IMPLEMENTATION OF STEREOSCOPIC 3D SYSTEMS ON CABLE

David K. Broberg, Mark Francisco
Cable Television Laboratories, Inc., Comcast Cable Communications, Inc

Abstract

Three-dimensional television is already upon is. Most of the major TV brands are introducing new models this year and some are already shipping. A tremendous effort has been made to pave the way for the introduction of stereoscopic 3D services over cable systems. This paper describes some of the implementation challenges that have already been overcome as well as some that remain.

INTRODUCTION

This paper divides the discussion of the impacts of deploying stereoscopic 3D (S3D) along the lines of the signal flow. Considerations for the formatting and ingest of S3D signals will be provided along with implications on encoders and headend processing. The importance of proper signal identification and solutions for downstream processing will be explained. Next the implications on the set-top box (STB) will be explored including impacts on firmware, on-screen-graphics and menus as well as closed caption decoding. Finally a summary of TV interface issues will be described.

STEREOSCOPIC VIDEO FORMATS

Frame-Compatible Formats

Stereoscopic video is captured from two identical cameras, one for each eye. To transmit the two pictures would typically require two identical signal paths, two wires, two separate video streams or two channels. Such a system is inefficient and many alternate approaches are available that exploit various aspects of the redundancy between the two images, reducing bandwidth demands by eliminating duplicate components while maintaining sync. Spatial multiplexing is one such mechanism that results in a frame-compatible signal.

A frame-compatible delivery of stereoscopic 3D video makes use of existing MPEG, AVC or VC1 coding standards [1]-[4]. It also uses various spatial filtering techniques to pack separate left and right images into a single frame or stream which can then be delivered through existing systems using only one channel or video stream. A frame-compatible approach delivers the stereoscopic content as if it were a regular 2D video stream and is otherwise compliant with 2D video standards.

Frame-compatible stereoscopic 3D delivery is technically possible using any type of multiplexing technique that is able to repackage the separate left-eye and right-eye images into the space and format normally used for 2D video transmission. Separate left and right images can be spatially reduced horizontally or vertically and squeezed to fit into a single video frame as a side-by-side image or as a top/bottom image, as illustrated below in figure 1 and figure 2.

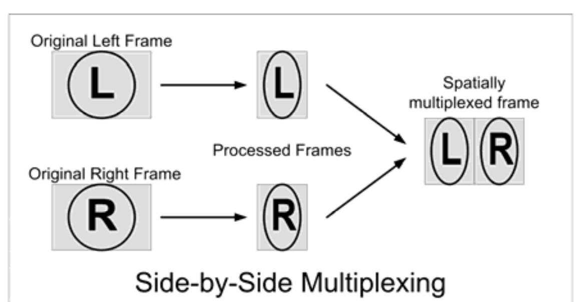

Fig. 1. Side-by-side frame packing is one example of a spatial multiplexing technique used to transport stereoscopic video using existing encoding systems.

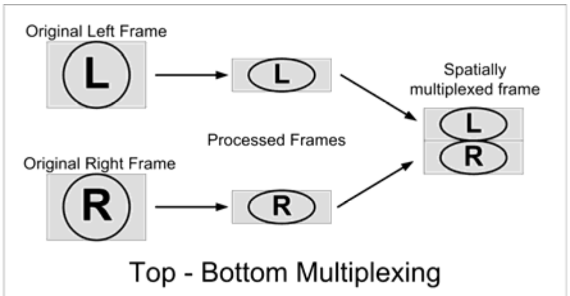

Fig. 2. Top and bottom frame packing is another example of a spatial multiplexing technique used to transport stereoscopic video using existing encoding systems.

The separate left and right images also can be spatially multiplexed on line-by-line, column-by-column or even in checkerboard patterns to be interleaved into a single frame. Different techniques and algorithms can be applied to the spatial reduction filtering with different levels of performance. Each of these methods has the consequence of reducing the spatial resolution overall. This loss of spatial resolution is exchanged for the inclusion of stereoscopic depth, which is conveyed as the horizontal disparity between the left and right images.

There are numerous variations within each of these format subgroups. Variations are created by unique subsampling or spatial reduction filtering. More variations are possible based on whether the left image is first or the right image is first, and whether the reduced images are flipped, mirrored or inverted. Illustrations in Figures 1 and 2 show just two examples of the many possible spatial multiplexing techniques.

In each of these spatially multiplexed variations a processor reduces the frame size of the original left and right video signals from the stereoscopic source so that both images may be packed or combined into a single video frame. The new spatially-multiplexed frame includes both the left and right views.

Separate left and right video sequences also can be temporally multiplexed into a common video stream by alternating frames or fields in a left-right-left-right frame-sequential pattern. Such a method would have the advantage of preserving the full spatial resolution at the expense of compromising the temporal resolution. While left and right stereoscopic signals are used as in the examples, frame-compatible techniques can be applied equally to 2D plus depth or 2D plus difference signals.

Any of these frame-compatible solutions can be compressed and encoded as if they were an ordinary 2D video frame. However, some of these systems are better able to survive encoding and decoding processes without the introduction of new errors or modifications needed to accommodate the new format.

Once these frame-compatible signals reach the 3D display, a stereoscopic processor must demultiplex the combined frame using an algorithm that complements or at least approximates the one used in the encoding process. This stereo-demultiplexing process restores the original left and right views.

Without any effort to narrow the number of usable choices, the market could see some programmers choosing one format while others chose a different format. Choices might even be made on a program-by-program basis finding reasons to favor the merits of one technique over another based on the nature of the content itself. Business arrangements might also influence these choices, as a number of the potential methods may need to be restricted to license agreements with IPR holders.

If so many variations of frame-compatible delivery coexist in the market the complexity of the stereoscopic demultiplexer increases dramatically with nearly unlimited multiplexing variations. To achieve a successful 3D delivery system, the number of choices needs to be dramatically reduced while preserving the flexibility to work with a variety of existing equipment, content types and video formats. Current plans for cable have limited these choices to just three video formats and two frame-compatible systems:

1. 1280x720p60 – Top-and-Bottom formatting
2. 1920x1080p24 – Top-and-Bottom formatting
3. 1920x1080i60 – Side-by-Side formatting

Full-Frame Stereoscopic Formats

The delivery of two full-resolution frames requires other mechanisms to optimize the transmission and eliminate redundancy between the separate stereo source streams.

Using AVC multi-view coding standards optimized for the carriage of stereoscopic signals [5]-[9] would seem to be a logical choice for delivery over cable TV systems. However, delivery of stereoscopic 3D content using this approach comes with the cost and deployment delays associated with the introduction of new equipment and systems designed for these signals.

For cable operators the cost of replacement STBs and other headend equipment is a sizable consideration. If the demand for stereoscopic 3D content in the home develops rapidly, it is more likely that the cable operators could justify the cost of deploying new MVC solutions. However, without a low-cost interim method of delivery such as the frame-compatible approach that market may never develop.

Within the scope of the MVC coding standard there are also numerous possible options for the delivery of stereoscopic content. These various competing systems must be evaluated as part of the road-map to bring full resolution stereoscopic content for both eyes.

(1.) Discrete Left and Right Signals

The MVC coding system is designed to support a primary or base view along with a secondary (or non-base) view. The base view can be the left or the right view while the secondary view can be the opposite view for stereoscopic delivery. (The MVC extension to H.264/AVC also supports free-view and multi-view image coding with more than two secondary views, but for the purpose of this paper, the analysis is limited to the two-view stereoscopic coding.) The advantage of such a system is its simplicity. Existing AVC decoders that lack the ability to decode more than one stream simultaneously can receive the 2D compatible stream in the main channel (base view) and simply ignore or discard the secondary or alternate view [10]. New receivers would be designed with the ability to decode simultaneously two HD streams (MVC) so that they are able to decode both the base layer and the secondary layer, producing a stereoscopic output.

(2.) 2D plus delta or 2D plus difference

Another variation of MVC coding makes use of a pre-coding subtraction algorithm in order to reduce information in the secondary stream. The main signal is simply the left-eye view while the delta or difference signal is left-eye view minus the right-eye view (L + L-R) [11]. The presumed advantage of such a system is the reduced information content in the secondary stream.

(3.) 2D plus depth, 2D plus depth, occlusion and transparency (DOT).

This variation of MVC coding was primarily developed for the support of multi-view and free-view displays rather than stereoscopic displays. While separate left and right viewpoints can be derived from 2D plus depth and 2D plus DOT successfully, the processing requirements are much greater in the receiver and performance may be lower.

(4.) Frame-compatible with enhancement signals

It is also possible to encode frame-compatible formats using MVC. In this case the primary spatially-multiplexed frame is encoded as the base view while a complementary spatially-multiplexed frame or enhancement signal designed to restore the missing resolution is encoded as the secondary stream or view [12].

In this case a receiver needs to decode both the base view and the secondary stream in order to produce a full-resolution 2D or 3D view since the base view only includes ½ of the available resolution for the left-eye and right-eye views. However, such an approach can still provide a frame-compatible stereoscopic signal to legacy receivers unable to decode the secondary stream. Such a choice enables a more gradual migration if the cable operator uses frame-compatible delivery initially.

3D FORMAT SIGNALING

Signaling and detection of the specific 3D format transmitted to the receiving device are necessary to avoid manual configuration to view 3D. They are also necessary due to the potential for reconfiguration when content changes format between programs in the case of channel changes or within a program in the case of commercials. Providing identification of 3D format within the stream can aid in set-top graphics formatting, television configuration and allowing features such as EAS and closed-captioning to function properly while operating in stereo. MPEG4 Part 10 specifies Supplemental Enhancement Information (SEI) that includes multiplex descriptions for 3D content [13]. A method of extending the SEI to MPEG-2 is being proposed that carries the equivalent descriptors as user data. The descriptors are intended to convey the frame packing arrangement of the content, which is expected to be one of the following:

Side by Side (type 3)
Top Bottom (type 4)
Checkerboard (type 0)

Note checkerboard is not likely to be used due to issues with the 4:2:0 chroma subsampling[1] used in most MPEG profiles, but it is included to align with the HDMI supported formats. Each frame configuration can map to a supported HDMI signaling should the content pass through the STB unaltered.

STB CONSIDERATIONS

The addition of SEI or MPEG-2 signaling in 3D content may allow the STB to format graphics accordingly, scale video appropriately and signal the television operational mode without user interaction. This can enable automatic adaptation to stereoscopic content, and in-program switching if interstitial 2D content is present. Note the term STB (set-top box) is used to represent any device that receives MPEG-2 transport streams and provides a digital display interface compatible with HDMI.

3D Signal Detection & Mode Switching

An MPEG section filter is required to detect 3D signaling within the video bitstream. Once implemented, a program change due to channel change, program boundary or interstitial, may trigger a mode change within the STB. This is likely to occur without user interaction, but may involve on-screen or front-panel messaging. The resultant mode change may depend upon the type of television detected. If a 3D TV is detected that supports HDMI 1.4 or 1.4a, automatic switching will occur. If no 3D capabilities are detected, an on-screen message can be displayed to request the consumer to change the TV mode appropriately. Options to allow 2D viewing

[1] Since the color space is half the resolution of the luminance, quincunx, or checkerboard sampling results in 3D image degradation.

of the content can be presented dependent upon STB capability. Upon automatic or manual progression to 3D operation, the STB will format subsequent graphics appropriate for the 3D frame packing arrangement.

3D Graphics Rendering

In order for STB-generated graphics to appear properly when viewing 3D content, the graphics must be formatted with an equivalent frame multiplex. For example, if top bottom video is being received, the set-top generated graphics should be formatted top bottom prior to bit blending in the destination video buffer. Additionally, it may be desired to shift the images horizontally to have the graphics appear slightly in front of the video plane, providing a natural viewing experience. This may be particularly important when affecting captions, which tend to stay on-screen for longer viewing durations.

3D Video Scaling

Video scaling is sometimes used to allow the currently viewed program to remain on screen while viewing a program guide or other interactive features. Scaling 3D video involves a choice of associated television mode and maintenance of 3D imagery of the video. It may be desirable to revert the television from 3D viewing to allow a simple 2D projection of scaled video and guide. This may be required due to the complexity of the video scaling, such as multiple thumbnails projected, or due to limitations in the STB processing and memory. Alternatively, a single eye image of the video may be scaled linearly and copied to the appropriate frame multiplex format. A second alternative maintains a 3D projection of the scaled video, which requires the left and right components of the frame to be scaled linearly and copied to the appropriate locations for the frame multiplex. Video scaling options are illustrated in Figure 3.

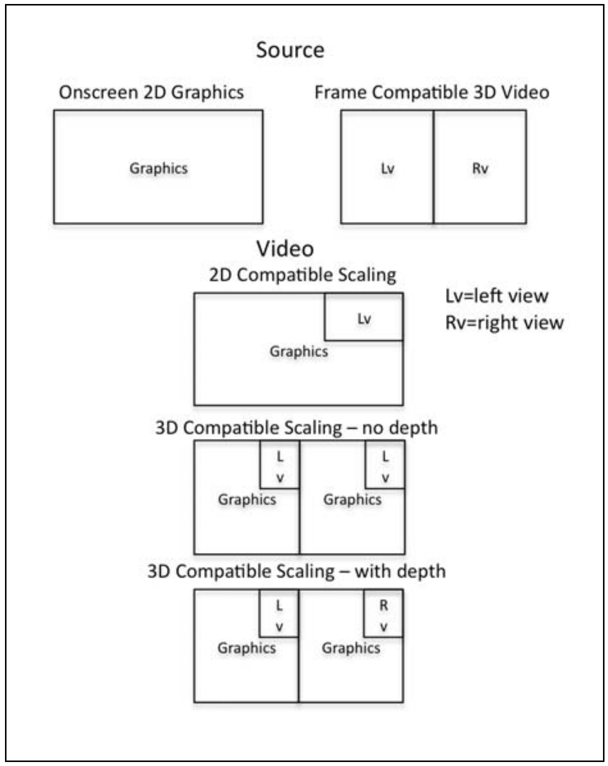

Fig. 3. Methods of combining graphics and video in a scaled window for 3D presentation

Television Signaling

Televisions supporting HDMI 1.4 and 1.4a 3D format signaling are currently available. Testing has shown the response to 3D signaling of currently available models to be without perceivable delay. This indicates the ability to support 3D mode switching at program start and stop, channel changes and within program boundaries. The ability to add interstitial 2D elements is possible without consumer perception of a program disruption. 3D eyewear may behave differently dependent upon implementation. It is generally a better experience to deliver a uniform program format to avoid the variation in light transmission that occurs when active 3D glasses temporarily stop shuttering. Formatting 2D advertising in a frame-compatible format may provide the best overall experience.

IPTV Considerations

While the methods described in this paper are primarily focused around delivery of 3D assets on QAM-based multi-program transport streams, application to IP and Internet delivery systems is readily achieved. IP encapsulation of MPEG-2 transport streams is typically used for network transport between source and edge QAMs, DOCSIS® delivery is readily achieved as delivery increasingly moves to IP oriented Content Delivery Network (CDN) architectures. Alternatively file-based delivery is possible although equivalents to the defined user data and SEI structures of MPEG are needed. One advantage of IP delivery is the ability to independently carry left and right eye images for adaptive edge multiplexing. Care must be taken to maintain synchronization of left, right and audio streams due to possibilities for propagation variation across IP networks.

Output Formatting and Rescaling

Another important consideration in the STB is the output formatting or rescaler. For 2DTV, the typical operation of the STB has been to adjust the various transmission video formats into a single video format preferred by the display through the use of a rescaler. For example HD video signals delivered as 1280x720p60 may be rescaled by the STB into a 1920x1080i60 format to drive the monitor. Conversely, depending upon the monitor, 1920x1080i60 HD content may be downscaled to 1280x720 and de-interlaced before being sent to the display.

For 3D signals such conversions can be far more detrimental to the 3D picture performance. Since the frame-compatible signals are already spatially reduced either horizontally or vertically, further spatial rescaling may add cumulative losses to the resolution, as well as destroy certain pixel relationships necessary to be decoded accurately. The interlace-progressive relationships must also be carefully preserved to assure the optimal 3D representation.

For these reasons it is essential that the STB operate in a video pass-through mode where the STB does not rescale when 3D video is being delivered. When a 3D signal is delivered as 1280x720p60 it must be output as 1280x720p60 by the STB and signals received as 1920x1080i60 must be output as 1920x1080i60 accordingly.

This video by-pass mode was not always possible with 2D televisions because many were not equipped to support a wide range of video input formats and scan rates. However this is not the case with the modern 3DTVs, and virtually all of the new 3DTVs are able to handle this wider range of video formats as an input signal.

3DTV INTERFACES

Legacy HDMI

Today's deployed HD STBs include the HDMI interface based upon the previous version 1.3 or older specifications. These specifications included no reference or provision for stereoscopic 3D video.

STBs that support this interface are still able to deliver stereoscopic 3D content, with certain limitations. First, the formats must be frame-compatible and use the same exact timing and signaling as any 2D video signal. Second, there is no direct provision for any automated detection or switching from 2D to 3D with such a system.

HDMI v1.4

The version 1.4 of the HDMI specification was release in June of 2009 [14] and added specific support for the carriage of stereoscopic 3D formats. It also added additional signaling to enable the discovery and identification of stereoscopic capabilities as well as 2D and 3D signal identification.

Unfortunately, the HDMI v1.4 standard failed to mandate that 3DTVs support the frame-compatible formats necessary for cable delivery. This initial version also failed to identify or specify the needed Top-and-Bottom format. Without these necessary provision in the standard the market place uncertainty would make it very difficult to reliably deliver 3D video services to a wide range of products and models.

HDMI v1.4a

With the update to version 1.4a of the HDMI standard [15] the Top-and-Bottom format was added along with mandatory support for the three needed frame-compatible formats by 3D displays.

Another important change was also added at the same time as a change to the license restrictions. This change enabled the STBs that are limited to legacy HDMI v1.3 or older implementations to be able to selectively support the frame-compatible modes along with the format signaling and self-discovery features, without any obligation to support the higher bit rate full-resolution 3D formats mandated by version 1.4 [16].

This important change paved the way for firmware updates to be possible in existing STBs so that fully automated 3D support could be enabled.

Analog Component

Most new televisions, including the new 3DTV models, are still equipped with analog component interfaces. Many subscribers continue to use systems connected using these outdated analog connections. There is a risk that some who use these systems will upgrade to a new 3DTV and reconnect the existing analog component interfaces out of habit or to avoid the cost of upgrading to HDMI.

Most of the new 3DTVs don't offer true 3D viewing from the analog components. However, some include built-in 2D to 3D converters that can be operated on the analog component inputs. Since there is no bi-directional hand-shaking on this interface it is impossible for the STB to recognize the presence of the 3D capable TV. There is also no provision for 3D signal identification, so at best the experience would require a totally manual 2D/3D switching function.

3D-capable systems that are connected this way will only lead to disappointment and customer service issues and should be avoided.

RF or Direct QAM

Many of the new 3DTVs entering the market this year continue to include support for "clear-QAM". These sets don't include the CableCARD™ slot and are not fully qualified UDCPs. Nonetheless they are often able to decode SCTE-07 compliant QAM modulation when no conditional access encryption has been applied (clear-QAM). However there will be additional challenges using this approach for 3D delivery or reception beyond the usual problems, such as channel mapping and EAS support that plague 2D Clear-QAM sets.

Some of these sets can support the frame-compatible (broadcast) formats even from the tuner input, but are limited to MPEG2 video decoding and are not built compliant with SCTE standards for decoding AVC/H.264 or VC1. These sets will not be able to detect any supplemental data used to identify the 3D signals or format types, forcing them into a manual operational mode for 2D/3D switching at best. Some may actually stumble when the 3D signals are received by failing to properly ignore the supplementary data signals.

To avoid disappointment and customer complaints the clear-QAM connections should be avoided for 3D services.

IR Signaling

Finally we can't overlook the need for infrared (IR) signaling between the 3DTV and the electronic shutter glasses. While this is not an interface that is provided by the cable operator or the STB, it does use a shared physical media with the IR-remote control.

Presently the market for 3DTVs predominantly uses IR signaling to activate and sync the 3D glasses. These systems are non-standardized and a variety of techniques, protocols and formats are used. Some use a subcarrier or pulse modulation, while others use base-band signaling. Some are broadband and others are narrowband filtered.

The risk of this chaotic range of non-standard implementations is to interference with the control of existing STBs. This interference can potentially be in either direction. For example, the STB remote could cause sync disruptions to the 3D eyewear or, more likely, the 3D sync signals from the TV could disrupt the IR remote operation of the STB.

Until standards are fully developed in this area, it is likely that some updates, patches or other field fixes may be necessary in the STBs, the 3DTVs, or both, to avoid these problems.

CONCLUSION

Cable can deliver 3DTV programming today in formats compatible with the latest generation of 3DTVs. This paper has described a variety of technical challenges that must be overcome to ensure success, maximum performance and easy operation. Proposed transmission standards are presented here that may provide a more seamless 2D to 3D transition for customers. Progress has been made and will continue so that the 3DTV experience of cable customers in the home can offer a rich new dimension in TV viewing never seen before.

REFERENCES

[1] ISO/IEC 13818-2, International standard (2000), MPEG-2 video.
[2] ANSI/SCTE 128 2008 AVC video systems and transport constraints for cable television.
[3] ANSI/SCTE 157 2008 VC-1 video systems and transport constraints for cable television.
[4] ANSI/SCTE 43 2005 Digital video systems characteristics standard for cable television" section 5.1.2, table 3, p3.
[5] ITU-T Rec. & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," March 2009.
[6] Yip, P.Y.; Malcolm, J.A.; Fernando, W.A.C.; Loo, K.K.; Arachchi, H.K., "Joint source and channel coding for H.264 compliant stereoscopic video transmission," Canadian Conference on Electrical and Computer Engineering, 2005., vol., no., pp.188-191, 1-4 May 2005.
[7] Martinian, E.; Behrens, A.; Jun Xin; Vetro, A.; Huifang Sun, "Extensions of H.264/AVC for multiview video compression," Image Processing, 2006 IEEE International Conference on, vol., no., pp.2981-2984, 8-11 Oct. 2006.
[8] Wenxian Yang; Ngan King Ngi, "MPEG-4 based stereoscopic video sequences encoder," Acoustics, Speech, and Signal Processing, 2004 Proceedings. (ICASSP '04). IEEE International Conference on , vol.3, no., pp. iii-741-4 vol.3, 17-21 May 2004.
[9] Yang, W.; Ngan, K.N.; Cai, J., "An MPEG-4-compatible stereoscopic/multiview video coding scheme," Circuits and Systems for Video Technology, IEEE Transactions on , vol.16, no.2, pp. 286-290, Feb. 2006.
[10] Ying Chen, Ye-Kui Wang, Kemal Ugur, Miska M. Hannuksela, Jani Lainema, and Moncef Gabbouj, “The emerging MVC standard for 3D video services,” EURASIP Journal on Advances in Signal Processing, vol. 2009, Article ID 786015, 13 pages, 2009.
[11] Ethan Schur, "Digital Stereoscopic Convergence Where Video Games and Movies for the Home User Meet," HDdailies.com, Mar. 2009.
[12] Walt Husak, "Dolby 3D White Paper", unpublished.

[13] Amendment to ITU-T Rec. H.264 | ISO/IEC 14496-10: "Constrained baseline profile and supplemental enhancement information," in-progress.
[14] HDMI Licensing, LLC: "High definition multimedia interface specification version 1.4", June 5, 2009.
[15] HDMI Licensing, LLC: "High-Definition Multimedia Interface Specification Version 1.4a Extraction of 3D Signaling Portion", Mar 4, 2010.
[16] HDMI Licensing, LLC "HDMI Licensing, LLC communicates further detail on 3D requirements within the HDMI specification version 1.4" Sunnyvale, California, December 23, 2009.

ABOUT THE AUTHORS

David Broberg
Vice President, Consumer Video Technology
CableLabs

Mr. Broberg has been with CableLabs® since 1999 and was a principal developer of the OpenCable™ hardware architecture. As a member of the System Architecture & Design Group, he is currently responsible for identifying, analyzing and developing key strategic video technology and related specifications and standards. Mr. Broberg has more than 30 years of technical and business experience in the television industry including broadcast, cable, satellite and consumer electronics and holds four awarded patents in the field. Prior to joining CableLabs, Mr. Broberg served in several key strategic engineering management and product development roles for Mitsubishi Electric including digital broadcasting and consumer electronics.

In 2009, Mr. Broberg received the Excellence in Standards Award from the Society of Cable Telecommunications Engineers. Mr. Broberg is a member of the SMPTE Task Force on 3D, the CEA 3D Task Force and chairs the SCTE Digital Video Subcommittee 3D Ad Hoc Group. Mr. Broberg is a Sr. Member of IEEE and a veteran of the US Air Force where he began his career in video technology and obtained an Associate in Applied Arts & Sciences, Electronic Technology from Mountain View College in 1980. Mr. Broberg is also a member of the International Stereoscopic Union and enjoys stereoscopic photography and video as a hobby.

Mark Francisco
Fellow, Office of the CTO
Comcast

Mr. Francisco has been with Comcast since 2001 and has been developing premises devices and services for Comcast's video, Internet and telephone customers. As a member of the Office of the CTO, he is currently responsible for video product architecture supporting broadcast, switched broadcast, on demand and internet delivery networks. Mr. Francisco has over 25 years of experience in design and systems integration with RF communications systems and holds three patents in the field, with four applications pending. Prior to joining Comcast, Mr. Francisco led systems integration and test activities for Motorola's New Jersey digital mobile phone design center and led the development of telemetry receivers and transmitters for Lockheed Martin Astro-Space Division. Mr. Francisco received a Masters of Science Degree in Electrical Engineering from Drexel University in Philadelphia, PA, and a Bachelors of Science Degree in Electrical Engineering from Rutgers College of Engineering in Piscataway, NJ.