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

Frame Compatible Formats for 3D Video Distribution

Anthony Vetro

TR2010-099   November 2010

Abstract

Stereoscopic video will soon be delivered to the home through various channels. To make this feasible for some channels, the representation of the stereo video is modified to accommodate certain constraints on legacy systems. Among the various constraints that must be considered include the capabilities of production equipment and transmission infrastructure, as well as existing receivers and uncompressed digital interfaces between devices within the home. This paper outlines the typical constraints that are encountered in these domains and provides an overview of the various frame-compatible formats that are being considered for distribution of 3D video through such legacy systems. The benefits and drawbacks of these formats are discussed and the current status in various industry forums is reviewed.

FRAME COMPATIBLE FORMATS FOR 3D VIDEO DISTRIBUTION

Anthony Vetro

Mitsubishi Electric Research Laboratories (MERL)
201 Broadway, 8th Floor, Cambridge, MA 02139 USA
Email: avetro@merl.com

ABSTRACT

Stereoscopic video will soon be delivered to the home through various channels. To make this feasible for some channels, the representation of the stereo video is modified to accommodate certain constraints on legacy systems. Among the various constraints that must be considered include the capabilities of production equipment and transmission infrastructure, as well as existing receivers and uncompressed digital interfaces between devices within the home. This paper outlines the typical constraints that are encountered in these domains and provides an overview of the various frame-compatible formats that are being considered for distribution of 3D video through such legacy systems. The benefits and drawbacks of these formats are discussed and the current status in various industry forums is reviewed.

Index Terms— frame-compatible, stereo interleaving, spatial multiplexing, temporal multiplexing, 3D video, distribution

1. INTRODUCTION

There is a growing interest in delivery of 3D content to the home. Production of 3D cinema content is steadily increasing, and there are already devices supporting stereoscopic display available to the consumer. To facilitate interoperable 3D services to the home, standards for production, distribution and digital interfaces are being developed or amended.

The current class of televisions that support stereoscopic video are referred to as 3D-Ready TVs. These devices can identify uncompressed content in a standard 3D image or video format, then properly display it. As the 3D market matures and the TV platform evolves, a new class of devices that will be referred to as 3D-Capable TVs will emerge; these devices will be able to identify compressed content in a standard 3D distribution format, then properly decode and display it.

Given the existence of these display devices, a very significant issue is the means by which 3D content is delivered to the home through legacy systems. One option is to consider a complete upgrade to the related equipment and infrastructure so that an additional view could be accommodated. However, this is very costly for some distribution scenarios and takes time. One exception is packaged media, such as Blu-ray Disc, which can more easily introduce new 3D players into the market and leverage the capabilities of existing 2D players for 3D. In fact, the Blu-ray Disc Association has adopted the Multiview Video Coding (MVC) standard, which is an extension of H.264/AVC. Blu-ray discs will offer high-definition for both left and right views. The storage constraints are satisfied with the high compression capabilities of MVC, while also providing compatibility with existing 2D players.

The above model does not work so well for services such as cable, where cable operators carry the cost of the set-top box and have a large installed customer base. Replacement of set-top boxes is costly and could not be accomplished in a short time. Therefore, it is of greater interest in the near-term to utilize the capabilities of the existing distribution infrastructure and equipment. Frame-compatible formats offer a solution to introduce 3D services under such constrained environments, which also include the need to broadcast live events.

The remainder of this paper is organized as follows. In the next section, the typical constraints that are encountered at various points in the production and delivery chain are presented. An overview of the various frame-compatible formats that are being considered for distribution of 3D video through such legacy systems is provided in Section 3 including a review of the signalling that has been recently standardized as part of the H.264/AVC standard. The paper concludes with a discussion on the benefits and drawbacks of the different frame-compatible formats and the current industry status regarding the deployment and likely use of these formats.

2. SYSTEM CONSTRAINTS

Television and home entertainment have experienced many upgrades throughout history. Color television was introduced in the 1950s through a compatible extension of black-and-white transmission standards. The last decade has witness a conversion from analog to digital video services. Also, existing standard-definition (SD) video is is being upgraded to high-definition (HD). Industry is now considering a similar type of upgrade to 3D, and must be mindful of constraints in the production and delivery chain.
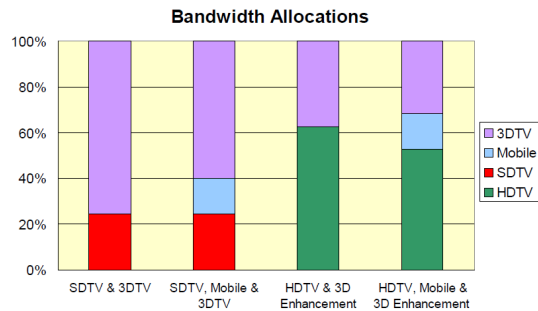
Fig. 1. Bandwidth allocation for terrestrial broadcast with 3DTV services.

2.1. Production

The main approaches to creating 3D content include camera capture, computer generated, and conversion from 2D video. Most 3D video that is captured use stereo cameras and there have been several recent products in the professional domain that capture stereo video in HD formats. However, even for setups that can capture this full-resolution stereo, there is still an issue of encoding it, transmitting it back to the network center or a local station, performing any necessary edits or program insertions and pushing the content back out to the consumer. The production workflow is substantially changed with the introduction of a second view. Other equipment upgrades would be required as well. With computer generated content, or 3D content converted from a 2D version, these same constraints may not exist.

An important element of the production domain is a master format. Whether the content is a 3D cinema production or a live event, the master format specifies a common image format along with high level metadata that are required to make sense of the data and prepare the data for distribution. The format is generally independent of any specific delivery channel. The Society of Motion Picture and Television Engineers (SMPTE) will specify a *3D Home Master* which would essentially be an uncompressed and high-definition stereo image format, i.e., 1920×1080 pixel resolution at 60Hz per eye [1]. The mastering format will also specify metadata, e.g., signaling of left and right image frames, as well as scene information such as the maximum and minimum depth of a scene. The master format is also expected to include provisions to associate supplementary data such as pixel-level depth maps. Derivatives of this master format, including frame-compatible formats discussed in the next section, could be created for each individual distribution channels.

2.2. Transmission

Cable operators have been actively considering the options for delivery of 3D video [2]. While bandwidth is not a major issue in the cable infrastructure, the set-top boxes to decode and format the content for display is a concern. A 3D format that is compatible with existing set-top boxes would enable faster deployment of new 3D services. It has bene recognized that a frame-compatible format could be useful for this purpose, while new boxes that support full-resolution formats may be introduced into the market at a later stage. The Society of Cable Telecommunications Engineers (SCTE), which is the standards organization that is responsible for cable services, is considering this roadmap.

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, so there ae legacy format issues to content with that limit the channel bandwidth that could be used for new services. A sample bandwidth allocation considering the presence of HD, SD and mobile services is shown in Fig. 1. 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. Broadcast of 3D video is likely to lag behind other distribution channels for these reasons.

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 TV, or the PC must be able to decode and have a suitable interface with the TV.

2.3. Interfaces & Displays

Since there are a number of displays already on the market that use different formats, the interface from distribution formats to native displays formats is a major issue. There is currently a strong need to standardize the signaling and data format to be transmitted between the various devices in the home.

On the TV side, there are a few issues that need to be addressed to ensure that legacy devices that only support their native display capability could still be utilized for new 3D services. In order for these TVs to operate in a 3D mode, the source material must be delivered in the native display format. This could be accomplished by either ensuring that service (or source device) provides a 3D format that exactly matches the display capabilities, or by performing the necessary conversion prior to the display. The former might be impossible to achieve in practice since the distribution format is generally different than the native display format. The latter is more practical, but would likely require an external conversion box as an interface between the source device and 3D-Ready TV. It is important to note that when the two formats have different sub-sampling structures, the quality of the conversion needs to be considered.
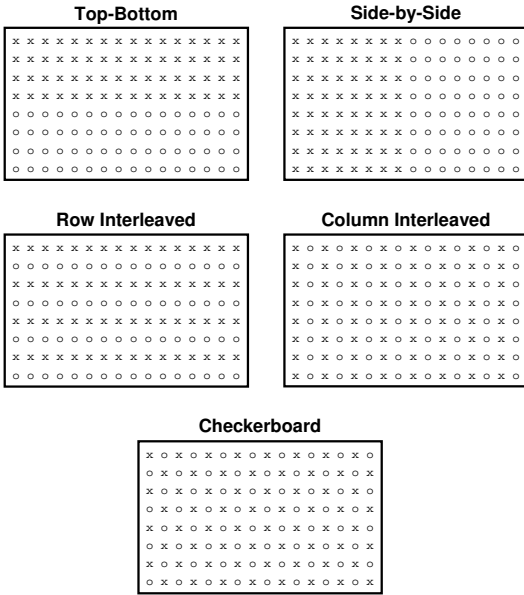
Top-Bottom	Side-by-Side

```
x x x x x x x x x x x x x x    x x x x x x x o o o o o o o o
x x x x x x x x x x x x x x    x x x x x x x o o o o o o o o
x x x x x x x x x x x x x x    x x x x x x x o o o o o o o o
x x x x x x x x x x x x x x    x x x x x x x o o o o o o o o
o o o o o o o o o o o o o o    x x x x x x x o o o o o o o o
o o o o o o o o o o o o o o    x x x x x x x o o o o o o o o
o o o o o o o o o o o o o o    x x x x x x x o o o o o o o o
o o o o o o o o o o o o o o    x x x x x x x o o o o o o o o
```

Row Interleaved	Column Interleaved

```
x x x x x x x x x x x x x x    x o x o x o x o x o x o x o
o o o o o o o o o o o o o o    x o x o x o x o x o x o x o
x x x x x x x x x x x x x x    x o x o x o x o x o x o x o
o o o o o o o o o o o o o o    x o x o x o x o x o x o x o
x x x x x x x x x x x x x x    x o x o x o x o x o x o x o
o o o o o o o o o o o o o o    x o x o x o x o x o x o x o
x x x x x x x x x x x x x x    x o x o x o x o x o x o x o
o o o o o o o o o o o o o o    x o x o x o x o x o x o x o
```

Checkerboard

```
x o x o x o x o x o x o x o
o x o x o x o x o x o x o x
x o x o x o x o x o x o x o
o x o x o x o x o x o x o x
x o x o x o x o x o x o x o
o x o x o x o x o x o x o x
x o x o x o x o x o x o x o
o x o x o x o x o x o x o x
```

Fig. 2. Common frame-compatible formats, where x represents the samples from one view and o represents samples from the another view.

Regarding the interface to the TV, HDMI v1.4 has recently been announced and includes support for a number of uncompressed 3D formats including both full-resolution and frame-compatible formats [3]. Efforts are underway to also update other digital interface specifications including those specified by the Consumer Electronics Associations (CEA). There are also new initiatives within CEA to standardize the specification of 3D glasses, as well as the interface between display devices and active glasses [4]. Another major constraint of existing 3D-Ready TVs is that they typically support an older version of the interface that was not specifically designed for 3D, e.g., HDMI v1.3. While such interfaces are capable of supporting the required bandwidth for a wide variety of 3D formats, there is no signaling in place to identify the format being sent. Therefore, upgrades need to be made so that existing devices could be identify the format of the content and display it correctly.

3. FRAME-COMPATIBLE FORMATS

Frame compatible formats refer to a class of formats in which the stereo signal is essentially a multiplex of the two views into a single frame or sequence of frames. Some common formats are shown in Figure 2. Other common names include stereo interleaving or spatial/temporal multiplexing formats. In the following, a general overview of these formats along with the key benefits and drawbacks are discussed. A standardized signalling for these formats is also described.

3.1. Overview

With a frame-compatible format, 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, a quincunx sampling may be applied to each view and the two views interleaved with alternating samples in both horizontal and vertical dimensions. Alternatively, the two views may be decimated horizontally or vertically and stored in a side-by side or top-bottom format, respectively.

Time 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 a that of a single view.

Frame-compatible formats have received considerable attention from industry since they facilitate the introduction of stereoscopic services through existing infrastructure and equipment. Representing the stereo video in such a way that is compatible with existing codecs and delivery infrastructure is the major advantage of this format. As a result, the video 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 can be quickly deployed to capable displays, which are already in the market.

The 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. An additional issue with frame-compatible formats is distinguishing the left and right views. To perform the de-interleaving, some out-of-band signaling is necessary. Since this signalling may not be understood by legacy receivers, it may not possible for such devices to extract, decode and display a 2D version of the 3D program. This might not be so problematic though. For one, it is not always the case that a 2D version of the content should be extracted from a 3D stream. The production may be different; also 2D and 3D versions may be edited differently. Second, the firmware on some devices, such as cable set-top boxes, could be upgraded to understand the new signaling that describes the video format. The same is not necessarily true for broadcast receivers and all types of equipment.

3.2. Signaling

The signalling for a complete set of frame-compatible formats has been standardized within the H.264/MPEG-4 AVC standard as Supplementary Enhancement Information (SEI). In general, SEI messages provide useful information to a decoder, but are not a normative part of the decoding process. However, a decoder that understands the SEI message can interpret the format of the decoded video and display the stereo content appropriately.

An earlier edition of the standard that was completed in 2004 specified a Stereo SEI message that identifies the left view and right view. More specifically, it was able to indicate a line interleaving of views that would be represented as individual fields of a video frame or a temporal multiplexing of views where the left and right views would alternate in time. The Stereo SEI message also has the capability of indicating whether the encoding of a particular view is self-contained, i.e., frame or field corresponding to the left view are only predicted from other frames or fields in the left view. Inter-view prediction for stereo is possible when the self-contained flag is disabled.

The functionality of the Stereo SEI message has recently been combined with the additional signaling for the various spatially multiplexed formats described above as a new SEI message. The new SEI message is referred to as the Frame Packing Arrangement SEI message, and has been specified as an amendment of the AVC standard [5]. This new SEI message is able to signal all the different frame packing arrangements shown in Fig. 2. With the side-by-side and top-bottom arrangements, it is also possible to signal whether one of the views has been flipped so as to create a mirror image in the horizontal or vertical direction, respectively. Independent of the frame packing arrangement, the SEI message also indicates whether the left and right views have been subject to a quincunx (checkerboard) sampling. For instance, it is possible to apply a quincunx filter and subsampling process, but then rearrange the pixels into a side-by-side format. The benefits of this will be discussed in the next section. Finally, the SEI message indicates whether the upper-left sample of a packed frame is the left or right view and additional offset values to indicate the grid position of samples for the left and right views, up to a precision of one sixteenth of the luma sample grid spacing.

There are two addition points that are worth noting. First, the arrangement of samples from left and right views into a single frame does not imply subsampling. The SEI message does not assume that the source format prior to encoding is known. While the left and right views in packed arrangement may be sub-sampled, it is also possible that the left and right views are not subsampled. Second, there is additional information carried in the Video Usability Information (VUI) to indicate whether any further resizing may be needed. The sample aspect ratio (SAR) syntax describes the intended horizontal distance between the columns and the intended vertical distance between the rows of the decoded frames.

4. DISCUSSION

Industry is now preparing for the introduction of new 3D services. With the exception of Blu-ray Discs, which will offer a full-resolution stereo format with HD resolution for each view, the majority of services will start this year based on frame-compatible formats. Some benefits and drawbacks of the various formats are discussed below.

In the production and distribution domains, side-by-side and top-bottom formats are most favored. Relative to row or column interleaving, and the checkerboard format, the quality of the reconstructed stereo signal after compression can be better maintained. Such formats introduce significant high frequency to the frame-compatible signal thereby requiring higher bit-rate. Also, the interleaving and compression process has the possibility to create cross-talk artifacts and color bleeding.

From the pure sampling perspective, however, there have been studies that discuss the benefits of quincunx sampling. In particular, quincunx sampling preserves more of the original signal and its frequency-domain representation is aligned with that of the human visual system. So, while it may not be a distribution-friendly format, quincunx sampling followed by a rearrangement to side-by-side or top-bottom format could potentially lead to higher quality compared to direct horizontal or vertical decimation of the left and right views by a factor of two.

Another issue to consider regarding frame-compatible formats is whether the source material is interlaced. Since the top-bottom format incurs a loss in the vertical dimension and an interlaced field is already half the resolution of the frame, the top-bottom format should not be used with interlaced content.

Since there will be displays in the market that support interleaved formats as their native display format, such as checkerboard for DLP televisions and row interleaving for some LCD-based displays, it is likely that the distribution formats will be converted to these display formats prior to reaching the display. Therefore, the signaling of these formats over the interface would be necessary along with the signaling of the various distribution formats.

The SEI message that has been specified in the latest version of the AVC standard supports a broad set of possible frame-compatible formats. It is expected to be used throughout the delivery chain from production to distribution, through the receiving devices, and all the way to the display.

5. REFERENCES

[1] SMPTE, "Report of SMPTE Task Force on 3D to the Home," 2009.

[2] D.K. Broberg, "Considerations for stereoscopic 3D video delivery on cable," in *IEEE International Conference on Consumer Electronics*, Las Vegas, NV, 2010.

[3] HDMI Licensing, LLC., "HDMI Specification 1.4," 2009.

[4] M.W. Stockfisch, "Prospective standards for in-home 3D entertainment products," in *IEEE International Conference on Consumer Electronics*, Las Vegas, NV, 2010.

[5] G.J. Sullivan, A.M. Tourapis, T. Yamakage, and C.S. Lim, "Draft AVC amendment text to specify Constrained Baseline profile, Stereo High profile, and frame packing SEI message," in *Joint Video Team, Doc. JVT-AE204*, London, UK, 2009.