Chapter 4
Recent Advances on 3D Video Coding Technology: HEVC Standardization Framework

Dragorad A. Milovanovic, Dragan Kukolj, and Zoran S. Bojkovic

Abstract 3D video is emerging media extension of conventional 2D video into third dimension adding depth sensation and resolving 2D viewing ambiguity. Primary usage scenario for 3D video is to support 3D video applications, where 3D depth perception of a visual scene is provided by a 3D display system. Multiview-plus-depth (MVD) is visual representation and coding format which takes 3D geometry information of acquisition system in the form of distance information. Applications require transmission of jointly encoded multiple synchronized video signals that show the same 3D scene from different viewpoints. Advances in multi-camera arrays and display technology enable new applications for 3D video. It is clear that these applications need to be based on well-defined and -documented technical standards. Recent advances and challenges in development of the 3D video formats and associated coding technologies are summarized in this chapter with focus on undergoing MPEG/ITU standardization framework for 3D extensions of HEVC high-efficiency video encoder. Research on coding efficiency improvement and complexity reduction of 3D-HEVC reference encoder implementation are outlined.

4.1 Introduction

Over the past 25 years, significant progress has been made in the coding and transmission of digital video. Three-dimensional digital video (3DV) signal processing technology has significantly affected the multimedia on Internet. The MPEG (*Moving Picture Coding Experts Group*) was established in January 1988 with the mandate to develop international standards for compression, decompression, processing, and coded representation of moving pictures, audio, and their combination, in order to satisfy a wide variety of applications. The ISO standards produced

---

D.A. Milovanovic • Z.S. Bojkovic
University of Belgrade, Studentski Trg 1, Belgrade 11000, Serbia
e-mail: dragoam@gmail.com

D. Kukolj (✉)
University of Novi Sad, Trg Dositeja Obradovica 6, Novi Sad 21000, Serbia
e-mail: dragan.kukolj@rt-rk.uns.ac.rs

A. Kondoz, T. Dagiuklas (eds.), *Connected Media in the Future Internet Era*,
DOI 10.1007/978-1-4939-4026-4_4

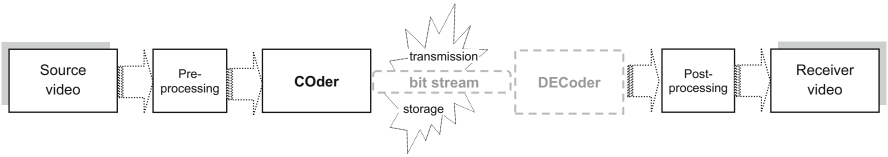


Fig. 4.1 Standardization scope of digital video codec is indicated by *dashed boxes*: only the syntax and semantics of the bit stream and its decoding are defined

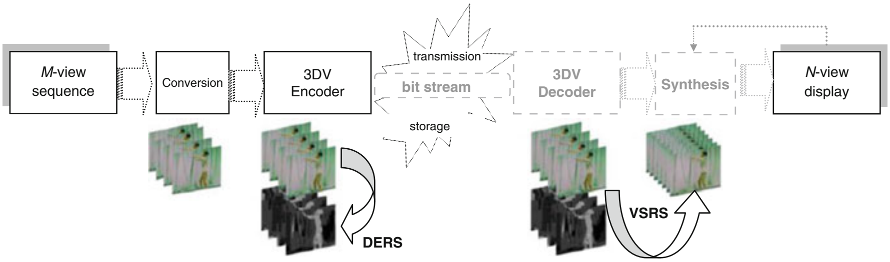


Fig. 4.2 Standardization scope of 3D video codec is indicated by *dashed boxes*: MPEG software DERS (Depth Estimation Reference Software) generates multiview+depth sequence, MPEG software VSRS (View Synthesis Reference Software) reconstructs multiview+depth and synthesizes N-views to display ($N \geq M$)

by MPEG are published in the last stage of a long process that starts with the proposal of new work within a committee, and continue through *competitive* and *cooperative* phases (Fig. 4.1). The evaluation of coding techniques is performed based on their performance (both objectively and by formal subjective testing), efficiency with respect to software/hardware implementation, and feasibility of system architectures.

ISO/IEC JTC1/SC29/WG11 *Moving Picture Experts Group* (**MPEG**) and ITU-T Study Group 16 *Video Coding Experts Group* (**VCEG**) are committees responsible for the development of video coding standards. These committees have jointly developed the widely deployed advanced video coding (**AVC**) and high-efficiency video coding (**HEVC**) standards. They are working on 3D extensions of these standards under the *Joint Collaborative Team on 3D Video* (**JCT-3V**), which was established in July 2012. The 3D video extensions (3D-HEVC, MV-HEVC, 3D-AVC, MVC+D) support the improved coding of stereoscopic and multiview video and facilitate advanced 3D capabilities such as view rendering through the use of depth maps (Fig. 4.2). Support for multiview enables representation of video content with multiple camera views and optional auxiliary information. Support for 3D enables ***joint*** representation of *video content* and *depth information* with multiple camera views.

The 3DV format targets two specific application scenarios:

- Enabling stereo devices to cope with varying display types and sizes, and different viewing preferences. This includes the ability to vary the baseline distance for stereo video to adjust the depth perception, which could help to avoid fatigue and other viewing discomforts.
- Support for high-quality auto-stereoscopic displays, in such a way that the new format enables the generation of many high-quality views from a limited amount of input data, e.g., stereoscopic video and respective depth maps.

Requirements for 3DV data format are as follows:

- **Video data.** The uncompressed data format shall support stereo video, including samples from left and right views as input and output. The source video data should be rectified to avoid misalignment of camera geometry and colors. Other input and output configurations beyond stereo should also be supported.
- **Supplementary data**. Supplementary data shall be supported in the data format to facilitate high-quality intermediate view generation. Examples of supplementary data include depth maps, reliability/confidence of depth maps, segmentation information, transparency or specular reflection, occlusion data, etc. Supplementary data can be obtained by any means.

  - **Metadata.** Metadata shall be supported in the data format. Examples of metadata include extrinsic and intrinsic camera parameters, scene data, such as near and far plane, and others.

Requirements for compression of 3DV data format are as follows:

- **Compression efficiency.** Video and supplementary data should not exceed twice the bit rate of state-of-the-art compressed single video. It should also be more efficient than state-of-the-art coding of multiple views with comparable level of rendering capability and quality.
- **Synthesis accuracy.** The impact of compressing the data format should introduce minimal visual distortion on the visual quality of synthesized views. The compression shall support mechanisms to control overall bitrate with proportional changes in synthesis accuracy.
- **Backward compatibility.** The compressed data format shall include a mode which is backward compatible with existing MPEG coding standards that support stereo and mono video. In particular, it should be backward compatible with MVC.
- **Stereo/mono compatibility**. The compressed data format shall enable the simple extraction of bit streams for stereo and mono output, and support high-fidelity reconstruction of samples from the left and right views of the stereo video.

Requirements for rendering of 3DV data format areas are as follows:

- **Rendering capability.** The data format should support improved rendering capability and quality. The rendering range should be adjustable.

- **Low complexity.** The data format shall allow real-time decoding and synthesis of views, required by any N-view display, with computational and memory power available to devices at the consumer electronics level.
- **Display types.** The data format shall be display independent. Various types and sizes of displays, e.g., stereo and auto-stereoscopic N-view displays of different sizes with different number of views, shall be supported. The data format shall be adaptable to the associated display interfaces.
- **Variable baseline.** The data format shall support rendering of stereo views with a variable baseline.
- **Depth range.** The data format should support an appropriate depth range.
- **Adjustable depth location.** The data format should support display-specific shift of depth location, i.e., whether the perceived 3D scene (or parts of it) is behind or in front of the screen.

Therefore, new coding methods are required for 3DV coding, which decouple the production and coding format from the display format. The primary goal of coding method is to *optimize* coding efficiency. *Coding efficiency* is the ability to minimize the bit rate necessary for representation of video content to reach a given level of video quality or, as alternatively formulated, to maximize the video quality achievable within a given available bit rate (Fig. 4.3).

The 3DV extensions based on the HEVC are developed jointly by MPEG and ITU-T for multiview video data with associated depth maps (MVD) coding for the highest compression efficiency. The 3D-HEVC base view is fully compatible with HEVC in order to extract monoscopic video, while the coding of dependent views and depth maps utilizes additional tools. HEVC *video coding layer* design is based on conventional block-based motion-compensated hybrid video coding concepts (Fig. 4.4). In HEVC, the main goal was to achieve a compression gain higher when compared to the second-generation video coding standard AVC at the same video quality. HEVC is targeted at next-generation ultra-HD (4/8K pixels per line) displays.

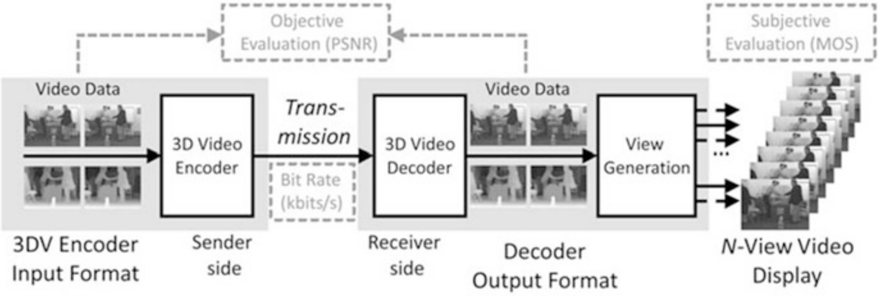


Fig. 4.3 Overall evaluation of different 3DV coding methods with compression and view generation methods (quality and data rate measurements are indicated by *dashed boxes*)

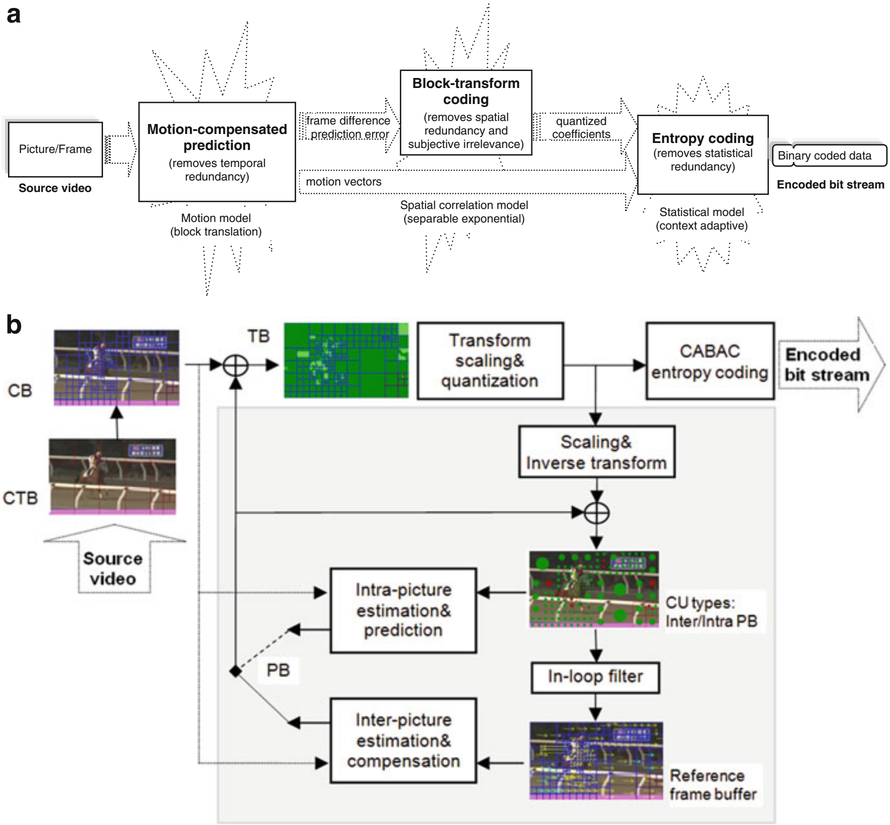


Fig. 4.4 Principles of digital video coding: (**a**) models exploit statistical redundancy of image and subjective irrelevance of viewer, (**b**) HEVC block-based hybrid MC prediction + TC transform coding

4.2 Three-Dimensional Video Formats and Associated Compression Technology

Efficient representation of three-dimensional (3D) video data is very closely involved with the other components of a system: content production, transmission, rendering, and display. It also has a significant impact on the overall performance of the system, including bandwidth requirement and end-user visual quality, as well as constraints such as backward compatibility with display equipment and transmission infrastructure. In this context, standardization is the key to guarantee interoperability and support mass deployment.

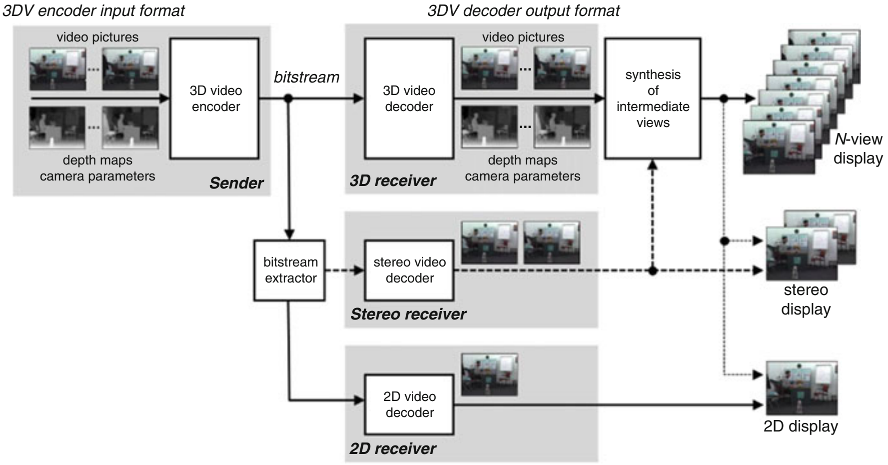


Fig. 4.5 Overview of the system structure and the data formats for encoding, transmission, decoding, rendering, and display of multiview video and associated depth maps

A variety of 3D video representations are available in the current ecosystem (Fig. 4.5):

- **Stereoscopic 3D** (S3D) video is the simplest and most widely used representation. It is based on the principle of *stereopsis*, in which two 2D views (L, R) with a disparity (D) are, respectively, received by the left and right eyes of an observer. The resulting binocular disparity is then exploited by the human visual system (HVS) to create a perception of depth in 3D scene.
- **Multiview video** (MVV) is a straightforward extension of the S3D representation; several texture videos are acquired in a *synchronous* manner by a system of cameras.
- **Multiview video-plus-depth** (MVD) is augmented with an extra channel conveying depth information. Depth maps (M) result in a display-independent representation that enables synthesis of a N ($N>M$) of views. The two sequences, *video texture* and *depth maps*, can then be encoded and transmitted independently. Alternatively, texture and depth can be jointly encoded, to exploit the redundancies between them, resulting in better coding performance.

The most important 3D video standardized codecs and associated formats are as follows:

- **Simulcast** is the simple independent coding (AVC/HEVC) and transmission of views. In addition, no synchronization between views is required. However, simulcast is not optimal in terms of rate-distortion efficiency because the correlation between cameras is not exploited.
- Multiview video coding (**MVC**) is AVC extension that exploits redundancy between views using inter-camera prediction to reduce required bit rate.

- Multiview video + depth coding (**3DV**) is in the current focus of MPEG standardization (MVC extension MVC+D, AVC-compatible extension 3D-AVC; HEVC extensions MV-HEVC and 3D-HEVC). Two major objectives are targeted: to support advanced stereoscopic display processing and to improve support for high-quality auto-stereoscopic multiview displays. It disconnects the video representation/coding from the captured video representation, and the displayed video representation.

For the sake of completeness, the other standardized 3D video formats are listed as follows:

- MPEG-2 Multiview Profile (MVP) uses scalable coding tools in transmission of two stereoscopic video signals inside an MPEG-2 transport stream, and guarantees backward compatibility with the MPEG-2 main profile.
- MPEG-C Part 3 specifies high-level syntax that allows an MPEG-2/AVC decoder to interpret two video streams correctly as texture and depth data inside an MPEG-2 transport stream.
- MPEG-4 Part 2 Multiple Auxiliary Components (MAC) specify a tool for coding video-plus-depth data.
- MPEG-4 Part 10 MVC multi-resolution frame-compatible stereoscopic video coding (MFC) specifies stereo interleaving (spatial/temporal multiplexing) formats, SEI (supplemental enhancement metadata) signaling massages for frame packing, as well as MFC+D enhancement for stereoscopic video coding with depth information.

4.2.1 3D-HEVC System Structure

3DV extensions based on the HEVC are developed jointly by MPEG and ITU-T for multiview video data with associated depth maps (MVD) coding for the highest compression efficiency. The 3D-HEVC base view is fully compatible with HEVC in order to extract monoscopic video, while the coding of dependent views and depth maps utilizes additional tools (Fig. 4.6). A subset of this 3DV coding extension includes MV-HEVC simple multiview extension, utilizing the same design principles of MVC in the AVC framework (providing backward compatibility for monoscopic decoding). MV-HEVC and 3D-HEVC extension are available as a final standards by mid-2014 and 2015, respectively. Additionally, it is planned to develop a suite of tools for scalable coding, where both view scalability and spatial scalability would allow for backward-compatible extensions for more views.

The system structure of 3D-HEVC is described as follows. The video pictures and depth maps are coded by **access units** as illustrated in Fig. 4.7. An access unit includes all video pictures and depth maps at the *same* time instant. The video picture and depth map corresponding to a particular camera position are indicated by a view identifier (`viewId`). The view identifier is also used for specifying the coding order. The view with view identifier equal to 0 is also referred to as the **base**

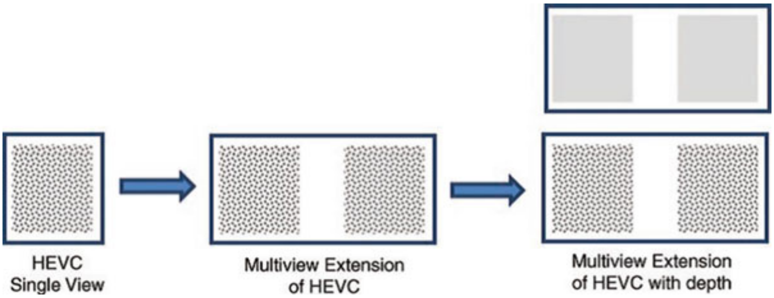


Fig. 4.6 3DVC extensions of HEVC coding standard: MV-HEVC supports MVV format, and 3D-HEVC supports MVD format

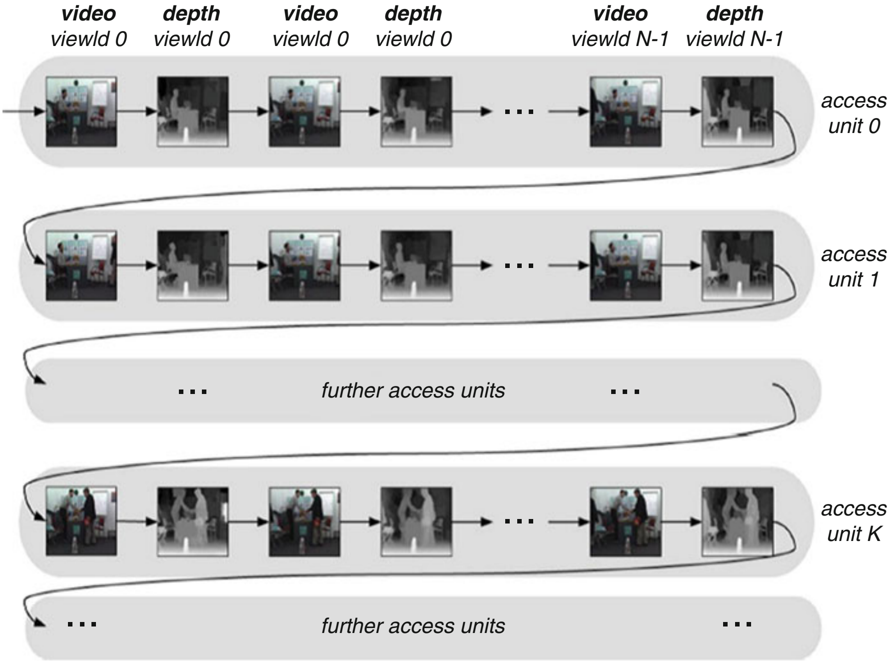


Fig. 4.7 3D-HEVC access unit structure

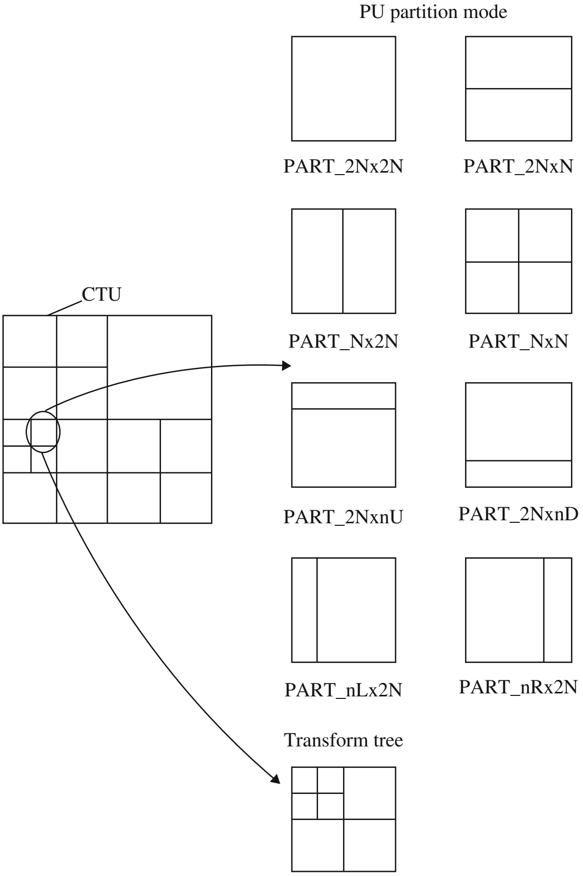
view or the **independent view** and is coded independently of the other views using a conventional HEVC video coder. The other views are referred to as **dependent views** and they can be coded with additional coding tools in 3D-HEVC.

4.2.2 3D-HEVC Encoding Process

The coding structure in 3D-HEVC includes three basic units, identical to that in HEVC: coding unit (**CU**), prediction unit (**PU**), and transform unit (**TU**). A picture is divided into a set of coding tree units (**CTU**s). The CTU is equivalent to a macroblock in H.264/AVC.

- The CU is represented as the leaf node of a *quadtree partitioning* of the CTU. It is a basic unit with a square shape which is associated with a **prediction mode**: intra, inter, or SKIP. A CTU may contain only one CU or may be split into four smaller CUs, and each CU could be recursively split into smaller CUs until the predefined splitting limitation is reached.
- A PU is a basic unit for prediction and has its root at the CU level. The shape of a PU is not necessarily square. Each CU may contain one, two, or four PUs according to the partition mode. The eight partition modes that can be used for an inter-coded CU are shown in Fig. 4.8. Only the PART_2Nx2N and PART_NxN

Fig. 4.8 Quadtree structure of a CTU and TU and possible PU partition modes



partition modes are used for an intra-coded CU. For both inter-coded CU and intra-coded CU, the partition mode PART_NxN can be allowed only when the corresponding CU size is equal to the minimum CU size.

- A TU is another basic unit with a square shape for transform and quantization. Multiple TUs within a CU form a quadtree structure called Residual QuadTree (RQT).

The 3D-HEVC encoder tests all the **coding modes** (up to 20 different modes, i.e., inter/merge/skip2N_2N, inter/merge 2N_N, inter/mergeN_2N, inter/merge N_N, inter/merge 2N_nU, inter/merge2N_nD, inter/mergenL_2N, inter/mergenR_2N, intra2N_2N, intraN_N, and intra PCM) for each CU and selects the mode with the least **RD cost**. Furthermore, each CU could be recursively split into four sub-CUs and the coding mode of each sub-CU is again determined by examining the RD cost of all the coding modes. Whether the CU should be further split or not is also decided by comparing the RD cost of the CU to the summation of the RD costs of the four sub-CUs. The motion estimation (**ME**) and the computation of the RD cost for each CU are the most computationally intensive parts.

The independent view, which is also referred to as the **base view**, is coded by a conventional HEVC codec. For dependent views, additional tools exploiting *inter-component* correlations have been integrated into 3D-HEVC (Fig. 4.9):

- To share the previously encoded texture information of reference views, the **disparity-compensated** prediction (**DCP**) has been added as an alternative to motion-compensated prediction (MCP).
- The **inter-view motion** prediction is employed to predict the motion information for the current block from the previously encoded motion information in the reference views.
- The **residual signal** of the current block can also be predicted from the residual signal of the corresponding block in the reference views.
- Backward **view synthesis prediction** (VSP) is a technique that exploits inter-view redundancies, in which a synthesized signal is used as a reference to predict the current picture.
- For the depth component, among all the above additional tools, only DCP is enabled. However, some new intra-prediction **depth modeling modes** (DMC) and the **motion parameter inheritance** (MPI) mode are added.

4.3 HEVC Standardization Framework

High-efficiency video coding (HEVC) is the current joint video coding standardization project of the ITU-T Video Coding Experts Group (ITU-T Q.6/SG 16) and ISO/IEC Moving Picture Experts Group (ISO/IEC JTC 1/SC 29/WG 11). The *Joint Collaborative Team on Video Coding* (JCT-VC) was established to work on this project. The scope of this group was extended to continue working on format range extensions (RExt), scalable HEVC (SHVC), and screen content coding (SCC) as

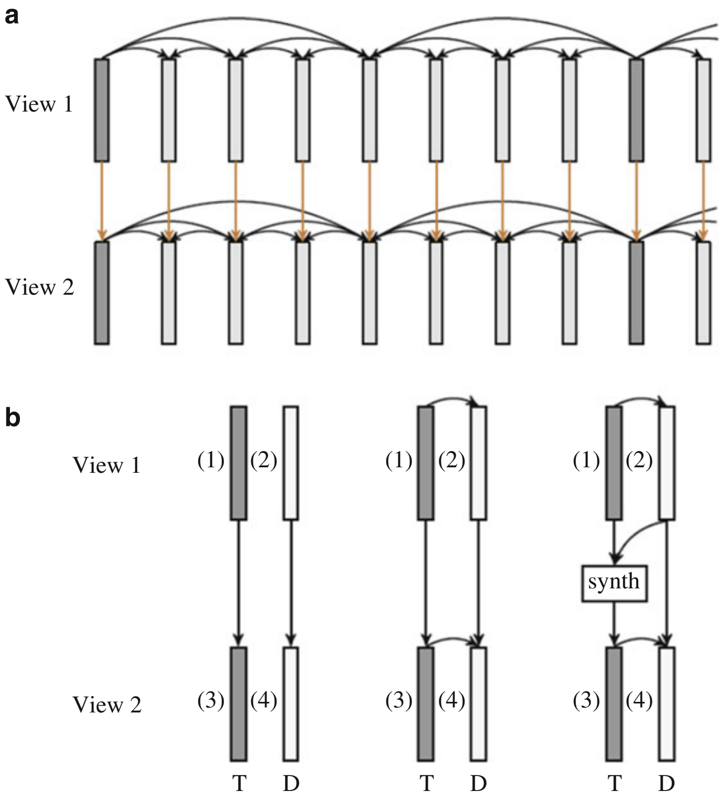


Fig. 4.9 Examples of 3D video prediction structures: (**a**) MVC inter-view prediction (view 1 is base/independent view), (**b**) MV-HEVC-independent coding of video texture (T) and depth maps (D), inter-component prediction within the same view, and BVSP view-synthesis prediction

extensions of HEVC. The *Joint Collaborative Team on 3D Video Coding Extension Development* (JCT-3V) was established to work on multiview and 3D video coding extensions of HEVC.

The main steps of HEVC technical developments are organized in four phases:

1. The HEVC first base specification finalized in 2013.
2. *Fidelity range extensions* (FRExt), *scalable video coding* (SHVC), and *multi-view video coding* (MV-HEVC) extensions finalized in 2014.
3. 3D video coding (3D-HEVC) extension finalized in 2015.
4. *SCC* extensions will be included in the fourth version of HEVC, which is expected to be finalized in 2016.

Where as the first three developments mainly targeted compression performance for consumer and professional uses, SHVC and MV/3D video coding have provided additional functionality such as variable rate access at the bit stream level and support for multiple camera inputs in combination with efficient compression.

After finalization of the HEVC base specification, JCT-VC continued to work on extensions.

1. The *format range extension* (RExt) provides tools to support 4:0:0, 4:2:2, and 4:4:4 color spaces and additional bit depths. RExt is included in the second version of HEVC, which has been finalized in October 2014.
2. Already during the initial phase of HEVC, *multi-layer extensions* were planned and the proper hooks were included into the base specification. The *scalability extension* of HEVC (SHVC) provides support for spatial, SNR, and color gamut scalability. It has been designed as a high-level syntax-only extension to allow reuse of existing decoder components. SHVC is included in the second version of HEVC, which has been finalized in October 2014.

   The JCT-3V was established to work on multiview and 3D video coding extensions of HEVC and other video coding standards. The *multiview extension* of HEVC (MV-HEVC) provides support for coding multiple views with inter-layer prediction. It was designed as a high-level syntax-only extension to allow reuse of existing decoder components. MV-HEVC is included in the second version of HEVC, which was finalized in October 2014.
3. The *3D extension* of HEVC (3D-HEVC) provides increased coding efficiency by *joint coding* of texture and depth for advanced 3D displays. 3D-HEVC is included in the third version of HEVC, which was finalized in February 2015.
4. The SCC extensions will improve compression capability for video containing a significant portion of rendered (moving or static) graphics, text, or animation rather than (or in addition to) camera-captured video scenes. Example applications include wireless displays, remote computer desktop access, and real-time screen sharing for videoconferencing. SCC will be included in the fourth version of HEVC, which is expected to be finalized in February 2016.

JCT-VC adopted an open standardization approach in the development of specifications. All inputs and contributions to a JCT-3V meeting are made by documents which are registered in a publicly accessible document repository. A set of deliverables, which turn to become normative or remain to be supplemental in their final form, are also publicly accessible. These comprise the **specification text** itself, the **reference software**, a **conformance specification**, and the **test model**. Furthermore, a **verification report** is produced which documents and demonstrates the achieved performance.

- **Draft specification** is developed as a *working draft document* or *draft amendment*, depending on the state of the working phase. Since this document represents the current state of the main deliverable of the group, it has highest priority. A new version of the draft text is released after every meeting, integrating the adoptions of the meeting. While the specification of the first version of 3D-HEVC has been finalized, ongoing JCT-3V work on maintenance and extensions is reflected in corresponding specification drafts. Depending on the scale of the introduced changes, they may be published as an *amendment* or as a *new edition* of the standard. While amendments only include the applicable changes and extensions of the specification, a new edition would imply the publication of a complete integrated version of the specification text.

- **Test model document** is maintained aligned with draft specification. In distinction from the original HM reference software for HEVC, the 3D-HEVC reference software is referred to as **HTM**. The text describes the encoder control and algorithms implemented in the reference software which implements the reference decoder and a rate-distortion optimized encoder. The document aids analysis of the reference software, including the integrated normative tools. By describing the encoder decisions for application of the specified tools, the test model text serves as a tutorial example on how to implement an encoder control for the tool set in the specification.
- **Reference software** implements the decoding process as specified in the (draft) specification and an example encoder which generates bit streams complying to the specification. A new version of the **HTM software** is released after every meeting, integrating the adoptions of the meeting. In the development phase, the reference software specifically serves as the platform to test and analyze proposed tool changes. *Simulations* which are performed using the reference software confirm the expected rate-distortion performance along the development of the specification draft. The software **reference decoder** can be used by encoder manufacturers for testing if their encoded bit streams comply to the specification. Since the reference software does not necessarily include all restrictions specified in the text, successful decoding of a bit stream by this software may give a good indication but not a final proof for compliance. The reference software is maintained and developed to meet the goal of compliance as closely as possible. The **reference encoder** provides a rate-distortion optimized implementation, which aids in comparing the performance impact of tools in the context of the reference model. It should be noted that the reference software commonly does not include sophisticated rate control for real encoding tasks nor does it include significant error concealment in the decoder in the case that, e.g., corrupted bit streams are fed to the decoder. Such tools are up to the implementers of encoders and decoders for their respective target applications.
- **Conformance specification** is developed to provide means to manufacturers of encoders and decoders to test their product for compliance to the specification text. An important means for conformance testing of decoders is a suite of conformance bit streams which are generated by JCT-3V. These bit streams are designed to include a test set as complete as possible for all tools included in the specification. With the approval of the final version of the specification text, the design task for the conformance specification is to approach completeness as much as possible.
- **Core experiment** is the regular process for a tool to be adopted into the draft specification. While the proponent reports the test performance results of the addition or modification of coding tool, the most important task is on the core experiment participants who provide a *cross-check* of the proposed technology. Conceptually, a successful core experiment can be considered as the last step before adopting a proposal into the draft specification. However, the successful

Table 4.1 Publication of ITU-T Rec. H.265 and ISO/IEC international standard MPEG-H

ITU-T Rec. H.265.1 *High efficiency video coding*	13.04.2013
Annex A *Profiles, tiers and levels* (3 profiles)	
Annex B *Byte stream format*	
Annex C *Hypothetical reference decoder*	
Annex D *Supplemental enhancement information*	
Annex E *Video usability information*	
ITU-T Rec. H.265.2 *Reference software for ITU-T H.265 high efficiency video coding*	10/2014
ITU-T Rec. H.265.3 *Conformance specification for ITU-T H.265 high efficiency video coding*	10/2014
ISO/IEC 23008-2 :2013 *High efficiency video coding*	01.12.2013
ISO/IEC 23008-5:2013 *Reference software for HEVC*	16.10.2013
ISO/IEC 23008-8:2013 *Conformance specification for HEVC*	16.10.2013
ITU-T Rec. H.265 *High efficiency video coding*	29.10.2014
Annex F *Common syntax, semantics and decoding process for multi-layer video coding extensions*	
Annex G *Multiview coding* (multiview main profile)	
Annex H *Scalable high efficiency video coding*	
H.265.2 *Reference software for HEVC v2* (*RExt, SHVC, MV-HEVC*)	10/2015
H.265.3 *Conformance specification for HEVC v2* (*RExt, SHVC, MV-HEVC*)	10/2015
ISO/IEC 23008-2 :2015 *High efficiency video coding*	01.05.2015
ISO/IEC 23008-5:2015 *Reference software for HEVC*	15.04.2015
ISO/IEC 23008-8:2015 *Conformance specification for HEVC*	15.04.2015
ITU-T Rec. H.265 *High efficiency video coding*	29.04.2015
Annex I *3D High efficiency video coding* (3D main profile)	
ITU-T Rec. H.265.2 *Reference software for HEVC v3* (*3D-HEVC*)	10/2015
ITU-T Rec. H.265.3 *Conformance specification for HEVC v3* (*3D-HEVC*)	10/2015

conduction of a core experiment does not imply guaranteed adoption. Studies of changes in structures above the coding layer (the high-level syntax) do not easily allow for verification of the benefit of proposed changes. In such cases, assessment by qualified experts is obligatory.

VCEG and MPEG jointly developed the three versions of high-efficiency video coding specifications and published Recommendation ITU-T H.265 and ISO/IEC 23008-2 International Standard in a technically aligned manner (Table 4.1):

- The first edition refers to the first approved 04/2013 version of this Recommendation | International Standard. *Annex A* specifies profiles, tiers, and levels as restrictions on the bit streams and hence limits on the capabilities needed to decode the bit streams.
- The second edition approved 10/2014 refers to the integrated text containing format range extensions, scalability extensions, multiview extensions, and additional supplement enhancement information. *Annex G* specifies syntax, semantics, and decoding processes for multiview high-efficiency video coding (MV-HEVC). This annex also specifies profiles (Multiview Main), tiers, and levels for multiview high-efficiency video coding.

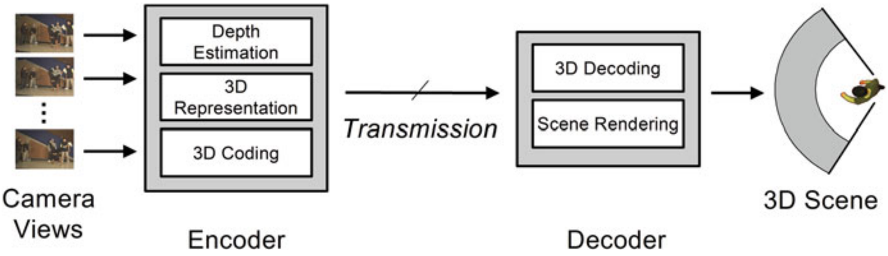


Fig. 4.10 Multiview + depth video processing chain

- The third edition approved 04/2015 refers to the integrated text containing 3D extensions. *Annex I* contains support for *3D high-efficiency video coding* (3D-HEVC), specifying a syntax and associated decoding process for efficient coding of video textures and depth maps for 3D video applications. One additional profile is defined in this revision, the 3D Main profile.

4.3.1 Competition Phase of Experimental Framework

Development of 3DV HEVC extensions is based on an experimental framework and multiview video-plus-depth (MVD) format. At the encoder side a real-world 3D scene is captured by multiple cameras, and a MVD representation is extracted from this input (Fig. 4.10). Once the **depth maps** are obtained, new views can be synthesized by interpolating the pixel values from nearby images. The *depth* of a 3D scene is expressed relative to the camera position or an origin in the 3D space. The *disparity* estimation is the correspondence between pixels in the left and right images. At the decoder side the decoder receives a coded representation (bit stream) of the data, which is then decoded and used for multiview rendering of the 3D scene.

The MPEG standardization adopted three steps of development-based formal subjective assessment of the 3D video quality:

- *Call for Evidence* (**CoE**) purpose is to explore *in house* whether the coding efficiency and 3DV functionality of the current version of HEVC standard can be further improved for MVD content.
- *Call for Proposals* (**CfP**) on 3D video coding technology is open to external parties (04/2011) with primary goal to define a 3DV data format and associated compression technology to enable the high-quality reconstruction of synthesized views for 3D displays. To evaluate the proposed technologies, formal subjective tests are performed. Results of these tests are made public (12/2011).
- *Verification tests* for 3D video coding technology include test conditions, evaluation methodology, and timeline to assess the improvement of the coding performance (10/2015) (Table 4.2).

Table 4.2 MPEG documents in competition phase of 3DV standardization

Doc. N10357 *Vision on 3D video*	Feb. 2009
Doc. N10359 *Call for 3D test material: depth maps & supplementary information*	Feb. 2009
Doc. N11631 *Report on experimental framework for 3D video coding*	Oct. 2010
Doc. N12035 *Applications and requirements on 3D video coding technology*	Mar. 2011
Doc. N12036 *Call for proposals on 3D video coding technology*	Mar. 2011
Doc. N12347 *Report of subjective test results from the call for proposals on 3D video coding tech*	Dec. 2011
Doc. N12348 *Overview of 3DV coding tools proposed in the CfP*	Dec. 2011

The *Call for Proposals on 3D video coding technology* represented the start of standardization of depth-based 3D formats, among which MVD was the first priority.

- In the CfP, two classes of test sequences (MVD format) were used as test materials. The individual sequences in each set were 8 or 10 s long.
- Two test scenarios were defined and refer to the 2-view input configuration and 3-view input configuration.
- Two test categories were defined in the CfP: AVC-compatible, and HEVC-compatible and unconstrained. For the AVC-compatible test, anchors for the objective and subjective measurements were generated using an MVC encoder (JMVC version 8.3.1) to encode the test sequences. For the HEVC compatibility test, anchors for the objective and subjective measurements were generated using an HEVC encoder (HM version 2.0) to encode the test sequences. For the AVC compatibility test, MVC was applied separately to texture data and depth data. For the HEVC compatibility test, HEVC simulcasting was used for each view of texture data and depth data. To calculate the objective rate-distortion (RD) performance and provide appropriate materials for subjective evaluation, four rate points (R1, R2, R3, and R4) were determined for each test sequence, for each test scenario, and for each test category.

Twenty-two proposals were submitted for the CfP. The submitted test materials were subjectively assessed in 12 test laboratories (18 naive viewers per test sequence) around the world. The subjective evaluations showed that, for most test sequences, the subjective quality of R3 of the best-performing proposal was better than R1 of the anchor. This suggests a significant improvement in coding efficiency compared to the anchor. In terms of objective performance, more than 25 and 55 % bitrate saving was reported by best proposals, Nokia in AVC test category and HHI in HEVC test category, respectively.

New coding tools proposed in CfP improve coding efficiency taking into account the unique *functionality* or *statistical* properties of depth data, as well as exploiting the *coherence* between texture and depth signals:

- **Texture-coding-dependent views that are independent of depth.** This involves coding the texture images of the side view. A **side view** is any view other than the first view in the coding order. The first view (also called the **base view**) is expected to be fully compatible with AVC or HEVC; the side view only uses inter-view texture information. Tools in this category include *motion parameter prediction and coding*, and *inter-view residual prediction*.

- **Texture-coding-dependent views that are dependent on depth.** This is applicable to side-view texture, in which original or reconstructed depth information is used to further exploit the correlation between texture images and associated depth maps. Tools in this category include *view synthesis prediction* for texture and *depth-assisted in-loop filtering* of texture.
- **Depth coding that is independent of texture.** Inter-view depth information or neighboring reconstructed depth values are used to compress the current macroblock in the depth map. Tools in this category include *depth intra-coding*, *synthesis-based inter-view prediction*, *inter-sample prediction*, and *in-loop filtering* for depth.
- **Depth coding that depends on texture.** Original or reconstructed texture information is used to further exploit the correlation between texture images and associated depth maps. Tools in this category include *prediction parameter coding*, *intra-sample prediction*, and *coding of quantization parameters*.
- **Encoder control optimization.** Tools in this category include *rate-distortion optimization* (RDO) techniques for depth, and texture encoding. They do not affect syntax or semantics.

4.3.2 Collaboration Phase of Experimental Framework

System structure of the best CfP proposals and coding tools from other proposals are included in the test model under consideration (TMuC) for HEVC-based 3D video coding. TMuC simulation software includes several applications and libraries for encoding, decoding, and view synthesis (Table 4.3).

The development of 3D extensions for HEVC and AVC is based on a set of core experiments (**CE**) that specifies tools under investigation and timeline of simulation and cross-check reports. Common test conditions (**CTC**) for 3DV experimentation specify test scenarios under consideration, test sequences, basic encoder configuration, and objective/subjective evaluation of visual quality (Table 4.4).

The standardization track of 3D extensions for AVC and HEVC is shown in Table 4.5.

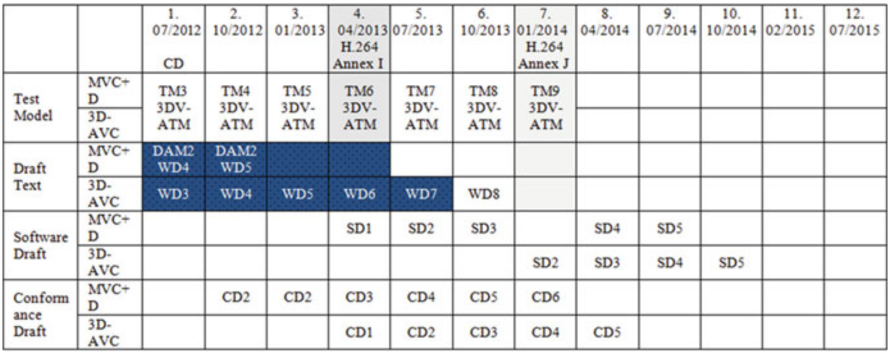
Table 4.3 MPEG documents in the start of collaboration phase of 3DV standardization

Doc. N12350 *Test model under consideration for HEVC based 3D video coding*	Dec. 2011
Doc. N12352 *Common test conditions for 3DV experimentation* (CTC)	Dec. 2011
Doc. N12353 *Description of core experiments in AVC based 3D video coding*	Dec. 2011
Doc. N12354 *Description of core experiments in HEVC based 3D video coding*	Dec. 2011
Doc. N12434 *Standardization tracks considered in 3D video coding*	Dec. 2011

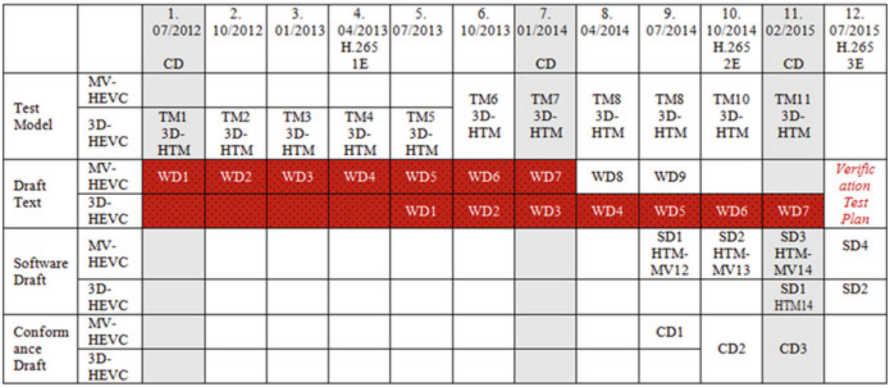
Table 4.4 Description of core experiments in AVC/HEVC 3D video coding (Dec. 2011)

CE	AVC compatible	HEVC compatible
CE1	View synthesis prediction for texture and depth	View synthesis-based prediction for texture
CE2	Depth-based prediction	View synthesis-based prediction for depth
CE3	Depth representation and coding	Motion parameter prediction and coding (independent of depth)
CE4	Depth intra-prediction without inter-component prediction	Transform coding for depth
CE5	Depth range compensation for inter/interview prediction	In-loop filtering for depth
CE6	In-loop depth resampling	Prediction parameter coding (motion data and intra-pred. mode)
CE7	RD optimization through view synthesis distortion	Coding of quantization parameters
CE8	Global depth-and-view prediction	Component extraction
CE9	Texture-based prediction for depth coding	Prediction structures for inter-view prediction
CE10	Depth in-loop filtering	Modified distortion measure for depth coding
CE11		View synthesis

Table 4.5 JCT-3V standardization track: (a) MVC+D (*multiview and depth video coding*), 3D-AVC (*multiview and depth video with enhanced non-base view coding*), (b) MV-HEVC (*multiview high-efficiency video coding*), 3D-HEVC (*3D high-efficiency video coding*)



		1. 07/2012 CD	2. 10/2012	3. 01/2013	4. 04/2013 H.264 Annex I	5. 07/2013	6. 10/2013	7. 01/2014 H.264 Annex J	8. 04/2014	9. 07/2014	10. 10/2014	11. 02/2015	12. 07/2015
Test Model	MVC+D	TM3 3DV-ATM	TM4 3DV-ATM	TM5 3DV-ATM	TM6 3DV-ATM	TM7 3DV-ATM	TM8 3DV-ATM	TM9 3DV-ATM					
	3D-AVC												
Draft Text	MVC+D	DAM2 WD4	DAM2 WD5										
	3D-AVC	WD3	WD4	WD5	WD6	WD7	WD8						
Software Draft	MVC+D				SD1	SD2	SD3		SD4	SD5			
	3D-AVC							SD2	SD3	SD4	SD5		
Conformance Draft	MVC+D		CD2	CD2	CD3	CD4	CD5	CD6					
	3D-AVC				CD1	CD2	CD3	CD4	CD5				



		1. 07/2012 CD	2. 10/2012	3. 01/2013	4. 04/2013 H.265 1E	5. 07/2013	6. 10/2013	7. 01/2014 CD	8. 04/2014	9. 07/2014	10. 10/2014 H.265 2E	11. 02/2015 CD	12. 07/2015 H.265 3E
Test Model	MV-HEVC						TM6 3D-HTM	TM7 3D-HTM	TM8 3D-HTM	TM8 3D-HTM	TM10 3D-HTM	TM11 3D-HTM	
	3D-HEVC	TM1 3D-HTM	TM2 3D-HTM	TM3 3D-HTM	TM4 3D-HTM	TM5 3D-HTM							
Draft Text	MV-HEVC	WD1	WD2	WD3	WD4	WD5	WD6	WD7	WD8	WD9			*Verification Test Plan*
	3D-HEVC					WD1	WD2	WD3	WD4	WD5	WD6	WD7	
Software Draft	MV-HEVC									SD1 HTM-MV12	SD2 HTM-MV13	SD3 HTM-MV14	SD4
	3D-HEVC											SD1 HTM14	SD2
Conformance Draft	MV-HEVC									CD1	CD2	CD3	
	3D-HEVC												

- MVC-compatible extension including depth MVC+D (no block-level changes to AVC/MVC syntax and decoding processes; add high-level syntax to enable efficient coding of depth data), FDAM 10/2012 (Final Draft Amendment).
- AVC-compatible extension plus depth 3D-AVC (change syntax and decoding process for non-base texture view and depth maps at block level), FDAM 07/2013.
- HEVC 3D extensions: MV-HEVC multiview extension (no change to the CU-level syntax, semantics, and decoding processes of HEVC), and 3D-HEVC (advanced multiview and 3D extension for higher compression efficiency by jointly compressing texture and depth data).

JCT-3V group developed a new data format and associated compression technology to enable the high-quality reconstruction of synthesized views for 3D displays in HEVC-based coding frameworks. As part of this work, two amendments of the HEVC standard have been developed as outlined below.

- Multiview extension (**MV-HEVC**): The main target of this extension is to enable coding multiview video sequences. Depth maps can be coupled with multiview video stream using auxiliary pictures, which are one of the features in the range extension of HEVC. There are no change to the CU-level syntax, semantics, and decoding processes of HEVC. The specification of this extension (ISO/IEC 23008-2:201x) has included in the second edition of HEVC, which has reached FDIS status in October 2014.

- 3D video extension (**3D-HEVC**): This extension has been developed that aims for higher compression efficiency by jointly compressing texture and depth data. The specification of this extension (ISO/IEC 23008-2:2013/Amd.4) has reached FDAM status in February 2015.

As the standardization of both specifications is completed, verification tests are planned to assess the improvement of the coding performance. MV-HEVC is planned to be compared with simulcast coding of HEVC as well as MVC in terms of stereo video coding. 3D-HEVC will be compared to MV-HEVC. The test conditions and evaluation procedure are based on CTC common test conditions for 3DV experimentation.

The timeline in verification test plan is as follows [Doc. N15441 July 2015]:

- Preparing viewing materials and bit streams with various bit rates (07/2015).
- Decide target bit rate for testing, perform expert viewing test with at least nine experts, and prepare the report (10/2015).

4.3.3 An Overview of 3D Video Coding Tools

4.3.3.1 MV-HEVC Coding Tools

MV-HEVC specification follows the same design principles of the MVC extension in the AVC framework. The coding schemes enable *inter-view prediction* based on **disparity-compensated prediction** (Fig. 4.11)**.** A block-based disparity shift between the reference view and the current views is determined and used in prediction. This is similar to the motion-compensated prediction used in conventional video coding, but it is based on pictures with different viewpoints rather than pictures at different time instances. MV-HEVC extends the high-level syntax so that the appropriate signaling of view identifiers and their references is supported and defines a process by which decoded pictures of other views can be used to predict a current picture in another view.

In order to support depth map coding, MV-HEVC enables *auxiliary picture* syntax. The auxiliary picture decoding process would be the same for video or multiview video. In AVC framework, an *independent* second stream is specified for the representation of depth as well as high-level syntax signaling of the necessary information to express the interpretation of the depth data and its association with the video data. This approach does not involve macroblock-level changes to the AVC or MVC syntax, semantics, and decoding processes. The corresponding 3D video codec is referred to as **MVC + D**.

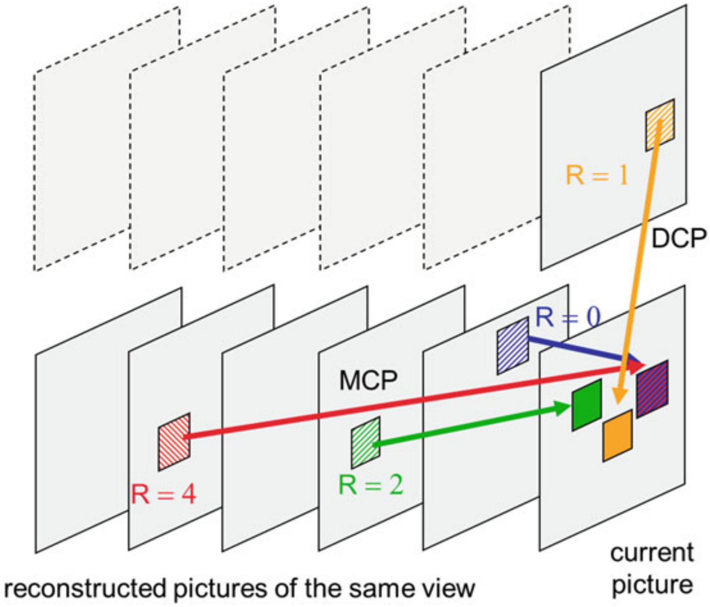


Fig. 4.11 Disparity-compensated prediction (DCP) as an alternative to motion-compensated prediction (MCP)

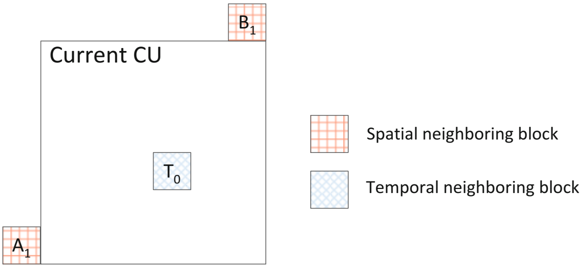
4.3.3.2 3D-HEVC Coding Tools for Texture

To achieve higher coding efficiency, researchers have studied and evaluated advanced coding tools that better exploit inter-view redundancy. In contrast to MV-HEVC, block-level changes to the syntax and decoding process are considered to maximize the possible coding gain. In the AVC framework, the **3D-AVC** extension supports new block-level coding tools for texture views.

Neighboring block-based disparity vector derivation (NBDV): This tool derives a disparity vector for a current block using an available disparity motion vector of spatial and temporal neighboring blocks. The derivation principle is the same in both 3D-AVC and 3D-HEVC, but the location of neighboring blocks differs slightly (Fig. 4.12). The main benefit of this technique is that disparity vectors to be used for inter-view prediction can be directly derived without additional bits and independent of an associated depth picture. Disparity information can also be derived from the decoded depth picture when camera parameters are available.

Inter-view motion prediction: The motion information between views exhibits a high degree of correlation, and inferring it from one view to another view leads to notable gains in coding efficiency because good predictions generally reduce the bit rate required to send such information. To achieve this, the disparity, such as that derived by the NBDV process, is used to establish a correspondence between the blocks in each view. The concept of inter-view motion prediction is supported

Fig. 4.12 3D-HEVC location of spatial and temporal neighbor blocks



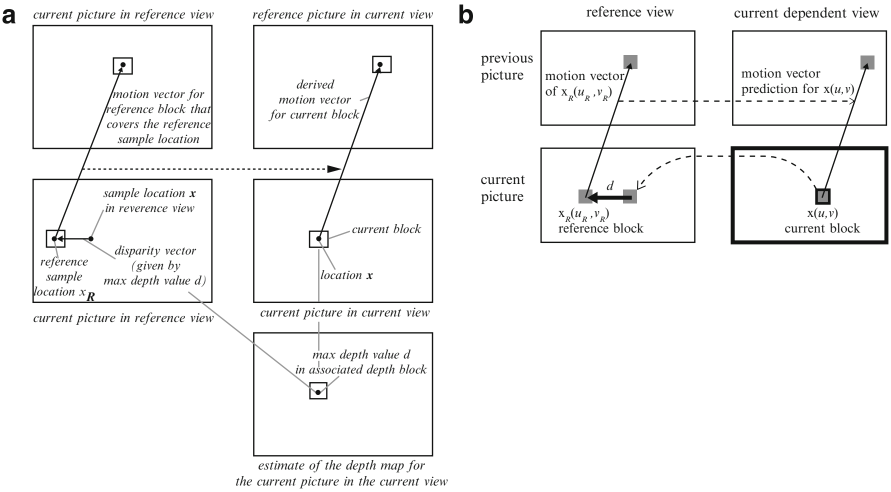


Fig. 4.13 (**a**) Basic principle of deriving motion parameters for a block in a current picture based on motion parameters in an already coded reference view and an estimate of the depth map for the current picture. (**b**) Motion vector correspondences between a block in a current picture of a dependent view and an already coded reference view, using the disparity vector *d* from a depth map estimate

in both the 3D-AVC and 3D-HEVC, but the designs differ. In 3D-AVC, interview motion prediction is realized with a new prediction mode, whereas in 3D-HEVC, it is realized by leveraging the syntax and decoding processes of the merge and advance motion vector prediction (AMVP) modes that were newly introduced by the HEVC standard (Fig. 4.13).

View synthesis prediction (VSP): This tool uses the depth information to warp texture data from a reference view to the current view in order to generate a predictor for the current view. Although depth is often available with pixel-level precision, a block-based VSP scheme has been specified in both 3D-AVC and 3D-HEVC to align this type of prediction with existing modules for motion compensation. To perform VSP, the depth information of the current block is used to determine the corresponding pixels in the inter-view reference picture (Fig. 4.14). Because texture is typically coded prior to depth, the depth of the current block can be

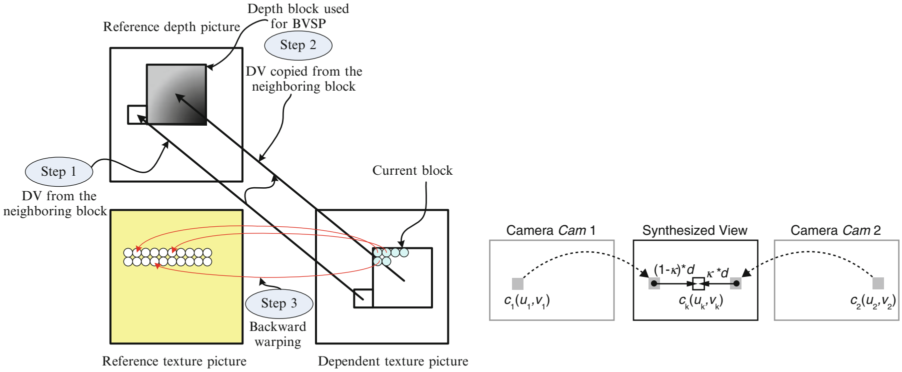
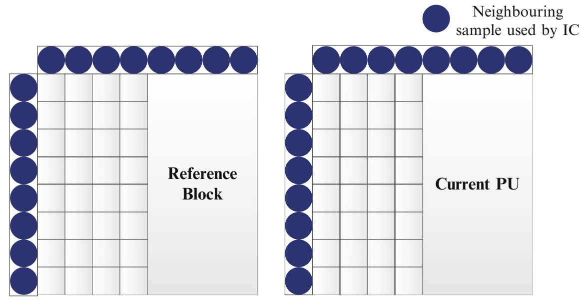


Fig. 4.14 (**a**) Illustration of the VSP scheme with the neighboring block disparity vector (DV), (**b**) view synthesis principle with horizontal disparity-based shift from original data (Cam1, Cam2) to new position in synthesized view

Fig. 4.15 Neighboring samples for the derivation of illumination compensation parameters



Reference block and its neighbouring samples in the reference view identified by a disparity vector

Current CU and its neighbouring samples

estimated using the NBDV process. In 3D-AVC, it is also possible to code depth prior to texture and hence obtain the depth information directly. As with inter-view motion prediction, the same VSP concept is supported in both 3D-AVC and 3D-HEVC, but the designs differ significantly. VSP is supported in 3D-AVC with a high-level *syntax flag* that determines whether the reference picture to be used for prediction is an inter-view reference picture or a synthesized reference picture as well as a low-level *syntax flag* to indicate when skip/direct mode prediction is relative to a synthesized reference picture. In 3D-HEVC, the VSP design is realized by extensions of the merge mode, whereby the disparity and inter-view reference picture corresponding to the VSP operation are added to the merge candidate list.

Illumination compensation (IC): This tool improves the coding efficiency for blocks predicted from inter-view reference pictures in case when prediction fails due to not calibrated cameras capturing the same scene or by lighting effects. This mode only applies to blocks that are predicted by an interview reference picture (Fig. 4.15).

Inter-view residual prediction: Advanced residual prediction (**ARP**) takes advantage of the correlation between the motion-compensated residual signal of two views. ARP mode only supported in 3D-HEVC increases the accuracy of the residual predictor. In ARP, the motion vector is aligned for the current block and the reference block, so the similarity between the residual predictor and the residual signal of the current block is much higher, and the remaining energy after ARP is significantly reduced. Two types of ARP designs exist: temporal ARP and inter-view ARP. In temporal ARP, the residual predictor is calculated as a difference between the reference block (Base) and its reference block (BaseRef). With inter-view ARP, an inter-view residual is calculated from the temporal reference block in a different view (BaseRef) and its inter-view reference block, hypothetically generated by the disparity (DMV) that is signaled for the current block (Fig. 4.16).

4.3.3.3 3D-HEVC Coding Tools for Depth Maps

To achieve higher compression efficiency, new coding tools have been adopted in 3D-HEVC for the coding of depth views. Depth views in 3D-AVC are coded similar to MVC+D, and no block-level changes for depth coding have been introduced.

Depth motion prediction: Similar to motion prediction in texture coding, depth motion prediction is achieved by adding new candidates into the merge candidate list. The additional candidates include interview merge candidate, subblock motion parameter inheritance candidate, and disparity-derived depth candidate.

Partition-based depth intra coding: To better represent the particular characteristics of depth, each depth block may be geometrically partitioned and more efficiently represented. In 3D-HEVC, these *nonrectangular* partitions are collectively referred to as depth modeling modes (**DMMs**), e.g., only coding the average value or predicting a planar function from already coded neighboring blocks without residual data. Two types of partitioning patterns are applied: wedglet pattern, which segments the depth block with a straight line, and contour pattern, which can support two irregular partitions.

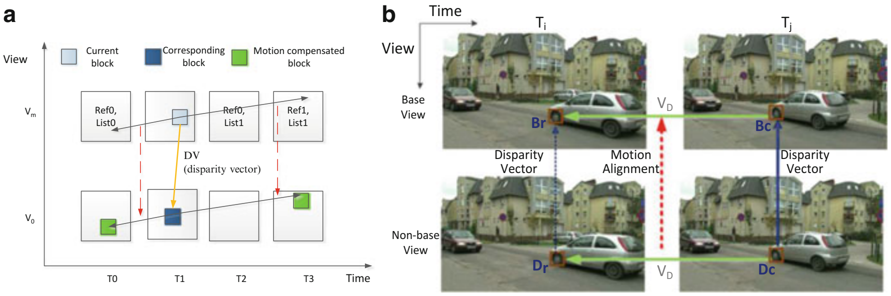


Fig. 4.16 (**a**) Relationship among current block, reference block, and motion-compensated block, (**b**) prediction structure of advanced residual prediction

Segment-wise DC coding (SDC): This coding mode enables the transform and quantization process to be skipped so that depth prediction residuals are directly coded. It also supports a depth look-up table **(DLT)** to convert the depth values to a reduced dynamic range. SDC can be applied to both intra- and inter-prediction, including the new DMM modes. When the SDC mode is applied, only one DC predictor is derived for each partition, and based on that, only one DC difference value is coded as the residual for the whole partition.

4.4 3D-HEVC Efficiency in Joint Coding-Dependent Views and Depth Data

3D-HEVC enables application that requires a high compression efficiency, such as transmitting 3D 4K content for stereoscopic as well as auto-stereoscopic multiview displays. 3D-HEVC extension targets multiview video and depth data coding with the best coding performance. To evaluate the compression efficiency of coding tools, simulations were conducted using the reference software and experimental evaluation methodology (Fig. 4.17). In the experimental framework, multiview video and corresponding depth are provided as input, while the decoded views and additional views synthesized at selected positions are generated as output. Common test conditions (CTC) for experimentation specify basic encoder configuration, and objective/subjective evaluation of decoded/synthesized views.

New 3D-HEVC added tools for joint coding the dependent views and depth data can be clustered, according to their redundancy reduction principles: **inter-view prediction** under consideration of depth, as well as **inter-component prediction** between texture and depth pictures.

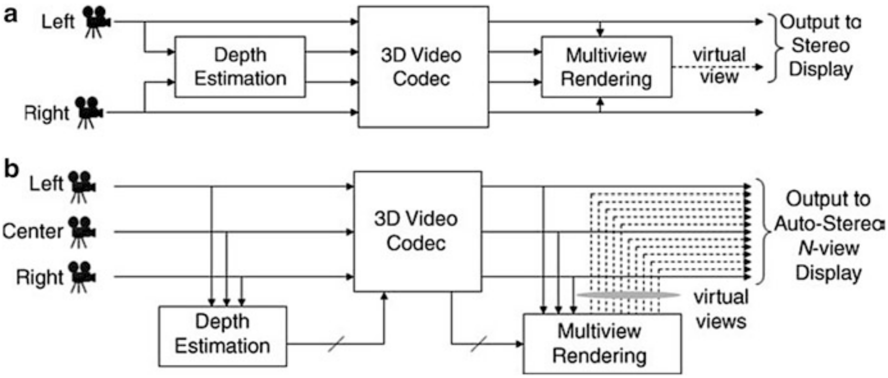


Fig. 4.17 (**a**) Advanced stereoscopic processing with two-view configuration, (**b**) auto-stereoscopic output with three-view configuration

Inter-view prediction: Similar to the compression of dependent views in MV-HEVC, the redundancy reduction across different views is one of the most important aspects for efficient coding. In addition to ***disparity-compensated prediction***, 3D-HEVC uses further tools for inter-view prediction. The first tool is ***view synthesis prediction***, which uses depth-based rendering to warp pixels from a reference view to a dependent view, while DCP uses one linear vector for a block. The second tool is inter-view ***motion parameter prediction***. Also, motion vectors for the same content in the different views can be similar, such that they can be predicted across views, using again the depth/disparity information. Third, inter-view ***residual prediction*** is used. Again, also the residual data in different views is similar for a certain amount of blocks, such that prediction across views can gain coding efficiency.

Inter-component prediction: Coding tools for reducing redundancies between the video and *co-located* depth component of each view were also developed for 3D-HEVC. One ***depth coding mode*** **DMM4** uses texture information for depth coding. Next, the ***motion parameter inheritance*** checks the partitioning and motion data from the texture information, whether it can be used for efficient coding of the current depth block. Also, tools for ***block partitioning prediction*** can be applied, e.g., quadtree prediction, where subdivision information of the texture is used to restrict the subdivision of a co-located depth block. This assumes that the texture is finer structured than depth, such that a depth block is never subdivided further than the texture.

Encoder control: 3D-HEVC uses a **joint rate-distortion optimization (JRDO)** for the depth data. For video data, the classical rate-distortion optimization (RDO) is used, when the optimal coding mode is sought. Here, the *Lagrangian* cost function is used, a weighted sum of video rate, and video distortion in terms of mean squared error (MSE) between original and reconstructed video data. In contrast, reconstructed depth maps are only used for the synthesis of intermediate views and not directly viewed. Therefore, the coding efficiency in 3D-HEVC is improved by applying a cost function that considers the distortion in synthesized intermediate views. This **view synthesis optimization** (VSO) modified the distortion measure for the mode decision process for depth maps in a way that a weighted average of the synthesized view distortion and the depth map distortion. To obtain a measure of the synthesized view distortion, two different metrics are applied in JRDO. The distortion measurement is designed based on the fact that the same depth distortions generally cause higher synthesis errors in highly textured regions than in textureless regions.

The results obtained showed that a 3D-HEVC achieved higher coding efficiency by ***optimizing*** existing coding tools and ***adding*** new methods. In particular, an improved inter-view prediction, new methods of inter-component parameter prediction, special depth coding modes, and an encoder optimization for depth data coding towards the synthesized views were applied for optimally encoding 3DV data and synthesizing multiview video data for different 3D displays from the decoded bit stream.

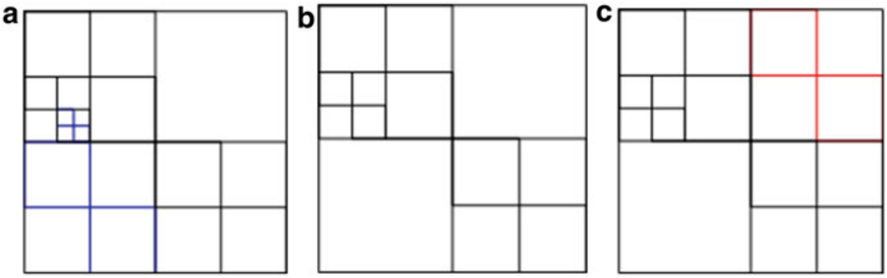


Fig. 4.18 (**a**) Example of a CTB QT partitioning for the texture, (**b**) allowed, and (**c**) disallowed collocated depth CTB QT partitioning

However, ***complexity reduction*** of an encoder is becoming a critical problem in implementations for specific application. The improvement of the 3D-HEVC coding efficiency is obtained at the expense of a computational complexity increase. The most computationally intensive parts are test all the coding modes and computation of the RD cost in recursive splitting of coding units.

Depth quadtree limitation: This tool prevents the encoder from making full investigation of every possible QT configuration for the depth coding. Based on RD optimized decisions, a given CTB is split into smaller CUs in the encoding process. A corresponding quadtree (QT) is obtained for the texture, and another one for the depth. The tool forces the encoder to limit the partitioning of the depth at the same level as the partitioning of the texture. For a given CTU, the quadtree of the depth is linked to the collocated CTB quadtree in the texture, so that a given CU of the depth cannot be split more than its collocated CU in the texture (Fig. 4.18).

Early decision algorithms: Two algorithms accelerate encoder decision by exploiting inter-view correlations in dependent texture view coding: early merge mode decision algorithm, and early CU splitting termination algorithm. Experimental results show that the proposed algorithm can achieve 47.1 % encoding timesaving with overall 0.1 % BD-rate reduction compared to 3D-HEVC test model version 7 under the common test condition (CTC). Both of the two strategies have been adopted into the 3D-HEVC reference software and enabled as a default encoding process under CTC.

4.5 Conclusion

Development of 3D video technologies is a challenging task. The current status is maturing of standardized 3D HEVC extensions and associated MVD formats. Current research issues are operational optimization of reference encoder configuration and performance improvements based on scalability extensions. New standardization activity is the next-generation video coding beyond HEVC for support of advanced 3D holoscopic representation beyond binocular cues.

Bibliography

Books

1. Assuncao P, Pinto L, Faria S (2014) Chapter 2: 3D media representation and coding. In: Kondoz A, Dagiuklas T (eds) 3D Future internet media, Springer, pp 9–38
2. Faria SMM, Debono CJ, Nunes P, Rodrigues NMM (2015) Chapter3: 3D video representation and coding. In: Kondoz A, Dagiuklas T (eds) Novel 3D media technologies. Springer, pp. 25–48
3. Kondoz A, Dagiuklas T (eds) (2016) Connected 3D media in the Internet era, Springer
4. Rao KR, Bojković Z, Milovanović D (2002) Multimedia communication systems: techniques, standards and networks. Prentice Hall PTR (Pearson Education)
5. Rao KR, Bojković Z, Milovanović D (2005) Introduction to multimedia communication: applications, middleware, networking. Wiley
6. Leonardo C (ed) (2012) The MPEG representation of digital media. Springer
7. Wien M (2015) High efficiency video coding: coding tools and specification. Springer Signals Commun Technol. (Chapter 12 extensions to HEVC, pp 291–308)
8. Möller S, Raake A (2014) Quality of experience: advanced concepts, applications and methods. Springer. (Lebreton P, Barkowsky M, Raake A, Le Callet P. Chapter 20: 3D video, pp 299–313)
9. Zhu C, Zhao Y, Yu L, Tanimoto M (2013) 3D-TV system with depth-image-based rendering: architectures, techniques and challenges, Springer. (Müller K, Merkle P, Tech G. Chapter 8: 3D video compression, pp 223–248)
10. Dufaux F, Pesquet-Popesu B, Cagnazzo M (eds) (2013) Emerging technologies for 3D video: creation, coding, transmission and rendering. Wiley. (Cagnazzo M, Pesquet-Popescu B, Dufaux F. Chapter 6: 3D video representation and formats, pp 102–120) (Vetro A, Müller K. Chapter 8: depth-based 3D video formats and coding technology, pp 139–161)
11. Zatt B, Shafique M, Bampi S, Henkel J (2013) 3D video coding for embedded devices: energy efficient algorithms and architectures. Springer
12. Ozaktas HM, Onural L (eds) (2007) Three-dimensional television: capture, transmission, and display. Springer. (Smolic A, Merkle P, Müller K, Fehn C, Kauff P, Wiegand T. Chapter 9: compression of multi-view video and associated data, pp 313–350)
13. Schreer O, Kauff P, Sikora T (eds) (2005) 3D video communication algorithms, concepts and real-time systems in human centered communication. Wiley. (Smolic A, Sikora T. Chapter 11: coding and standardization, pp 193–216)

Journals

14. (2014) Advances in 3D video processing. J Vis Comm Image Represent 25(4)
15. (2014) Special issue on QoE in 2D/3D video systems. J Vis Comm Image Represent 25(3)
16. (2013) Special section on 3D video representation, compression, and rendering. IEEE Trans Image Process 22(9)
17. (2011) Special issue 3D media and displays. Proc IEEE 99(4)
18. (2007) Special issue on 3DTV. Signal Process Image Comm. Elsevier
19. (2007) Special issue in Multiview imaging and 3DTV. IEEE Signal Process Mag
20. (2007) Special issue on Multiview video coding and 3DTV. IEEE Tran CSVT
21. (2008) Special issue on 3DTV: capture, transmission and display of 3D video. EURASIP J Adv Signal Process

B1. Introduction

22. Sullivan G, Wiegand T (2005) Video compression—from concepts to H.264/AVC standard. Proc IEEE 93(1):18–31
23. ISO/IEC JTC1/SC29/WG11 Doc. N13364 (2013) White paper on state of the art in compression and transmission of 3D video. MPEG 103. Meeting, Geneva CH

B2. Three-Dimensional Video Formats and Associated Compression Technology

24. Müller K, Merkle P, Wiegand T (2011) 3-D video representation using depth maps. Proc IEEE 99(4):643–656
25. Vetro A, Wiegand T, Sullivan G (2011) Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. Proc IEEE 99(4):626–642
26. Wiegand T, Sullivan GJ, Bjøntegaard G, Luthra A (2003) Overview of the H.264/AVC video coding standard. IEEE Trans Circ Syst Video Technol 13(7):560–576
27. Sullivan G, Ohm J-R, Han W-J, Wiegand T (2012) Overview of the high efficiency video coding (HEVC) standard. IEEE Trans CSVT 22(12):1649–1668
28. Ohm JR, Sullivan GJ, Schwarz H, Tan TK, Wiegand T (2012) Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC). IEEE Trans Circ Syst Video Technol 22(12):1669–1684
29. Bossen F, Bros B, Suhring K, Flynn D (2012) HEVC complexity and implementation analysis. IEEE Trans Circ Syst Video Technol 22(12):1685–1696
30. Sullivan GJ, Boyce J, Chen Y, Ohm J-R (2013) Standardized extensions of high efficiency video coding (HEVC). IEEE J Selected Topics Signal Process 7:1001–1016

B3. HEVC Standardization Framework

31. Chen Y, Vetro A (2014) Next-generation 3D formats with depth map support. IEEE Multimedia 21(2):90–94
32. Hannuksela MM, Rusanovskyy D, Su W, Chen L, Li R, Aflaki P, Lan D, Joachimiak M, Li H, Gabbouj M (2013) Multiview-video-plus-depth coding based on the advanced video coding standard. IEEE Trans Image Process 22(9):3449–3458
33. Müller K, Schwarz H, Marpe D, Bartnik C, Bosse S, Brust H, Hinz T, Lakshman H, Merkle P, Rhee FH, Tech G, Winken M, Wiegand T (2013) 3D high-efficiency video coding for multiview video and depth data. IEEE Trans Image Process 22(9):3366–3378
34. Domanski M, Stankiewicz O, Wegner K, Kurc M, Konieczny J, Siast J, Stankowski J, Ratajczak R, Grajek T (2013) High efficiency 3D video coding using new tools based on view synthesis. IEEE Trans Image Process 22(9):3517–3527

B4. 3D-HEVC Efficiency in Joint Coding of Dependents View and Depth Data

35. Müller K (2014) 3D extensions for high-efficiency video coding. IEEE Comsoc MMTC E-Lett 9(1)
36. Zhang N, Zhao D, Chen Y-W, Lin J-L, Gao W (2014) Fast encoder decision for texture coding in 3D-HEVC. Signal Process Image Commun 29(9):951–961
37. Zhang Y, Kwong S, Xu L, Hu S, Jiang G, Kuo C-CJ (2013) Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model. IEEE Trans Image Process 22(9):3497–3512