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

3D-TV Content Storage and Transmission

Vetro, A.; Tourapis, A.M.; Muller, K.; Chen, T.

TR2011-023 January 2011

Abstract

There exist a variety of ways to represent 3D content, including stereo and multiview video, as well as frame compatible and depth-based video formats. There are also a number of compression architectures and techniques that have been introduced in recent years. This paper provides an overview of relevant 3D representation and compression formats. It also analyzes some of the merits and drawbacks of these formats considering the application requirements and constraints imposed by different storage and transmission systems.

IEEE Transactions on Broadcasting

3D-TV Content Storage and Transmission

Anthony Vetro, *Fellow, IEEE*, Alexis M. Tourapis, *Senior Member, IEEE*, Karsten Müller, *Senior Member, IEEE*, and Tao Chen, *Member, IEEE*

Abstract—There exist a variety of ways to represent 3D content, including stereo and multiview video, as well as frame-compatible and depth-based video formats. There are also a number of compression architectures and techniques that have been introduced in recent years. This paper provides an overview of relevant 3D representation and compression formats. It also analyzes some of the merits and drawbacks of these formats considering the application requirements and constraints imposed by different storage and transmission systems.

***Index Terms*—3D video, compression, depth, digital television, frame-compatible, multiview, stereo.**

I. Introduction

It has recently become feasible to offer a compelling 3D video experience on consumer electronics platforms due to advances in display technology, signal processing, and circuit design. Production of 3D content and consumer interest in 3D has been steadily increasing, and we are now witnessing a global roll-out of services and equipment to support 3D video through packaged media such as Blu-ray Disc and through other broadcast channels such as cable, terrestrial channels, and the Internet.

A central issue in the storage and transmission of 3D content is the representation format and compression technology that is utilized. A number of factors must be considered in the selection of a distribution format. These factors include available storage capacity or bandwidth, player and receiver capabilities, backward compatibility, minimum acceptable quality, and provisioning for future services. Each distribution path to the home has its own unique requirements. This paper will review the available options for 3D content representation and coding, and discuss their use and applicability in several distribution channels of interest.

The rest of this paper is organized as follows. Section II describes 3D representation formats. Section III describes various architectures and techniques to compress these different representation formats, with performance evaluation given in Section IV. In Section V, the distribution of 3D content through packaged media and transmission will be discussed. Concluding remarks are provided in Section VI.

Manuscript received October 1, 2010.

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

A.M. Tourapis is with Dolby Laboratories, Burbank, CA 91505 USA (email: alexis.tourapis@dolby.com).

K. Müller is with the Fraunhofer Institute for Telecommunications – Heinrich Hertz Institute (HHI), Einsteinufer 37, 10587 Berlin, Germany (email: Karsten.Mueller@hhi.fraunhofer.de).

T. Chen is with Panasonic Hollywood Labs, Universal City, CA 91608 USA (e-mail: chent@research.panasonic.com).

II. 3D Representation Formats

This section describes the various representation formats for 3D video and discusses the merits and limitations of each in the context of stereo and multiview systems. A comparative analysis of these different formats is provided.

A. Full-Resolution Stereo and Multiview Representations

Stereo and multiview videos are typically acquired at common HD resolutions (e.g., 1920x1080 or 1280x720) for a distinct set of viewpoints. In this paper, we refer to such video signals as full-resolution formats. Full-resolution multiview representations can be considered as a reference relative to representation formats that have a reduced spatial or temporal resolution, e.g., to satisfy distribution constraints, or representation formats that have a reduced view resolution, e.g., due to production constraints. It is noted that there are certain cameras that capture left and right images at half of the typical HD resolutions. Such video would not be considered full-resolution for the purpose of this paper.

In the case of stereo, the full-resolution representation (Fig. 1) basically doubles the raw data rate of conventional single view video. For multiview, there is an N-fold increase in the raw data rate for N-view video. Efficient compression of such data is a key issue and will be discussed further in Section III.B.

B. Frame-Compatible Representations

To facilitate the introduction of stereoscopic services through the existing infrastructure and equipment, frame-compatible formats have been introduced. With such formats, the stereo signal is essentially a multiplex of the two views into a single frame or sequence of frames. Typically, the left and right views are sub-sampled and interleaved into a single frame.

There are a variety of options for both the sub-sampling and interleaving. For instance, the two views may be filtered and decimated horizontally or vertically and stored in a side-by-side or top-and-bottom format, respectively. Temporal multiplexing is also possible. In this way, the left and right views would be interleaved as alternating frames or fields. These formats are often referred to as frame sequential and field sequential. The frame rate of each view may be reduced so that the amount of data is equivalent to that of a single view.

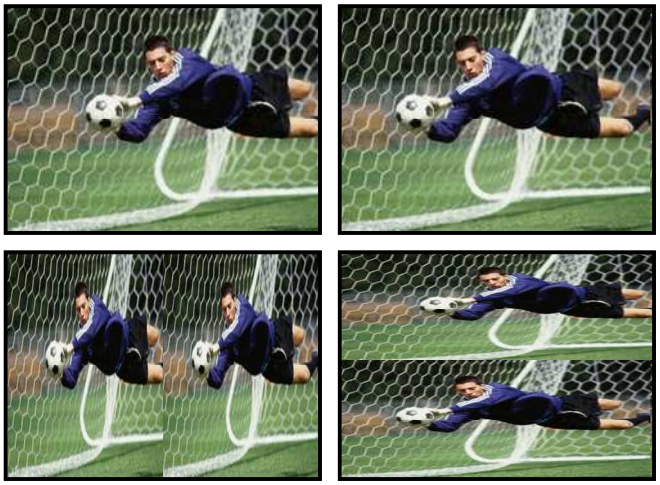

Fig. 1: Full Resolution (top) and Frame Compatible (bottom) representations of stereoscopic videos.

Frame-compatible video formats can be compressed with existing encoders, transmitted through existing channels, and decoded by existing receivers and players. This format essentially tunnels the stereo video through existing hardware and delivery channels. Due to these minimal changes, stereo services can be quickly deployed to capable displays, which are already in the market. The corresponding signaling that describes the particular arrangement and other attributes of a frame-compatible format are discussed further in Section III.A.

The obvious drawback of representing the stereo signal in this way is that spatial or temporal resolution would be lost. However, the impact on the 3D perception may be limited and acceptable for initial services. Techniques to extend frame-compatible video formats to full resolution have also recently been presented [13], [14] and are briefly reviewed in section III.B.

C. Depth-based Representations

Depth-based representations are another important class of 3D formats. As described by several researchers [15]-[17], depth-based formats enable the generation of virtual views through depth-based image rendering (DBIR) techniques. The depth information may be extracted from a stereo pair by solving for stereo correspondences [18] or obtained directly through special range cameras [19]; it may also be an inherent part of the content, such as with computer generated imagery. These formats are attractive since the inclusion of depth enables a display-independent solution for 3D that supports generation of an increased number of views, which may be required by different 3D displays. In principle, this format is able to support both stereo and multiview displays, and also allows adjustment of depth perception in stereo displays according to viewing characteristics such as display size and viewing distance.

ISO/IEC 23002-3 (also referred to as MPEG-C Part 3) specifies the representation of auxiliary video and supplemental information. In particular, it enables signaling for depth map streams to support 3D video applications. Specifically, the well-known 2D plus depth format as illustrated in Fig. 2 is specified by this standard. It is noted that this standard does not specify the means by which the depth information is coded, nor does it specify the means by which the 2D video is coded. In this way, backward compatibility to legacy devices can be provided.

Fig. 2: 2D plus depth representation.

The main drawback of the 2D plus depth format is that it is only capable of rendering a limited depth range and was not specifically designed to handle occlusions. Also, stereo signals are not easily accessible by this format, i.e., receivers would be required to generate the second view to drive a stereo display, which is not the convention in existing displays.

To overcome the drawbacks of the 2D plus depth format, while still maintaining some of its key merits, MPEG is now in the process of exploring alternative representation formats and is considering a new phase of standardization. The targets of this new initiative are discussed in [20]. The objectives are:

- Enable 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 so that the depth perception experienced by the viewer is within a comfortable range. Such a feature could help to avoid fatigue and other viewing discomforts.
- Facilitate support for high-quality auto-stereoscopic displays. Since directly providing all the necessary views for these displays is not practical due to production and transmission constraints, the new format aims to enable the generation of many high-quality views from a limited amount of input data, e.g. stereo and depth.

A key feature of this new 3D video (3DV) data format is to decouple the content creation from the display requirements, while still working within the constraints imposed by production and transmission. The 3DV format aims to enhance 3D rendering capabilities beyond 2D plus depth. Also, this new format should substantially reduce the rate requirements relative to sending multiple views directly. These requirements are outlined in [21].

III. 3D Compression Formats

The different coding formats that are being deployed or are under development for storage and transmission systems are reviewed in this section. This includes formats that make use of existing 2D video codecs, as well as formats with a base view dependency. Finally, depth-based coding techniques are

also covered with a review of coding techniques specific to depth data, as well as joint video/depth coding schemes.

A. 2D Video Codecs with Signaling

1) Simulcast of Stereo/Multiview

The natural means to compress stereo or multiview video is to encode each view independently of the other, e.g., using a state-of-the-art video coder such as H.264/AVC [1]. This solution, which is also referred to as simulcast, keeps computation and processing delay to a minimum since dependencies between views are not exploited. It also enables one of the views to be decoded for legacy 2D displays.

The main drawback of a simulcast solution is that coding efficiency is not maximized since redundancy between views, i.e., inter-view redundancy, is not considered. However, prior studies on asymmetrical coding of stereo, whereby one of the views is encoded with less quality, suggest that substantial savings in bit rate for the second view could be achieved. In this way, one of the views can be low pass filtered, more coarsely quantized than the other view [8], or coded with a reduced spatial resolution [9], yielding an imperceptible impact on the stereo quality. However, eye fatigue could be a concern when viewing asymmetrically coded video for long periods of time due to unequal quality to each eye. It has been proposed in [10], [11] to switch the asymmetrical coding quality between the left-eye and right-eye views when a scene change happens to overcome this problem. Further study is needed to understand how asymmetric coding applies to multiview video.

2) Frame-Compatible Coding with SEI Message

Frame-compatible signals can work seamlessly within existing infrastructures and already deployed video decoders. In an effort to better facilitate and encourage their adoption, the H.264/AVC standard introduced a new Supplemental Enhancement Information (SEI) message [1] that enables signaling of the frame packing arrangement used. Within this SEI message one may signal not only the frame-packing format, but also other information such as the sampling relationship between the two views and the view order among others. By detecting this SEI message, a decoder can immediately recognize the format and perform suitable processing, such as scaling, denoising, or color-format conversion, according to the frame-compatible format specified. Furthermore, this information can be used to automatically inform a subsequent device, e.g. a display or a receiver, of the frame-compatible format used by appropriately signaling this format through supported interfaces such as the High-Definition Multimedia Interface (HDMI) [10].

B. Stereo/Multiview Video Coding

1) 2D Video as a Base View

To improve coding efficiency of multiview video, both temporal redundancy and redundancy between views, i.e., inter-view redundancy, should be exploited. In this way, pictures are not only predicted from temporal reference pictures, but also from inter-view reference pictures as shown in Fig. 3. The concept of inter-view prediction, or disparity-compensated prediction, was first developed in the 1980s [2] and subsequently supported in amendments of the MPEG-2 standard [3]-[6]. Most recently, the H.264/AVC standard has been amended to support Multiview Video Coding (MVC) [1]. A few highlights of the MVC standard are given below, while a more in-depth overview of the standard can be found in [7].

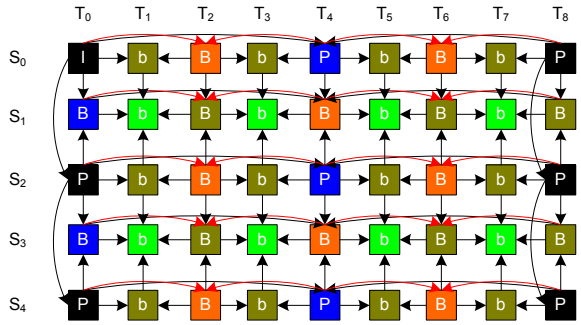

Fig. 3: Typical MVC picture coding structure

In the context of MVC, inter-view prediction is enabled through flexible reference picture management that is supported by the standard, where decoded pictures from other views are essentially made available in the reference picture lists. Block-level coding decisions are adaptive, so a block in a particular view may be predicted by a temporal reference, while another block in the same view can be predicted by an inter-view reference. With this design, decoding modules are not necessarily aware of whether a reference picture is a temporal reference or an inter-view reference picture.

Another important feature of the MVC design is the mandatory inclusion of a base view in the compressed multiview stream that could be easily extracted and decoded for 2D viewing; this base layer stream is identified by the NAL unit type syntax in H.264/AVC. In terms of syntax, the standard only requires small changes to high-level syntax, e.g., view dependency needs to be known for decoding. Since the standard does not require any changes to lower-level syntax, implementations are not expected to require significant design changes in hardware relative to single-view AVC decoding.

As with simulcast, non-uniform rate allocation could also be considered across the different views with MVC. Subjective quality of this type of coding is reported in section IV.A.

2) Frame-Compatible Video as a Base View

As mentioned in section II.B, although frame-compatible methods can facilitate easy deployment of 3D services to the home, they still suffer from a reduced resolution, and therefore reduced 3D quality perception. Recently, several methods that can extend frame-compatible signals to full resolution have been proposed. These schemes ensure backwards compatibility with already deployed frame-compatible 3D services, while permitting a migration to full-resolution 3D services.

One of the most straightforward methods to achieve this is

by leveraging existing capabilities of the Scalable Video Coding (SVC) extension of H.264/AVC [1]. For example, spatial scalability coding tools can be used to scale the lower resolution frame-compatible signal to full resolution. This method, using the side-by-side arrangement as an example, is shown in Fig. 4. An alternative method, also based on SVC, utilizes a combination of both spatial and temporal scalability coding tools. Instead of using the entire frame for spatial scalability, only half of the frame relating to a single view, i.e., $view_0$, is upconverted using region-of-interest based spatial scalability. Then, the full resolution second view can be encoded as a temporal enhancement layer (Fig. 5).

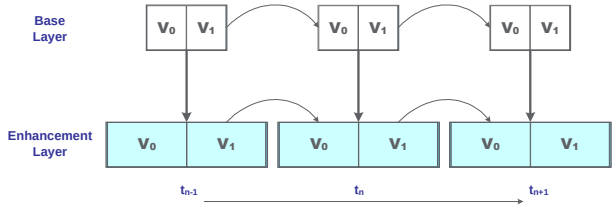

Fig. 4: Full resolution frame-compatible delivery using SVC and spatial scalability.

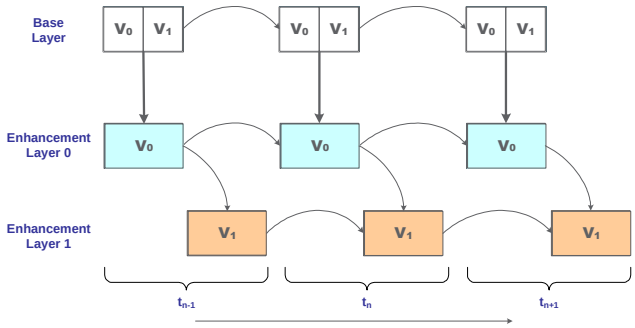

Fig. 5: Full resolution frame-compatible delivery using SVC and a combination of spatial and temporal scalability.

This second method somewhat resembles the coding process used in the MVC extension since the second view is able to exploit both temporal and inter-view redundancy. However, the same view is not able to exploit the redundancies that may exist in the lower resolution base layer. This method essentially sacrifices exploiting spatial correlation in favor of inter-view correlation. Both of these methods have the limitation that they may not be effective for more complicated frame-compatible formats such as side-by-side formats based on quincunx sampling or checkerboard formats.

MVC could also be used, to some extent, to enhance a frame-compatible signal to full resolution. In particular, instead of low-pass filtering the two views prior to decimation and then creating a frame-compatible image, one may apply a low-pass filter at a higher cut-off frequency or not apply any filtering at all. Although this may introduce some minor aliasing in the base layer, this provides the ability to enhance the signal to a full or near-full resolution with an enhancement layer consisting of the complementary samples relative to those of the base layer. These samples may have been similarly filtered and are packed using the same frame-compatible packing arrangement as the base layer.

The advantage of this method is that one can additionally exploit the spatial redundancies that may now exist between the base and enhancement layer signals, resulting in very high compression efficiency for the enhancement layer coding. Furthermore, existing implementations of MVC hardware could easily be repurposed for this application with minor modifications in the post-decoding stage.

An improvement over this method that tries to further exploit the correlation between the base and enhancement layer, was presented in [13]. Instead of directly considering the base layer frame-compatible images as a reference of the enhancement layer, a new process is introduced that first pre-filters the base layer picture given additional information that is provided within the bitstream (Fig. 6). This process generates a new reference from the base layer that has much higher correlation with the pictures in the enhancement layer.

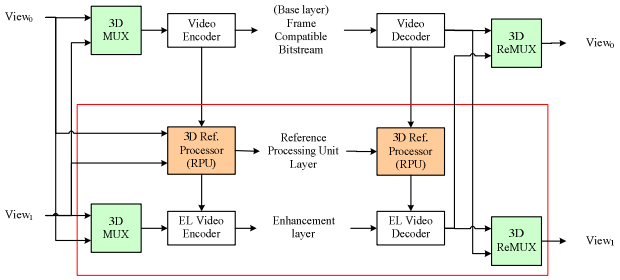

Fig. 6 Enhanced MVC architecture with reference processing, optimized for frame-compatible coding.

A final category for the enhancement of frame-compatible signals to full resolution considers filter-bank like methods [13]. Essentially, the base and enhancement layers contain the low and high frequency information, respectively. The separation is done using appropriate analysis filters in the encoder, whereas the analogous synthesis filters can be used during reconstruction at the decoder.

All of these methods have clear benefits and drawbacks and it is not yet clear which method will be finally adopted by the industry. The coding efficiency of these different methods will be analyzed in section IV.B.

C. Depth-based 3D Video Coding

In this subsection, advanced techniques for coding depth information are discussed. Methods that consider coding depth and video information jointly or in a dependent way are also considered.

1) Advanced Depth Coding

For monoscopic and stereoscopic video content, highly optimized coding methods have been developed, as reported in the previous subsections. For depth-enhanced 3D video formats, specific coding methods for depth data that yield high compression efficiency are still in the early stages of investigation. Here, the different characteristics of depth in comparison to video data must be considered. A depth signal

mainly consists of larger homogeneous areas inside scene objects and sharp transitions along boundaries between objects at different depth values. Therefore, in the frequency spectrum of a depth map, low and very high frequencies are dominant. Video compression algorithms are typically designed to preserve low frequencies and image blurring occurs in the reconstructed video at high compression rates. In contrast to video data, depth maps are not reconstructed for direct display but rather for intermediate view synthesis of the video data. A depth sample represents a shift value for color samples from original views. Thus, coding errors in depth maps result in wrong pixel shifts in synthesized views. Especially along visible object boundaries, annoying artifacts may occur. Therefore, a depth compression algorithm needs to preserve depth edges much better than current coding methods such as MVC.

Nevertheless, initial coding schemes for depth-enhanced 3D video formats used conventional coding schemes, such as AVC and MVC, to code the depth [24]. However, such schemes did not limit their consideration of coding quality to the depth data only when applying rate-distortion optimization principles, but also on the quality of the final, synthesized views. Such methods can also be combined with edge-aware synthesis algorithms, which are able to suppress some of the displacement errors caused by depth coding with MVC [27], [32]. In order to keep a higher quality for the depth maps at the same data rate, down-sampling before MVC encoding was introduced in [29]. After decoding, a non-linear up-sampling process is applied that filters and refines edges based on the object contours in the video data. Thus, important edge information in the depth maps is preserved. A similar process is also followed in [33] and [25], where wavelet decompositions are applied. For block-based coding methods, platelet coding was introduced for depth compression [26]. Here, occurrences of foreground/background boundaries are analyzed block-wise and approximated by simpler linear functions. This can be investigated hierarchically, i.e., starting with a linear approximation of boundaries in larger blocks and refining the approximation by subdividing a block using a quadtree structure. Finally, each block with a boundary contains two areas, one that represents the foreground depth and the other that represents the background depth. These areas can then be handled separately and the approximated depth edge information is preserved.

In contrast to pixel-based depth compression methods, a conversion of the scene geometry into computer graphics based meshes and the application of mesh-based compression technology was described in [23].

2) Joint Video/Depth Coding

Besides the adaptation of compression algorithms to the individual video and depth data, some of the block-level information, such as motion vectors, may be similar for both and thus can be shared. An example is given in [28]. In addition, mechanisms used in scalable video coding can be applied, where a base layer was originally used for a lower quality version of the 2D video and a number of enhancement layers were used to provide improved quality versions of the video. In the context of multiview coding, a reference view is encoded as the base layer. Adjacent views are first warped onto the position of the reference view and the residual between both is encoded in further enhancement layers.

Other methods for joint video and depth coding with partially data sharing, as well as special coding techniques for depth data, are expected to be available soon in order to provide improved compression in the context of the new 3D video format that is anticipated. The new format will not only require high coding efficiency, but it must also enable good subjective quality for synthesized views that could be used on a wide range of 3D displays.

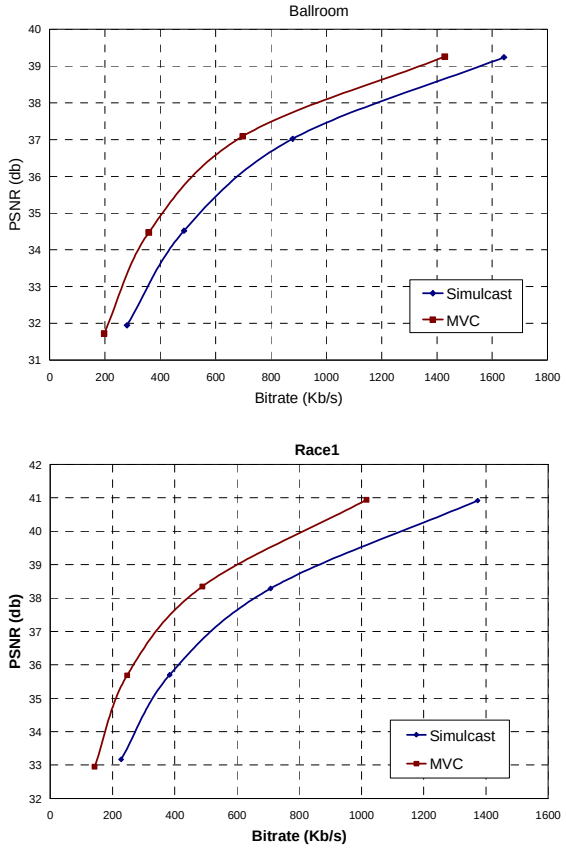

Fig. 7: Sample coding results for Ballroom and Race1 sequences; each sequence includes 8 views at VGA resolution.

IV. Performance Comparisons & Evaluation

A. MVC versus Simulcast

It has been shown that coding multiview video with inter-view prediction can give significantly better results compared to independent coding [33]. A comprehensive set of results for multiview video coding over a broad range of test material was also presented in [34]. This study used the common test conditions and test sequences specified in [35], which were used throughout the MVC development. For multiview video with up to 8 views, an average of 20%

reduction in bit rate relative to the total simulcast bit rate was reported with equal quality for each view. All of the results were based on the Bjontegaard delta measurements [36]. Fig. 7 shows sample rate-distortion (RD) curves comparing the performance of simulcast coding with the performance of the MVC reference software. In other studies [37], an average bit-rate reduction for the second (dependent) view of typical HD stereo movie content of approximately 20-30% was reported, with a peak reduction up to 43%. It is noted that the compression gains achieved by MVC using the stereoscopic movie content, which are considered professional HD quality and representative of entertainment quality video, are consistent with gains reported earlier on the MVC test set [35].

A recent study of subjective picture quality for the MVC Stereo High Profile targeting full-resolution HD stereo video applications was presented in [38]. For this study, different types of 3D video content were selected (see Table 1) with each clip running 25-40 seconds. In the MVC simulations, the left-eye and right-eye pictures were encoded as the base-view and dependent-view, respectively. The base-view was encoded at 12Mbps and 16Mbps. The dependent view was coded at a wide range of bit rates, from 5% to 50% of the base-view bit rate (see Table 2). As a result, the combined bit rates range from 12.6Mbps to 24Mbps. AVC simulcast with symmetric quality was selected as the reference. Constant bit rate (CBR) compression was used in all the simulations with configuration settings similar to those that would be used in actual HD video applications, such as Blu-ray systems.

Table 1: 3D video content used in the evaluation.

Clip A	1080p @ 23.98fps	Live action, drama
Clip B	1080p @ 23.98fps	Animation movie
Clip C	1080p @ 23.98fps	Live action, drama
Clip D	1080p @ 23.98fps	Animation movie
Clip E	720p @ 59.94fps	Live action, beach volleyball
Clip F	1080i @ 29.97fps	Live action, documentary
Clip G	1080i @ 29.97fps	Live action, mixture of sports
Clip H	1080i @ 29.97fps	Live action, tennis
Clip I	1080i @ 29.97fps	Live action, Formula 1 racing

Table 2: Bitrate configuration.

Test cases	Base-view bit rate	Dependent-view bit rate		Combined bit rate
12L_5Pct	12 Mbps	0.6 Mbps	5%	12.6 Mbps
12L_10Pct	12 Mbps	1.2 Mbps	10%	13.2 Mbps
12L_15Pct	12 Mbps	1.8 Mbps	15%	13.8 Mbps
12L_20Pct	12 Mbps	2.4 Mbps	20%	14.4 Mbps
12L_25Pct	12 Mbps	3.0 Mbps	25%	15.0 Mbps
12L_35Pct	12 Mbps	4.2 Mbps	35%	16.2 Mbps
12L_50Pct	12 Mbps	6.0 Mbps	50%	18.0 Mbps
16L_5Pct	16 Mbps	0.8 Mbps	5%	16.8 Mbps
16L_10Pct	16 Mbps	1.6 Mbps	10%	17.6 Mbps
16L_15Pct	16 Mbps	2.4 Mbps	15%	18.4 Mbps
16L_20Pct	16 Mbps	3.2 Mbps	20%	19.2 Mbps
16L_25Pct	16 Mbps	4.0 Mbps	25%	20.0 Mbps
16L_35Pct	16 Mbps	5.6 Mbps	35%	21.6 Mbps
16L_50Pct	16 Mbps	8.0 Mbps	50%	24.0 Mbps

At each base-view bit rate, there are 9 test cases for each clip, which include the 7 MVC coded results, the AVC simulcast result, and the original video. The display order of the 9 test cases was random and different for each clip. Viewers were asked to give a numeric value based on a scale of 1 to 5 scale, with 5 being excellent and 1 very poor. 15 non-expert viewers participated in the evaluation.

The subjective picture quality evaluation was conducted in a dark room. A 103-inch Panasonic 3D plasma TV with native display resolution of 1920x1080 pixels and active shutter glasses were used in the setup. Viewers were seated at a distance between 2.5 and 3.5 times the display height.

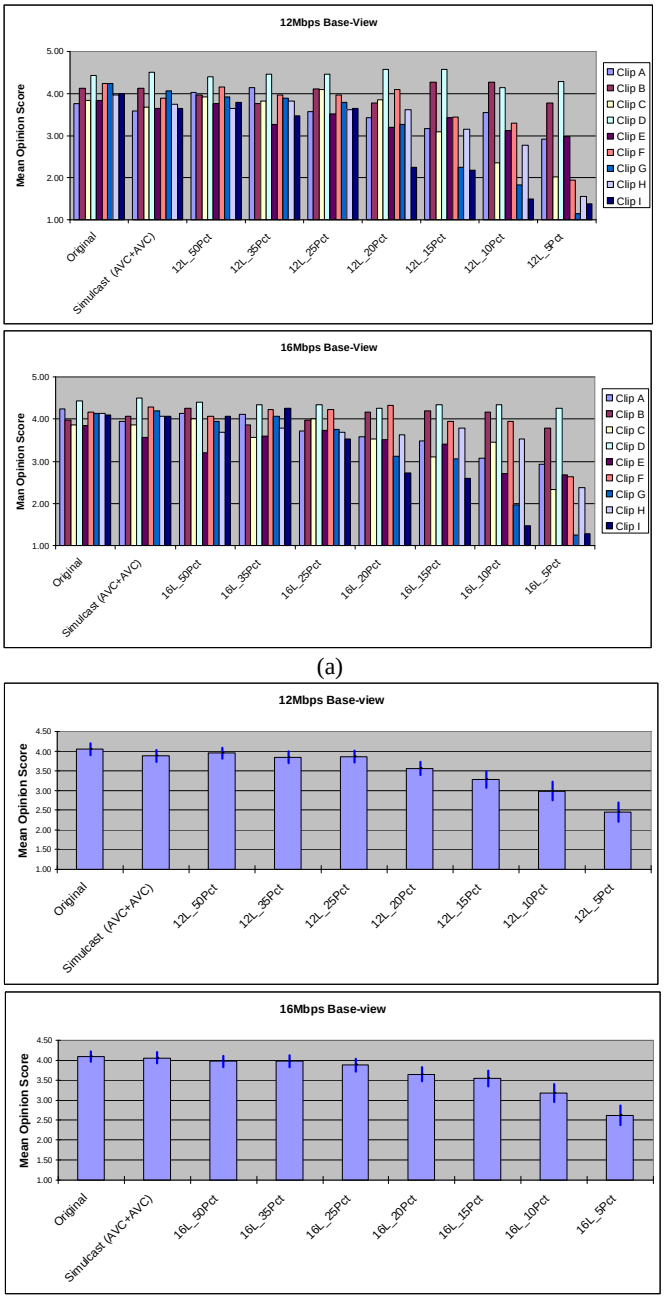

Fig. 8: Subjective picture quality evaluation results: (a) clip-wise MOS; (b) average MOS and its 95% confidence intervals.

The mean opinion score (MOS) of each clip is shown in Fig. 8(a). It is clear that the animation clips receive fair or better scores even when the dependent-view is encoded at 5% of the base-view bit rate. When the dependent-view bit rate drops below 20% of the base-view bit rate, the MVC encoded interlaced content starts to receive unsatisfactory scores. Fig. 8(b) presents the average MOS of all the clips. In the bar charts, each short line segment indicates a 95% confidence interval. The average MOS and 95% confidence intervals show the reliability of the scoring in the evaluation. Overall, when the dependent-view bit rate is no less than 25% of the base-view bit rate, the MVC compression can reproduce the subjective picture quality comparable to that of the AVC simulcast case. It is noted that long-term viewing effects such as eye fatigue were not considered as part of this study.

Given a total bandwidth, there is a trade-off in choosing the base-view bit rate. A lower base-view bit rate would leave more bits to the dependent-view, and the 3D effect and convergence could be better preserved. Both of the cases of 12L_50Pct and 16L_15Pct result in combined bit rates around 18Mbps. From Fig. 8(b), it is obvious that 12L_50Pct was favoured over 16L_15Pct in terms of 3D video quality, especially for live action shots. However, this is achieved at the cost of an inferior base-view picture quality as compared to the case of higher base-view bit rate. It is also important to maintain the base-view picture quality because many people may choose to watch a program on conventional 2D TVs.

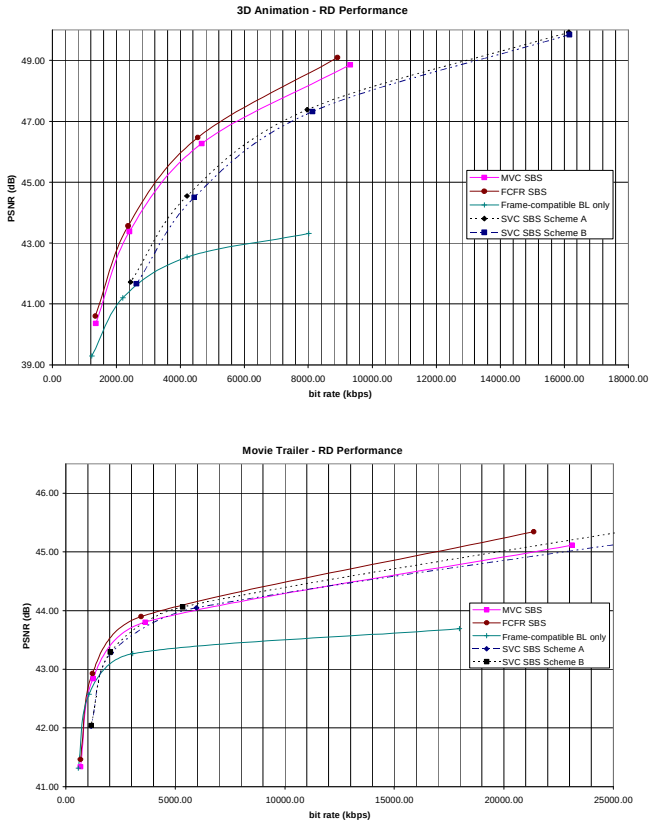

Fig. 9: Performance evaluation of different frame-compatible full resolution methods.

B. Evaluation of Frame-Compatible Video as a Base View

An evaluation of the performance of different frame-compatible, full resolution methods was presented in [39] using primarily the side-by-side format. In particular, the methods presented in Section III.B, including the spatial SVC method (SVC SBS Scheme A), the spatio-temporal SVC method (SVC SBS Scheme B) as well as the frame-compatible MVC method (MVC SBS) and its extension that includes the base layer reference processing step (FCFR SBS) were considered. In addition, basic upscaling of the half resolution frame compatible signal was also evaluated in this test. Commonly used test conditions within MPEG were considered, whereas the evaluation focused on a variety of 1080p sequences, including animated and movie content. The RD curves of two such sequences are presented in Fig. 9.

Fig. 9 suggests that the FCFR SBS method is superior to all other methods and especially compared to the two SVC schemes in terms of coding performance. In some cases, a performance improvement of over 30% can be achieved. Performance improvement over the MVC SBS is smaller, but still not insignificant (>10%). However, all of these methods can provide an improved quality experience with a relatively small overhead in bit rate compared to simple upscaling of the frame-compatible base layer.

C. Evaluation of Depth-based Formats

Several advanced coding methods for joint video and depth coding, including algorithms adaptive to the characteristics of depth data, are currently under development. One important aspect for the design of such coding methods is the quality optimization for all synthesized views. In contrast to conventional coding measures used for 2D video data, in which a decoded picture is compared against an uncoded reference and the quality was evaluated using an objective distortion measure such as peak signal-to-noise ratio (PSNR), the new 3D video format with video and depth data requires that synthesized views at new spatial positions must also look good.

It is often the case that there is no original reference image available to measure the quality. Therefore, a comprehensive subjective evaluation has to be carried out in order to judge the reconstruction quality of the 3D video data. This is important as new types of errors may occur for 3D video in addition to the classic 2D video reconstruction errors such as quantization or blurring. Examples of such 3D errors include wrong pixel shifts, frayed object boundaries at depth edges, or parts of an object appearing at the wrong depth. Nevertheless, an objective quality measure is still highly desirable in order to carry out automatic coding optimization. For this, high quality depth data as well as a robust view synthesis are required in order to provide an uncoded reference. The quality of the reference should ideally be indistinguishable from that of the original views. In the experiments that follow, coded synthesized views are compared with such uncoded reference views based on PSNR. An example is shown in Fig. 10 for two different bit rate distributions between video and depth data.

In these plots, "C30D30" stands for color quantization parameter (QP) 30 and depth QP 30. A lower QP value represents more bit rate and thus better quality. For the curve "C30D30", equal quantization for color and depth was applied. For the second curve "C24D40", the video bit rate was increased at the expense of the depth bit rate. Therefore, better reconstruction results are achieved for "C24D40" at original positions 2.0, 3.0 and 4.0, where no depth data is required. For all intermediate positions, "C24D40" performs worse than "C30D30"as the lower quality of coded depth data causes degrading displacement errors in all intermediate views. It is noted that both curves have the same overall bit rate of 1200 Kbps.

The view synthesis algorithm that was used generates intermediate views between each pair of original views. The two original views are warped to an intermediate position using the depth information. Then, view-dependent weighting is applied to the view interpolation in order to provide seamless navigation across the viewing range. Fig. 10 shows, that lower quality values are especially obtained for the middle positions 2.5 and 3.5. This also represents the furthest distance from any original view and aligns with subjective viewing tests. Consequently, new 3D video coding and synthesis methods need to pay special attention to the synthesized views around the middle positions.

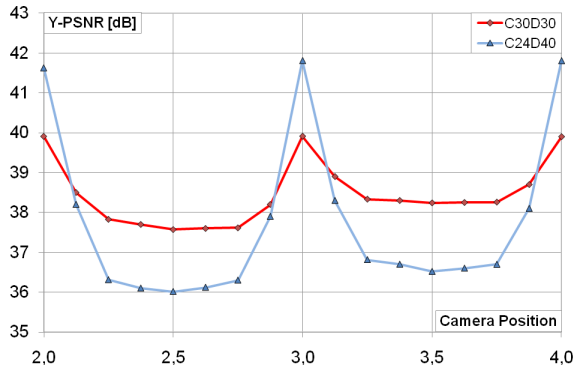

Fig. 10: PSNR curves across the viewing range of original cameras 2, 3, and 4 for two different bit rate distributions between video and depth data for the *Ballet* test set.

V. Distribution of 3D

This section discusses the requirements and constraints on typical storage and transmission systems (e.g., backward compatibility needs, bandwidth limitations, set-top box constraints). We focus our discussion on Blu-ray Disc (BD), cable, and terrestrial channels as exemplary systems. The suitability for the various coding formats for each of these channels is discussed. We also discuss feasible options for future support of auto-stereoscopic displays.

A. Storage Systems

The Blu-ray Disc Association (BDA) finalized a Blu-ray 3D specification [38] in December 2009. As a packaged media application, Blu-ray 3D considered the following factors during its development:

a) picture quality and resolution
b) 3D video compression efficiency
c) backward compatibility with legacy BD players
d) interference among 3D video, 3D subtitles, and 3D menu

As discussed in the prior section, frame-compatible formats have the benefit of being able to use existing 2D devices for 3D applications, but suffer from a loss of resolution that cannot be completely recovered without some enhancement information. To satisfy picture quality and resolution requirements, a frame sequential full-resolution stereo video format was considered as the primary candidate for standardization. In 2009, BDA conducted a series of subjective video quality evaluations to validate picture quality and compression efficiency. The evaluation results eventually led to the inclusion of MVC Stereo High Profile as the mandatory 3D video codec in the Blu-ray 3D specification.

With the introduction of Blu-ray 3D, backward compatibility with legacy 2D players was one of the crucial concerns from consumer and studio perspectives. One possible solution for delivering MVC encoded bitstreams on a Blu-ray disc is to multiplex both the base and dependent-view streams in one MPEG-2 transport stream (TS). In this scenario, a 2D player can read and decode only the base-view data, while discarding the dependent-view data. However, this solution is severely affected by the bandwidth limitations of legacy BD players. In particular, the total video rate in this scenario is restricted to a maximum bit rate of only 40Mbps, implying that the base-view picture may not be allocated the maximum possible bit rate that may have been allocated if the same video was coded as a single view.

Instead, a preferred solution was to consider the use of two transport streams: a main-TS for the base-view and associated audio needed for 3D playback, and a sub-TS for the dependent-view and other elementary streams associated with 3D playback such as the depth of 3D subtitles. In this case, the maximum video rate of stereo video is 60Mbps while the maximum video rate of each view is 40Mbps.

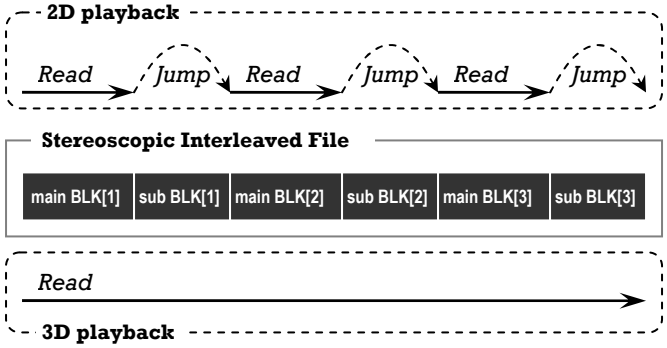

Fig. 11: Data allocation of 2D compatible TS and 3D extended TS in Blu-ray 3D.

The playback of stereo video requires continuous reading of streams from a disc. Therefore, the main-TS and sub-TS are interleaved and stored on a 3D disc. When a 3D disc is played

in a 2D player, the sub-TS is skipped by jump reading since the bandwidth is limited in legacy BD players. In optical disc I/O, a jump reading operation imposes a minimum waiting time before it initiates a new reading operation. The minimum waiting time is much longer than the playback duration of one frame. As a result, stream interleaving at a frame level is prohibited. In Blu-ray 3D, the two TSs are divided into blocks, and typically each block contains a few seconds of AV data. The blocks of main-TS and sub-TS are interleaved and stored on a Blu-ray 3D disc. In this case, the jump distance (i.e., the size of each sub-TS block) is carefully designed to satisfy the BD-ROM drive performance in legacy 2D players. Fig. 11 illustrates the data storage on a 3D disc and the operations in the 2D and 3D playback cases. The Stereoscopic Interleaved File is used to record the interleaved blocks from the main-TS and sub-TS. A Blu-ray 3D disc can be played in a 3D player using either the 2D Output Mode or Stereoscopic Output Mode for 2D and 3D viewing, respectively.

In Blu-ray 3D, both single-TS and dual-TS solutions are applicable. A single TS is used when a 3D bonus video is encoded at a lower bit rate, or when a 2D video clip is encoded using MVC to avoid switching between AVC and MVC decode modes. In the latter case, the dependent-view stream consists of skipped blocks and the bit rate is extremely low. Without padding zero-bytes in the dependent-view stream, it is not suitable to use the block interleaving of two TSs as described above. Padding zero-bytes certainly increases the data size, which quite often is not desirable due to limited disc capacity and the overwhelming amount of extra data that may have been added to the disc.

B. Transmission Systems

Different transmission systems are characterized by their own constraints. In the following, we consider delivery of 3D over cable and terrestrial channels.

The cable infrastructure is not necessarily constrained by bandwidth. However, for rapid deployment of 3D services, existing set-top boxes that decode and format the content for display would need to be leveraged. Consequently, cable operators have recently started delivery of 3D video based on frame-compatible formats. It is expected that video-on-demand (VOD) and pay-per-view (PPV) services could serve as a good business model in the early stages. The frame-compatible video format is carried as a single stream, so there is very little change at the TS level. There is new signaling in the TS to indicate the presence of the frame-compatible format and corresponding SEI message signaling. The TS may also need to carry updated caption and subtitle streams that are appropriate for the 3D playback. New boxes that support full-resolution formats may be introduced into the market later depending on market demand and initial trials. The Society of Cable Telecommunications Engineers (SCTE), which is the standards organization that is responsible for cable services, is considering this roadmap and the available options.

Terrestrial broadcast is perhaps the most constrained distribution method. Most countries around the world have defined their digital broadcast services based on MPEG-2, which is often a mandatory format in each broadcast channel. Therefore, there are legacy format issues to contend with that limit the channel bandwidth that could be used for new services. A sample bandwidth allocation considering the presence of high-definition (HD), standard-definition (SD) and mobile services is shown in Fig. 12. This figure indicates that that there are significant bandwidth limitations for new 3D services when an existing HD video service is delivered in the same terrestrial broadcast channel. The presence of a mobile broadcast service would further limit the available bandwidth to introduce 3D. Besides this, there are also costs associated with upgrading broadcast infrastructure and the lack of a clear business model on the part of the broadcasters to introduce 3D services. Terrestrial broadcast of 3D video is lagging behind other distribution channels for these reasons.

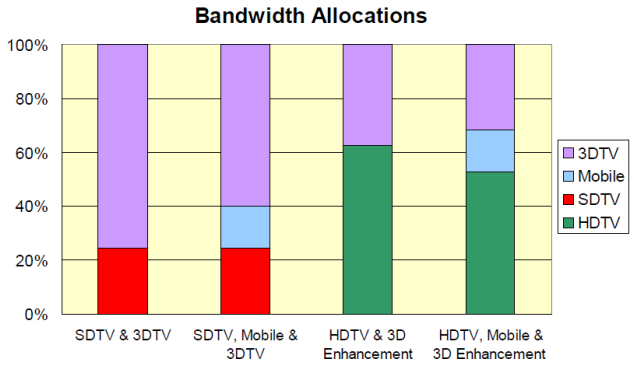

Fig. 12: Bandwidth allocation for terrestrial broadcast with 3D-TV services.

It is also worth noting that with increased broadband connectivity in the home, access to 3D content from web servers is likely to be a dominant source of content. Sufficient bandwidth and reliable streaming would be necessary; download and offline playback of 3D content would be another option. To support the playback of such content, the networking and decode capabilities must be integrated into the particular receiving devices (e.g., TV, PC, gaming platform, optical disc player) and these devices must have a suitable interface to the rendering device.

C. Supporting Auto-Stereoscopic Displays

As shown in section II.C, an important feature of advanced 3D TV technology is the new 3D video format, which can support any 3D display and especially high-quality auto-stereoscopic (glasses-free) displays. Currently, glasses-based stereo displays are used for multi-user applications, e.g., 3D cinema. However, for applications like mobile 3D TV, where single users are targeted, stereoscopic displays without glasses can be used. Glasses-free displays are also desirable for 3D home entertainment. In this case, multi-view displays have to be used; however the desired resolution of these displays is not yet sufficient. Current stereoscopic displays still show a benefit since they only need to share the total screen resolution among the two stereo views, yielding half the resolution per view. For

multi-view displays, the screen resolution needs to be distributed across all N views, only leaving $1/N$ of the total resolution for each view. This limitation also restricts the total number of views to between 5 and 9 views based on current display technology, and therefore the viewing angle for each repetition of the views is rather small.

These disadvantages of multi-view displays are expected to be overcome by novel ultra high-resolution displays, where a much larger number of views, e.g., on the order of 50, with good resolution per view can be realized. In addition to the benefit of glasses-free 3D TV entertainment, such multi-view displays will offer correct dynamic 3D viewing, i.e., different viewing pairs with slightly changing viewing angle, while a user moves horizontally. This leads to the expected "look-around" effect, where occluded background in one viewing position is revealed besides a foreground object in another viewing position. In contrast, stereo displays only show two views from fixed positions and in the case of horizontal head movement, background objects seem to move in the opposite direction. This is known as the parallax effect.

Since the new depth-based 3D video format aims to support both existing and future 3D displays, it is expected that multiple services from mobile to home entertainment, as well as support for single or multiple users, will be enabled. A key challenge will be to design and integrate the new 3D format into existing 3D distribution systems discussed earlier in this section.

VI. Concluding Remarks

Distribution of high-quality stereoscopic 3D content through packaged media and broadcast channels is now underway. This article reviewed a number of 3D representation formats and also a variety of coding architectures and techniques for efficient compression of these formats. Furthermore, specific application requirements and constraints for different systems have been discussed. Frame-compatible coding with SEI message signaling has been selected as the delivery format for initial phases of broadcast, while full-resolution coding of stereo with inter-view prediction based on MVC has been adopted for distribution of 3D on Blu-ray Disc.

The 3D market is still in its infancy and it may take further time to declare this new media a success with consumers in the home. Various business models are being tested, e.g., video-on-demand, and there needs to be strong consumer interest to justify further investment in the technology. In anticipation of these next steps, the roadmap for 3D delivery formats is beginning to take shape. In the broadcast space, there is strong consideration for the next phase of deployment beyond frame-compatible formats. Coding formats that enhance the frame-compatible signal provide a graceful means to migrate to a full-resolution format, while still maintaining compatibility with earlier services. Beyond full-resolution stereo, the next major leap would be towards services that support auto-stereoscopic displays. Although the display technology is not yet mature, it is believed that this technology will eventually become feasible and that depth-based 3D video formats will enable such services.

Acknowledgment

We would like to thank the Interactive Visual Media Group of Microsoft Research for providing the *Ballet* data set.

References

[1] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.

[2] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets," *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, vol. 1, pp. 521-524, Tokyo, Japan, 1986.

[3] ITU-T and ISO/IEC JTC 1, "Final Draft Amendment 3," Amendment 3 to ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), MPEG document N1366, Sept. 1996.

[4] A. Puri, R. V. Kollarits, and B. G. Haskell. "Stereoscopic video compression using temporal scalability," *Proc. SPIE Conf. Visual Communications and Image Processing*, vol. 2501, pp. 745–756, 1995.

[5] X. Chen and A. Luthra, "MPEG-2 multi-view profile and its application in 3DTV," *Proc. SPIE IS&T Multimedia Hardware Architectures*, San Diego, USA, Vol. 3021, pp. 212-223, Feb. 1997.

[6] J.-R. Ohm, "Stereo/Multiview Video Encoding Using the MPEG Family of Standards," *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems VI*, San Jose, CA, Jan. 1999.

[7] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/AVC Standard," *Proceedings of the IEEE*, 2011.

[8] L. Stelmach and W. J. Tam, " Stereoscopic Image Coding: Effect of Disparate Image-Quality in Left- and Right-Eye Views," *Signal Processing: Image Communication* , Vol. 14, pp. 111-117, 1998

[9] L. Stelmach, W.J. Tam; D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 10, No. 2, pp. 188-193, Mar. 2000.

[10] W. J. Tam, L. B. Stelmach, F. Speranza and R. Renaud, "Cross-switching in asymmetrical coding for stereoscopic video," *Stereoscopic Displays and Virtual Reality Systems IX*, Vol. 4660, pp. 95-104, 2002.

[11] W. J. Tam, L. B. Stelmach, and S. Subramaniam, "Stereoscopic video: Asymmetrical coding with temporal interleaving," *Stereoscopic Displays and Virtual Reality Systems VIII*, Vol. 4297, pp. 299-306, 2001.

[12] HDMI Licensing, LLC., "High Definition Multimedia Interface: Specification Version 1.4a," May 2009

[13] A.M. Tourapis, P. Pahalawatta, A. Leontaris, Y. He, Y. Ye, K. Stec, and W. Husak, "A Frame Compatible System for 3D Delivery," ISO/IEC JTC1/SC29/WG11 Doc. M17925, Geneva, Switzerland, Jul. 2010.

[14] Video and Requirements Group, "Problem statement for scalable resolution enhancement of frame-compatible stereoscopic 3D video," ISO/IEC JTC1/SC29/WG11 Doc. N11526, Geneva, Switzerland, Jul. 2010.

[15] K. Müller, P. Merkle, T. Wiegand, "3D video representation using depth maps," *Proceedings of the IEEE*, 2011.

[16] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TVm" *Proc. SPIE Conference on Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, San Jose, CA, USA, Jan. 2004.

[17] A. Vetro, S. Yea, and A. Smolic, "Towards a 3D video format for auto-stereoscopic displays," *Proc. SPIE Conference on Applications of Digital Image Processing XXXI*, San Diego, CA, Aug. 2008.

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

[19] E.-K. Lee, Y.-K. Jung, and Y.-S. Ho, "3D Video Generation Using Foreground Separation And Disocclusion Detection", *Proc. IEEE 3DTV Conference*, Tampere, Finland, June 2010.

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

[21] Video and Requirements Group, "Applications and Requirements on 3D Video Coding," ISO/IEC JTC1/SC29/WG11 Doc. N11550, Geneva, Switzerland, Jul. 2010.
[22] I. Daribo, C. Tillier, and B. Pesquet-Popescu, "Adaptive wavelet coding of the depth map for stereoscopic view synthesis", *Proc. IEEE International Workshop on Multimedia Signal Processing* (MMSP'08), Cairns, Australia, pp. 34-39, Oct. 2008.
[23] S.-Y. Kim and Y.-S. Ho, "Mesh-Based Depth Coding for 3D Video using Hierarchical Decomposition of Depth Maps," *Proc. IEEE International Conference on Image Processing* (ICIP'07), San Antonio, USA, pp. V117–V120, Sept. 2007.
[24] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map coding with distortion estimation of rendered view", *Visual Information Processing and Communication, Proceedings of the SPIE*, vol. 7543, 2010.
[25] M. Maitre and M. N. Do, "Shape-adaptive Wavelet Encoding of Depth Maps", *Proc. Picture Coding Symposium* (PCS'09), Chicago, USA, May 2009.
[26] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P.H.N. de With, and T. Wiegand, "The Effects of Multiview Depth Video Compression on Multiview Rendering", *Signal Processing: Image Communication*, vol. 24, is. 1+2, pp. 73-88, Jan. 2009.
[27] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View Synthesis for Advanced 3D Video Systems", EURASIP Journal on Image and Video Processing, Special Issue on 3D Image and Video Processing, vol. 2008, Article ID 438148, 11 pages, 2008. doi:10.1155/2008/438148.
[28] H. Oh and Y.-S. Ho, "H.264-Based Depth Map Sequence Coding Using Motion Information of Corresponding Texture Video", *Springer Berlin/Heidelberg, Advances in Image and Video Technology*, vol. 4319, 2006.
[29] K.-J. Oh, S. Yea, A. Vetro, and Y.-S. Ho, "Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3-D Video", *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 747-750, Sept. 2009.
[30] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, and Y. Yashima, "View Scalable Multi-view Video Coding Using 3-D Warping with Depth Map," *IEEE Trans. Circuits Systems for Video Technology*, vol.17, no.11, pp.1485-1495, Nov. 2007.
[31] S. Yea and A. Vetro, "View Synthesis Prediction for Multiview Video Coding", *Signal Processing: Image Communication*, vol. 24, is. 1+2, pp. 89-100, Jan. 2009.
[32] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-Quality Video View Interpolation Using a Layered Representation", *ACM SIGGRAPH and ACM Trans. on Graphics*, Los Angeles, CA, USA, Aug. 2004.
[33] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, Nov. 2007.
[34] D. Tian, P. Pandit, P. Yin, and C. Gomila, "Study of MVC coding tools," Joint Video Team, Doc. JVT-Y044, Shenzhen, China, Oct. 2007.
[35] Y. Su, A. Vetro, and A. Smolic, "Common Test Conditions for Multiview Video Coding," Joint Video Team, Doc. JVT-U211, Hangzhou, China, Oct. 2006.
[36] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q.6, Doc. VCEG-M033, Austin, TX, April 2001.
[37] T. Chen, Y. Kashiwagi, C.S. Lim, and T. Nishi, "Coding performance of Stereo High Profile for movie sequences," Joint Video Team, Doc. JVT-AE022, London, United Kingdom, July 2009.
[38] T. Chen and Y. Kashiwagi, "Subjective Picture Quality Evaluation of MVC Stereo High Profile for Full-Resolution Stereoscopic High-Definition 3D Video Applications," *Proc. IASTED Conference on Signal and Image Processing*, Maui, HI, Aug. 2010.
[39] A. Leontaris, P. Pahalawatta, Y. He, Y. Ye, A.M. Tourapis, J. Le Tanou, P.J. Warren, and W. Husak, "Frame Compatible Full Resolution 3D Delivery: Performance Evaluation," ISO/IEC JTC1/SC29/WG11 Doc. M17927, Geneva, Switzerland, Jul. 2010
[40] Blu-ray Disc Association, "System Description: Blu-ray Disc Read-Only Format Part3: Audio Visual Basic Specifications", Dec. 2009.

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, where he is currently a Group Manager responsible for research and standardization on video coding, as well as work on display processing, information security, speech processing, and radar imaging. He has published more than 150 papers in these areas. He has also 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 and specifications. Most recently, he was a key contributor to the Multiview Video Coding extension of the H.264/MPEG-4 AVC standard. He also serves as Vice-Chair of the U.S. delegation to MPEG.

Dr. Vetro is also active in various IEEE conferences, technical committees, and editorial boards. He currently serves on the Editorial Boards of IEEE Signal Processing Magazine and IEEE MultiMedia, and as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and IEEE TRANSACTIONS ON IMAGE 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 Signal Processing Magazine (2006–2007), as Conference Chair for ICCE 2006, Tutorials Chair for ICME 2006, and as a member of the Publications Committee of the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS (2002–2008). He is a member of the Technical Committees on Visual Signal Processing & Communications, and Multimedia Systems & Applications of the IEEE Circuits and Systems Society. He has also received several awards for his work on transcoding, including the 2003 IEEE Circuits and Systems CSVT Transactions Best Paper Award.

Alexis M. Tourapis (S'99-M'01-SM'07) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 1995 and the Ph.D. degree in electrical engineering from the Hong Kong University of Science & Technology, HK, in 2001. During his Ph.D. years, Dr. Tourapis made several contributions to MPEG standards on the topic of Motion Estimation.

Dr. Tourapis joined Microsoft Research Asia in 2002 as a Visiting Researcher, where he worked on next generation video coding technologies and was an active participant in the H.264/MPEG-4 AVC standardization process. From 2003 to 2004, he worked as a Senior Member of the Technical Staff for Thomson Corporate Research in Princeton, NJ, on a variety of video compression and processing topics. He later joined DoCoMo Labs USA, as a Visiting Researcher, where he continued working on next generation video coding technologies. Since 2005, Dr. Tourapis has been with the Image Technology Research Group at Dolby Laboratories where he currently manages a team of engineers focused on multimedia signal processing and compression.

In 2000, Dr. Tourapis received the IEEE HK section best postgraduate student paper award and in 2006 he was acknowledged as one of 10 most outstanding reviewers by the IEEE Transactions on Image Processing. Dr. Tourapis currently holds 6 US patents and has more than 80 US and international patents pending. He has made several contributions to several video coding standards on a variety of topics, such as motion estimation and compensation, rate distortion optimization, rate control and others. Dr. Tourapis currently serves as a co-chair of the development activity on the H.264 Joint Model (JM) reference software.

Karsten Müller (M'98-SM'07) is heading the 3D Coding group within the Image Processing department of the Fraunhofer Institute for Telecommunications - Heinrich Hertz Institute, Berlin, Germany. He 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 and coordinates 3D Video and 3D Coding related international projects. His research interests are in the field of representation, coding and reconstruction of 3D scenes in Free Viewpoint Video scenarios and Coding, immersive and interactive multi-view technology, and combined 2D/3D similarity analysis. He has been involved in ISO-MPEG standardization activities in 3D video coding and content description.

Tao Chen (S'99–M'01) received the B.Eng. in computer engineering from Southeast University, China, M.Eng. in information engineering from Nanyang Technological University, Singapore, and Ph.D. in computer science from Monash University, Australia. Since 2003, he has been with Panasonic Hollywood Laboratory in Universal City, CA, where he is currently Manager of Advanced Image Processing Group. Prior to joining Panasonic, he was a Member of Technical Staff with Sarnoff Corporation in Princeton, NJ. His research interests include image and video compression and processing.

Dr. Chen has served as session chairs and has been on technical committees for a number of international conferences. He was appointed vice-chair of a technical group for video codec evaluation in the Blu-ray Disc Association (BDA) in 2009.

Dr. Chen was a recipient of an Emmy Engineering Award in 2008. He received Silver Awards from the Panasonic Technology Symposium in 2004 and 2009. In 2002, he received the Most Outstanding Ph.D. Thesis Award from the Computer Science Association of Australia and a Mollie Holman Doctoral Medal from Monash University.