore, in H.264/AVC, to enable motion compensation with other pictures, a buffer with available reference pictures called the Decoded Picture Buffer (DPB) was maintained. By indexing in this buffer the reference picture lists are filled either implicitly with a default filling procedure or explicitly with RPLM commands. Because the indexing happened relative to the DPB position, a mismatch in encoder and decoder caused by packet loss could remain undetected. On the contrary, in the RPSs used in HEVC, only explicit signaling of reference frames is incorporated. This is done by including for every reference list an enumeration of the Picture Order Counts (POC) of the referenced frames. The POCs in the RPS are relative to the current POC making it easier to reuse the RPS for frames with similar prediction structure. As a result, it is very convenient for an HEVC decoder to detect frame losses while decoding the video stream. The reference picture lists filled by means of RPS signaling will be utilized on a lower level as described next.

HEVC is still a block based video compression scheme, but partitioning of the picture comes with increased flexibility compared to H.264/AVC [12][13]. On a high level, the concept of slice partitioning is kept similar to H.264/AVC. It is on the lower levels that the partitioning differences proliferate. Instead of dividing the slice in macroblocks of 16x16 pixels, Largest Coding Units (LCU) of 64x64 pixels form the basis for a first subdivision. These LCUs are processed in a raster scan order within the slice. On every LCU, a quadtree partitioning is applied dividing the LCU in Coding Units (CU) as small as 8x8 in size. All different sized CUs in the LCU are signaled in a Z-scan order traversing the formed quadtree. On a CU level, the encoder can choose to skip, intra predict, or inter predict the considered block.

Starting on the CU size, information about the way the block is predicted is made clear by means of Prediction Unit (PU) information. A CU can be predicted entirely at once, resulting in a PU size equal to the CU size or it can be split in smaller rectangular or square PUs. On the PU level, motion information consisting of the chosen reference frames, the motion vectors, and the motion vector predictors is indicated. It is on this level that a reference frame can be chosen with an index in the reference picture lists. These reference picture

lists are created by means of RPS signaling as described earlier.

After prediction of the CU, a transform step is performed described on the Transform Unit (TU) level [14]. Again, a recursive tree can be created on this level reducing the transform to a 4x4 size.

For random access purposes, HEVC offers similar functionality as H.264/AVC by means of IDR pictures. Additionally, Clean Random Access (CRA) pictures are included as well. These pictures fulfill the same functionality as open-GOP I-pictures but in an explicit standardized way.

With this subset of features of HEVC a multiview HEVC adaptation is proposed as described next.

IV. HEVC-BASED MULTIVIEW COMPRESSION

In the proposed Multiview HEVC-based Coding (MHC), single view HEVC is adapted such that it matches the features of MVC as closely as possible. As a consequence, the prediction structure used for MVC as described in Fig. 1 will still be applicable to MHC.

The biggest change in the realization of MVC was enabling a certain view to be inter-view predicted from an earlier decoded view. In MHC, this concept was enabled similarly to MVC by giving the possibility to include pictures from an earlier decoded view to the reference picture lists. The change needed to enable this feature can be found in the RPS signaling. In the RPS, reference frames are indicated with a POC difference relative to the current POC. To access another view at the same time instance, a POC difference of zero is enabled in the RPS. Because different views at the same time instance can occur, an additional view index must be signaled. In our proposed approach, only the closest view is used for inter-view prediction because of rate-distortion considerations. Consequently, the additional view index does not need to be utilized in the remaining of this paper.

Specifically, in the RPS, it is signaled to first include the lower view from the same temporal unit before surrounding pictures are included as well. So, except from the first reference picture being the lower view, no modifications were made to the reference picture list filling algorithm. The size of the reference picture list was restricted to four after evaluating complexity against compression efficiency.

By making inter-view prediction possible on a reference picture level, adaptive inter-view prediction on PU level results. More specifically, each PU indicates the chosen reference frame by means of an index in the reference picture lists. In these lists one or several lower views occur, making it possible to choose for inter-view prediction instead of temporal prediction. On a PU level, the encoder can choose in a Rate Distortion (RD) optimal way if inter-view prediction or inter-frame prediction should be applied.

With the proposed multiview compression scheme, forward compatibility with HEVC is guaranteed similarly to the forward compatibility provided by MVC. The same remark can be made for the view scalability aspect of MHC.

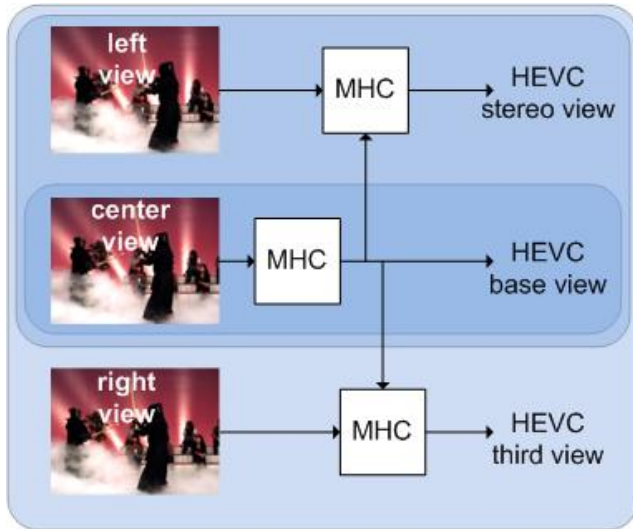


Fig. 2 scalable three view MHC configuration

Additionally, in the proposed compression, the complexity restriction limiting inter-view prediction only for within the same time instance is forced as well.

For random access purposes, the inclusion of the extra CRA within HEVC results in small modifications for anchor picture signaling. Anchor pictures associated with a CRA picture instead of an IDR picture should imply the restrictions of the CRA.

When graphically representing the proposed multiview HEVC compression, Fig. 2 can be obtained. From this figure, it is clear that the center view is configured as the forward compatible HEVC view. When the target application only supports stereo vision, the left and center view can be easily extracted. For the stereo vision scenario, the left view is arbitrarily chosen. From the dependency arrows it can be observed that a stereo pair of center and right is equally possible.

Overall the encoding speed of multiview compared to simulcast remains unchanged because the number of reference frames is not altered.

At the decoder, it could be stated that the dependencies would restrict the decoder to sequential decoding of every frame in the different views. This is not the case, because by pipelining the decoding process, only the first two frames are to be decoded sequentially.

V. FORWARD COMPATIBLE MULTIVIEW COMPRESSION

A new video coding standard can offer opportunities for more efficient video systems. However, migrating a hardware based environment to a new video codec brings about a huge investment providers or end users are not willing to pay for. A gradual upgrade where investments could be spread in time could offer the solution. For such a gradual upgrade to work, a forward compatible design is proposed here.

Nowadays, the majority of hardware solutions provide compatibility with H.264/AVC compression. Therefore, this

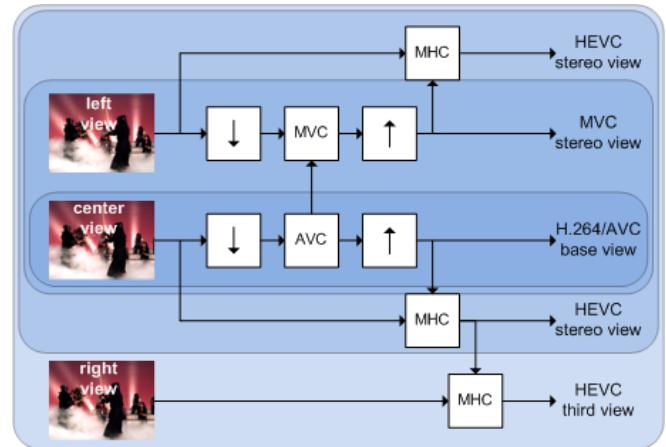


Fig. 3 H.264/AVC compatible multiview compression based on HEVC

standard is chosen as the basis to start from. In Fig. 3, an overview of the proposed forward compatible solution is provided. In the middle of the figure, it can be noticed that the center view acts as the H.264/AVC compatible base view. As a result, legacy decoders that only provide compatibility with H.264/AVC will decode a 2D representation of the scene, shot with the center camera. Additional to the H.264/AVC compression, a down sampling and up sampling filter are added to the base view compression scheme. Before applying the H.264/AVC compression, a horizontal downsampling with a factor two is performed on the input sequence. The downsampled sequence is then H.264/AVC encoded to a bitstream. After decoding the bitstream, a horizontal upsampling is applied to regenerate the original resolution. The down and upsampling operation is included to offer forward compatibility with MVC based stereo 3D.

In the proposed compression scheme, stereo 3D compatibility is offered with an MVC enhancement based on the left camera view. Similar to the base view, horizontal down and upsampling with a factor two is also applied on this video information. In combination with the H.264/AVC base view, a stereo pair is created with each view downsampled to half the horizontal resolution. Conceptually, a side-by-side MVC compatible stereo pair is created. This combination is chosen because of the large adoption of side-by-side stereo 3D in the market.

For newer systems, which provide HEVC compatibility, a full resolution enhancement is made of both views of the stereo pair. This is accomplished by multiview encoding the full resolution with the aid of the half resolution version of the same view. As an example, for the full resolution HEVC encoded center view, the upsampled half resolution H.264/AVC encoded center view acts as the dependent view.

The last step needed to provide autostereoscopic functionality is encoding the third view in an efficient way. Because a full resolution version of the center and left stereo pair is available, a multiview compression of the right view depending on the center view is utilized. As a result, with this

scheme, a three view HEVC-based 3D video compression scheme is obtained.

It must be noted that this compression scheme is only relevant for texture information. In the proposed scenarios, depth maps are only needed when synthesizing the three input textures to N views needed by the autostereoscopic display. As a result, for depth maps, the HEVC-based multiview compression as described in the previous section can be applied.

VI. RESULTS

In this paper, two 3D video compression schemes based on HEVC are described. In this section, for each of the schemes, a relevant test configuration will be described which covers a broad range of application scenarios. Possible applications covered by the test results vary from IPTV to broadcast TV, Personal Video Recording (PVR), and video on optical storage. Outside the scope of these test results are video conferencing and similar scenarios.

Corresponding to the description order in this paper, multiview compression entirely based on HEVC will be evaluated first. Because of the technological relevance of compressing two and three views, our proposed multiview compression is evaluated for these cases. The implementation of our proposed multiview compression scheme is based on HEVC Model (HM) 3.0 as developed in JCT-VC during HEVC standardization. The sequences on which the tests are run, correspond to the test sequences used for the Call for Proposals (CfP) of 3D Video [15] issued by MPEG. These sequences are named in Tables I and II. The first four sequences are 1080p in resolution with 30fps. The last four sequences are 720p at 25fps. To simulate realistic scenarios as described before, a random access period of 16 frames is configured. This approximately corresponds with a random access of 0.5s up to 0.7s. As a tradeoff between compression efficiency and complexity, four reference frames and a GOP size of eight are used. To challenge the proposed multiview scheme, hierarchical coding is enabled. Furthermore, an internal bit depth of 10 is used in combination with CABAC and rate distortion optimized quantization (RDOQ). This results in a highly efficient configuration of HEVC. Finally, to generate results in a 30 to 40 dB PSNR range, Quantization Parameters (QP) between 35 and 47 are selected in steps of 4.

Additional to these decisions, it can also be observed that a very compression efficient tool for texture coding did not perform well for depth map coding. This filtering tool, called Adaptive Loop Filter (ALF) only reduced the rate with 0.3% while keeping the quality constant. Therefore, it was decided to disable this tool for depth maps and obtain a significant faster encoding and decoding speed.

Bjøntegaard Delta (BD) measurements for two views with these configurations can be found in Table I. The BD-rate gains stated in the table are calculated from rate distortion graphs resulting from the average PSNR of both views and the sum of the bitrates of both views. The first three columns show

TABLE I
BD-RATE GAIN OF MULTIVIEW COMPRESSION COMPARED TO SIMULCAST WHEN ENCODING TWO VIEWS [%]

Sequence	Y	U	V	Depth
poznan_hall2	-23.2	-23.5	-24.2	4.9
poznan_street	-34.6	-34.5	-34.0	0.4
undo_dancer	-36.0	-37.8	-37.4	-25.5
gt_fly	-37.7	-38.1	-38.1	-28.1
kendo	-21.5	-20.0	-21.3	2.1
balloons	-25.4	-24.1	-25.0	2.2
lovebird1	-31.7	-30.6	-30.9	-4.8
newspaper	-26.6	-27.9	-26.5	5.1
average	-29.6	-29.5	-29.7	-5.5

TABLE II
BD-RATE GAIN OF MULTIVIEW COMPRESSION COMPARED TO SIMULCAST WHEN ENCODING THREE VIEWS [%]

Sequence	Y	U	V	Depth
poznan_hall2	-26.1	-27.2	-28.2	8.3
poznan_street	-45.8	-46.9	-44.8	0.3
undo_dancer	-44.7	-47.1	-46.7	-30.3
gt_fly	-47.2	-48.2	-48.2	-35.6
kendo	-28.1	-29.0	-27.7	3.4
balloons	-33.0	-31.3	-32.3	4.4
lovebird1	-41.1	-38.0	-38.9	1.0
newspaper	-31.5	-31.5	-32.1	6.4
average	-37.2	-37.4	-37.4	-5.3

BD-rate gains for the luma and chroma components of the multiview video stream. The last column indicates BD-rate gains for compression of the depth map.

From Table I, it can be observed that, for two views, a joint compression of the texture results in 29.6% average bitrate gain. For the three view case, an overall gain of 37.2% is shown in Table II. From these numbers, it can be derived that, on average, a BD-rate reduction of more than 55% can be obtained for every enhancement view individually. By jointly coding texture views, every enhancement texture takes less than half the bitrate when compared to simulcast.

From both tables, it can be observed that multiview compression of depth maps is less efficient than simulcast except for synthetically generated depth maps like *gt_fly* and *undo_dancer*. This is because other depth maps contain a lot more noise.

For the evaluation of the forward compatible multiview compression, the H.264/AVC component of the compression scheme was implemented using Joint reference Model (JM) version 18.0. According to performed visual tests, the QPs used in different views were changed depending on the relevance of each view for the view synthesis process. To

TABLE III
FOUR TEST RATE POINTS INDICATING THE MAXIMUM BITRATE FOR EVERY SEQUENCE IN A 2-VIEW SCENARIO.

Seq ID	Sequence	Rate 1 [kbps]	Rate 2 [kbps]	Rate 3 [kbps]	Rate 4 [kbps]
S01	poznan_hall2	500	700	1000	1500
S02	poznan_street	500	700	1000	1250
S03	undo_dancer	1000	1300	1700	2200
S04	gt_fly	1200	1700	2100	2900
S05	kendo	400	500	800	1300
S06	balloons	320	430	600	940
S07	lovebird1	375	500	750	1250
S08	newspaper	400	525	800	1300

TABLE IV
FOUR TEST RATE POINTS INDICATING THE MAXIMUM BITRATE FOR EVERY SEQUENCE IN A 3-VIEW SCENARIO.

Seq ID	Sequence	Rate 1 [kbps]	Rate 2 [kbps]	Rate 3 [kbps]	Rate 4 [kbps]
S01	poznan_hall2	750	900	1300	2300
S02	poznan_street	750	1100	1800	4000
S03	undo_dancer	1380	1750	2300	2900
S04	gt_fly	2000	2380	2900	4000
S05	kendo	800	1000	1300	1900
S06	balloons	500	600	800	1250
S07	lovebird1	500	800	1250	2000
S08	newspaper	500	700	1000	1350

simplify this quality weighting process, a reference QP was chosen where all other QPs are derived from. In general, it can be stated that when the encoded view is depending on another view the QP parameter is increased with a delta of three. Furthermore, the QP of the depth information is increased with a QP delta of three relative to the corresponding view.

Additional to these design decisions, restrictions are applied as described for the evaluation of the 3D Video CfP from MPEG. One of these restrictions states that encoded video streams should be encoded at four rate points and at each point the total video bitrate should not exceed the rates in Table III and Table IV.

Additionally, a random access restriction is applied stating random access should be possible within 0.5 seconds. Because the evaluated sequences are 25 and 30 frames per second, a random access period of 12 frames is applied. Consequently, it is most efficient to apply a GOP period of 12 frames as well. Also, 28 synthesized views must be generated from every 3-view compressed video stream to perform the visualization on an autostereoscopic display. For the synthesization of the three texture views and depth maps to these 28 views the MPEG View Synthesis Reference Software (VSRS) version 3.5 [16] is used.

Objective results from the odd numbered sequences are included in Fig. 4 to Fig. 11. For every sequence first the results in a 2-view scenario are presented followed by the results from the 3-view scenario. From these graphs it can be concluded that the proposed HEVC-based solutions always outperforms MVC. When comparing the forward compatible solution as proposed in this paper with simulcast HEVC, in Fig. 4 and Fig. 8, a small performance drop can be noticed. On the other hand, for the undo_dancer sequence (Fig. 6) and lovebird1 sequence (Fig. 10), the proposed forward compatible solution outperforms simulcast HEVC. For the 3-view scenario, similar or improved performance can be noticed for all sequences. Some caution should be exercised when comparing these methods because there is only a small overlap between both curves making a comparison difficult. When comparing the proposed multiview HEVC compression with simulcast HEVC, similar performance gains as stated in the previous comparison can be found.

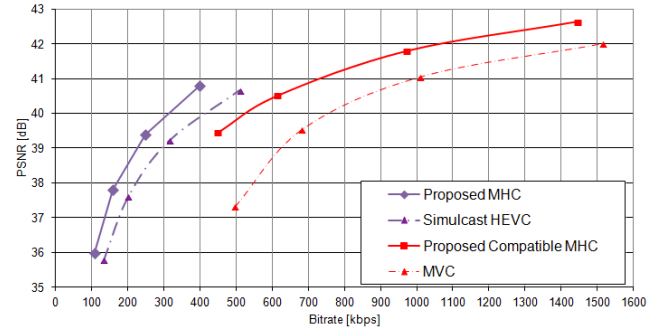


Fig. 4 Both proposed 2-view 3D compression schemes compared to simulcast HEVC and MVC for poznan_hall2.

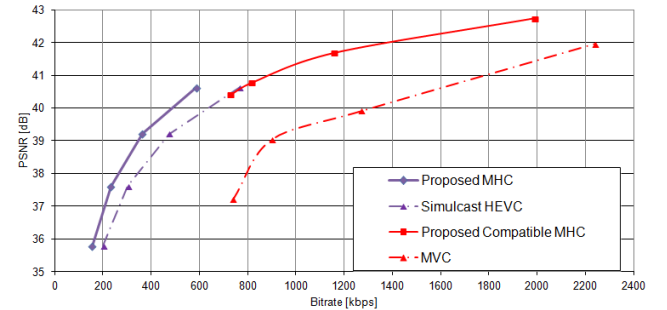


Fig. 5 Both proposed 3-view 3D compression schemes compared to simulcast HEVC and MVC for poznan_hall2.

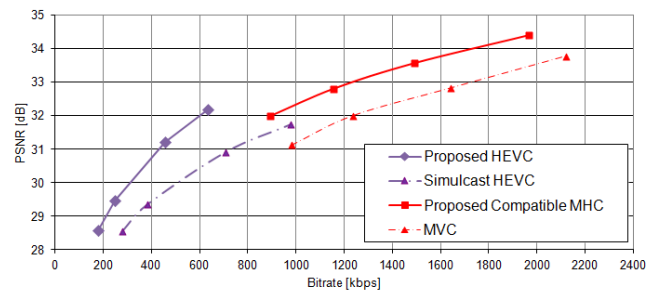


Fig. 6 Both proposed 2-view 3D compression schemes compared to simulcast HEVC and MVC for undo_dancer.

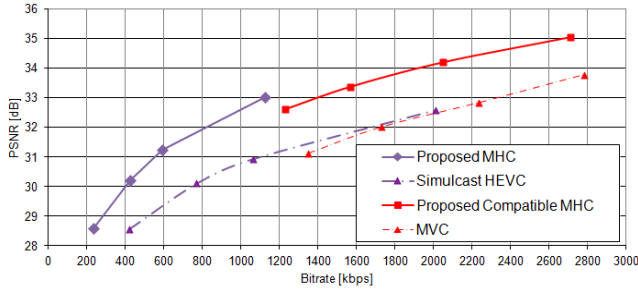


Fig. 7 Both proposed 3-view 3D compression schemes compared to simulcast HEVC and MVC for undo_dancer.

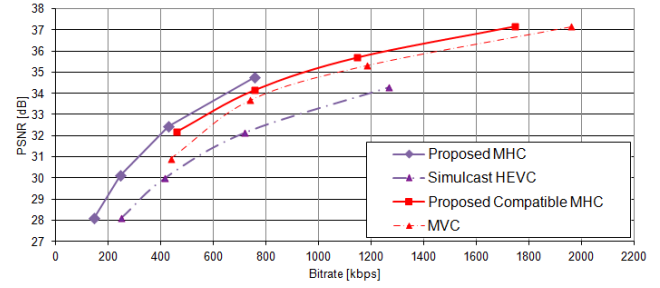


Fig. 11 Both proposed 3-view 3D compression schemes compared to simulcast HEVC and MVC for lovebird1.

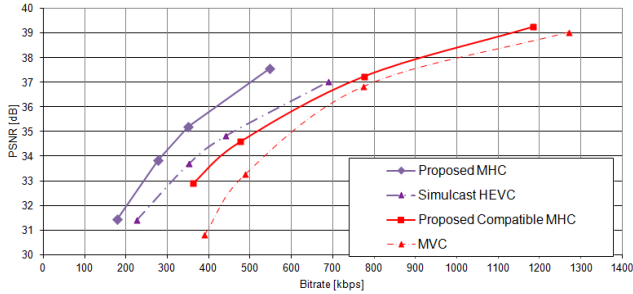


Fig. 8 Both proposed 2-view 3D compression schemes compared to simulcast HEVC and MVC for kendo.

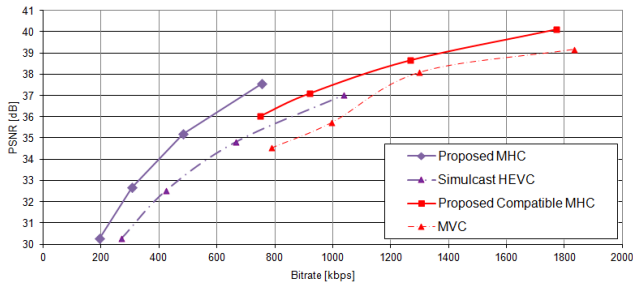


Fig. 9 Both proposed 3-view 3D compression schemes compared to simulcast HEVC and MVC for kendo.

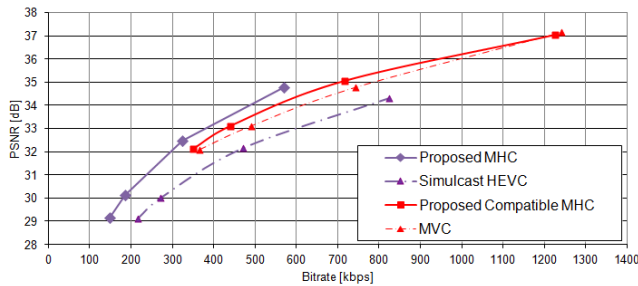


Fig. 10 Both proposed 2-view 3D compression schemes compared to simulcast HEVC and MVC for lovebird1.

To get more insight in how the bits are distributed between different components of the bitstream, an overview is given in Table V to Table VIII. For both the 2-view (Table V) and 3-view (Table VII) scenario, the ratio of MVC compatible data compared to the total bitstream is listed. It can be observed that although the MVC data only represents two views at half the horizontal resolution, it still constitutes a large amount of the total bitstream data. Additionally, the amount of bitstream data taken up by the depth information is enumerated in Table VI and Table VIII. From this information it can be concluded that depth information takes up a maximum of 32% of the total bitstream. It can also be observed that there is a large variation between different sequences. This is mainly caused by the capturing quality of the depth information and the information content it provides.

Finally, the proposed forward compatible HEVC-based multiview solution was provided as a response to the 3D Video CfP of MPEG. During the assessment of the responses, a subjective evaluation of all submitted techniques has been carried out by the MPEG Test Group with the help of several specialized labs from all over the world. Out of the 12 contributions in the AVC-compatible category to the 3D Video CfP, the proposed forward compatible technique was subjectively rated as the second best.

VII. CONCLUSION

In this paper, a multiview compression algorithm entirely based on HEVC and a solution which is forward compatible with H.264/AVC and stereo MVC is proposed. With the purely HEVC-based solution, we show that even with a compression efficient single view codec like HEVC, gains from multiview encoding can still be significant. For texture data, the gains are certainly worth the view dependability. However, for depth maps, compression performance is largely influenced by the capture quality of the depth map. With the forward compatible solution, evidence is provided showing significant improvements compared to MVC based compression. Furthermore, our proposed forward compatible solution is also able to compete with simulcast HEVC compression.

TABLE V
RATIO OF MVC COMPATIBLE DATA COMPARED TO THE TOTAL
BITSTREAM FOR 2-VIEW SCENARIO

Rate	S01	S02	S03	S04	S05	S06	S07	S08
Rate 1	57%	68%	58%	46%	46%	53%	64%	59%
Rate 2	56%	67%	59%	45%	45%	53%	66%	60%
Rate 3	55%	66%	59%	43%	46%	53%	66%	60%
Rate 4	55%	65%	58%	42%	45%	52%	63%	58%

TABLE VI
RATIO OF DEPTH DATA COMPARED TO THE TOTAL BITSTREAM FOR
2-VIEW SCENARIO

Rate	S01	S02	S03	S04	S05	S06	S07	S08
Rate 1	16%	12%	8%	12%	28%	22%	18%	21%
Rate 2	15%	11%	7%	12%	29%	22%	17%	21%
Rate 3	14%	12%	7%	12%	31%	23%	17%	22%
Rate 4	14%	12%	6%	11%	32%	24%	15%	23%

TABLE VII
RATIO OF DEPTH DATA COMPARED TO THE TOTAL BITSTREAM FOR
3-VIEW SCENARIO

Rate	S01	S02	S03	S04	S05	S06	S07	S08
Rate 1	55%	66%	60%	50%	49%	54%	65%	59%
Rate 2	55%	65%	60%	49%	49%	53%	67%	61%
Rate 3	53%	63%	60%	47%	49%	54%	64%	62%
Rate 4	51%	56%	59%	45%	48%	53%	60%	61%

TABLE VIII
RATIO OF DEPTH DATA COMPARED TO THE TOTAL BITSTREAM FOR
3-VIEW SCENARIO

Rate	S01	S02	S03	S04	S05	S06	S07	S08
Rate 1	15%	11%	7%	13%	26%	20%	17%	21%
Rate 2	14%	11%	7%	13%	27%	20%	16%	21%
Rate 3	14%	11%	6%	12%	28%	20%	15%	22%
Rate 4	13%	11%	5%	12%	29%	21%	13%	23%

REFERENCES

- [1] J. Konrad and M. Halle, "3-D Displays and Signal Processing – An Answer to 3-D Ills?", *IEEE Signal Processing Magazine*, Vol. 24, No. 6, Nov. 2007.
- [2] P. Merkle, K. Müller, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Trans. Consum. Electron.*, vol.56, no.2, pp.946-950, May 2010.
- [3] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D", *ACM SIGGRAPH 2010*, New York, NY, USA, Article 75, 10 pages.
- [4] Y.-m. Feng, D.-x. Li, K. Luo, and M. Zhang, "Asymmetric bidirectional view synthesis for free viewpoint and three-dimensional video," *IEEE Trans. Consum. Electron.*, vol.55, no.4, pp.2349-2355, Nov. 2009
- [5] K. Müller, P. Merkle, and T. Wiegand, "3-D Video Representation Using Depth Maps," *Proceedings of the IEEE*, vol.99, no.4, pp.643-656, Apr. 2011
- [6] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [7] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC standard", *Proceedings of the IEEE*, Special Issue on "3D Media and Displays", vol. 99, issue 4, pp. 626-642, Apr. 2011.
- [8] T. Wiegand, B. Bross, W.-J. Han, J.-R. Ohm, and G. J. Sullivan, "WD3: Working Draft 3 of High-Efficiency Video Coding", 5th JCT-VC Meeting, Geneva, CH, 16-23 March, 2011.
- [9] B. Li, G. J. Sullivan, and J. Xu, "Comparison of Compression Performance of HEVC Working Draft 4 with AVC High Profile," JCTVC-G399, Geneva, Switzerland, Nov. 2011.
- [10] M. Flierl, A. Mavliankar, and B. Girod, "Motion and Disparity Compensated Coding for Multiview Video," *IEEE Trans. Circuits Syst. Video Technol.*, vol.17, no.11, pp.1474-1484, Nov. 2007
- [11] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol.17, no.11, pp.1461-1473, Nov. 2007
- [12] M. Karczewicz, P. Chen, R. Joshi, X. Wang, W.-J. Chien, R. Panchal et al., "A Hybrid Video Coder Based on Extended Macroblock Sizes, Improved Interpolation, and Flexible Motion Representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1698–1708, Dec. 2010.
- [13] D. Marpe, H. Schwarz, S. Bosse, B. Bross, P. Helle, T. Hinz et al., "Video Compression Using Nested Quadtree Structures, Leaf Merging, and Improved Techniques for Motion Representation and Entropy Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1676–1687, Dec. 2010.
- [14] F. Bossen, V. Drugeon, E. Francois, J. Jung, S. Kanumuri, M. Narroschke et al., "Video Coding Using a Simplified Block Structure and Advanced Coding Techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1667–1675, Dec. 2010.
- [15] ISO/IEC JTC1/SC29/WG11 (MPEG), "Call for Proposals on 3D Video Coding Technology," Geneva, Switzerland, March 2011.
- [16] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference Softwares for Depth Estimation and View Synthesis," ISO/IEC JTC1/SC29/WG11 MPEG2008/M15377, Archamps, France, Apr. 2008.

BIOGRAPHIES



Glenn Van Wallendael obtained the M.Sc. degree in Applied Engineering from the University College of Antwerp, Antwerp, Belgium, in 2006 and the M.Sc. degree in Engineering from Ghent University, Ghent, Belgium in 2008. Since then, he is working towards a Ph.D. at Multimedia Lab, Ghent University, with the financial support of the Agency for Innovation by Science and Technology (IWT). Main topics of interest are video compression including scalable video compression and transcoding.



Sebastiaan Van Leuven received the M.Sc. degree in Applied Engineering from the University College of Antwerp, Antwerp, Belgium, in 2006 and the M.Sc. degree in Computer Science Engineering from Ghent University, Ghent, Belgium in 2008. Currently, he is with Multimedia Lab, Ghent University, where he is working towards a Ph.D., with the financial support of the Agency for Innovation by Science and Technology (IWT). His main research topic is video coding, including scalable video coding, adaptation of video streams, transcoding and next generation video compression.



Jan De Cock obtained the M.S. and Ph.D. degrees in Engineering from Ghent University, Belgium, in 2004 and 2009, respectively. Since 2004 he has been working at Multimedia Lab, Ghent University, and the Interdisciplinary Institute for Broadband Technology (IBBT). His research interests include video compression and transcoding, scalable video coding, and multimedia applications.



Fons Bruls received his degree in electrical engineering at the “Hogere Technische School Heerlen”, Heerlen, The Netherlands. Currently he works for the Philips Research Laboratories in Eindhoven in the Video Image Processing group. He contributed to the realization of HDTV video storage based on D1 recorders, the DV camcorder standard, video compression algorithms for MPEG encoding and transcoding, the DVD+RW recording standard, DVD recorders, the realization of single chip MPEG encoders and codecs, the MPEG 3D standard based on video plus depth, and one of the first commercial available 3D-TV's with variable depth adjustment for stereo input signals. As a principle scientist, he currently focuses on 3DTV video compression and standardization. He received twice the ICCE Best Paper Award. In 2003, he obtained two company Invention Awards. His work has resulted in over 100+ patents and patent applications, and various commercially available ICs.



Rik Van de Walle received his M.Sc. and PhD degrees in Engineering from Ghent University, Belgium in 1994 and 1998 respectively. After a visiting scholarship at the University of Arizona (Tucson, USA), he returned to Ghent University, where he became professor of multimedia systems and applications, and head of the Multimedia Lab. His current research interests include multimedia content delivery, presentation and archiving, coding and description of multimedia data, content adaptation, and interactive (mobile) multimedia applications.