MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Overview of the Multiview and 3D Extensions of High Efficiency Video Coding

Tech, G.; Chen, Y.; Mueller, K.; Ohm, J.-R.; Vetro, A.; Wang, Y.-K.

TR2015-125 September 2015

Abstract

The High Efficiency Video Coding standard has recently been extended to support efficient representation of multiview video and depth-based 3D video formats. The multiview extension, MV-HEVC, allows efficient coding of multiple camera views and associated auxiliary pictures, and can be implemented by reusing single-layer decoders without changing the block-level processing modules since block-level syntax and decoding processes remain unchanged. Bit rate savings compared to HEVC simulcast are achieved by enabling the use of inter-view references in motion-compensated prediction. The more advanced 3D video extension, 3D-HEVC, targets a coded representation consisting of multiple views and associated depth maps, as required for generating additional intermediate views in advanced 3D displays. Additional bit rate reduction compared to MV-HEVC is achieved by specifying new block-level video coding tools, which explicitly exploit statistical dependencies between video texture and depth, and specifically adapt to the properties of depth maps. The technical concepts and features of both extensions are presented in this paper.

2015 IEEE Transactions on Circuits and Systems for Video Technology

Overview of the Multiview and 3D Extensions of High Efficiency Video Coding

Gerhard Tech, Ying Chen, *Senior Member, IEEE*, Karsten Müller, *Senior Member, IEEE*, Jens-Rainer Ohm, *Member, IEEE*, Anthony Vetro, *Fellow, IEEE*, and Ye-Kui Wang

***Abstract*—The High Efficiency Video Coding standard has recently been extended to support efficient representation of multiview video and depth-based 3D video formats. The multiview extension, MV-HEVC, allows efficient coding of multiple camera views and associated auxiliary pictures, and can be implemented by reusing single-layer decoders without changing the block-level processing modules since block-level syntax and decoding processes remain unchanged. Bit rate savings compared to HEVC simulcast are achieved by enabling the use of inter-view references in motion-compensated prediction. The more advanced 3D video extension, 3D-HEVC, targets a coded representation consisting of multiple views and associated depth maps, as required for generating additional intermediate views in advanced 3D displays. Additional bit rate reduction compared to MV-HEVC is achieved by specifying new block-level video coding tools, which explicitly exploit statistical dependencies between video texture and depth, and specifically adapt to the properties of depth maps. The technical concepts and features of both extensions are presented in this paper.**

***Index Terms*—High Efficiency Video Coding (HEVC), Multiview HEVC (MV-HEVC), 3D-HEVC, Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V), Moving Picture Experts Group (MPEG), Video Coding Experts Group (VCEG), standards, video compression.**

I. Introduction

3D applications have attracted wide interest in recent years. While much of the emphasis has been on stereoscopic displays, which require glasses to enable depth perception, a new generation of auto-stereoscopic displays, which emit different pictures depending on the position of the observer's eyes and do not require glasses for viewing, are starting to emerge and become commercially available [1][2]. The latter often employ depth-based image rendering techniques to generate a dense set of views to the scene [3]. In order to render these views with acceptable quality, it is desirable to use high-quality depth maps, which need to be represented and coded along with the texture. Depth maps can be estimated from a stereo or multi-camera setup using stereo correspondence techniques [4]. They could also be acquired by a special depth camera; this particular area has seen notable advances in recent years with designs based on structured light [5] or time-of-flight based imaging [6]. Finally, depth information is an integral part of computer-generated imagery, which is popular in many cinema productions.

To address the above needs and leverage the state-of-the-art compression capabilities offered by the High Efficiency Video Coding (HEVC) standard [7][8], a vision for the next-generation 3D video format was published by MPEG [9] with the aim to develop a 3D video format that could facilitate the generation of intermediate views with high compression capabilities in order to support advanced stereoscopic display functionality and emerging auto-stereoscopic displays. Following this, a reference framework that utilized depth-based image rendering was prepared so that candidate technology could be evaluated. A key challenge was generating high-quality depth maps for the available multiview video sequences and preparing anchor material with sufficiently high quality. It was also critically important to define an appropriate evaluation procedure, as no well-defined process for evaluating the impact of depth coding and rendering results existed. It was ultimately decided to measure the PSNR of both coded and synthesized views as well as subjectively assess the quality on stereoscopic and autostereoscopic multiview displays.

In 2011, a Call for Proposals (CfP) was issued based on a specified set of requirements and the defined evaluation procedure [10], which solicited technology contributions for both the AVC (Advanced Video Coding) and HEVC coding frameworks. The responses demonstrated that substantial benefit over existing standards could be achieved. As a result, the ISO/IEC MPEG and ITU-T VCEG standardization bodies established the Joint Collaborative Team on 3D Video Coding Extensions Development (JCT-3V) in July 2012, mandated to develop next-generation 3D coding standards with more advanced compression capability and support for synthesis of additional perspective views, covering both AVC and HEVC based extensions. For the HEVC development, a first version of the reference software was contributed by proponents of top-performing responses to the CfP [11]. Based on this software platform, which also included tools for view synthesis

Manuscript received December 15, 2014; revised June 11, 2015; accepted August 14, 2015. Date of current version August 26, 2015.

G. Tech and K. Müller are with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, 10587 Berlin, Germany (e-mail: {gerhard.tech, karsten.mueller}@hhi.fraunhofer.de)

Y. Chen and Y.-K. Wang are with Qualcomm Incorporated, 5775 Morehouse Drive, San Diego, CA 92122, USA (e-mail: {cheny, yekuiw}@qti.qualcomm.com)

J.-R. Ohm is with the Institute of Communication Engineering, RWTH Aachen University, Aachen 52056, Germany (e-mail: ohm@ient.rwth-aachen.de)

A. Vetro is with Mitsubishi Electric Research Labs, Cambridge, MA 02139 (e-mail: avetro@merl.com)

and synthesized view distortion based rate-distortion optimization [12], a range of core experiments were conducted over a period of three years in order to develop all major aspects of the specifications that are described in this paper. As part of this, the JCT-3V has developed two extensions for HEVC, namely the Multiview HEVC (MV-HEVC) [13], which is integrated in the second edition of the standard [14], and 3D-HEVC [15], which was completed in February 2015 and will be part of the third edition.

MV-HEVC comprises only high-level syntax (HLS) additions and can thus be implemented using existing single-layer decoding cores. Higher compression (compared to simulcast) is achieved by exploiting redundancy between different camera views of the same scene. 3D-HEVC aims to compress the video-plus-depth format more efficiently by introducing new compression tools that a) explicitly address the unique characteristics of depth maps, and b) exploit dependencies between multiple views as well as between video texture and depth. Due to this, 3D-HEVC provides further benefits primarily in application scenarios requiring video texture and depth as e.g. more advanced 3D displays.

It is noted that MV-HEVC follows the same design principle as multiview video coding (MVC), the multiview extension of H.264/MPEG-4 AVC [16][17]. Moreover, since MV-HEVC and 3D-HEVC were developed in parallel with the scalable extension of HEVC (SHVC [18]), all extensions share a basic inter-layer prediction design utilizing almost the same HLS. The common design enables a single texture base view to be extracted from MV-HEVC, SHVC and 3D-HEVC bitstreams, which is decodable by a Main profile compliant HEVC decoder. Also, a 3D-HEVC encoder can generate a bitstream that allows the stereoscopic texture views to be decoded by an MV-HEVC decoder. Further aspects of these designs will be explained in the following sections.

The rest of this paper is organized as follows. In the next section, basic concepts of multi-layer coding in HEVC are explained. Section III outlines the specific aspects of the HLS design for MV-HEVC and 3D-HEVC. Section IV describes the new coding tools that are specified in 3D-HEVC. Section V provides definitions of conformance points, i.e., profiles defined for MV-HEVC and 3D-HEVC. Section VI reports the compression performance of the two extensions. Conclusions and outlook are given in Section VII. Note that only the MV-HEVC and 3D-HEVC extension parts of HEVC are discussed in this paper, while a description of the first edition of HEVC [7] can be found in [8].

II. Multi-Layer Coding Design

MV- and 3D-HEVC, as well as SHVC, employ a multi-layer approach where different HEVC-coded representations of video sequences, called layers, are multiplexed into one bitstream and can depend on each other. Dependencies are created by inter-layer prediction to achieve increased compression performance by exploiting similarities among different layers.

In MV- and 3D-HEVC a layer can represent texture, depth or other auxiliary information of a scene related to a particular camera perspective. All layers belonging to the same camera perspective are denoted as a view; whereas layers carrying the same type of information (e.g., texture or depth) are usually called components in the scope of 3D video (and should not be mistaken in the following with the color components composing a picture as defined in HEVC [7]).

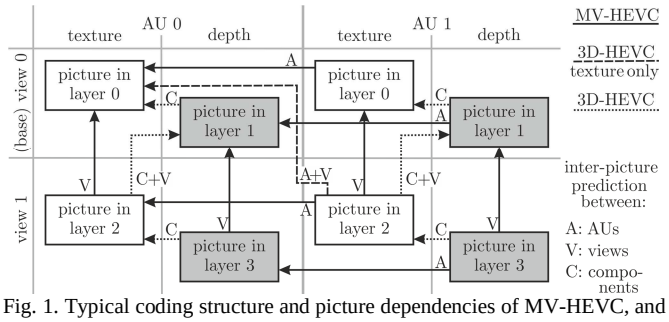


Fig. 1. Typical coding structure and picture dependencies of MV-HEVC, and additional dependencies for 3D-HEVC texture-only coding and 3D-HEVC texture-and-depth coding.

Fig. 1 shows a typical coding structure for pictures, including four layers of two views and two components (texture and depth) for each of the shown two time instances: By design choice, all pictures associated with the same capturing or display time instance are contained in one access unit (AU) and have the same picture order count (POC). The layer of the first picture within an AU is usually denoted as the base layer. Unless the base layer is external (e.g., when using hybrid codec scalability as described in Section III.A.7), it is required to conform to an HEVC single-layer profile, and hence to be the texture component of the base view. The layers of the pictures following the base layer picture in an AU are denoted as enhancement layers or non-base layers, and the views other than the base view are denoted as enhancement views or non-base views. In an AU the order of views is required to be the same for all components. To facilitate combined coding, it is further required in 3D-HEVC that the depth component of a particular view immediately follows its texture component. An overview of dependencies between pictures in different layers and AUs is depicted in Fig. 1 and further discussed below. Note that enhancement-layer random access point pictures are usually coded using inter-layer prediction, thus are not necessarily only intra-picture predicted.

A. MV-HEVC Inter-Layer Prediction

A key benefit of the MV-HEVC architecture is that it does not change the syntax or decoding process required for HEVC single-layer coding below the slice level. This allows re-use of existing implementations without major changes for building MV-HEVC decoders.

Beyond conventional temporal inter-picture prediction (marked (A) in Fig. 1), using pictures of the same view and component however in different AUs, MV-HEVC allows prediction from pictures in the same AU and component but in different views, in the following denoted as inter-view prediction (V). For this, the decoded pictures from other views are inserted into one or both of the reference picture lists of a current picture. As in MVC [16], a concept is followed where temporal reference pictures of the same view and inter-view reference pictures of the same time instance can appear in any

positions of the reference picture list. However, MV-HEVC uses a more flexible reference picture management design, as discussed in Section III.

This way, the motion vectors (MVs) may be actual temporal motion vectors (subsequently denoted as TMVs) when related to temporal reference pictures of the same view, or may be disparity MVs (DMVs) when related to inter-view reference pictures. Existing block-level HEVC motion compensation modules can be used which operate the same way regardless of whether an MV is a TMV or a DMV.

In HEVC single-layer coding, motion information (MV and reference index) for a current prediction block (PB) can be coded in merge mode or using advanced motion vector prediction (AMVP). In both modes a list of candidates is created from the motion information of spatial or temporal neighboring PBs. In this process MVs from neighboring blocks may be temporally scaled, by 1) multiplying the POC difference between the picture of the current PB and its reference picture; and 2) dividing the POC difference between the picture of the neighboring PB and its reference picture.

In MV-HEVC, pictures in the same AU are coded with the same POC, therefore the POC of an inter-view reference picture is the same as the current picture and the MV scaling might fail due to a division or multiplication by zero. To avoid this, inter-view reference pictures in MV-HEVC are always marked as long-term reference pictures. This way, a mechanism of the block-level design in HEVC version 1 (referred to as motion hook) can be used, which disables scaling of MVs associated with long-term reference pictures [19]. Moreover, since DMVs are related to spatial distance and TMVs to temporal distance, prediction between DMVs and TMVs is inefficient. Therefore, an MV of a neighboring PB can be a candidate for the current PB in HEVC single-layer coding only when the related reference pictures of the current and the neighboring PB are both marked as long-term reference pictures or are both marked as short-term reference pictures. The motion hook as described above allows MV-HEVC to efficiently reuse motion prediction in HEVC single-layer coding with the HLS-changes-only design.

B. 3D-HEVC Inter-Layer Prediction

For increased compression performance, 3D-HEVC extends MV-HEVC by allowing new types of inter-layer prediction. As indicated in Fig. 1, the new prediction types are: a) combined temporal and inter-view prediction (A+V), referring to a picture in the same component but in a different AU and a different view; b) inter-component prediction (C), referring to pictures in the same AU and view but in a different component; and c) combined inter-component and inter-view prediction (C+V), referring to pictures in the same AU but in a different view and component. A further design change compared to MV-HEVC is that, besides sample and motion information, residual, disparity and partitioning information can also be predicted or inferred. A detailed overview of dependencies and reference and predicted information of 3D-HEVC coding techniques is provided in Table I.

C. Limitations of Inter-Layer Prediction

Since inter-view prediction in both MV- and 3D-HEVC is achieved through block-based disparity compensation (in contrast to full epipolar geometric transformations), the coding tools described in this paper are most efficient when the view signals are aligned in a one-dimension linear and coplanar arrangement. This can be achieved through camera setup or pre-processing of the sequences through a rectification process. While the standard does not impose any limitations regarding the arrangement of multiview video sequences, the coding efficiency can be expected to decrease when there is significant misalignment, similar as with MVC [20].

A second assumption is that texture and depth pictures in the same AU and view are spatially aligned (or have been appropriately registered), such that samples at equal positions represent the same point of the depicted scene. If they are not aligned, the effectiveness of coding tools with a dependency between texture and depth components decreases.

III. MV- AND 3D-HEVC HIGH-LEVEL SYNTAX

High-level syntax (HLS) is an integral part of a video codec. An important part of it is the network abstraction layer (NAL), providing a (generic) interface of a video codec to (various) networks/systems. HEVC (single-layer coding) HLS was designed with significant consideration of extensibility mechanisms. These are also referred to as hooks, which basically allow future extensions to be backward compatible to earlier versions of the standard. Important HLS hooks in HEVC include: a) Inclusion of layer identifier (ID) in the NAL unit header, whereby the same NAL unit header syntax applies to both HEVC single-layer coding and its multi-layer extensions; b) Introduction of the video parameter set (VPS), which was introduced mainly for use with multi-layer extensions, as VPS contains cross-layer information; c) Introduction of the layer set concept and the associated signaling of multi-layer hypothetical reference decoder (HRD) parameters; d) Addition of extensibility for all types of parameter sets and slice header, which allows the same syntax structures to be used for both the base layer and enhancement layers without defining new NAL unit types and to be further extended in the future when needed.

A common HLS framework has been jointly developed for MV-HEVC (which is largely applicable to 3D-HEVC as well) and SHVC. This section focuses on the new HLS features developed for the three multi-layer HEVC extensions compared to HEVC single layer coding HLS, for which an overview can be found in [21]. After the discussion of the common HLS, deviations of the 3D-HEVC HLS from MV-HEVC HLS are highlighted.

A. Common HLS for Layered HEVC Extensions

1) Parameter Set and Slice Segment Header Extensions: The VPS has been extended by adding the VPS extension structure to the end, which mainly includes information on: a) Scalability type and mapping of NAL unit header layer ID to scalability IDs; b) Layer dependency, dependency type, and independent layers; c) Layer sets and output layer sets; d) Sub-

layers and inter-layer dependency of sub-layers; e) Profile, tier, and level (PTL); f) Representation format (resolution, bit depth, color format, etc.); g) Decoded picture buffer (DPB) size; h) Cross-layer video usability information (VUI), which includes information on cross-layer picture type alignment, cross-layer intra random access point (IRAP) picture alignment, bit rate and picture rate of layer sets, video signal format (color primaries, transfer characteristics, etc.), usage of tiles and wavefronts and other enabled parallel processing capabilities, and additional HRD parameters.

It should be noted that the VPS applies to all layers, while in the AU decoding order dimension it applies from the first AU where it is activated up to the AU when it is deactivated. Different layers (including the base layer and a non-base layer) may either share the same sequence parameter set (SPS) or use different SPSs. Pictures of different layers or AUs can also share the same picture parameter set (PPS) or use different PPSs. To enable sharing of SPS and PPS, all SPSs share the same value space of their SPS IDs, regardless of the layer ID values in their NAL unit headers; the same is true for PPSs.

Among other smaller extensions, the slice segment header has been extended in a backward compatible manner by adding the following information: a) The discardable flag that indicates whether the picture is used for at least one of temporal inter-picture prediction and inter-layer prediction or neither (when neither applies the picture can be discarded without affecting the decoding of any other pictures in the same layer or other layers); b) A flag that indicates, whether an instantaneous decoder refresh (IDR) picture is a bitstream splicing point (if yes, then pictures from earlier AUs would be unavailable as references for pictures of any layer starting from the current AU); c) Information on lower-layer pictures used by the current picture for inter-layer prediction; and d) POC resetting and POC most significant bits (MSB) information. The latter two sets of information are used as the basis for derivation of the inter-layer reference picture set (RPS) and for guaranteeing cross-layer POC alignment, both of which are discussed later.

2) Layer and Scalability Identification: Each layer is associated with a unique layer ID, which must be increasing across pictures of different layers in decoding order within an AU. In addition, a layer is associated with scalability IDs specifying its content, which are derived from the VPS extension and denoted as view order index (VOI) and auxiliary ID.

All layers of a view have the same VOI. The VOI is required to be increasing in decoding order of views. Furthermore, a view ID value is signaled for each VOI, which can be chosen without constraints, but should indicate the view's camera position (e.g., in a linear setup).

The auxiliary ID signals whether a layer is an auxiliary picture layer carrying depth, alpha or other user defined auxiliary data. By design choice, auxiliary picture layers have no normative impact on the decoding of non-auxiliary picture layers (denoted as primary picture layers).

3) Layer Sets: The concept of layer sets was already introduced in HEVC version 1. A layer set is a set of independently decodable layers that conventionally contains the base layer. Layer sets are signaled in the base part of the VPS. During the development of the common multi-layer HLS, two related concepts, namely output layer sets (OLSs) and additional layer sets, were further introduced. An OLS is a layer set or an additional layer set for which the target output layers are specified. Non-target-output layers are for example those layers that are used only for inter-layer prediction but not for output/display. An OLS can have two layers for output (e.g., stereoscopic viewing) but contain three layers. An HEVC single-layer decoder would only process one target output layer, i.e., the base layer, regardless of how many layers the layer set contains. This is the reason why the concept of OLSs was not needed in HEVC version 1.

An additional layer set is a set of independently decodable layers that does not contain the base layer. For example, if a bitstream contains two simulcast (i.e., independently coded) layers, then the non-base layer itself can be included in an additional layer set. This concept can also be used for signaling the PTL for auxiliary picture layers, which are usually coded independently from the primary picture layers. For example, a depth or alpha (i.e., transparency) auxiliary picture layer can be included in an additional layer set and indicated to conform to the Monochrome (8 bit) profile, regardless of which single-layer profile the base (primary picture) layer conforms to. Without such a design, many more profiles would need to be defined to handle all combinations of auxiliary picture layers with single-layer profiles. To realize the benefits of this design, a hypothetical independent non-base layer rewriting process was specified, which "transcodes" independent non-base layers to a bitstream that conforms to a single-layer profile.

By design choice, an additional layer set is allowed to contain more than one layer, e.g., three layers with layer ID values equal to 3, 4, and 5, where the layer with layer ID equal to 3 is an independent non-base layer. Along with this, a bitstream extraction process for additional layer sets was specified. While the extracted sub-bitstream does not contain a base layer, it is still a conforming bitstream, i.e., the multi-layer extensions of HEVC allow for a conforming multi-layer bitstream to not contain the base layer, and compliant decoding of the bitstream may not involve the base layer at all.

4) Profile, Tier, and Level (PTL): Compared to earlier multi-layer video coding standards, a fundamentally different approach was taken for MV-HEVC and SHVC for the specification and signaling of interoperability points (i.e., PTL in the context of HEVC and its extensions). Rather than specifying PTL for an operation point that contains a set of layers, PTL is specified and signaled in a layer-specific manner in MV-HEVC and SHVC. Consequently, a decoder that is able to decode two-layer bitstreams with 1080p@30fps at the base layer and 1080p@60fps at the enhancement layer should express its capability as a list of two PTLs equivalent to {Main profile Main tier Level 4, Multiview Main profile Main tier Level 4.1}. A key advantage of this design is that it facilitates easy decoding of multiple layers by reusing single-layer decoders. If PTL was specified for the two layers together, then the decoder would need to be able to decode the two-layer

bitstreams with both the base and enhancement layers of 1080p@60fps, causing overprovisioning of resources.

5) RPS and Reference Picture List Construction: In addition to the five RPS lists (RefPicSetStCurrBefore, RefPicSetStCurrAfter, RefPicSetStFoll, RefPicSetLtCurr, and RefPicSetLtFoll) defined in HEVC version 1, two more RPS lists, RefPicSetInterLayer0 and RefPicSetInterLayer1 (denoted as RpsIL0 and RpsIL1, respectively), were introduced to contain inter-layer reference pictures. Given a current picture, those inter-layer reference pictures are included into two sets depending on whether they have view ID values greater or smaller than the current picture. If the base view has a greater view ID than the current picture, then those with greater view IDs are included into RpsIL0 and those with smaller view IDs into RpsIL1, and vice versa. The derivation of RpsIL0 and RpsIL1 is based on VPS extension signaling (of layer dependency and inter-layer dependency of sub-layers) as well as slice header signaling (of lower-layer pictures used by the current picture for inter-layer prediction).

When constructing the initial reference picture list 0 (i.e., RefPicListTemp0), pictures in RpsIL0 are inserted immediately after pictures in RefPicSetStCurrBefore, and pictures in RpsIL1 are inserted last, after pictures in RefPicSetLtCurr. When constructing the initial reference picture list 1 (i.e., RefPicListTemp1), pictures in RpsIL1 are inserted immediately after pictures in RefPicSetStCurrAfter, and pictures in RpsIL0 are inserted last, after pictures in RefPicSetLtCurr. Otherwise the reference picture list construction process stays the same as for HEVC single-layer coding.

6) Random Access, Layer Switching, and Bitstream Splicing: Compared to AVC, HEVC provides more flexible and convenient random access and splicing operations, by allowing conforming bitstreams to start with a clean random access (CRA) or broken link access (BLA) picture. In addition, MV-HEVC and SHVC support 1) non-cross-layer aligned IRAP pictures, i.e., it is allowed in an AU to have IRAP pictures at some layers and non-IRAP pictures at other layers, and 2) a conforming bitstream can start with any type of IRAP AU, including an IRAP AU where the base layer picture is an IRAP picture while (some of) the enhancement layer pictures are non-IRAP pictures. The latter allows easy splicing of multi-layer bitstreams at any type of IRAP AU and random accessing from such AUs. Non-cross-layer aligned IRAP pictures also allow for flexible layer switching.

To support non-cross-layer aligned IRAP pictures, the multi-layer POC design needs to ensure that all pictures in an AU have the same POC value. The design principle is referred to as cross-layer POC alignment and is required to enable a correct in-layer RPS derivation and a correct output order of pictures of target output layers.

The multi-layer HEVC design allows extremely flexible layering structures. Basically, a picture of any layer may be absent at any AU. For example, the highest layer ID value can vary from AU to AU, which was disallowed in SVC and MVC. Such flexibilities imposed a great challenge on the multi-layer POC design. In addition, although a bitstream after layer or sub-layer switching is not required to be conforming, the design should still enable a conforming decoding behavior to work with layer and sub-layer switching, including cascaded switching behavior. This is achieved by a POC resetting approach.

The basic idea of POC resetting is to reset the POC value when decoding a non-IRAP picture (as determined by the POC derivation process in HEVC version 1), such that the final POC values of pictures of all layers of the AU are identical. In addition, to ensure that POC values of pictures in earlier AUs are also cross-layer aligned and that POC delta values of pictures within each layer remain proportional to the associated presentation time delta values, POC values of pictures in earlier AUs are reduced by a specified amount [22].

To work with all possible layering structures as well as some picture loss situations, the POC resetting period is specified based on a POC resetting period ID that is optionally signaled in the slice header [23]. Each non-IRAP picture that belongs to an AU that contains at least one IRAP picture must be the start of a POC resetting period in the layer containing the non-IRAP picture. In that AU, each picture would be the start of a POC resetting period in the layer containing the picture. POC resetting and the decreasing of POC values of same-layer pictures are applied only for the first picture within each POC resetting period, such that these operations would not be performed more than necessary; otherwise POC values would be corrupted.

7) Hybrid Codec Scalability and Multiview Support: The HEVC multi-layer extensions support the base layer being coded by other codecs, e.g., AVC. A simple approach was taken for this functionality by specifying the necessary elements of a conceptual interface by which the base layer may be provided by the system environment in some manner that is not specified within the HEVC standard. Basically, except for information on the representation format and whether the base layer is a target output layer as signaled in the VPS extension, no other information about the base layer is included in the bitstream (as input to the enhancement-layer decoder).

8) Hypothetical Reference Decoder (HRD): The main new developments of the HRD compared to HEVC version 1 include the following three aspects relevant for MV- and 3D-HEVC. Firstly, the bitstream conformance tests specified for HEVC version 1 are classified into two sets and a third set is additionally specified. The first set of tests is for testing the conformance of the entire bitstream and its temporal subsets. The second set of bitstream conformance tests is for testing the conformance of the layer sets specified by the active VPS and their temporal subsets. For the first and second sets of tests, only the base layer pictures are decoded and other pictures are ignored by the decoder. The third set of tests is for testing the conformance of the OLSs specified in the VPS extension and their temporal subsets.

The second aspect is the introduction of bitstream partition (BP) specific coded picture buffer (CPB) operations, wherein each BP contains one or more layers, and CPB parameters for each BP can be signaled and applied. These parameters can be utilized by transport systems that transmit different sets of layers in different physical or logical channels; one extreme

TABLE I
OVERVIEW OF 3D-HEVC TEXTURE AND DEPTH CODING TOOLS, DEPENDENCIES, AND REFERENCE AND PREDICTED INFORMATION

	Technique	Abbr.	Description	Dependency	Ref.	Pred.	Sec.
Texture	Neighboring block disparity vector	NBDV	Derives predicted disparity information (PDI) for a CU	V, A, I	M, D	D	IV-A1
	Extended TMVP[1] for merge	-	Extends TMVP to also operate for inter-view prediction	V, A	M	M	IV-C1
	Inter-view motion prediction[2, 3]	IV, IV_S	Uses PDI for inter-view prediction of merge candidates	V	M	M	IV-C4
	Disparity information merge cand.[3]	DI, DI_S	Uses PDI directly as a merge candidate	-	-	M	IV-C2
	Residual prediction[3]	RP	Predicts residual from a different view or AU	V, A, A+V	S, M[4]	S, R	IV-D1
	Illumination compensation	IC	Adapts an inter-view sample prediction to the current view	V, I	S	S	IV-D2
	Depth refinement[3]	DR	Improves PDI using disparity from a depth map	C+V	S	D	IV-A2
	View synthesis prediction[2, 3]	VSP	Derives merge candidates from samples of a depth block	C+V, I	S	M	IV-C6
	Depth based block partitioning[3]	DBBP	Subdivides an inter-pic. predicted CB based on a depth block	C+V, V, A	S	S	IV-D3
Depth	Inter-view motion prediction	IV	Uses a default disparity value to predict a merge candidate	V	M	M	IV-C4
	Full sample motion accuracy	-	Reduces ringing artifacts and complexity	-	-	-	IV-D4
	Intra_Wedge mode	-	Subdivides a PB by a straight line and predicts DCs	I	S	S	IV-E1
	Intra-pic. skip / Intra_Single mode	-	Signals an angular mode or a single value for the entire CB	I	S	S	IV-E2
	DC offsets / DC-only mode	-	Codes residual DC explicitly / skips transform coefficients	-	-	-	IV-F1
	Depth look-up table	DLT	Quantizes the residual DC non-linearly	-	-	-	IV-F2
	Quadtree limitation	QTL	Derives a depth CB partitioning from texture	C	P	P	IV-B
	Texture merge candidate[2]	T	Derives a merge candidate from texture	C	M	M	IV-C5
	Intra_Contour mode	-	Predicts a PB subdivision from texture and predicts DCs	C, I	S	S	IV-E1

Dependency from picture in different: V) View; C) Component A) AU; A+V) AU and View; C+V) Component and View; I) Intra-picture.
Reference and predicted information: M) Motion; D) PDI; S) Sample; R) Residual; P) Partitioning syntax.
[1]Temporal motion vector prediction. [2]Using sub-block partition (SBP) motion accuracy (IV-C3). [3]Depends on NBDV. [4]For RP from a different AU only.

example is one channel for each layer. The layer specific CPB parameters are also a basis for defining the semantics of layer specific PTL. The third aspect is the layer specific DPB management operations, where each layer exclusively uses its own sub-DPB. To ensure the design works with (cascaded) layer switching behavior, sharing of a particular memory unit across layers is disallowed.

9) SEI Messages: SEI messages in HEVC version 1 have been adapted to be applicable in the multi-layer contexts, in a backward compatible fashion, some of them with significant semantics changes. In addition, some new SEI messages are specified that apply to all multi-layer HEVC extensions

Furthermore, the following new SEI messages are specified for MV-HEVC and 3D-HEVC: the 3D reference displays information SEI message, the depth representation information SEI message, the multiview scene information SEI message, the multiview acquisition information SEI message, and the multiview view position SEI message. The latter three correspond to the SEI messages of the same name in MVC.

B. 3D-HEVC Specific HLS

The MV-HEVC HLS provides generic support for multi-layer extensions, therefore only a few additional HLS features have been introduced in 3D-HEVC to support the signaling of depth layers, additional reference layers, tool parameters and a new SEI message, as described in the following.

In MV-HEVC, the auxiliary ID can be used to signal that a layer is carrying depth. In 3D-HEVC a new scalability ID element called depth flag has been introduced. In contrast to layers indicating depth by the auxiliary ID, layers enabling the depth flag can use the new 3D-HEVC coding tools.

Reference layers additionally required for new inter-layer prediction methods are signaled in the VPS as in MV-HEVC. However, when a reference picture list is constructed, only pictures from the current component are included, such that inter-component sample prediction is avoided.

Enabling flags for several of the tools shown in Table I are signaled in an additional SPS extension. Moreover, camera parameters can be present in a VPS extension (when constant) or the slice header (when varying over time). Camera parameters allow the conversion of values of a depth picture to disparities by scaling and offsetting and are required by VSP (Section IV.C.6) and depth refinement (Section IV.A.2). A depth look-up table (DLT), utilized as described in Section IV.F.2, can be signaled in a PPS extension.

Lastly, the alternative depth information SEI message provides information required for alternative rendering techniques, based on global depth maps or warping.

IV. 3D-HEVC VIDEO CODING TECHNIQUES

An overview of the 3D-HEVC texture and depth coding tools is provided in Table I. Texture coding tools provide increased compression performance by applying new inter-view prediction techniques, or enhancing existing ones. Some of the texture coding tools derive disparity for inter-view prediction, or segmentation information from samples of an already decoded depth layer. These depth-dependent techniques can be disabled when texture-only coding is performed.

Improved coding of depth maps has also been introduced into 3D-HEVC. Since depth maps typically contain homogeneous areas separated by sharp edges, new intra-picture prediction and residual coding methods have been specified to account for these unique signal characteristics. Additionally, new depth coding tools that allow for inter-view prediction of motion or the prediction of motion and partitioning information from texture layers have also been specified.

Some of the new prediction techniques allow prediction with higher accuracy by introducing sub-block partitioning (SBP), which in some cases can also sub-divide a prediction block into two parts with a non-rectangular shape.

In the remainder of this section the 3D-HEVC decoding processes as listed in Table I are discussed in detail. A new module, which forms the basis for several 3D-HEVC tools, is disparity derivation (IV.A). Further techniques modify or extend existing HEVC single-layer coding processes for block

partitioning (IV.B), motion prediction (IV.C), inter-picture sample prediction (IV.D), intra-picture sample prediction (IV.E), and residual coding (IV.F). By design choice, several core elements of HEVC such as entropy coding, deblocking, sample adaptive offset (SAO), coding of quantized transform coefficients, the transform tree (except conditions for its presence), and AMVP have not been modified.

A. Disparity Derivation

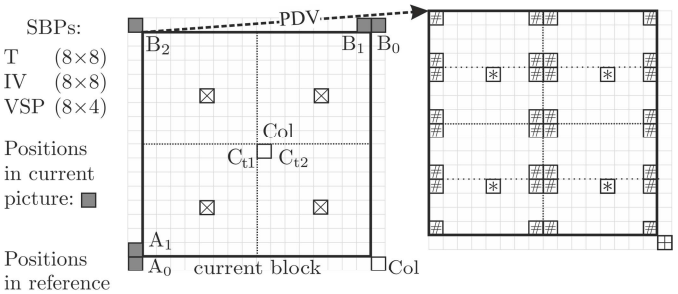
The majority of 3D-HEVC coding techniques are based on inter-view prediction, wherein sample values, prediction residuals, sub-partitioning, or motion information of a block in a picture of the current view are predicted from a reference block in a picture of a different view. To find a reference block, the disparity derivation process is invoked at the coding unit (CU) level to provide a VOI of a reference view (RV), to be used for inter-view prediction, and a predicted disparity vector (PDV). The PDV indicates the spatial displacement of the reference block in the RV relative to the position of the coding block (CB) in the current picture. In the following, the reference view order index (RVOI) and the associated PDV are referred to as predicted disparity information (PDI). The PDI for texture layers is derived as described in the following sections. For simplicity, the PDI for depth layers is constant for a slice and correspond to an available reference view and a disparity vector derived from a depth value of 128 (for 8 bit sample precision).

1) Neighboring Block Disparity Vector (NBDV): NBDV operates without referring to depth layers to allow the prediction of PDI for applications in which only the texture information is of interest (referred to as texture-only coding). As such, the PDI is derived from motion information of temporally and spatially neighboring blocks [24].

The temporally neighboring blocks are located in two different pictures. The first picture is the collocated reference picture, which is the picture signaled for temporal motion vector prediction (TMVP) [8]. The second picture is chosen among the temporal reference pictures, where the one which may have blocks more likely to be coded using DMVs is selected [22]. The two temporally neighboring blocks cover the positions C_{t1} and C_{t2} that correspond to the center position of the current CB in both pictures, respectively, as depicted in Fig. 2. The spatial neighboring blocks in the current picture cover the positions A_1 and B_1, adjacent to the bottom-left and the top-right sample, respectively, of the current PB. Blocks at A_1, B_1, and C_{t1} are required to be accessed in merge mode, AMVP, and TMVP for motion information. Thus, in comparison to HEVC single layer coding, additional memory accesses are only required when referring to the block at C_{t2}.

To determine the PDI for the current CU, blocks at the neighboring positions are searched in order $\{C_{t1}, C_{t2}, A_1, B_1\}$. If a neighboring block is associated with a DMV related to an available inter-view reference picture, the search from other neighbors is terminated by setting the PDV to the DMV and the RVOI to the VOI of that reference picture.

When the PDI cannot be derived from motion information of neighboring blocks as described above, a second search is applied, in which the stored PDI of the neighboring CUs from



in	Col C_{t1}		C_{t2}	T	IV	IV_S	VSP
AU	other	cur.	other	cur.	cur.	cur.	cur.
View	cur.	other	cur.	cur.	RV	RV	RV
Comp.	cur.	cur.	cur.	text.	cur.	cur.	depth

Fig. 2. Positions accessed in NBDV or merge mode for e.g. a 16×16 block.

positions A_1 and B_1 (in that order) are considered. In particular, the condition for inheritance is that the neighboring CU is coded in skip mode and using an inter-view predicted MV, derived as described in Section IV.C.4. The PDI of the first neighboring CU that fulfills the condition is chosen.

If derivation from neighboring blocks or CUs fails, the VOI of an inter-view reference picture of the current picture is used as RVOI and the PDV is set equal to the zero vector.

2) Depth Refinement: Although NBDV can operate without referring to depth layers, the accuracy of the PDI can be increased by exploiting the additional depth information. Since the texture of a view is coded first in 3D-HEVC, the depth map of the current view is not available when coding texture. For this, the PDI derived by NBDV is used to identify the block that corresponds to a current CB in the already decoded depth map of the RV. The PDV of the CU is then replaced by a refined disparity vector, derived from the maximum of the four corner sample values of the depth block [25].

B. Partitioning Syntax Prediction

A new tool in 3D-HEVC, called *quadtree limitation* (QTL), predicts the partitioning of a depth CU from syntax elements of a collocated texture CU. By design choice, QTL is not available in I slices and IRAP pictures.

When a region of a texture picture includes only low frequencies, such that a coarse split in CBs is applied by an encoder, it can be assumed that high-frequency signal parts in the collocated region of the associated depth picture are either also not present or irrelevant for view synthesis. Therefore, when a depth block has the same position and size as a CB in the corresponding texture picture, the flag indicating a further split is not present in the depth coding quadtree and the depth block cannot be split into smaller CBs. Though this could have been implemented as an encoder only restriction, the additional bit rate saving was considered beneficial.

For the same reasons, the PB partitioning of a depth CB having limited size is restricted by the texture CB. When the texture CB consists of a single PB, the same applies for the depth CB without additional signaling. When the texture CB is split horizontally or vertically, only splits in the same direction can be signaled. In the case that the texture CB is associated with four PBs, no restrictions apply for the depth CB.

TABLE II
OVERVIEW OF THE EXTENDED MERGE CANDIDATE LIST

	Candidate	T	IV	A_1	B_1	VSP	B_0	DI	A_0	B_2	IV_S	DI_S	*rem.*
	Order	0	1	2	3	4	5	6	7	8	9	10	11..
Availability	Texture	−	s^1	x^2 IV	x IV	s^3	x	x A_1B_1	x	x	x^4 IV	x^5	x
	Depth	s	x T	x T	x T	−	x	−	x	x	−	−	x

−: Candidate is not available; *s*, *x*: Candidate may be available; *s*: SBP motion information; To avoid redundancies, a candidate is not included when a candidate listed below *x* has the same motion information (when SBPs are used, motion information of the default SBP is used for comparison).

[1]*x* when DBBP is used, − when IC is used; [2]When the PU at position A_1 uses VSP and A_1 is selected, VSP is also applied for the current PU; [3]− when DBBP, RP, or IC are used or when A_1 uses VSP; [4]− when IC is used; [5]− when DI is not available or IV_S is available.

C. Motion Prediction

3D-HEVC specifies an extended candidate list for the merge mode. The list includes the conventional HEVC single-layer coding candidates (although the temporal MV candidate is predicted by a modified process), and several new candidates. Some of the new candidates are based on inter-layer prediction of motion information in SBP granularity, such that the granularity in the reference layer can be taken into account.

The derivation of merge candidates is performed in two separate steps. In the first step, an initial merge candidate list is derived as specified for HEVC single-layer coding, including the removal of redundant entries, but using the modified TMVP derivation as described below. In the second step, an extended merge candidate list is constructed from the initial list and additional candidates. To limit worst-case complexity, the second step is not applied for PUs with luma PB sizes 8×4 or 4×8. The candidates, their order in the extended list, and conditions for inclusion into the list are provided in Table II and discussed in the following. Fig. 2 shows positions, which are accessed for their derivation.

Candidates from positions A_1, B_1, B_0, A_0 and B_2 are only available when they are included in the initial candidate list. They are interleaved with the additional 3D-HEVC specific candidates, in the order indicated in Table II. Remaining candidates of the initial list (collocated, combined, zero) are inserted at the end of the extended list, if the maximum number of merge candidates, which is increased by one compared to HEVC single-layer coding, is not exceeded.

Additional candidates are the texture candidate (T), the inter-view (IV) and shifted inter-view (IV_S) candidates, the view synthesis prediction (VSP) candidate, and the disparity information (DI) and shifted disparity information (DI_S) candidates.

1) Extended Temporal Motion Vector Prediction (TMVP) for Merge: In MV-HEVC (and single-layer HEVC) the reference index is always zero for a TMVP merge candidate (also denoted as collocated (Col) candidate). Therefore, the reference indices of the collocated block and the TMVP candidate may indicate reference pictures of different types (one temporal, the other inter-view), such that the TMVP candidate is not available. However, when this case occurs in 3D-HEVC, the reference picture index of the TMVP candidate is changed to an alternative value, which indicates an available reference picture having the same type as the reference picture of the collocated block [19]. Hence, the same type of prediction (either temporal or inter-view) is indicated by both the candidate and the collocated block, such that the MV of the TMVP candidate can be predicted from the MV of the collocated block. For MV prediction, scaling might be applied. If the collocated block refers to a temporal reference picture, the MV is scaled based on POC values, as described in Section II.A. Otherwise, when the collocated block uses inter-view prediction, scaling based on the view ID values, which correspond to spatial camera positions, is performed.

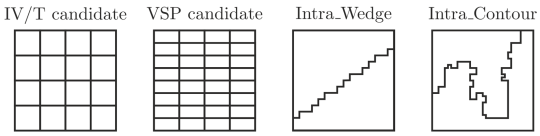


Fig. 3. Examples of possible sub-block partitions (SBPs) of a 32×32 PB.

2) Disparity Information Candidates: The PDI derived for the current CU can be used to identify a reference block for inter-view sample prediction. When a reference picture list of the current picture includes a picture from the RV, the DI candidate is available and its motion information related to this list is given by the PDV (with vertical component set to zero) and the reference index to the picture [26].

The DI_S candidate is derived from the DI candidate by adding an offset of 1 in units of luma samples to the horizontal MV components of the DI candidate. The primary motivation for the DI_S candidate is that the PDV (predicted as described in Section IV.A) used to derive the DI candidate may not always match the actual disparity, such that offering an additional choice can improve performance.

3) Sub-Block Motion Prediction: Conventionally, a merge candidate (e.g., an initial list candidate) consists of a single set of motion information (up to two MVs and their reference picture indices, and a reference picture list indication), which might be used for inter-picture prediction of the entire current PB. In 3D-HEVC, a PB can be further subdivided into rectangular SBPs (e.g., as depicted in Fig. 3), when the T, the IV or the VSP candidate is selected. For this purpose, these candidates consist of multiple sets of motion information each for inter-picture prediction of one SBP of the current PB. Thus, motion information can be inherited with a finer granularity.

3D-HEVC supports the smallest bi-predictive SBP size of 8×8 in luma samples and the smallest uni-predictive SBP size of either 4×8 or 8×4 in luma samples, hence the same as in single layer HEVC. The division of a PB depends on the candidate and component the PB belongs to. For PUs in depth layers selecting the T candidate or PUs in texture layers selecting the IV merge candidate, the luma PB is partitioned in luma SBPs of size *S*×*S*, where *S* is signaled in the SPS and can be equal to 8, 16, 32 or 64. When a partitioning in SBPs of size *S*×*S* is not possible without remainder, the luma PB is only associated with a single luma SBP having the luma PB size. To reduce complexity by disabling SBP motion accuracy, the same applies for PUs in depth layers selecting the IV candidate. When the VSP candidate is selected, all associated luma SBPs have either the size 8×4 or the size 4×8, as described in Section IV.C.6. In cases of all candidates, chroma PBs are divided according to the luma PBs (i.e. half size horizontally

and vertically for 4:2:0 chroma sampling).

4) Inter-View Candidates: Prediction of MVs from other views has been proposed, e.g., in [27]. Based on this idea, the IV candidate inherits motion information from a picture included in the same AU and the RV. An example for a 16×16 PB split in four SBPs is shown in Fig. 2. The derivation for a current SBP of the IV candidate operates as follows [28]:

First, the position of the corresponding block in the picture of the RV is identified by adding the PDV to the center position of the current SBP. When the corresponding block is coded using inter-picture prediction, its MVs might be inherited for the current SBP. The condition for inheritance is that a reference picture list of the current picture includes a picture with the same POC as the reference picture of the corresponding block. When such a picture is available, the motion information for the current SBP is set equal to the MV of the corresponding block and the index of the picture in the reference picture list of the current picture. Since the condition ensures that the temporal distance between pictures containing: a) the corresponding block and its reference block; and b) the current SBP and its reference block; are equal, POC based MV scaling is not necessary. When the corresponding MV cannot be inherited for the current SBP, it inherits the motion information of the default SBP, which is the SBP whose top-left corner sample is closest to the center of the current PB. In the case that motion information is also not available for the default SBP, the IV candidate is not available.

In some cases, the PDV can be inaccurate, such that the prediction error of the IV candidate is high. Here, the IV_S candidate can be an alternative, as it is based on another disparity assumption: To derive the IV_S candidate the same method as for the IV candidate is applied, but with a single SBP (equal to the PB), and additional horizontal and vertical offsets (equal to the half width and half height of the PB, respectively) added to the PDV.

5) Texture Candidate: The derivation process of the T candidate is similar to that of the IV candidate. However, whereas the IV candidate is inherited from blocks of the same component in the RV, the T candidate is derived from the texture component in the same view. Hence, instead of a corresponding block, a collocated block is identified at the center position (marked with "×" in Fig. 2) of the current SBP [29]. Moreover, the condition for inheritance is that the reference picture list of the current picture includes a picture with the same POC and view ID as the reference picture of the corresponding block, such that POC or view ID based scaling can be avoided.

6) View Synthesis Prediction: In view synthesis, a picture is conventionally rendered by shifting samples of a texture picture by disparities obtained from a depth map. For 3D-HEVC sample prediction, the same principle can be applied in coarser granularity by disparity compensation of a reference block in a texture picture. More specifically, when the VSP candidate is selected, inter-view sample prediction is performed for the SBPs of the current PB using MVs obtained from its corresponding depth block [30]. As for disparity refinement, the corresponding depth block is identified in the depth picture of the RV using the PDV derived by NBDV.

The VSP candidate can be chosen by signaling its index in the merge candidate list, or, in case that the PB at A_1 uses VSP, by signaling the index of A_1. When the VSP candidate is selected, the current PB is first divided, such that all its SBPs have the same size of either 8×4 or 4×8 in luma samples, as follows: If only one partitioning pattern is possible (e.g., 4×8 for a 4×16 PB), it is chosen. If multiple partitioning patterns are possible, the partitioning is determined according to the gradient direction in the corresponding depth block as estimated from its four corner samples. This is done to decrease prediction granularity in the gradient direction. As such, the 8×4 pattern is chosen when the gradient is higher in the vertical direction, while the 4×8 pattern is used when it is higher in the horizontal direction.

After partitioning, the motion information of the SBPs is derived. For complexity reduction, all SBPs use uni-prediction from the texture picture of the RV. Moreover, the vertical component of MVs is always set equal to zero. The horizontal component is derived for each SBP by converting the maximum value of the four corner samples (marked with " # " in Fig. 2) of its corresponding depth SBP within the corresponding depth block.

D.Inter-Picture Sample Prediction

3D-HEVC extends inter-picture sample prediction in texture layers by three techniques: Residual prediction exploits the correlation of sample prediction errors in different views or AUs. Adaptive weighting of an inter-view sample prediction has been enabled by illumination compensation. Depth-based block partitioning (DBBP) combines two predictions for a texture PB according to a sub-partitioning derived from a corresponding depth CB. For depth coding, motion compensation has been simplified.

1) Residual Prediction (RP): In texture layers, the energy of the residue of the current PB may be reduced by performing additional motion compensation either in the RV or a different AU. The concept of RP is to reuse the MVs of the current PB to predict the residual signal. The predicted residual is calculated and added on top of the motion compensated signal derived with the MVs of the current PB for each used prediction direction. To accommodate possible quantization differences, a CU level weighting factor *w* being equal to 1 or 1/2 can be chosen by the encoder to be applied to the residual signal.

Since the additional motion compensation to derive the residual signal would require a significant increase of memory bandwidth and calculations, the following design trade-offs were made: 1) Only CUs associated with a single PU may have RP enabled, thus the smallest luma PB size with RP enabled is 8×8; 2) RP is not used for chroma components of PBs of luma size 8×8. 3) bi-linear interpolation is used for the motion compensation in RP enabled CUs to calculate the residual signal as well as the motion compensation between the current PB and the reference block identified by its MV.

Examples of residual prediction are shown in Fig. 4, where an MV (denoted as ***m***) of the current PB (*PB*) provides the displacement of a reference block (*RB*) in its reference picture, similar to conventional motion compensation. Depending on

whether ***m*** is a TMV or a DMV, *RB* is either a temporal reference block, or an inter-view reference block. In addition, ***m*** provides the displacement between two more blocks, corresponding to *PB* and *RB*, either in a different view or a different AU. The corresponding blocks are denoted as *PB′* and *RB′*. The residual prediction is calculated as the difference between the samples of *PB′* and *RB′*. The type of ***m*** (being a TMV or DMV) determines where the residual signal is predicted from. Consequently, RP can be categorized as inter-view residual prediction [31] or temporal residual prediction [32].

a) *Inter-View Residual Prediction:* As illustrated in Fig. 4 (a), ***m*** is a TMV and identifies a temporal reference block *RB*. To identify *PB′* and *RB′* a disparity vector ***d*** is needed and set equal to the PDV (without Disparity Refinement). *PB′* and *RB′* are located both in the RV but in different pictures. *PB′* is an inter-view reference block of *PB* with a displacement of ***d***, while *RB′* is a temporal inter-view reference block with a displacement being the sum of ***m*** and ***d***. The inter-view predicted residual signal is calculated as the difference between *RB′* and *PB′*. Afterwards, the combined prediction signal of *PB* is computed by adding the weighted inter-view predicted residual to the motion compensated signal of *RB*.

b) *Temporal Residual Prediction:* As illustrated in Fig. 4 (b), ***m*** is a DMV and identifies an inter-view reference block *RB* in an inter-view reference picture. To identify *PB′* and *RB′*, a TMV (denoted as ***t***) is derived from the MVs of *RB*. If a TMV is unavailable (e.g. only DMVs are available or *RB* is intra-picture predicted), ***t*** is set equal to the zero vector. *PB′* and *RB′* are both located in a reference AU. *PB′* is a temporal reference block of *PB* with a displacement of ***t***, while the *RB′* is a temporal inter-view reference block with a displacement being the sum of ***m*** and ***t***. Similar as in inter-view residual prediction, the temporal predicted residual signal is calculated as the difference between *PB′* and *RB′*, and weighted before generating the final prediction signal of *PB*.

c) *Further Constraints:* The signaled (***m***) or derived (***t***) TMVs are scaled before the relevant blocks are identified so that they point to pictures in the same AU for each prediction direction. Moreover, when temporal RP applies to both prediction directions, a unique vector ***t*** (and its associated picture) is used to avoid additional memory accesses.

A bi-directionally predicted PB, with the two MVs, denoted as $\boldsymbol{m}_0$ and $\boldsymbol{m}_1$, can also use a different type of residual prediction for each prediction direction. Without loss of generality, assume that $\boldsymbol{m}_1$ is a TMV, $\boldsymbol{m}_0$ is a DMV, and the vector pairs, as shown in Fig. 4 (a) and (b), for the two prediction directions are ($\boldsymbol{m}_1$, ***d***) and ($\boldsymbol{m}_0$, ***t***), respectively. In this case, to reduce the number of block accesses, ***t*** is set equal to $\boldsymbol{m}_1$, such that only one additional block access is needed.

2) Illumination Compensation: The purpose of illumination compensation (IC) is to improve inter-layer prediction when there are illumination mismatches between views. This is done by applying a scale factor and offset to the prediction samples. IC is separately applied for each unidirectional inter-view sample prediction of a PB (hence twice for bi-prediction). For complexity reduction, IC can be signaled only in CUs associated with a single PU not using RP.

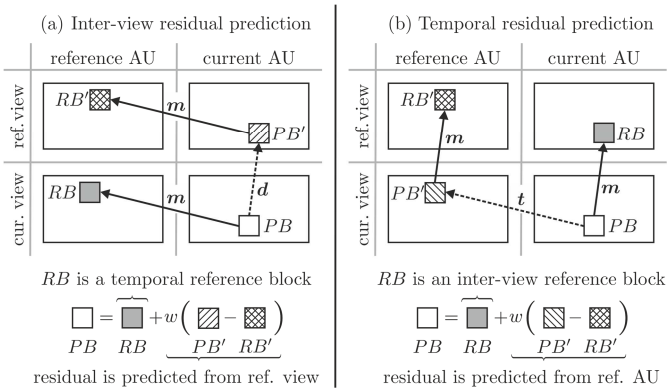


Fig. 4. Residual Prediction cases for one prediction direction.

The scale and offset values are calculated by matching a set of samples in the reference picture to a set of samples in the current picture [33]. The set in the current picture is given by the samples spatially adjacent to the top and left of the current PB, where only every second sample is used to keep complexity low. Accordingly, the set of the reference picture includes every second sample adjacent to the top and left of the block used as reference for inter-view prediction. A linear least square solution is approximated using a look-up table to avoid a division operation. For chroma PBs, IC is even simpler since it only derives and applies the offset.

3) Depth Based Block Partitioning (DBBP): DBBP predicts segmentation information from an already decoded depth map to improve the compression of a dependent texture [34]. It is invoked by a flag that is present with luma CBs of size 16×16 or larger, when the PB partitioning is either in PART_N×2N or PART_2N×N mode. When DBBP is used, the PUs are required to signal unidirectional prediction and the CB is reconstructed as follows: First, NBDV is used to identify the depth block in a picture of the current AU and the RV that likely corresponds to the position of the current CB. Then, a threshold is derived as the mean of the four corner samples of that depth block and a binary pattern is generated by classifying the samples that are above or below the threshold. Inter-picture sample prediction with the motion information of the two signaled PUs is performed for the entire CB, such that two prediction signals with the same CB size are obtained. Then, the two prediction signals are merged by using samples of one prediction for positions with a value of 0 in the binary pattern, and samples of the other prediction for remaining positions. Finally, a vertical or horizontal 3-tap filter is applied to the merged prediction at positions corresponding to the edge positions in the binary pattern, where the direction of the filtering is perpendicular to the signaled (but not used) split direction of the CB.

This way, if the depth map bears reliable information about the position of an object boundary, different motion compensation modes could be used at both sides of that boundary, and the prediction can be improved without using small PB sizes.

4) Full Sample Motion: Fractional sample interpolation at sharp edges in depth maps can create ringing artifacts. For this, 3D-HEVC only supports full sample motion accuracy for depth layers. Further benefits of this approach are a reduced bit rate for MV signaling, and a reduced complexity for mo-

tion compensation.

E. Depth Intra-Picture Prediction

In addition to intra-picture prediction modes provided by HEVC single-layer coding, which are unchanged in 3D-HEVC, a new skip mode and three new prediction modes, called Intra_Single, Intra_Wedge, and Intra_Contour mode, are available for intra-picture coding in depth layers. The new skip mode allows an early signaling of frequently used intra-picture prediction modes. The three prediction modes have been introduced for efficient representation of sharp edges and homogeneous areas, which are typical in depth maps. The Intra_Single mode signals a single boundary sample value as prediction for the entire PB and can only be applied together with the intra-picture skip mode. In the Intra_Wedge and Intra_Contour modes a subdivision of the PB into two SBPs is signaled or derived. The SBPs are not required to have a rectangular shape (as e.g. depicted in Fig. 3). For each SBP, a DC value is predicted from decoded boundary samples of adjacent blocks. The new intra-picture prediction modes have two common differences in comparison to the conventional intra-picture prediction modes: First, boundary value smoothing is not applied; and second, PBs using the new modes are unavailable in the derivation of most probable modes (MPMs).

1) Intra_Wedge and Intra_Contour: The Intra_Wedge and the Intra_Contour modes [35] are signaled at the PU level by two flags. The first flag indicates that one of the two new modes is used instead of a conventional intra-picture prediction mode. The other flag indicates which mode is used.

When a PU uses the Intra_Wedge mode, the sub-partitioning is explicitly signaled by an index value referring to a set of binary patterns denoted as wedgelets. The set of wedgelets contains patterns resulting from a segmentation of the PB with straight lines. The number of wedgelets in the set depends on the PB size. The wedgelets can be either created on-the-fly when decoding the PU, or be created and stored in advance.

In the Intra_Contour mode, the sub-partitioning of the PB is inter-component predicted from a collocated block in the texture picture of the same view and AU. For segmentation, first a threshold is derived by averaging the four corner sample values of the collocated texture block, and then a binary pattern representing the SBPs is derived by comparing sample values of the texture block with the threshold value. Due to the thresholding, the area of an SBP can also be disjunct.

After generation of the binary pattern defining the two SBPs, the same sample prediction process is applied in Intra_Wedge and Intra_Contour modes. The process is invoked for each of the two SBPs and provides a DC prediction from a subset of the decoded boundary samples (depicted in Fig. 5 (a)) of blocks adjacent to the current PB. How the prediction DC is calculated depends on which of the neighboring samples l_a, t_a, l_c and t_c of the current PB are also neighboring samples of the current SBP as shown in Fig. 5 (b).

A special case occurs when none of the four samples is a neighbor of the current SBP (as for the SBP marked with " * " in Fig. 5 (a)). Then, it is assumed that the SBP boundary within the PB is aligned with an edge, which intersects the boundary of a neighboring block either between positions of l_c and l_d or between positions of t_c and t_d. If the absolute difference of t_a and t_d is greater than the absolute difference of l_a and l_d, the latter intersection is assumed and the prediction is given by t_d, as indicated in Fig. 5 (c) [36]. Otherwise, samples values of the SBP are set equal to l_d.

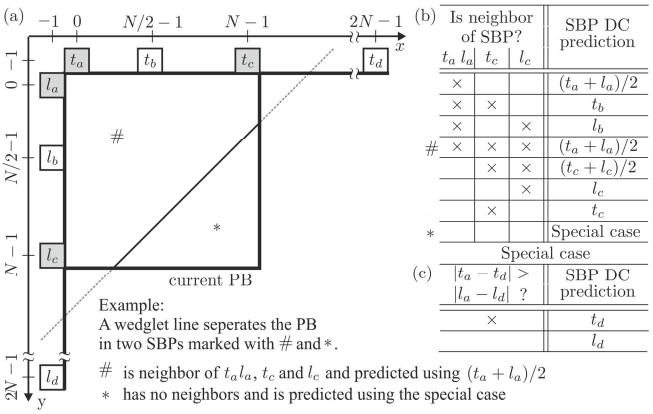


Fig. 5. Intra_Wedge and Intra_Contour DC prediction. (a) t_a, t_b, t_c, l_a, l_b, l_c, are neighbor samples of the current PB; (b) Prediction depending on whether t_a, l_a, l_c, t_c, are also neighbors of the current SBP; (c) Special case.

2) Intra-Picture Skip and Intra_Single: The intra-picture skip mode can be applied in depth layers for CUs, which are not using the conventional (inter-picture) skip mode [7]. When the intra-picture skip mode is used, the CU contains only three syntax elements: The skip flag being equal to 0, a flag indicating the intra-picture skip mode being equal to 1, and an index, which selects the prediction mode. Similar to the (inter-picture) skip mode, other syntax elements are not present and the CU is associated with a single PB.

Depending on the signaled index [37] the prediction is derived by the horizontal or vertical Intra_Angular mode [7] or the Intra_Single mode. In the Intra_Single mode [38][37] the prediction for the whole PB is chosen again depending on the signaled index as the value of the decoded boundary sample adjacent to position *(N/2, 0)* or *(0, N/2)* in the PB, with *N* denoting the PB size (in Fig. 5 right to t_b or below l_b).

F. Depth Residual Coding

As described in Section IV.B, high frequency components of blocks in a depth map can be irrelevant for view synthesis, such that the depth DC becomes more important. To efficiently preserve the DC component of the prediction residual (denoted as DC offset) in 3D-HEVC, it can be signaled explicitly in addition to, or as alternative to, quantized transform coefficients, where the latter is referred to as DC-only mode. DC offset coding has furthermore been extended by a depth lookup table (DLT) technique, which exploits the fact that the value range of depth samples is often only sparsely used.

1) DC(-Only) Coding: The DC-only mode can be enabled by a flag present in CUs associated with a single PU and not coded in skip or intra-picture skip mode. When enabled together with intra- or inter-picture prediction modes as already specified in single layer HEVC, one DC offset is signaled and added on top of the intra-picture predicted [39] or motion compensated signal. In the Intra_Contour or the Intra_Wedge modes, one DC offset is present for each of the two SBPs.

Moreover, when a PU is coded in Intra_Wedge or Intra_Contour mode and not using DC-only coding, it can be assumed that it contains an edge, which is relevant for view synthesis. For better preservation such PUs can signal both DC offsets and quantized transform coefficient [40].

A flag in the PPS indicates if the DC offsets are directly added to each sample of the prediction, or if they are scaled before in a non-linear process using the DLT.

2) Depth Look-Up Table (DLT): Samples of a depth map are typically represented with 8 bits of precision, although only a small set of distinct depth values, potentially non-uniformly distributed over the value range, might be used. To map a compressed range of consecutive index values to such a set of distinct depth values, a DLT can be transmitted in the PPS [41]. When the DLT is present, coding performance is increased by signaling DC offsets in the compressed index range instead of DC offsets with higher magnitude in the depth sample range. For this, an encoder derives the index offset Δi as difference of the two index values representing the DCs of the original samples $\bar{s}_O$ and the predicted samples $\bar{s}_P$, as follows, with IDLT denoting a look-up in an inversed version of the DLT:

$$\Delta i = \mathrm{IDLT}(\bar{s}_O) - \mathrm{IDLT}(\bar{s}_P)$$

In Intra_Wedge or Intra_Contour mode, $\bar{s}_P$ is directly available for an SBP as described in Section IV.E.1. For other modes $\bar{s}_P$ is estimated as the mean value of the prediction of the four corner samples of the PB. At the decoder, the residual DC $\bar{s}_R$ is then computed using the inverse mapping before addition to each sample of the prediction signal:

$$\bar{s}_R = \mathrm{DLT}(\mathrm{IDLT}(\bar{s}_P) + \Delta i) - \bar{s}_P$$

V. Profiles

The second edition of HEVC specifies one profile for MV-HEVC, which is the Multiview Main profile, or simply MV Main profile. For backward compatibility, the MV Main profile requires the base layer to conform to the Main profile. Moreover, block-level coding tools of enhancement layers are similarly constrained to enable reusing legacy decoder hardware below the slice level. Hence, only 4:2:0 chroma and a sample precision of 8 bits are supported. A constraint introduced to limit complexity of inter-view prediction is that the number of reference layers (including indirect reference layers) used by a layer must not be greater than 4.

For 3D-HEVC the 3D Main profile specifies a superset of the capabilities of the MV Main profile, such that a 3D Main profile conforming decoder is able to decode MV Main profile conforming bitstreams. The base layer is required to conform to the Main profile. New low-level coding techniques of 3D-HEVC are supported only by enhancement layers that are not auxiliary picture layers. Texture layers support only 4:2:0 chroma and depth layers support only monochrome. For both components the sample precision is restricted to 8 bits.

Furthermore, both profiles disallow inter-layer prediction between layers that use different picture sizes.

TABLE III
BIT RATE SAVINGS OF MV-HEVC (MV) AND 3D-HEVC (3D)

1) Two-view texture (stereo)		2) Three-view texture and depth		
Test set	MV vs. Sim.[1]	Test set	3D vs. Sim.[1]	3D vs. MV
		PoznanHall	−40.5% [4]	−19.3% [4]
		PoznanStreet	−41.5% [4]	−11.9% [4]
UndoDancer	−37.4% [2]	*UndoDancer*	−50.9% [4]	−19.2% [4]
GTFly	−42.6% [2]	*GTFly*	−55,3% [4]	−18.5% [4]
Bmx	−29.0% [2]	*Balloons*	−40.3% [4]	−19.3% [4]
Band06	−34.7% [2]	*Newspaper*	−41.0% [4]	−19.9% [4]
Musicians	−25.3% [2]	*Kendo*	−40.0% [4]	−21.4% [4]
Poker	−24.9% [2]	*Shark*	−59,2% [4]	−24.4% [4]
Average	−32.3% [2]	Average	−46.1% [4]	−19.3% [4]
Enh. Only	−70.8% [3]	Enh. Only	−72.8% [5]	−34.3% [5]

[1]Simulcast coding with single-layer HEVC. Savings calculated based on:
[2]T_0, T_1 (PSNR and bit rate); [3]T_1 (PSNR and bit rate);
[4]six synthesized views (PSNR) and T_0, D_0, T_1, D_1, T_2, D_2 (bit rate);
[5]six synthesized views (PSNR) and D_0, T_1, D_1, T_2, D_2 (bit rate).

VI. Compression Performance

To evaluate the compression efficiency of the different extensions, simulations were conducted using the reference software HTM [42], and experimental evaluation methodology that has been developed and is being used by the standardization community [43][44]. In that framework, multiview texture video and the corresponding depth can be provided as input, while the decoded views and additional views synthesized at selected positions can be generated as output. For evaluation, two setups have been used as shown in Table III.

The first setup evaluates the typical use case foreseen for MV-HEVC, which is the coding of stereo video without depth (hence of two texture layers denoted as T_0 and T_1). For this, the reported results were obtained by averaging the bit rate savings of six test sequences suitable for stereoscopic displays. Savings for each sequence have been calculated based on total bit rate and averaged PSNRs of both layers (average) as well as of the enhancement layer T_1 only (enh. only). The total results for MV-HEVC compared to simulcast coding are about 32%. Regarding the enhancement texture T_1 only, which benefits from inter-view prediction, bit rate savings of about 71% have been achieved.

Although not shown in Table III, it is worth mentioning that only a modest bit rate saving of about 6% on average (26% enh. only) can be achieved by 3D-HEVC compared to MV-HEVC for the stereo case. However, the target application of 3D-HEVC is coding of data suitable for view synthesis at auto-stereoscopic displays. For this, a set of eight sequences, each comprising texture (T_0, T_1, T_2) and depth (D_0, D_1, D_2) of three views, have been coded in the second setup. Compared to the stereo sequences used in the first setup, original camera distances (and thus disparities) between views are approximately doubled, such that views are less correlated. For each sequence, rate savings have been calculated based on averaged PSNRs of six synthesized intermediate views. Averaged over all sequences, total bit rate savings of about 46% and 19% are achieved when comparing to simulcast and to MV-HEVC using auxiliary pictures, respectively. When again neglecting the rate of the base layer T_0, which cannot use the new tools, savings are about 73% and 34%, respectively.

VII. Conclusion

Experts of ITU-T VCEG and ISO/IEC MPEG have jointly developed the multiview and 3D extensions of HEVC. Both extensions allow the transmission of texture, depth, and auxiliary data for advanced 3D displays. An increased compression performance compared to simulcast HEVC is achieved by inter-layer prediction. In contrast to 3D-HEVC, MV-HEVC can be implemented without block-level changes. However, thanks to advanced coding techniques a higher coding efficiency can be provided by 3D-HEVC in particular for cases where depth maps have to be coded.

As both extensions have been developed to support stereoscopic and autostereoscopic displays, they have not been specifically designed to handle arrangements with a very large number of views or arbitrary view positioning. Accordingly, improved capabilities for supporting such configurations may be a subject for future standardization.

Acknowledgements

The authors would like to thank all the experts who have contributed to the development of MV-HEVC and 3D-HEVC.

References

[1] H. Urey, K.V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540-555, Apr. 2011.

[2] N. A. Dodgson, "Autostereoscopic 3D Displays," *IEEE Computer*, vol. 38, no. 8, pp. 31-36, Aug. 2005.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE 5291, Stereoscopic Displays and Virtual Reality Syst. XI*, San Jose, May 2004.

[4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. of Computer Vision*, vol. 47, no. 1, pp. 7-42, May 2002.

[5] J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, no. 4, pp. 827-849, Apr. 2004.

[6] S. Foix, G. Alenya, and C. Torras, "Lock-in Time-of-Flight (ToF) Cameras: A Survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917-1926, Sep. 2011.

[7] *High Efficiency Video Coding*, Rec. ITU-T H.265 and ISO/IEC 23008-2, Jan. 2013.

[8] 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.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[9] MPEG Video and Requirements, "Vision on 3D Video Coding," ISO/IEC JTC1/SC29/WG11 Doc. N10357, Lausanne, Switzerland, Feb. 2009.

[10] MPEG Video and Requirements, "Call for Proposals on 3D Video Coding Technology," ISO/IEC JTC1/SC29/WG11 Doc. N12036, Geneva, Switzerland, Mar. 2011.

[11] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3D High Efficiency Video Coding for Multi-View Video and Depth Data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366-3378, Sep. 2013.

[12] G. Tech, H. Schwarz, K. Müller, and T. Wiegand, "3D Video Coding using the Synthesized View Distortion Change," *Picture Coding Symp. (PCS), 2012*, Krakow, Poland, pp. 25–28, May 2012.

[13] G. Tech, K. Wegner, Y. Chen, M. M. Hannuksela, and J. Boyce, "MV-HEVC Draft Text 9," JCT-3V Doc., JCT3V-I1002, Sapporo, JP, Jul. 2014.

[14] *High Efficiency Video Coding*, Rec. ITU-T H.265 and ISO/IEC 23008-2, Oct. 2014.

[15] G. Tech, K. Wegner, Y. Chen, and S. Yea, "3D-HEVC Draft Text 7," JCT-3V Doc., JCT3V-K1001, Geneva, CH, Feb. 2015.

[16] 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," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[17] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The Emerging MVC Standard for 3D Video Services," *EURASIP J. Advances Signal Process*, vol. 2009, no. 1, Jan. 2009.

[18] J. Boyce, Y. Yan, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, July 2015, to be published.

[19] Y. Chen, L. Zhang, V. Seregin, and Y.-K. Wang, "Motion Hooks for the Multiview Extension of HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2090–2098, Dec. 2014.

[20] S. Bouyagoub, A. Sheikh Akbari, D. Bull, and N. Canagarajah, "Impact of camera separation on performance of H.264/AVC-based stereoscopic video codec," *Electron. Lett.*, vol.46, no.5, pp.345–346, Mar. 2010.

[21] R. Sjöberg, Y. Chen, A. Fujibayashi, M. M. Hannuksela, J. Samuelsson, T. K. Tan, Y.-K. Wang, and S. Wenger, "Overview of HEVC High-Level Syntax and Reference Picture Management," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1858–1870, Dec. 2012.

[22] Y. Chen and Y.-K. Wang, "MV-HEVC/SHVC HLS: Cross-layer POC alignment," JCT-3V Doc., JCT3V-E0075, Vienna, AT, Jul. 2013.

[23] Hendry, A. K. Ramasubramonian, Y.-K. Wang, and Y. Chen, "MV-HEVC/SHVC HLS: On picture order count," JCT-3V Doc., JCT3V-G0031, San José, USA, Jan. 2014.

[24] J.-W. Kang, Y. Chen, L. Zhang, and M. Karczewicz, "Low Complexity Neighboring Block Based Disparity Vector Derivation in 3D-HEVC," *IEEE Int. Symp. on Circuits and Syst.(ISCAS)*, Melbourne, Jun. 2014, pp. 1921–1924.

[25] Y.-L. Chang, C.-L. Wu, Y.-P. Tsai, and S. Lei, "CE1.h: Depth-oriented Neighboring Block Disparity Vector (DoNBDV) with virtual depth retrieval," JCT-3V Doc., JCT3V-C0131, Geneva, CH, Jan. 2013.

[26] L. Zhang, Y. Chen, and L. Liu, "3D-CE5.h: Merge candidates derivation from disparity vector," JCT-3V Doc., JCT3V-B0048, Shanghai, CN, Oct. 2012.

[27] H.-S. Koo, Y.-J. Jeon, and B.-M. Jeon, "Motion information inferring scheme for multi-view video coding," *IEICE Trans. on Commun.*, E91-B(4), pp. 1247-1250, Mar. 2008.

[28] J. An, K. Zhang, J.-L. Lin, and S. Lei, "3D-CE3: Sub-PU level inter-view motion prediction," JCT-3V Doc., JCT3V-F0110, Geneva, CH, Oct. 2013.

[29] Y. Chen, H. Liu, and L. Zhang, "CE2: Sub-PU based MPI," JCT-3V Doc., JCT3V-G0119, San Jose, US, Jan. 2014.

[30] F. Zou, D. Tian, A. Vetro, H. Sun, O. C. Au, and S. Shimizu, "View Synthesis Prediction in the 3-D Video Coding Extensions of AVC and HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1696-1708, Oct. 2014.

[31] L. Zhang, Y. Chen, X. Li, and S. Xue, "Low-complexity advanced residual prediction design in 3D-HEVC," in *Proc. of IEEE Int. Symp. on Circuits and Syst. (ISCAS)*, Melbourne, Jun. 2014, pp. 13–16.

[32] L. Zhang and Y. Chen, "Advanced residual prediction enhancement for 3D-HEVC," in *IEEE Int. Conf. Multimedia and Expo Workshops (ICMEW)*, Chengdu, China, Jul. 2014.

[33] H. Liu, J. Jung, J. Sung, J. Jia, and S. Yea, "3D-CE2.h: Results of Illumination Compensation for Inter-View Prediction," JCT-3V Doc., JCT3V-B0045, Shanghai, CN, Oct. 2012.

[34] F. Jäger, "Depth-based block partitioning for 3D video coding," *Picture Coding Symp. (PCS), 2013*, San Jose, US, pp. 410–413 , Dec. 2013.

[35] P. Merkle, K. Müller, D Marpe, and T. Wiegand, "Depth Intra Coding for 3D Video based on Geometric Primitives," *IEEE Trans. Circuits Syst. Video Technol.*, Feb. 2015, to be published.

[36] X. Zhao, L. Zhang, Y. Chen, and M. Karczewicz, "CE6.h related: Simplified DC predictor for depth intra modes," JCT-3V Doc., JCT3V-D0183, Incheon, KR, Apr. 2013.

[37] J. Y. Lee, M. W. Park, and C. Kim, "3D-CE1: Depth intra skip (DIS) mode," JCT-3V Doc., JCT3V-K0033, Geneva, CH, Feb. 2015.

[38] Y.-W. Chen, J.-L. Lin, Y.-W. Huang, and S. Lei, "3D-CE2: Single depth intra mode for 3D-HEVC," JCT-3V Doc., JCT3V-I0095, Sapporo, JP, Jul. 2014.

[39] H. Liu and Y. Chen, "Generic Segment-Wise DC for 3D-HEVC Depth Intra Coding," *IEEE Int. Conf. on Image Process. (ICIP)*, Paris, Oct. 2014, pp. 3219–3222.

[40] F. Jäger, "Simplified depth map intra coding with an optional depth lookup table," in *Proc. of Int. Conf. on 3D Imaging IC3D '12*, Liège, Belgium, Dec. 2012.

[41] F. Jäger, M. Wien, and P. Kosse, "Model-based intra coding for depth maps in 3D video using a depth lookup table," in *Proc. of 3DTV Conf. 3DTV-CON '12*, Zurich, Switzerland, Oct. 2012.
[42] JCT-3V. (2015, May) MV- and 3D-HEVC reference software, HTM-14.1 [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-14.1/
[43] Y. Chen, G. Tech, K. Wegner, and S. Yea, "Test Model 11 of 3D-HEVC and MV-HEVC," JCT-3V Doc., JCT3V-J1003, Geneva, CH, Feb. 2015.
[44] K. Müller and A. Vetro, "Common Test Conditions of 3DV Core Experiments," JCT-3V Doc., JCT3V-G1100, San Jose, US, Jan. 2014.

Gerhard Tech received the Dipl.-Ing. degree in Electrical Engineering from RWTH Aachen University, Germany, in 2007.

He has been with Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany, since 2008, where he is currently Research Associate. His research interests are video coding and processing, including 3D representations.

He has contributed to the standardization activities of the JCT-3V of ITU-T VCEG and ISO/IEC MPEG. During the development of the multiview and 3D extensions of H.265/HEVC, he co-chaired several ad hoc groups of the JCT-3V. He has been appointed as an editor of H.265/HEVC Multiview extension, as editor of the H.265/HEVC 3D extension, and as software coordinator for the 3D- and MV-HEVC reference software HTM.

Ying Chen (M'05-SM'11) received a B.S. in Applied Mathematics and an M.S. in Electrical Engineering & Computer Science, both from Peking University, in 2001 and 2004 respectively. He received his PhD in Computing and Electrical Engineering from Tampere University of Technology (TUT), Finland, in 2010.

He is currently a Senior Staff Engineer/Manager at Qualcomm Incorporated, San Diego, CA, USA. Dr. Chen joined Qualcomm in Mar. 2009. His earlier working experiences include Researcher in TUT and Nokia Research Center, Finland from 2006 to Feb. 2009 and Research Engineer in Thomson Corporate Research, Beijing, from 2004 to 2006.

Dr. Chen has been actively contributing to MPEG, JVT, JCT-VC, and JCT-3V, on Scalable Video Coding, Multiview Video Coding (MVC), and 3D Video (3DV) Coding extensions of H.264/AVC, as well as HEVC and its multiview, scalable, screen content coding and 3DV extensions. Dr. Chen has also been involved in standardization activities of MPEG systems. Dr. Chen has served as an editor for several standard specifications, including 3DV extensions of AVC, multiview and 3D extensions of HEVC (MV-HEVC and 3D-HEVC). In addition, Dr. Chen has served as a software coordinator of MVC and a conformance test editor of MVC, MV-HEVC and 3D-HEVC. Dr. Chen has co-authored about 500 standardization contribution documents, over 50 academic papers and over 200 families of granted or pending patents in the fields of image processing, video coding, and video transmission. Dr. Chen is a member of the IEEE CAS Visual Signal Processing and Communications (VSPC) technical committee and a member of IEEE CAS Digital Signal Processing (DSP) technical committee.

Karsten Müller (M'98-SM'07) received the Dr.-Ing. degree in Electrical Engineering and Dipl.-Ing. degree from the Technical University of Berlin, Germany, in 2006 and 1997 respectively.

He has been with the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Berlin, since 1997, where he is currently head of the Image and Video Understanding Group. His research interests include representation, coding and reconstruction of 3D scenes in Free Viewpoint Video scenarios and Coding, multiview applications and combined 2D/3D similarity analysis.

He has been involved in international standardization activities, successfully contributing to the ISO/IEC Moving Picture Experts Group for work items on visual media content description, multiview, multi-texture and 3D video coding. Here, he co-chaired an adHoc group on 3D Video Coding from 2003 to 2012. Dr. Müller has served as chair and editor in IEEE conferences, and is an Associated Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Jens-Rainer Ohm (M'92) received the Dipl.-Ing., Dr.-Ing., and Habil. degrees from the Technical University of Berlin (TUB), Berlin, Germany, in 1985, 1990, and 1997, respectively. Since 2000, he is Full Professor and Chair of the Institute of Communication Engineering, RWTH Aachen University, Germany. His past and current research and teaching activities cover the areas of motion-compensated, stereoscopic and 3-D image processing, multimedia signal coding, transmission and content description, audio signal analysis, as well as fundamental topics of signal processing and digital communication systems. He has authored German and English textbooks on multimedia signal processing, analysis and coding, on basics communication engineering and signal transmission, as well as numerous papers in the fields mentioned above.

Dr. Ohm has been participating in the work of the Moving Picture Experts Group (MPEG) since 1998. He has been the Chair and Co-Chair of various standardization activities in video coding, namely, the MPEG Video Subgroup since 2002, the Joint Video Team of MPEG and ITU-T SG 16 VCEG from 2005 to 2009, and currently, the JCT-VC as well as the JCT-3V.

Anthony Vetro (S'92–M'96–SM'04–F'11) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Polytechnic University, Brooklyn, NY. He joined Mitsubishi Electric Research Labs, Cambridge, MA, in 1996. He is currently a Deputy Director of the lab and also manages a group that is responsible for research on video coding and image processing, information security, sensing technologies, and speech/audio processing. He has published more than 200 papers and has been an active member of the ISO/IEC and ITU-T standardization committees on video coding for many years, where he has served as an ad-hoc group chair and editor for several projects. He was a key contributor to 3D and multiview extensions of the AVC and HEVC standards, and served as Head of the U.S. delegation to MPEG.

Dr. Vetro is also active in various IEEE conferences, technical committees, and editorial boards. He currently serves as an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, and as a member of the Editorial Boards of IEEE MultiMedia and IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He served as Chair of the Technical Committee on Multimedia Signal Processing of the IEEE Signal Processing Society and on the steering committees for ICME and the IEEE TRANSACTIONS ON MULTIMEDIA. He served as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2010-2013) and IEEE Signal Processing Magazine (2006–2007) and later as a member of the Editorial Board (2009-2011). He also served as a member of the Publications Committee of the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS (2002–2008). He has also received several awards for his work on transcoding, including the 2003 IEEE Circuits and Systems CSVT Transactions Best Paper Award. He is a Fellow of IEEE.

Ye-Kui Wang received his BS degree in industrial automation in 1995 from Beijing Institute of Technology, Beijing, China, and his PhD degree in electrical engineering in 2001 from the Graduate School in Beijing, University of Science and Technology of China, Beijing, China. He is currently a Director of Technical Standards at Qualcomm, San Diego, CA, USA. His earlier working experiences and titles include Multimedia Standards Manager at Huawei Technologies, Bridgewater, NJ from Dec. 2008 to Aug. 2011, various positions, including Principal Member of Research Staff, at Nokia Corporation, Tampere, Finland from Feb. 2003 to Dec. 2008, and Senior Researcher at Tampere International Center for Signal Processing, Tampere University of Technology, Tampere, Finland from Jun. 2001 to Jan. 2003. His research interests include video coding and multimedia transport and systems.

Dr. Wang has been a contributor to various multimedia standards, including video codecs, file formats, RTP payload formats and video application systems, developed in ITU-T VCEG, ISO/IEC MPEG, JVT, JCT-VC, 3GPP SA4, IETF and AVS. He has been an editor for several standards, including HEVC, SHVC, HEVC file format, HEVC RTP payload format, SHVC/MV-HEVC file format, ITU-T H.271, Scalable Video Coding (SVC) file format, Multiview Video Coding (MVC), IETF RFC 6184, IETF RFC 6190, and 3GPP TR 26.906. He has co-authored over 500 standardization contributions, about 50 academic papers, and over 200 families of granted or pending patents.