Final Report of the ATSC Planning Team on 3D-TV

Doc. PT1-049r1
31 August 2011

Advanced Television Systems Committee
1776 K Street, N.W.
Suite 200
Washington, D.C. 20006
www.atsc.org

The Advanced Television Systems Committee, Inc., is an international, non-profit organization developing voluntary standards for digital television. The ATSC member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

Specifically, ATSC is working to coordinate television standards among different communications media focusing on digital television, interactive systems, and broadband multimedia communications. ATSC is also developing digital television implementation strategies and presenting educational seminars on the ATSC standards.

ATSC was formed in 1982 by the member organizations of the Joint Committee on InterSociety Coordination (JCIC): the Electronic Industries Association (EIA), the Institute of Electrical and Electronic Engineers (IEEE), the National Association of Broadcasters (NAB), the National Cable Telecommunications Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). Currently, there are approximately 140 members representing the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

ATSC Digital TV Standards include digital high definition television (HDTV), standard definition television (SDTV), data broadcasting, multichannel surround-sound audio, and satellite direct-to-home broadcasting.

Table of Contents

Index of Tables and Figures

Final Report of the ATSC Planning Team on 3D-TV

1. EXECUTIVE SUMMARY

ATSC Planning Team 1 (PT-1) has studied 3D television with a goal of providing a report to the ATSC on the benefits and limitations of a standard or a set of standards for terrestrial delivery of 3D-TV over an ATSC transmission system. This report covers three primary elements:

- Visual Sciences
- Technology
- Content.

1.1 Visual Sciences

The human visual system experiences the world in true three-dimensional space. 3D television simulates this visual experience through the presentation of two distinct views separately to the eyes. The advantage of 3D television over regular television is that it provides a significant enhancement to the perceived depth of 3D objects in a depicted scene. However, the simulation is not perfect and, depending on how the 3D content is created, processed and displayed, the end users' experience can either be enhanced or made uncomfortable and potentially painful.

Whereas the fundamental science of stereoscopic 3D displays is common across all platforms, an understanding of characteristics of the human visual system and its interaction with the visual environment is critical to insure an enhanced viewer experience. The report provides an overview of the human visual system and explains the differences between viewing in a real three-dimensional world to a virtual 3D world depicted using a two-dimensional screen. A number of physiological limitations associated with viewing stereoscopic 3D are explained. Important concepts of "disparities" in the images; "binocular fusion" for singleness of vision; and "comfort zone" for taking into account the unnatural dissociation between "focusing" of the lens of the eye and "vergence" eye movements are introduced for the proper display of stereoscopic 3D content. The report indicates that stereoscopic 3D content produce an enhanced viewing experience along more than one perceptual dimension, such as naturalness, sharpness, and presence. Importantly, viewers tend to prefer 3D content over 2D, but only when the 3D version is comfortable to view and any visible artifacts that arise from anywhere along the delivery chain are not annoying in the 3D presentation. The report also notes that research methodologies and findings on the perceptual impact of 3D content, relative to that of 2D content, for mobile and handheld devices (with small screens) are somewhat variable.

Substantial sections of the report deal with the much publicized issues of visual discomfort and fatigue associated with 3D content. Multiple sources along the production and delivery chain can contribute to a negative viewer experience. Many of the symptoms and contributing factors are described and explained and accommodated by insuring proper viewing distances from the screen as well as understanding the limitations of depicting stereoscopic 3D objects on a 2D display. Individual differences and general population characteristics are also mentioned as important factors to consider. Understanding the issues raised in these sections is vital to the creation of 3D content that is compelling for the viewer without distraction or discomfort. Despite these issues, enough is known to produce practical production and display guidelines. Furthermore, it is expected that 3D displays will allow end users to adjust the depth information contained in a scene to match individual's preference. Finally, the report brings up the importance of further studies that are required for examining the potential long term effects of viewing stereoscopic 3D contents. Since stereoscopic 3D content on 2D displays disassociates vergence and accommodation and since this natural association develops over time in young children, the question of the sensitive period in which potential disruption or negative impact of that development can occur needs to be addressed. For broadcasters, a solution for practical means to measure and monitor image and depth quality of 3D content is required.

In summary, there is no doubt that creating and displaying 3D content offers many benefits to increasing the viewer experience and enhancing revenue. The potential downside for broadcasters is that it will require more bandwidth, which is scarce, and if done improperly, can result in a negative and potential painful experience for their viewers. Clearly, technology can solve many of the issues but given the subjective nature of the impact of many of the factors, more information is needed to better understand and implement the services.

1.2 Technology

This report documents the benefits and limitations of various 3D data formats and encoding architectures to support the broadcast of 3D services to the home using the existing ATSC modulation system based on 8-VSB (with or without M/H service). Three broad types of 3D broadcast services are analyzed including real-time transmission to fixed receivers, real-time transmission to mobile receivers and non-real-time (NRT) delivery of 3D content.

Given recent market developments, there is interest in exploring standardization of 3D broadcast within ATSC. Since NRT delivery of 3D content is already within the scope of ongoing ATSC 2.0 work, a direct set of recommendations on formats to be specified for NRT delivery of 3D content is provided as input to the Technology Specialist Groups that are developing the suite of ATSC 2.0 specifications.

It is recognized that there are different regional needs and varying levels of potential 3D services. Therefore, it is expected that multiple 3D profiles will be defined as part of a 3D broadcast standard and that different transmission scenarios would be supported, including scenarios in which the 3D program is independent with the 2D program, as well as scenarios in which the 3D program is dependent on the 2D program. The profiles themselves should utilize advanced codecs given the significant benefits in terms of bandwidth efficiency and deployment feasibility. Various deployment strategies are also possible to enable service upgrades and transitions.

It is expected that this report be used as a reference in the development of 3D broadcast standards to understand the availability, capabilities and limitations of existing 3D technology as well as technology that is under development. This report is intended to help facilitate a strategy and develop requirements for the deployment of 3D services, and prompt further questions on the technology itself.

1.3 Content

Producing a program in 3D can be a very complex undertaking. The cameras are more complex, and suitable camera positions may be different from 2D. The “production grammar” for 3D differs from that in 2D because many of the techniques that make a 2D program more interesting (pans, zooms, scene cuts) can make the 3D audience uncomfortable. Live sporting events, in particular, are using dual productions in order to yield acceptable 2D and 3D versions; this greatly increases the cost to cover the event and makes finding a viable business model for 3D even more difficult.

Part I: Visual Sciences

1. SCOPE

The purpose of this report is to provide an overview of the human visual sciences that are relevant to the broadcast of stereoscopic three-dimensional television (3D-TV) services. This report summarizes:

1) Current knowledge of human stereoscopic vision and depth perception
2) Problems and issues related to human visual perception and the current technology and technology-in-development for broadcast delivery of stereoscopic 3D-TV
3) Existing knowledge of viewers' requirements, relevant to 3D-TV viewing for both fixed (e.g., home) and mobile receivers, with respect to stereoscopic 3D image quality, depth quality, visual comfort, and safety for the general public
4) Matters that are important and relevant to 3D-TV, but for which there is not a current knowledge base

2. INTRODUCTION

2.1 Stereoscopic 3D-TV Terrestrial Broadcasting

Stereoscopic 3D-TV terrestrial broadcasting refers to the delivery, over air using standard television channels, of additional information that would allow receivers to render different images for the two eyes. The different images provide two viewing perspectives, with one image designated for viewing by the left eye and the other by the right eye. The image pairs are intended to be viewed simultaneously or very nearly simultaneous.

2.2 Motivation: Benefits for Broadcasters and End Users

Stereoscopic 3D-TV broadcasting has potential benefits for both broadcasters and end users. For broadcasters, the potential benefits are: a) improved competitiveness against new multimedia services, b) an increase in market share, c) an increase in advertising revenue, d) an opportunity to charge viewers for a premium service, and e) opportunities for new programs and services[1]. For end users, the potential benefits are: a) an increased choice of programs and services, and b) enhanced visual experiences.

2.3 General Issues: Bandwidth and Visual Health Concerns

Other than implications of cost, two fundamental issues need to be considered in the development of a standard(s) for stereoscopic 3D-TV broadcast. One is the technical challenges of providing essentially twice the amount of information with minimal impact or change in the current standards and infrastructure for delivery of television video signals. The other is the oftentimes and much-conversed negative side effects of watching stereoscopic images, such as watery eyes, visual strain, headaches, and sometimes blurred or even double vision. In particular, visual safety issues of long-term effects of watching stereoscopic images are important concerns.

[1] New services could involve new channels or programs that exploit the capability to offer visualization of objects in stereoscopic 3D; e.g., new game shows and infomercials.

3. HUMAN VISUAL SYSTEM AND PROCESSES

3.1 Human Visual System, Visual Fixation, and Eye Movements

The human visual system consists of the two eyes, optic pathways, relay stations and a higher signal processing center in the brain called the visual cortex [1].

The image of a scene in the form of light enters each eye through the *pupil*. The image is then focused at the light-sensitive *retina*, with the aid of an adjustable *lens* through a process called *accommodation* (focus). A central zone of the retina, called the *fovea*, is where fine spatial details can be resolved. Rotational *eye movements* are required to position the image of an object of interest onto the fovea. This action of the eye is called *visual fixation* or, if the two eyes are considered, *binocular visual fixation*.

There are two types of fixational eye movements: *saccades* and *vergence* [2]. Saccades are for changing fixation from left to right or vice versa. Vergence movements are for changing fixation from near to far (*divergence*) or far to near (*convergence*). Latencies for vergence eye movements are about 160 ms [3] and for accommodation are about 290 ms [4]. A difference in latencies for vergence and accommodation might contribute to a reduction in visual comfort in the viewing of dynamic object movements in depth [5].

Both divergence and convergence are yoked or cross-coupled to accommodation responses. The stimulus for vergence eye movements is retinal disparity information (see Section 3.2), and the stimulus for accommodation is retinal image blur (see Section 3.2). Normally, in the real world, changes in fixation from near to far or far to near are initiated by congruent retinal disparity information and retinal image blur information. With current stereoscopic displays, the retinal disparity information is not accompanied by the appropriate retinal image blur. That is, the screen is at a fixed distance to the viewer no matter whether the image appears closer or farther. Also, any limitation in image sharpness of objects caused by the camera-capture process (e.g., depth of field, motion blur) cannot be compensated for by trying to change accommodation to focus at a camera-blurred object.

3.2 Fundamentals of Human Binocular/Stereoscopic Vision

The eyes are located horizontally apart in the head, thus, the visual system receives two slightly different views of a visual scene. The visual system processes the *horizontal disparities* contained in the two retinal images to produce single vision [6] and stereoscopic depth; e.g., [7] [8] of objects in the scene. This is termed *binocular fusion* and *stereoscopic vision*, respectively. (Sometimes, screen parallax is incorrectly equated to horizontal retinal disparity. They are different because the retinal disparity produced by a given screen parallax depends on the viewing distance.)

A binocularly fixated object is imaged on the same relative coordinates in the left-eye and right-eye views. The fixation point falls on the *horopter* [9] [10], a curved line or surface which contains all points that are at the same geometrical or perceived distance of the fixation point. Objects located on the horopter also give rise to a single fused percept.

Points located in front of or behind the horopter are imaged at different relative positions in the left-eye and right-eye images. They are said to have *horizontal retinal disparities*. The magnitude of the retinal disparity increases with the distance of the object from the horopter. Points in front of the horopter are said to have a *negative or crossed disparity*, and points behind it are said to have a *positive or uncrossed* retinal disparity. The human visual system uses these disparities to extract the relative depth of objects in the visual scene.

Objects that give rise to disparities produce disparate images on the left and right retinas. However, objects that are located within a small region in front of and behind the horopter still give rise to a single fused percept. The region, within which objects are fused binocularly despite having disparate images in the two eyes, is called *Panum's fusional area*. Objects located outside the Panum's area result in double vision (i.e., diplopia) but they might still be perceived in depth [11] [12]. The size of Panum's area is not fixed and its area depends on the spatial and temporal properties of the fixation target, such as exposure duration [13], spatial resolution [14], and temporal frequency of disparity variation [15].

When a viewer fixates an object with his or her two eyes, the image of the fixated object is focused on the retina. Points located closer or farther than the accommodation distance are no longer properly imaged on the retina and therefore subject to a degree of blur that increases with the distance away from the focused point. However, the visual system is tolerant of a small amount of blur. Points located within a small region around the focused point are perceived to be sharp. The size of this region, known as the *depth of field* (DOF), varies inversely with pupil diameter. Thus, a major determinant of the DOF of the human eye is the diameter of the pupil, which in turn varies with the level of available light. As such, the DOF, which is usually measured in diopters (D), can vary substantially [16]. The depth of field has a corresponding region straddling the retinal plane called the *depth of focus*.

Under normal conditions, changes in accommodation of the two eyes and the process of vergence occur in an integrated fashion; e.g., [17]. That is, changes in accommodation induce changes in vergence [18] and vice versa [19]. However, the two processes can conflict when watching stereoscopic displays, because the actual images being viewed remain at a fixed distance (involving focus, accommodation) while the depicted object can appear to be at a different distance (involving disparity, vergence). This mismatch is referred to as *vergence-accommodation conflict*.

3.3 Fundamentals of Human Visual Depth Perception

Aside from horizontal retinal disparities giving rise to depth perception, there are other visual cues to depth. *Pictorial depth cues* (also known as *monocular cues*) include occlusion, relative size, familiar size, texture gradient, atmospheric attenuation and color shift, linear perspective, height in the visual scene, lighting, and shading. These depth cues are often used in 2D prints, as well as standard 2D television and cinematic productions. Pictorial depth cues have also been exploited in many algorithms used for converting standard 2D image sequences to stereoscopic 3D images [20].

Motion parallax also offers a very effective depth cue in the form of relative motion [1]. It provides information on the relative depth of objects as a result of camera motion or self-motion, such as looking at a scene while walking or while seated in a moving vehicle. Objects that are closer are perceived to move past the viewer or vehicle much faster than those that are farther away. While current stereoscopic displays offer both disparity and pictorial depth information they do not offer motion parallax information. (Three-dimensional displays that provide look-around capability through presentation of multiple camera viewpoints of a given scene offer motion parallax information.) Vergence and accommodation response provide (non-retinal) depth information. They differ from the other depth cues in that they can theoretically provide absolute distance information. However, they are considered weak cues. They do not offer precise distance information and they are generally considered limited in their effective distances, roughly less than 2 m for both vergence [21] and accommodation.

3.4 Stereoacuity, Spatial, and Temporal Properties of Stereopsis

Fine stereopsis requires accurate accommodation and vergence otherwise stereoacuity is reduced [22]. Consistent with this assertion, it has been reported that when focus cues are correct or nearly correct in a 3D display: 1) the time required to identify a stereoscopic stimulus is reduced, 2) stereoacuity in a time-limited task is increased, and 3) viewer fatigue and discomfort are reduced [23]. Stereoacuity of most adults is 2 minutes of arc, but can be as fine as 2 to 6 seconds of arc. Furthermore, it can be as fine as 20 seconds of arc even when targets have a velocity of 2 deg of arc/sec [24]. It is also generally better for crossed disparity (~6 seconds of arc) than for uncrossed disparity (~15 seconds of arc). More comprehensive and detailed information can be found in [25].

Pastoor from HHI, investigated the limits of binocular depth perception with respect to bandwidth requirements and reported that the cut-off frequencies of depth perception is about 4 c/deg and that the temporal limit is 6 Hz [26]. However, a follow-up experiment using these spatial-temporal cut-off frequencies indicated that the above limits cannot generally be applied for bandwidth reduction because distortions were observed; the visibility depended on local luminance and depth contrast.

Temporal processing limits were determined for two types of stereoscopic percept associated with square wave disparity alternation: apparent depth motion and depth pulsation by Norcia and Tyler [27]. With dynamic random dot stereograms, which do not contain monocular cues for either target motion or disparity changes, they found that the limit for apparent depth motion was approximately 6 Hz. Above this frequency two pulsating depth planes were seen simultaneously. Depth pulsations were visible up to 14 Hz (i.e., up to 28 depth reversals/sec). Above 14 Hz two transparent planes were perceived without depth pulsation. The results indicate a higher temporal resolution for stereoscopic position change than has been reported in previous studies of apparent depth motion.

The results from another laboratory demonstrated that depth discrimination depends also on exposure duration and the target's spatial frequency [28]. Measurements of disparity threshold for depth discrimination with varying stimulus exposure duration between 0.05 and 2 s showed that disparity threshold decreased with an increase in exposure duration up to a certain duration, beyond which it was approximately constant. This critical duration was about 150 ms for gratings with low and middle spatial frequencies (0.23 and 0.94 c/deg) while the duration was about 750 ms for gratings with high spatial frequency (3.75 c/deg).

3.5 Critical Periods

According to Fawcett, Wang, and Birch [29] the following is known about the critical periods of binocular vision. It has been acknowledged that there are separate and distinct critical periods for the development and susceptibility of binocular vision. The *critical period* for the development of stereopsis in humans is well-defined. Onset occurs at approximately 3 months of age, followed by a rapid period of maturation until 8 to 18 months of age, and a more gradual improvement until at least 3 years of age. The critical period for susceptibility of the binocular visual system can be assessed by determining the period during which an anomalous binocular visual experience results in an anatomical or electrophysiological change. Monocular deprivation studies with cats and monkeys suggest that the critical period for susceptibility of stereopsis overlaps the critical period for development. For humans, retrospective medical chart reviews and clinical studies of stereoacuity outcomes after surgery for eye realignment to treat infantile esotropia (condition of an eye deviating inward) also indicate such an overlap. Whereas eye misalignment during infancy is associated with severe deficits in stereopsis, stereoacuity outcomes are better for those with eye misalignment after the age of two. Nevertheless, the critical period for susceptibility of stereopsis does not end during infancy because there is evidence to suggest that susceptibility of stereoacuity continues up to at least age five. An example cited was that children with late-onset accommodative esotropia (crossing of the eyes caused by farsightedness) can show deficiencies in both high-grade stereopsis and foveal fusion after as few as three months of constant eye misalignment after normal maturation of binocular vision [30] [31].

3.6 Stereo-Deficiencies

Not all viewers can benefit from the disparity information contained in stereoscopic images. While some of the stereo-deficient are stereoblind, others are stereo-anomalous, unable to discriminate depth when disparities are either crossed or uncrossed [32].

Studies to estimate the percentage of the general population who are stereo-deficient have led to widespread disagreements, ranging from 30% to 6%, but the percentage appears to be highly dependent on the method of testing [33]. In reviewing studies on the assessment of stereo-deficiencies and the duration of target presentation, Stelmach and Tam concluded that for moving images such as stereoscopic 3D programs on television, most individuals (>95%) should be able to benefit from the stereoscopic depth information because scene durations typically exceed one second between cuts or fades [34]. Consistent with this assessment, it has been shown that with dynamic scene contents, such as television programs, even individuals who have been classified as stereo-deficient with still images can make use of the disparity information [35].

3.7 Individual Differences

Stereoscopic depth perception depends on an individual's physical attributes and processing abilities. An individual's eye separation, also known as either *inter-pupillary distance* (IPD) or inter-ocular distance (IOD), is one important factor that determines the perceived depth and the actual disparity that is perceived by that individual. For a given screen parallax and viewing distance of a stereoscopic object, individuals with smaller IPDs would result in a larger perceived depth than individuals with larger IPDs. That is, viewers would perceive objects that in front of the screen to be closer and objects that are behind the screen to be farther away. The implication of this is that, for a given stereoscopic image, the vergence-accommodation conflict would be larger for individuals with smaller IPDs.

For adults the mean IPD is 63 mm. For broadcasters, consideration of the range is more important. Dodgson [36] published a useful review showing that the range of 40–80 mm is likely to include all adults as well as all children who are above four years old. Thus, the lower limit of the IPD range should be taken into consideration for television program production that are geared towards children.

Tam and Stelmach [34], tested 100 viewers in two depth discrimination tasks, and found large individual differences in their performances. Some individuals required relatively long display durations (>500 ms) to perform at criterion, while others were able to complete the task in as little as 20 ms. Individuals also differed in their performance depending on whether the disparity of the test stimuli consisted of crossed (object in front of screen) or uncrossed (object behind screen) disparity. That is, some performed better with crossed than with uncrossed disparities, or vice versa. This variability is in agreement with the results of other researchers; e.g., [32].

These findings suggest a possible need to allow viewers to adjust the range (depth volume of the scene) and the type of disparities (positioning of scene in front or behind the screen plane) in program material to suit their individual preferences, when they view stereoscopic image sequences at home or on a mobile display. This customized-disparity content could be achieved through the generation of newly rendered stereoscopic image pairs of the same scene at the display, utilizing depth maps and depth-based image rendering methods [37] [38].

4. PERCEPTUAL DIMENSIONS OF STEREOSCOPIC 3D IMAGES AND THEIR ASSESSMENT

It is well-established that the introduction of disparity information into standard 2D image sequences leads to an enhanced sensation of depth for most viewers. (See Section 3.4). However, stereoscopic image sequences produce an enhanced viewing experience along more than one perceptual dimension. The dimensions include image quality, naturalness, sharpness, and presence. However, improvements along these perceptual dimensions are not as robust as for depth quality.

As a result of the enhanced viewing experience, viewers prefer stereoscopic 3D image sequences over their 2D counterparts [39] [40] [41]. However, the addition of disparity information does not automatically lead viewers to choose stereoscopic images over their 2D counterparts. Depending on the composition of the images, viewers may actually prefer to view the 2D versions. Preference for 3D holds only if the stereoscopic image sequence is not accompanied by annoying coding artifacts, distortions created by excessive disparity, crosstalk, or conflicts between monoscopic and stereoscopic depth information [42].

4.1 Subjective Assessment

For subjective assessment of the various perceptual dimensions generated by stereoscopic image sequences, researchers have mainly adapted on the methods for assessing picture quality that are outlined in ITU-R Rec. BT. 500 [43], and described in Annex 2 of ITU-R Rec. BT. 1438 [44]. The Double Stimulus Continuous Quality Scale (DSCQS) is often chosen because of its reliability. The scale consists of a line that has been divided into five segments that are labelled “Excellent,” “Good,” “Fair,” “Poor,” and “Bad.” For the dimension under consideration, such as sharpness, viewers would use the scale to rate two versions of the same stereoscopic sequence: one would be a Test sequence whose sharpness has been reduced or changed, and the other would be a Reference (original) version of the same sequence. The results would be analyzed based on the difference between the ratings for the Reference and the Test sequences. Typically, statistical tests (such as an Analysis of Variance or ANOVA) would be conducted on the results to determine whether the findings are significantly different based on estimations of likelihood that the results are due to chance.

4.2 Image Quality

The image quality refers to one or more characteristics of an image, as compared to an original version, usually after it has been subjected to a process or treatment over a visual communication channel. The image quality of stereoscopic 3D image sequences is multi-dimensional. Thus, ratings of subjective image quality depend greatly on the instructions given to viewers prior to testing.

Pioneering studies by Yano and Yuyama [45] have shown higher ratings of image quality for 3D over 2D images. Interestingly, they found that the difference in quality was largest when the display screen subtended a viewing angle of 30 deg. of arc or more. Seuntiens [46] however, argues that the depth enhancement is not captured by ratings of image quality for stereoscopic 3D images. Instead, they propose that the added value from disparity information is best captured in the measurements of “viewing experience” and naturalness. Research on image quality of stereoscopic images is an ongoing topic of research.

The image quality of stereoscopic images in which one constituent member has been degraded as a result of coding or processing, such as for bandwidth reduction using mixed resolution coding [47] [48], is contingent on the type of artifacts introduced. For image filtering in which blur is introduced the binocular perception is weighted more in terms of image quality towards the higher quality image. In the case of quantization (blocky) artifacts the binocular percept is slightly below the average of the quality of the images presented to the two eyes [48] [49].

4.3 Depth Quality

Depth quality refers to the three-dimensional aspects of depicted objects in a visual scene. The range of the depth of a scene, the vividness of the depth of the scene, the sense of volume in the scene as well as the sense of distance between objects and within objects (such as the folds in clothing, facial features, etc.) are all assumed to contribute to the depth quality of stereoscopic 3D images.

Stereoscopic 3D images produce a reliable and consistent increase in the perceived depth of image sequences, compared to standard 2D images. Enhanced sensation of depth is maintained even if the stereoscopic images are blurred. Furthermore, depth quality is relatively robust to blocky coding artifacts [48] [49].

4.4 Naturalness

Naturalness refers to the perceived sense of depicted objects as being a good representation of what is observed in nature or in life. It appears to be another dimension of the perceptual experience associated with viewing stereo video sequences. This perceptual dimension was examined in a study [50] in which observers viewed stereo video image sequences and used a handheld slider to provide real-time ratings of naturalness. In separate sessions, viewers also rated perceived depth. A key finding was that ratings of perceived depth and naturalness varied independently. However, a more recent study argues that ratings of naturalness reflect depth quality [46]. As indicated in Section 4.2 the multi-dimensionality of the perception of stereoscopic images is a current research topic of interest.

4.5 Presence

Sense of presence denotes the perception of participants as being involved in and being part of the displayed scene and space. A study of visual presence in 3D-TV indicated that subjective presence ratings are subject to considerable temporal variation depending on the image content and camera techniques used. The results also indicated that an increase in depth can lead to an enhanced sense of presence, if depth is perceived as being natural [50].

4.6 Sharpness

Sharpness refers to the appearance of the clarity of objects, edges, and spatial details. It is reported that objects in 3D images appear to be sharper than those appearing in 2D images. Perceived image quality is strongly correlated with perceived sharpness, and weakly correlated with perceived depth. This suggests that the perceptual experience of a stereo video sequence can vary independently along the dimensions of perceived sharpness and depth; sharpness can increase while depth decreases, and vice-versa [42].

5. VISUAL DISCOMFORT AND VISUAL FATIGUE: SYMPTOMS

The symptoms indicative of visual fatigue from viewing stereoscopic images are watery eyes, visual strain, headaches, and sometimes blurred or even double vision [51]. There can also be indications of motion sickness [52]. Measurements of accommodation and vergence responses of adult viewers have also shown decreased performances after they have viewed half-hour of stereoscopic test stimuli [53] [54]. Subjective assessments have also shown that a multitude of factors can contribute to visual discomfort from viewing stereoscopic images. These will be described next.

6. VISUAL DISCOMFORT AND VISUAL FATIGUE: CONTRIBUTING FACTORS

6.1 Disparity Magnitude

Disparity magnitude is a major potential contributor to visual discomfort. This is consistent with the common knowledge that as an object is brought within arm's length and moved closer and closer towards the eyes the effort to maintain binocular fixation on the object increases and becomes stressful.

Disparity magnitude is often expressed as *screen parallax* which is the horizontal distance separating corresponding pixels of the left-eye and right-eye images of an object on a screen. However, disparity magnitude expressed as screen parallax depends on viewing distance; it gets smaller as viewing distance is increased. It is more appropriate to express disparity magnitude in angular terms (degrees or minutes of arc) because it takes viewing distance as well as IPD into account.

Disparity magnitude depicted on a stereoscopic display determines the ease of how the left and right eye view of an object can be perceptually fused into a single object in depth. The closer a stereoscopic object is depicted the larger is the disparity magnitude and the more difficult it is to fuse the stereoscopic images. Disparity magnitude of less than 1 degree of arc has been cited as comfortable to view for most viewers [5] [54] [55], although others have indicated a much safer value of 0.5 degree of arc [26]. It is useful to note that the width of a thumb with an extended arm subtends a visual angle of approximately 2 degrees of arc [56].

6.2 Vergence-Accommodation Conflict

Current stereoscopic display methods have a fundamental weakness. When stereoscopic objects out of the screen are depicted, they force viewers to dissociate the normal interaction between the vergence and accommodation systems. In natural viewing the convergence point and the focus point on an object are at the same distance. However, for current stereoscopic displays systems, the left-eye and right-eye images for stereoscopic objects are displayed on a screen at a constant viewing distance from the viewer. On the other hand, the positions of stereoscopic objects are typically positioned at various positions in front of or behind the screen plane. Thus, a viewer has to maintain accommodation (i.e., focus) on the images that are located at the screen distance but have to carry out different extents of convergence and diverge eye movements to fixate objects at different distances away from the screen plane. This mismatch between where to accommodate (focus) and where to converge (fixate) is known as *vergence-accommodation conflict*. Also, see Section 3.2.

The larger the disparity of an object, the larger is the vergence response. The change in vergence in turn will elicit an accommodation response which might cause the focal point to move away from the screen towards the point of convergence [17]. However, if accommodation moves away from the screen by an amount exceeding the depth of field, then the object, which is actually depicted on the screen, becomes blurred. These conflicting demands are often mentioned as a significant source of visual fatigue and discomfort [23]. Another way of looking at it, the accommodation-vergence conflict is reduced if the perceived depths of objects are bounded within the limits of the depth of field of the eye so that accommodation responses are minimized.

Measurements of how far the converged and the accommodated distance can be separated without leading to visual discomfort has been studied [17] [23] [57]. For the viewing conditions typical of television broadcast, researchers have assumed a depth of field between ±0.2D and ±0.3D [17] [23] [57]. The latter corresponds to a disparity magnitude of approximately ±1 degree of arc, a value that is the suggested maximum disparity magnitude for comfortable viewing [5] [54] [55]. (See Section 6.1.)

According to the specifications of standard organizations [58], the optimal viewing distance for the 1920 x 1080 HDTV signal is 3.1 times the picture height (3.1H). At the recommended viewing distance the normal viewer will optimize picture quality because the separation between adjacent pixels (~1 minute of arc subtended at the viewer's eye) roughly equals the acuity limit of the average viewer. Thus, the comfort zone for different screen sizes using the optimal viewing distance as a guideline can be calculated, as presented in [59]. As an example, consider a viewer focusing on a TV screen located 3 meters away. For a ±0.2D depth of field, the range of depth distances within which objects will appear to be in focus varies from 1.87 m to 7.5 m; for a ±0.3D depth of field, the range varies from 1.57 m to 30 m. The depth of field varies with distance. For the same ±0.2D depth of field, decreasing the viewing distance to 1.5 m will result in the range varying from 1.15 m to 2.14 m, whereas increasing it to 4.5 m will result in a range varying from 2.36 m to 45 m.

Another method for expressing the limits of a comfort zone is using a measure of the screen parallax, expressed as a percentage of the horizontal screen size. For cinema applications, values of 1% for crossed (negative) disparities and 2% for uncrossed (positive) disparities (for a total value of about 3%) have been suggested [60]. Recently, some broadcasters have advocated the use of similar limits for the broadcasting environment [61] as well. However, it has been noted that these limits might be too small for television considering that the latter is typically characterized by smaller screen sizes than cinema; on that basis larger values, possibly as high as ±3%, have been proposed [62].

It is an interesting question as to how aging of the accommodative power of the human lens might influence the effect of vergence-accommodation conflict. Presbyopia describes the condition whereby the amplitude of accommodation, or ability to focus on objects at near, decreases with increasing age. For comparison, the lens power for a child is about 20D (~2"), a 25-year old is about 10D (~4"), and a 50-year old is 1D (~40"). Lens hardening is considered an important factor in the development of presbyopia [63]. How presbyopia affects the visual experience of viewing stereoscopic 3D images is not entirely clear, because it is not just accommodation amplitude per se but the interaction between vergence and accommodation that is important.

6.3 Parallax Distribution

It has been reported that the parallax distribution of a stereoscopic image can affect visual comfort [64] [65]. If the range of parallaxes is distributed behind the screen, the image has been found to be more comfortable to view. On the other hand, objects depicted in front of the screen are more difficult to fuse and view. Importantly, research has also found that viewers rated scenes lower in visual comfort if they had a large amount of parallax or if they had large variations in the parallaxes.

In another study Nojiri, et al., [66] assessed not only the effect of parallax distribution but also the impact of abrupt shifts of parallax distribution (i.e., scene cuts) on visual comfort. They found that if the parallax distribution range and the amount of discontinuous temporal changes in the distribution were within 1 deg. of arc viewers found the stereoscopic images more comfortable to view.

6.4 Motion in Depth

Motion in depth (z-axis) is another potential major contributor to visual discomfort; this conclusion is based on comparison of measurements of accommodation responses before and after stereoscopic viewing [53] [57]. The frequency in which object moves in depth can be problematic [57]. The rate of change in disparity magnitude over time appears to be more detrimental to visual comfort than the absolute magnitude of the crossed and uncrossed disparities [5]. Motion in depth can be problematic probably because the dissociation between vergence and accommodation are stressed as the slower accommodation system tries to catch up with the vergence system during binocular tracking of stereoscopic objects. See Section 3.1 on response latencies for the vergence and accommodation systems.

6.5 Crosstalk

Current stereoscopic displays are based on the presentation of separate images to each of the two eyes of a viewer. However, leakage and visibility of an image that is intended for one eye can end up being visible in the other eye. This is called *crosstalk* [38] [67].

Crosstalk is often said to contribute to visual discomfort, but the experimental evidence to support that statement is meager. The main experimental evidence appears to rest on a study that indicates 5% crosstalk can lead to "slightly reduced viewing comfort [68]. Another study concluded ambiguously that "visual strain remained constant until 15% crosstalk" [69]. In contrast, there isstrong evidence that crosstalk can affect image quality [70] and perceived depth [71].

The visibility of crosstalk depends on the local contrast and the disparity of the depicted objects in the stereoscopic images [26]. If both are small, the crosstalk might be perceived as image blur. Crosstalk is typically in the range of 0.1% to 0.3% with polarization techniques, based on an older study [26], and can be as high as 8% for current stereoscopic systems using shutter glasses [72]. It has been reported that visibility of crosstalk is approximately 0.2% [62], and this level is consistent with the suggestion that crosstalk should be kept below 0.3% by Pastoor [26].

6.6 Inter-Ocular Mismatches

The human visual system is generally quite robust to mild inter-ocular geometrical misalignments and mismatches such as vertical disparity, keystone distortions, image size, image rotation, luminance and blur. Table 6.1, from Kooi and Toet [68] shows the ratings on visual comfort for the various types of inter-ocular misalignments and mismatches for short term presentations (5 sec). In general, their data suggest that the visual system is relatively robust to mild and moderate levels of binocular asymmetries for short term viewing. More studies are required for longer-term viewing.

Table 6.1 from Kooi and Toet [50] (see the legend at the bottom of the table)

No.	Image Manipulation	LQ	Median	UQ
	Rotations (1°)			
1	Out, symmetric	1.0	1.3	1.5
2	In, one eye	1.0	1.3	1.7
3	In, including border, one eye	1.0	1.4	2.0
4	Out, one eye	1.0	1.4	1.9
5	In, symmetric	1.0	1.5	2.0
	Other Distortions			
6	1.5% Overall magnification	1.1	1.3	1.8
7	1PD trapezoid	1.3	1.5	2.0
8	3% Meridional horizontal	1.4	1.9	2.3
9	2.5% Overall magnification	1.8	2.1	2.6
10	3% Meridional vertical	1.8	2.4	3.0
	Shifts			
11	2PD horizontal (converging)	1.5	1.6	2.6
12	1PD vertical	2.1	2.8	3.3
13	3PD horizontal (converging)	2.0	2.8	3.5
14	2PD vertical	4.0	4.5	4.9
	Stereo Images			
15	Normal stereo (1PD=6 cm)	1.2	2.0	2.5
16	2xHyperstereo (1PD=12 cm)	1.5	1.9	2.6
17	4xHyperstereo (1PD=24 cm)	2.0	2.5	3.8
	Crosstalk			
18	5% and 1PD horizontal shift	1.5	2.0	2.7
19	15% and 1PD horizontal shift	3.5	3.5	4.5
20	25% and 1PD horizontal shift	4.4	4.8	5.0
	Combined Manipulations			
21	Stereo and 5% crosstalk	1.0	1.0	1.5
22	5% Crosstalk and blur	2.0	2.5	3.0
23	Stereo and 1PD vertical shift	2.2	2.8	3.5
	Filter			
24	Higher contrast (+25%)	1.0	1.3	1.4
25	Overall luminance difference (-25%)	1.1	1.5	2.0
26	Black and white representation	1.0	1.5	3.0
27	Gaussian blur (sigma=1 pixel)	1.7	2.2	2.7
28	Lowered contrast (-50%)	2.0	2.5	3.1
29	Quantization difference (4 bi)	2.0	2.8	3.3

30	Gaussian blur (sigma=1.5 pixels)	2.3	3.0	3.5
31	Color asymmetry (red/green)	3.2	4.0	5.0
	Reference Image			
32	Reference image	1.0	1.0	1.0
The rating scores on various types of binocular asymmetries: 1 = viewing comfort not reduced; 2 = slightly reduced; 3 = reduced; 4 = considerably reduced; 5 = extremely reduced LQ, lower quartile, indicating that 25% of the subjects had a lower score. UQ, upper quartile indicating that 25% of the subjects had a higher score. 1 prismatic dioptre (PD) corresponds to the angle of 1 cm viewed from 1 m distance and which equals 0.57°				

7. VISUAL DISCOMFORT AND VISUAL FATIGUE: ASSESSMENT METHODS

7.1 Rating Scale

As with subjective image quality assessment, methods described in ITU-R Rec. BT 500 have been adapted for the subjective assessment of visual comfort or discomfort. Visual comfort is often measured based on a continuous rating scale that is divided into labeled segments; e.g., [5] [41]. A typical example of a rating scale is shown in Figure 7.1. The rating scale is accompanied by instructions on what viewers have to take into consideration when rating the reference and test images. Depending on the assessment procedure, viewers are aware of whether an image that is to be assessed is a reference or a test image. In another procedure, viewers are unaware of which image is the reference or test because they are randomized from trial to trial. At the end of a test, the mean rating for all viewers is calculated and used for comparative analysis across test image sequences and across experimental conditions. Statistical analyses, such as analysis of variance (ANOVA), are used to summarize and analyze the data and draw conclusions.

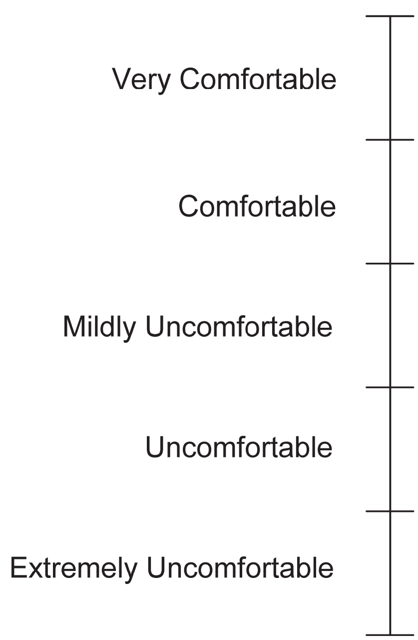

Figure 7.1 Rating scale for visual comfort. For each presentation trial, each viewer would put a horizontal mark where he/she rates the presented stereoscopic image.

The advantage of subjective assessment is that it is a more sensitive method than objective methods for measuring visual comfort. It is also a well-established method for subjective evaluation. The disadvantages of subjective assessments are that it requires a huge amount of resources to generate the test video sequences, recruit viewers, and to conduct the assessments. It is also quite time-consuming.

7.2 Survey/Questionnaire

Given that questionnaires have been designed to evaluate visual fatigue for 2D displays, it is not surprising that researchers have attempted to conduct studies using similar questionnaires for evaluating visual fatigue arising from viewing of stereoscopic images [73]. Lambooiji, et al., [74] also suggested the use of questionnaires originally intended for 2D images for stereoscopic images and asserted that the questionnaires should cover the list examined in [40]. Kuze and Ukai [74] started with a list of 28 questionnaire items and reduced them to 5 key factors that can be used to evaluate visual fatigue arising from viewing stereoscopic images: eye strain, general discomfort, nusea, focusing difficulty, and headache.

7.3 Accommodation/Vergence Response

Visual fatigue can be assessed objectively by measuring either accommodation or vergence responses or both [76]. Accommodation is measured using an optometer and vergence can be measured using one of several eye tracking methods. Measurements can be carried out in real-time during a viewing session to observe changes. Alternatively a baseline can be obtained through measurements before the testing period and then compared with post-test measurements.

Although this method provides objective measurements of the physiological changes underlying accommodation and vergence, the measurement equipment is relatively expensive and the restrictions imposed by the equipment upon the viewing and testing conditions are quite restrictive, such as requiring a chin rest. The advantage is that the measurements are objective. In particular, it allows direct measurements of the physiological changes involved with the vergence-accommodation conflict. As well, typically the pupil size information can also be tracked with such monitoring systems. Pupil size has an impact on the depth of field.

7.4 Critical Flicker Frequency

A method that is not often used for assessing visual discomfort arising from viewing of stereoscopic material is the measurement of flicker fusion frequency [77]. It refers to the frequency at which a light source (which could be a point source or a distributed source) is seen as a steady rather than a flickering light source or vice versa. That is, the critical flicker frequency (CFF) can be measured in two ways: increasing the frequency from 5 to 60 Hz until the subject perceived fusion or decreasing the frequency from 60 to 5 Hz until flicker is detected. The average frequency of the two tests is then taken as the CFF measure. A pre-test and a post-test measurement are conducted to assess the change that has occurred as a result of an intervening manipulation, such as viewing stereoscopic material for a specified period of time. A decrease in CFF threshold measured at the end of the stereoscopic viewing task would indicate an increase of visual fatigue.

8. 3D ON MOBILE/HANDHELD DISPLAYS

Terrestrial broadcast of 3D-TV signals will involve distribution of signals not only to the home but also to mobile and hand-held (M/H) portable receivers or devices. One distinguishing features of M/H devices is that, compared to household television sets, images will be displayed on much smaller screens that vary between approximately 7 cm x 4 cm and 30 cm x 23 cm [78]. The pixel resolution will be reduced by approximately one-fifth to one-quarter that of HDTV in the horizontal direction. Another feature is that 3D-capable M/H devices are mainly based on eyewear-free autostereoscopic displays; e.g., lenticular or barrier. Lenticular-based displays utilize a screen formed by micro-lens to refract partial images of multiplexed stereoscopic image pairs to the left and the right eyes separately. Barrier-based displays utilize a screen that selectively blocks certain pixels from each eye, resulting in completely separate images viewed by the left and right eyes. For autostereoscopic displays, the viewer has to be in a sweet spot to experience an optimal 3D effect.

The main implications of portable displays with small autostereoscopic screens are that the visual impact and requirements are not the same as with larger 3D displays that are typically found in the home environment. This is because stereoscopic effects are influenced by screen size and viewing distance. Based on geometrical considerations of the optics, the depth volume and the distances between objects in depth within a depicted scene are reduced when either screen size or viewing distance is reduced. Furthermore, Yano and Yuyama [45] reported that an improvement in subjective assessment of image quality for 3D images over 2D images was found for larger display sizes with horizontal viewing angles that were greater than 30 degrees.

Although the need for 3D might not be as obvious for small-screen applications, studies have been conducted to investigate possible improvements in perceptual effects from 3D for screen sizes typical of M/H devices. Early reports appear to indicate that with lower resolution and reduced screen sizes the differences in perceptual impact of 3D compared to 2D are not robust, varying with scene contents and the level of visible artifacts. Specifically, results indicating an improvement in image quality and sense of presence of 3D images over 2D images, that were presented with a M/H device,were reported by Shibata, et al. [79]. However, in another study a clear-cut improvement in ratings of image quality or sharpness for the 3D version, compared to the 2D, was not found [80]. In the latter study, viewers were asked to rate image quality, sharpness and sense of presence for both 3D and 2D versions of video sequences with an image size of 15.5 cm x 11.6 cm (352 pixels x 240 pixels). The test sequences were degraded at four quantization levels: with Q = 0, 32, 36, or 39. It was found that ratings of perceived image quality and sharpness were quite similar, decreasing as a function of quantization and showing little difference between the stereoscopic and non-stereoscopic sequences at higher levels of quantization. Ratings of sense of presence tended to be higher for stereoscopic than for non-stereoscopic sequences as long as the scenes did not appear “artificial”. A follow-up study [81] using the same display size as in the earlier study confirmed that sense of presence was enhanced for certain stereoscopic video sequences, compared to the 2D sequences. Shibata, et al., [79] also reported an enhanced sense of presence for 3D than for 2D versions of video that were displayed on M/H devices.

In a more recent study, Utriainen and Jumisko-Pykkő [82] reported two subjective assessment experiments comparing 3D against 2D with different scene contents, frame rates, and video and audio bitrates. A portable parallax barrier display was used with relatively low total bitrates that are relevant to broadcasting for mobile devices. It was found that the quality of experience for the 3D versions was lower than that for the 2D, and that a significant increase in bitrate or frame rate resources did not improve the visual quality of the 3D over 2D. The researchers suggested that compression artifacts and associated visual discomfort probably played a role in their findings. Another study from the same lab [83] compared the roles of depth range in the scene and compression artifacts on user acceptance. The findings indicate that increased compression artifacts had an effect on both lower-level components of spatial visual quality and higher level components of viewing experience (i.e., ease of viewing, overall quality and pleasantness of viewing). The results indicated that video compression artifact, rather than depth range, is a dominant factor that determines the quality of experience. In contrast to the earlier report, in this study they concluded that 3D provides a higher quality of experience over the 2D counterpart. Thus, taken as a whole, the aforementioned studies indicate mixed results with respect to the potential benefits of 3D over 2D contents when displayed on M/H devices. The results depend on both scene contents and the level of visible quantization artifacts.

Another potential issue with 3D-capable M/H devices is that of visual comfort. 3D contents produced for a large screen venue will be reduced in image size and resolution on M/H displays, and they will be viewed at a much closer distance than they were originally intended. As discussed in Section 6.2 on vergence-accommodation conflict, when the viewing distance is reduced the comfort zone (in which objects are bounded within the limits of the depth of field of the eye so that accommodation responses are minimized) is reduced (also see Figure 4 in [84]). Nevertheless, although displaying a typical 3D movie on M/H devices will produce a very shallow depth effect, the depth range will still be within the theoretical comfort zone. Interestingly, a recent study showed that re-purposing original high-resolution stereoscopic sequences to the viewing conditions and resolution of a smaller mobile display based on the theoretical comfort zone did not guarantee comfortable viewing [83]. The researchers suggested that the comfort zone is narrower for small M/H displays.

Considered overall, experimental research examining issues related to the display and viewing of 3D images on mobile devices is still in its infancy. Given the wide range of experimental methodologies and findings, more systematic studies are required to determine the exact conditions that give rise to enhancements in perceived depth, sharpness, presence and naturalness. Recent research work examining 3D mobile television and delivery optimization has gone further than simply evaluating the perceptual effects by proposing a holistic framework for user-centered evaluation of quality of experience (UC-QoE) for mobile systems [85]. The approach emphasizes quality evaluation of actual experience and use in its context, as opposed to testing only perceptual effects in tightly-controlled settings of laboratory environments. That is, drawbacks of the displays themselves, such as crosstalk that can lead to degradation in image quality and perceived depth (see Section 6.5) and possibly to visual discomfort for M/H devices [78], are also considered a component of the system under evaluation. Finally, it has been argued that creating good 3D movies for small devices will require a specific approach targeting small screens only and that “the experience of stereographers in this field is almost non-existent today and is still an experimental playground” [86].

9. ISSUES AND CONCERNS

9.1 Effects of Long Term Viewing

The potential effect(s) of long term viewing are a major concern for broadcasters and users alike. Effects of long term viewing would be less of a concern if viewers find viewing of stereoscopic material as comfortable as viewing 2D material. However, given that there are many factors and conditions that can give rise to visual discomfort and that television viewing can become daily routine, more studies and more scientific data on this issue would be beneficial.

Long term refers to the continuous viewing or the repeated watching of stereoscopic images over an extended period of time. Such studies are lacking, even though there have been a limited number of studies that have examined the performance of both the vergence and accommodative systems for viewing periods lasting one or more hours [53] [57] [73]. For these studies, the slightly reduced performances manifested at the end of the viewing period do return to their original state. Note that these data were obtained from adults. Unfortunately, for young children it is not known what the effects are on visual comfort for these durations. One difficult aspect in designing and conducting such studies concerns the ethics of having to submit viewers to conditions in which the potentially negative effect(s) that are being studied are not fully understood. Given this situation, a likely major limitation in the introduction of 3D-TV services at the early stages will be the need to moderate the use of large disparities. In other words, this will curtail the potential depth impact that can be achieved on viewers. On this cautious side, BSkyB has recently published guidelines [61] with relatively narrow parameters for stereoscopic program content and made explicit the target range of display sizes the 3D contents are intended for.

9.2 Effects on Young Children

Given that children have different ranges of IPDs than adults and the fact that their visual system might still be under development, it is important that their well-being and the impact of both short-term and long-term viewing of stereoscopic material be properly investigated.

Given that current stereoscopic displays rely on the unnatural dissociation of vergence and accommodation for sharp single vision of objects that we fixate, it is also desirable to study the potential impact of viewing stereoscopic image sequences on the development of binocular visual functions. So far, it is known that at 3 months of age, infants are able to dynamically change their accommodation and vergence required by natural targets in binocular viewing conditions [87]. From the same study, it is also known that the accommodation and vergence systems are also cross-coupled by this age. However, it is not until 7 to 10 years of age that monocular accommodative gains akin to adults are achieved [88]. This suggests that the accommodative system of children in this 7–10 year age group and under is still under development. Studies are needed to confirm and establish that these developmental years are not only for monocular vision but also for binocular vision.

On the cautious side, how to incorporate the delivery of 3D signals to the homes of families with young children and how to ensure that they are safeguarded are issues to consider.

10. DISCUSSION AND CONCLUSIONS

10.1 Benefits and Limitations

The potential benefits for end-users from stereoscopic television services are enhanced depth sensations, sense of presence and naturalness, as well as increased perceived sharpness (see Section 4). In general, viewers tend to prefer viewing of stereoscopic material over standard 2D material, with the caveat that the material has to be free of visual discomfort and annoying artifacts; visible artifacts appear to affect image quality much more for 3D than 2D image sequences (see Section 4.2].

As reviewed in various sections of this report, there are a multitude of factors that can give rise to viewers' visual comfort. Several factors, including disparity magnitude, vergence-accommodation conflict, inter-ocular mismatches, widespread display and viewing conditions, and individual differences were identified.

10.2 Visual Health and Safety

Aside from the value of adding stereoscopic depth information to terrestrial broadcast, an important concern is about visual health and safety with respect to the short-term and long-term viewing of stereoscopic 3D video contents. While factors contributing to short-term visual discomfort have been well-researched, there are practical guidelines that can be closely followed to provide reliable and visually comfortable stereoscopic viewing. The catch is that these guidelines are generally intended to be conservative, making it difficult to strike a balance between the opposing requirements for maximizing visual impact while minimizing visual discomfort.

10.3 Required Studies

Studies are required to look more closely at the spatial-temporal characteristics of motion in depth and visual discomfort.

Several researchers have concluded that even if the depth volume is restricted to lie within the frequently mentioned ±1 deg of arc of comfort zone (see Section 6.1), the visual comfort of viewers can be negatively affected by watching objects moving repeatedly and rapidly in depth. A better understanding of the factors and conditions that can lead to a reduction in visual comfort within this "safe zone" for vergence and accommodation interaction will be of practical use.

Studies are required to look at the impact of long-term viewing that takes into consideration the viewing habits of television owners; e.g., 2-4 hours per day over a week.

It is difficult to carry out a long-term viewing study on visual comfort. Aside from ethical issues (see Section 8.1), there are various questions that need to be tackled. What should the contents of the test sequences consist of? What should be the viewing conditions, such as ambient lighting, display size and viewing distance? How to ensure the safety of viewers who are being tested? In selecting viewers for the study what should be the viewer characteristics, such as age and previous 3D viewing experience? In addition, there are no standard ways to measure visual comfort for long-term viewing of stereoscopic images. In short, there are many questions that have to be carefully considered before a proper study can be undertaken.

What is the depth resolution requirement for 3D-TV?

Studies are required to determine the minimum depth information required for the efficient delivery of 3D-TV television signals. While many studies have been conducted on stereoacuity and the upper limits of stereopsis, almost all the studies have been conducted with simple stimuli under laboratory environments. It would be useful to determine human visual characteristics of depth perception using more complex stimuli that are more attuned to images encountered in typical television programs.

How can 3D image quality and depth quality be measured and monitored?

For ensuring quality of service in the delivery of 3D television signals, there is a need to develop objective methods for measuring and monitoring image quality and depth quality of stereoscopic images. It appears that image quality is a multi-dimensional space in which image quality can be manipulated and assessed independently from depth quality (see Section 4.2). More studies are required to identify the major components of this complex issue.

11. REFERENCES

[1] V. Bruce, P. R. Green, and M. A. Georgeson, *Visual Perception: Physiology, Psychology, and Ecology*, pp. 25–66, United Kingdom: Psychology Press, 1996.

[2] R.H.S. Carpenter, *Movements of the Eyes (2nd ed.).* London: Pion Ltd., 1988.

[3] C. Rashbass, and G. Westheimer, "Disjunctive eye movements." *Journal of Physiology (Lond.),* vol. 159, pp. 339–360, 1961.

[4] S. Kasthurirangan, and A., Glasser, "Age related changes in accommodative dynamics in humans," *Vision Research*, vol. 46, pp. 1507–1519, 2006.

[5] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," *SPIE Stereoscopic Displays and Virtual Reality Systems XIII*, vol. 6055, pp. 60550B-1–60550B-9, 2006.

[6] H. Ono, R. Angus, and P. Gregor, "Binocular single vision achieved by fusion and suppression," *Perception & Psychophysics*, vol. 21, pp. 513–521, 1977.

[7] C. Wheatstone, "Contributions to the Physiology of Vision.—Part the First. On some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838. Available at: http://www.stereoscopy.com/library/wheatstone-paper1838.html.

[8] I. P. Howard, *Seeing in Depth: Basic Mechanisms,* vol. 1, Toronto: Porteous Publishing, 2002.

[9] H. Helmholtz, "Helmholtz's treatise on physiological optics," 1925. In J. P. C. Shouthall (Ed. & Trans.), *Handbuch der Physiologischen Optik (Vol. 3).* New York: Optical Society of America. (Original work published 1909.)

[10] K. M. Schreiber, J. M. Hillis, H. R. Filippini, C. M. Schor, and M. S. Banks, "The surface of the empirical horopter," *Journal of Vision*, vol. 8(3):7, pp. 1–20, 2008.

[11] K. N. Ogle, *Researches in binocular vision*. Philadelphia: Saunders, 1950.

[12] K. N. Ogle, "Precision and validity in stereoscopic depth perception from double images," *Journal of the Optical Society of America*, vol. 43, pp. 906–913, 1953.

[13] G. C. S. Woo "The effect of exposure time on the foveal size of Panum's area," *Vision Research*, vol. 14(7), pp. 473–480, 1974.

[14] C. M. Schor, I. Wood, and J. Ogawa, "Binocular sensory fusion is limited by spatial resolution," *Vision Research*, vol. 24, pp. 661–665, 1984.

[15] C. M. Schor and C. W. Tyler, "Spatio-temporal properties of Panum's fusional area," *Vision Research*, vol. 21(5), pp. 683–692, 1981.

[16] S. Marcos, E. Moreno, and R. Navarro, "The depth-of-field of the human eye from objective and subjective measurements," *Vision Research*, vol. 39, no.12, pp. 2039–2049, 1999.

[17] N. Hiruma and T. Fukuda, "Accommodation response to binocular stereoscopic TV images and their viewing conditions," *SMPTE Journal,* vol. 102, pp. 1137–1144, 1993.

[18] B. C. Jiang, "Accommodative vergence is driven by the phasic component of the accommodative controller," *Vision Research*, vol. 36(1), pp. 97–102, 1996.

[19] M. L. Kung, T. L. Alvarez, and J. L. Semmlow, "Interaction of disparity and accommodative vergence," IEEE *29th Annual Bioengineering Conference*, pp. 29–30, 2003.

[20] W. J. Tam and L. Zhang, "3D-TV content generation: 2D-to-3D conversion," Special session of IEEE *International Conference on Multimedia & Expo's* "3D-TV: Primed for success?" pp. 1869–1872, 2006.

[21] A. Viguier, G. Clëment, Y. Trotter, "Distance perception within near visual space," *Perception*, vol. 30, pp. 115–124, 2001.

[22] C. Blakemore, "The range and scope of binocular depth discrimination in man," *Journal of Physiology*, vol. 211, pp. 599–622, 1970.

[23] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8(3), pp. 1–30, 2008.

[24] G. Westheimer and S. P. McKee, "Stereoscopic acuity for moving retinal images," *Journal of the Optical Society of America*, vol. 68(4), pp. 450–455, 1978.

[25] I. P. Howard, and B. J. Rogers, *Seeing in Depth*, vol. 2, pp. 151–173, Toronto: Porteous Publishing, 2002.

[26] S. Pastoor, "Human factors of 3-D imaging: Results of recent research at Heinrich-Hertz-Institut Berlin," *2nd Int. Display Workshop*, pp. 69–72, 1995.

[27] A. M. Norcia and C. W. Tyler, "Temporal frequency limits for stereoscopic apparent motion processes," *Vision Research*, vol. 24(5), pp. 395–401, 1984.

[28] S. Lee, S. Shioiri and H. Yaguchi, The effect of exposure duration on stereopsis and its dependency on spatial frequency," *Optical Review*, Volume 11, Number 4, 258–264, 2004.

[29] S. L. Fawcett, Y.-Z. Wang and E. E. Birch, "The critical period for susceptibility of human stereopsis," *Investigative Ophthalmology and Visual Science*, Vol. 46, pp. 521–525, 2005.

[30] M. E. Wilson, E. C. Bluestein, and M. M. Parks, "Binocularity in accommodative esotropia," *Journal of Pediatric Ophthalmology and Strabismus*, vol. 30, pp.233–236, 1993.

[31] S. Fawcett, J. Leffler, E. E. Birch, "Factors influencing stereoacuity in accommodative esotropia," *Journal of AAPOS*, vol.4, pp.15–20, 2000.

[32] W. Richards, "Stereopsis and stereoblindness," *Experimental Brain Research*, vol. 10(4), pp. 380–388, 1970.

[33] R. Patterson and R. Fox, "The effect of testing method on stereoanomaly," *Vision Research*, vol. 24, pp. 403–408, 1984.

[34] W. J. Tam and L. B. Stelmach, "Display duration and stereoscopic depth perception," *Canadian Journal of Experimental Psycho*logy, vol. 52(1), pp. 56–61, 1998.

[35] M. W. Rouse, J. S Tittle, and M. L. Braunstein, "Stereoscopic depth perception by static stereo-deficient observers in dynamic displays with constant and changing disparity," *Optometry and Vision Science*, 66(6), pp. 355–62, 1989.

[36] N. A. Dodgson, "Variation and extrema of human interpupillary distance," *SPIE Stereoscopic Displays and Virtual Reality Systems XI,* vol. 5291, pp. 40–75, 2004.

[37] G. Sun and N.S. Holliman, "Evaluating methods for controlling depth perception in stereoscopic cinematography," SPIE-IS&T *Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems XX*, vol.7237, pp. 72370I-1–72370I-12, 2009.

[38] S. Daly, R. T. Held, D. M. Hoffman, "Perceptual issues in stereoscopic signal processing," *IEEE Transactions on Broadcasting*, vol. 57(2):2, pp. 347–361, 2011.

[39] I. Yuyama and M. Okui, "Stereoscopic HDTV." In *Three-Dimensional Television, Video, and Display Technolologies* (Eds. B. Javidi, F. Okano), New York: Springer-Verlag, 2002.

[40] M. Wopking, "Subjective assessments of 3D versus 2D presentation of motion, pictures for large screen home television," *1st International Symposium On 3-D Images*, pp. 51–58, 1991.

[41] W. J. Tam, L. B. Stelmach, and P. Corriveau, "Psychovisual aspects of viewing stereoscopic sequences," *SPIE Stereoscopic Displays and Virtual Reality Systems V*, vol. 3295, pp. 226–235, 1998.

[42] L. B. Stelmach, W. J. Tam, and D. Meegan, "Perceptual basis of stereoscopic video," *SPIE Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639, pp. 260–265, 1999.

[43] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," *ITU-R Recommendation* BT. 500-10, 2010.

[44] ITU-R, "Subjective assessment of stereoscopic television pictures," *ITU-R Recommendation* BT.1438, 2000.

[45] S. Yano and I. Yuyama, *SMPTE Journal*, vol. 100, p. 14–18, 1991.

[46] P. J. H. Seuntïens, I. E. J. Heynderickx, W. A. IJsselsteijn, P. M. J. van den Avoort, J. Berentsen, I. J. Dalm, M. T. M. Lambooij, and W. Oosting, "Viewing experience and naturalness of 3D images," *SPIE Three-Dimensional Television, Video, and Display IV*, vol. 6016, 2005.

[47] M. G. Perkins, "Data compression of stereopairs," *Transactions on Communications*, vol. 40(4), pp. 684–696, 1992.

[48] W. J. Tam, L. B. Stelmach, D.V. Meegan, and A. Vincent, "Bandwidth reduction for stereoscopic video signals," *SPIE Stereoscopic Displays and Virtual Reality Systems VII*, vol. 3957, pp. 33–40, 2000.

[49] P. Seuntiens, L. Meesters, and W. IJsselsteijn, "Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Transactions on Applied Perception*, vol. 3(2), pp. 95–109, 2006.

[50] W.A. Ijsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and feeling of presence in 3DTV," *Displays*, vol. 18, pp. 207–214, 1998.

[51] J. E. Sheedy, J. Hayes, and J. Engle, "Is all asthenopia the same?" *Optometry and Vision Science*, vol. 80, pp.732–739, 2003.

[52] S. Yang, "Effects of 3D display on accommodative and vergent responses and subsequent visual discomfort and motion sickness," Stereoscopic Displays and Applications, South San Francisco, CA, USA, 23–27 January 2011.

[53] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, pp. 141–150, 2004.

[54] Y. Nojiri, H. Yamanoue, A. Hanazato, M. Emoto, and F. Okano, "Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing," *SPIE Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, pp. 303–313, 2004.

[55] M. Wopking, "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," *Journal of the Society for Information Display*, vol. 3, pp. 101–103, 1995.

[56] R. P. O'Shea, "Thumb's rule tested: visual angle of thumb's width is about 2 deg," *Perception*, vol. 20(3), pp. 415–418, 1991.

[57] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays*, vol. 23, pp. 191–201, 2002.

[58] ITU-R, "Guidelines on metrics to be used when tailoring television programmes to broadcasting applications at various image quality levels, display sizes and aspect ratios," *ITU-R Recommendation* BT.1845-1, 2010.

[59] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual comfort," *IEEE Transactions on Broadcasting*, vol. 57(2):2, pp. 335–346, 2011.

[60] B. Mendiburu, *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*, Oxford: Focal Press (Elsevier), 2009.

[61] BSkyB, "3D technical specification for PlanoStereocopic (3D) program content." Available at: http://introducingsky3d.sky.com/a/bskyb-3d-tech-spec/.

[62] B. Clark, "3D production and post," White Paper. Accessed Oct 22, 2010. http://www.itbroadcastanddigitalcinema.com/docs/3d_production_and_post.pdf.

[63] A. Glasser and M. C. W. Campbell, "Biometric, optical and physical changes in the isolated human crystalline lens with age in relation to presbyopia," *Vision Research*, vol. 39(11), pp.1991–2015, 1999.

[64] Y. Nojiri, H. Yamanoue, A. Hanazato, and F. Okano, "Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV," *SPIE Stereoscopic Displays and Virtual Reality Systems X*, vol. 5006, pp. 195–205, 2003.

[65] S. Ide, H. Yamanoue, M. Okui, and F. Okano, "Parallax distribution for ease of viewing in stereoscopic HDTV," *SPIE Stereoscopic Displays and Virtual Reality Systems IX*, vol. 4660, pp. 38–45, 2002.

[66] Y. Nojiri, H. Yamanoue, A. Hanazato, M. Emoto, and F. Okano, "Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing," *SPIE Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, pp. 303–313, 2004.

[67] A. J. Woods, "Understanding crosstalk in stereoscopic displays," 3DSA (Three-Dimensional Systems and Applications) conference, Tokyo, Japan, 19–21 May 2010.

[68] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, pp. 99–108, 2004.

[69] P. J. H. Seuntiëns, L. M. J. Meesters, and W. A. IJsselsteijn, "Perceptual attributes of crosstalk in 3D images," *Displays*, vol. 26, pp. 177–183, 2005.

[70] L. M. Wilcox and J. A. D. Stewart, "Determinants of perceived image quality: Ghosting vs. brightness," *SPIE Stereoscopic Displays and Virtual Reality Systems X*, vol. 5006, pp. 263–268, 2003.

[71] I. Tsirlin, L. M. Wilcox, and R. S. Allison, "The effect of crosstalk on the perceived depth from disparity and monocular occlusions," *IEEE Transactions on Broadcasting*, vol. 57(2):2, pp. 445–453, 2011.

[72] C. C. Pan, Y. R. Lee, K. F. Huang, and T. C. Huang, "Crosstalk evaluation of shutter-type stereoscopic 3D displays," *Society for Information Display*, vol.16, pp. 825–833, 2008.

[73] J. Kuze and K. Ukai, "Subjective evaluation of visual fatigue caused by motion images," *Displays*, vol. 29(2), pp. 159–166, 2008.

[74] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: a review," *Journal of Imaging Science and Technology*, vol. 53(3), pp. 030201–030201-14, 2009.

[75] J. Kuze and K. Ukai, "Subjective evaluation of visual fatigue caused by motion images," *Displays*, vol. 29(2), pp. 159–166, 2008.

[76] K. Ukai and P. A. Howarth, "Changes in performance can be reflected in the latency, amplitude, and velocity of the responses to stimulation of the accommodation and vergence systems Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations", *Displays*, vol. 29(2), pp.106–116, 2008.

[77] T. Mitsuhashi, "Evaluation of stereoscopic picture quality with CFF," *Ergonomics*, vol. 39, pp. 1344–1356, 1996.

[78] A. Boev and A. Gotchev, "Comparative study of autostereoscopic displays for mobile devices," *SPIE Multimedia on Mobile Devices 2011,* vol. 7881, 78810B-1–78810B-12, 2011.

[79] T. Shibata, S. Kurihara, T. Kawai, T. Takahashi, T. Shimizu, R. Kawada, A. Ito, J. Häkkinen, J. Takatalo, and G. Nyman, "Evaluation of stereoscopic image quality for mobile devices using interpretation based quality methodology," *SPIE Stereoscopic Displays and Applications XX*, vol. 7237, pp. 72371E-1–72371E-12, 2009.

[80] W. J. Tam, A. Vincent, R. Renaud, P. Blanchfield, and T. Martin, "Comparison of stereoscopic and non-stereoscopic video images for visual telephone systems," *SPIE Stereoscopic Displays and Virtual Reality Systems X*, vol. 5006, pp. 304–312, 2003.

[81] W. J. Tam, R. Renaud, A. Vincent, T. Martin, and P. Blanchfield, "Stereoscopic video telephony: Naturalness and sense of presence," *SPIE Three-Dimensional Television, Video, and Display II*, Vol. 5243, pp. 19–29, 2003.

[82] D. Utriainen and S. Jumisko-Pyykkö, "Experienced audiovisual quality for mobile 3D television," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4, 2010.

[83] S. Jumisko-Pyykkö, T. Haustola, A. Boev, and A. Gotchev, "Subjective evaluation of mobile 3D video content: Depth range versus compression artifacts," *SPIE Multimedia on Mobile Devices 2011*, vol. 7881, pp. 78810C-1–78810C-12, 2011.

[84] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," *SPIE Three Dimensional Imaging, Visualization, and Display*, vol. 7690, pp 76900B-1–76900B-12, 2010.

[85] A. Gotchev, A. Smolic, S. Jumisko-Pyykkö, D. Strohmeier., G. B. Akar, P. Merkle, and N. Daskalov, "Mobile 3D television: Development of core technological elements and user-centered evaluation methods toward an optimized system," *SPIE Multimedia on Mobile Devices 2009,* vol. 7256, pp. 72560J-1–72560J-15, 2009.

[86] B. Michel, "Production issues with 3D content targeting cinema, TV, and mobile devices," Accessed May 30, 2011, http://www.edcf.net/edcf_docs/BMI-production-issues-with-3D-content.pdf

[87] M. S. Banks, "Infant refraction and accommodation," *International Ophthalmology Clinics*, vol. 20, pp. 205–232, 1980.

[88] S. R. Bharadwaj, and T. R. Candy, "Cues for the control of ocular accommodation and vergence during postnatal human development," Journal of Vision, vol. 8(16):14, pp. 1–16, 2008.

Part II: Technology

1. SCOPE

The purpose of this report is to outline the benefits and limitations of existing and developing technology for 3D broadcast. The report considers three different types of broadcast services including real-time HD broadcast to fixed receivers, real-time broadcast to mobile receivers and non-real-time (NRT) broadcast.

The following factors are considered as part of the analysis for all service types:

- **2D/3D Program Dependency**: Indicate any dependencies between 2D and 3D programs. In particular, any dependencies that would impose constraints on the broadcast of 2D and 3D programs; e.g., if not possible to transmit 3D version of a program that is independent from the 2D program. Additionally, any dependencies that facilitate simple extraction of a 2D program from a 3D program.
- **Bandwidth Requirements**: Consider whether a particular coding architecture or format is capable of satisfying bandwidth limitations for real-time transmission. This is also closely related to picture quality. It is recognized that MPEG-2 Video is a mandatory broadcast format of at least SD (NTSC) quality. An HD service in the MPEG-2 Video format would have a bit rate in the range of 10–12 Mbps[2] per program. The presumption is that 3D would have to co-exist with an MPEG-2 HD service. For NRT, the bandwidth requirements translate to download time and storage.
- **Ease of Deployment**: Examine factors that would enable rapid deployment of 3D services with particular solutions as well as factors that may be a hindrance. This may include the existing workflows, existing capabilities of both emission and receiver equipments, as well as backwards and/or 2D compatibility considerations.
- **Standardization Status**: Indicate the needs for any further standardization or recommended practices that would be necessary to realize a particular architecture or format for 3D broadcast services. It is recognized that standardization of media formats for NRT is a current activity in ATSC; therefore the immediate subset of candidates for NRT focuses on currently defined standards.
- **Picture Quality**: Identify any inherent limitations in picture quality, either due to filtering and decimation steps or resulting from bandwidth limitations.
- **MVPD Compatibility**: Indicates compatibility with formats being deployed by, or anticipated to be deployed by, Multichannel Video Programming Distributors (MVPD) including cable, satellite and IPTV service providers that would carry terrestrial broadcast content.
- **Multiview Output**: Indicate whether the format includes data to enable the generation of multiple views beyond stereo, which could be used to better support 3D services beyond stereoscopic display such as auto-stereoscopic or multiview displays.

The report also indicates formats that are suitable for different types of display technologies; e.g., stereoscopic displays that require glasses for viewing 3D content versus auto-stereoscopic or multiview displays that do not require glasses to view the 3D content.

[2] Further analysis could be done assuming different bit rates or SD broadcast for the main 2D program if these conditions are more realistic and there is sufficient interest.

Finally, the report discusses different levels of service compatibility. For instance, the potential evolution from formats that support stereoscopic displays to ones that support auto-stereoscopic displays is discussed. The report also considers service compatibility among fixed and mobile receivers.

2. ORGANIZATION OF REPORT

This report considers both real-time and non-real-time transmission of 3D video over terrestrial broadcast channels. Since 3D production grammar is not necessarily the same as for 2D production, this report considers two major transmission scenarios:

- 3D program independent of the 2D program (sending two more views)
- 3D program dependent on 2D program (sending one more view)

The need to support both transmission scenarios in a time-multiplexed manner may also be considered to accommodate certain production or delivery issues.

A short description of different options within each transmission scenario is provided in Sections 3 and 4. It is important to note that the various options emphasize the architectural aspects of different solutions rather than particular encoding configurations.

Supplemental data such as depth maps may be also be considered as part of the emission format. This report considers a few candidate architectures that integrate depth information as part of the broadcast in Section 5.

For all options described in this report, it is assumed that the 3D program targets HD resolution for the real-time fixed and NRT services. For real-time mobile services, the 3D program is of lower resolution and in accordance with the video resolutions specified in A/153. Furthermore, in the case of NRT services, it is assumed that only the 3D program, or portion of the data that is transmitted to support 3D, is delivered via the NRT channel.

Sections 6, 7, and 8 provide an analysis of the different formats for each of the services types considering the factors identified above. The suitability of different formats for specific types of 3D displays is discussed, and the various types of 3D broadcast services that may be considered are also outlined.

There are other Standards Development Organizations (SDOs) working on technical specifications related to 3D formats and delivery. There are also benefits in some level of global harmonization of 3D standards. For instance, compatibility between different systems could be achieved, including high-quality interchange between regions. Also, reduced complexity and equipment can be realized. As a reference, Section 9 provides a brief review and outline of related standardization activities in other SDOs. Select broadcasting services and trials are also highlighted.

3. TRANSMISSION SCENARIO 1 – 3D PROGRAM INDEPENDENT OF 2D PROGRAM

This transmission scenario assumes that the 3D program is independent and different than the 2D program. The primary benefit of such a scenario is that it does not impose any constraints on the production of 2D and 3D programs. In the following sections, two main architectures are considered:

A) One that is based solely on MPEG-2

B) Another that considers the use of advanced codecs

3.1 MPEG-2 Based Architectures (Scenario 1, Option A)

With MPEG-2 based transmission of both the 2D and independent 3D program, two codec options are considered, which vary in the way that the 3D program is represented and ultimately coded.

3.1.1 MPEG-2 Dual for 3D (Scenario 1, Option A-1)

One option is to simply use existing profiles of MPEG-2 to encode both the left and right views, as illustrated in Figure 3.1. The Main Profile is the most widely deployed profile of MPEG-2 and could be used to independently encode both left and right views. Another option to consider is the Multiview Video Profile (MVP) of MPEG-2, which enables a limited form of inter-view prediction to potentially reduce the total bit rate[3].

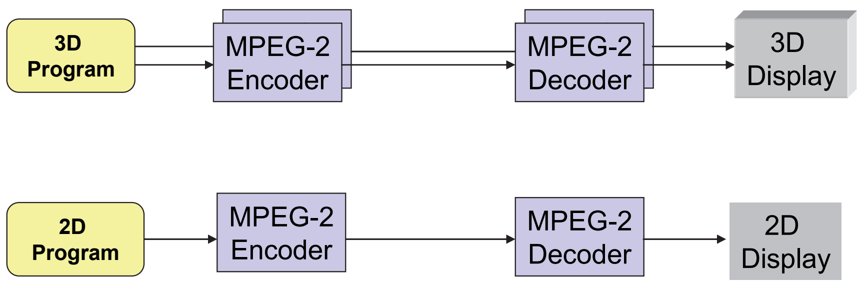

Figure 3.1 The MPEG-2 dual codec option.

3.1.2 MPEG-2 Frame-Compatible for 3D (Scenario 1, Option A-2)

To reduce bandwidth requirements and make maximal use of existing infrastructure, frame compatible coding of the stereoscopic video based on MPEG-2 is considered, as illustrated in Figure 3.2. In this option, the left and right views are decimated (e.g., by a factor of 2) and arranged into one of the common frame-compatible formats such as side-by-side or top-and-bottom. The resulting video is then encoded with the Main Profile of MPEG-2 and transmitted as an auxiliary stream along with the MPEG-2 bitstream for the 2D program.

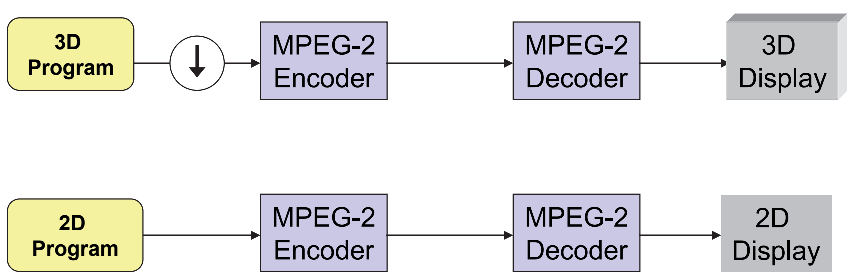

Figure 3.2 The MPEG-2 frame-compatible approach.

3.2 Architectures Based on Advanced Codecs (Scenario 1, Option B)

In the following, the use of advanced codecs is considered for transmission of the 3D program, while maintaining MPEG-2 coding for the 2D program in the main stream.

[3] The bit rate benefits of MPEG-2 MVP are not expected to be significant relative to independent encoding of the views. Given this and the limited interest in this profile, this report only considers the benefits and limitations of independent encoding of left and right views based on MPEG-2 Main Profile.

3.2.1 AVC Frame-Compatible for 3D (Scenario 1, Option B-1)

Similar to the option described in Section 3.1.2 (*Scenario 1, Option A-2*), frame-compatible coding of the stereoscopic video is considered in this approach. (See Figure 3.3) The main difference is that the High profile of H.264/MPEG-4 AVC is used for compression of the frame compatible video rather than MPEG-2.

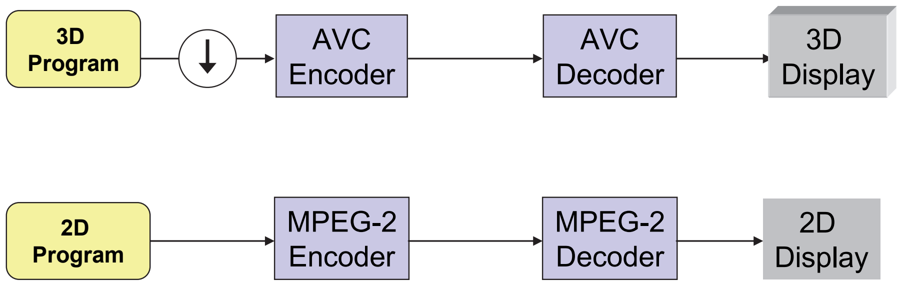

Figure 3.3 The AVC frame-compatible approach.

3.2.2 MVC for 3D (Scenario 1, Option B-2)

This option considers the encoding of both left and right views at full-resolution (Figure 3.4). The views may be encoded independently using an advanced codec such as the High Profile of H.264/MPEG-4 AVC. Another option is to utilize MVC with inter-view prediction to encode both views with lower bandwidth requirements. The Stereo High Profile of H.264/MPEG-4 AVC could be applied for this purpose.

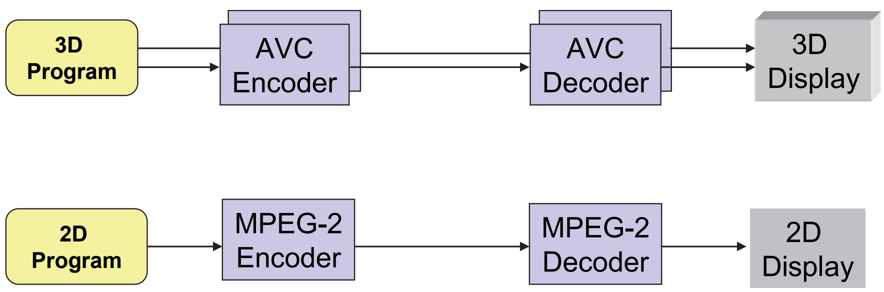

Figure 3.4 MVC for 3D approach.

3.2.3 AVC Frame-Compatible with Resolution Enhancement for 3D (Scenario 1, Option B-3)

In the option described in Section 3.2.2 (*Scenario 1, Option B-2*), the 3D program is encoded as a base layer representing one view and an enhancement layer representing the second (dependent) view. An alternative means for coding the full-resolution stereoscopic video is shown in Figure 3.5, where the base layer is in a frame-compatible format, and the enhancement layer provides resolution enhancement.

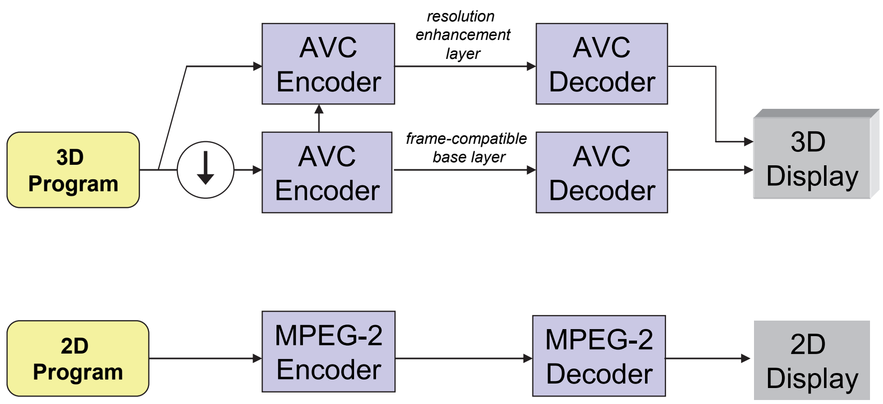

Figure 3.5 Frame-compatible enhancement approach.

Encoding the stereoscopic video in this way could be accomplished with existing profiles in the H.264/MPEG-4 family of specifications including AVC, MVC, or SVC. However, new coding schemes are also being studied and considered for standardization within the MPEG committee. An advantage of this scheme is that it could provide an enhancement of an initial 3D service that is based on the option described in Section 3.2.1 (*Scenario 1, Option B-1*).

3.2.4 AVC Full-Resolution Frame-Compatible for 3D (Scenario 1, Option B-4)

This option is a variant of *Scenario 1, Option B-1* in which a frame-packing arrangement is used to encode the left and right views of the stereoscopic video signal (Figure 3.6). The main difference is that no sub-sampling of the view is performed. Therefore, the input to the AVC encoder is double the typical HD resolution in either the horizontal or vertical dimension. In contrast to the MVC-based architecture described in *Scenario 1, Option B-2*, there is no prediction between the views which would results in some coding efficiency loss relative to that solution.

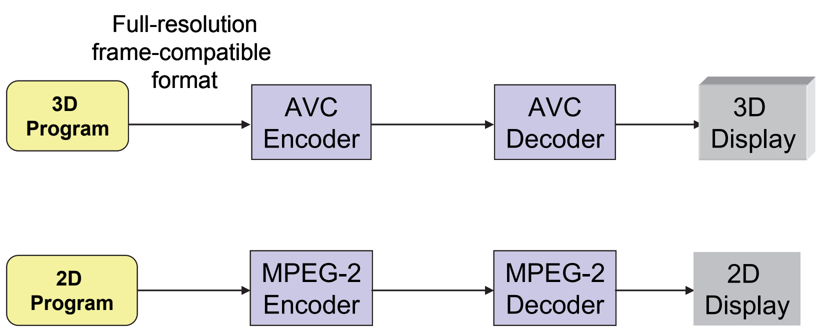

Figure 3.6 Full-resolution frame-compatible approach.

4. TRANSMISSION SCENARIO 2: 3D PROGRAM DEPENDENT ON 2D PROGRAM

This transmission scenario assumes that the 3D program is dependent on the 2D program; i.e., the 3D program shares a view with the 2D program. The primary benefit of such a scenario is that the bandwidth savings that could be achieved since one of the views is shared between both programs.

Two main architectures are considered:

A) One that is based solely on MPEG-2

B) Another that considers the use of advanced codecs

4.1 MPEG-2-Based Architectures (Scenario 2, Option A)

4.1.1 MPEG-2 for Second View (Scenario 2, Option A-1)

With this option, the one view of the shared 2D/3D program (e.g., the left view) is used for the 2D program. (See Figure 4.1) The 3D program is reconstructed by combining the 2D program with the second view (e.g., the right view). The 2D program is conventionally encoded by MPEG-2, while the right view that is used for the 3D program is also conventionally encoded using MPEG-2. The Multiview Video Profile (MVP) of MPEG-2, which enables a limited form of inter-view prediction to potentially reduce the total bit rate, could also be used.

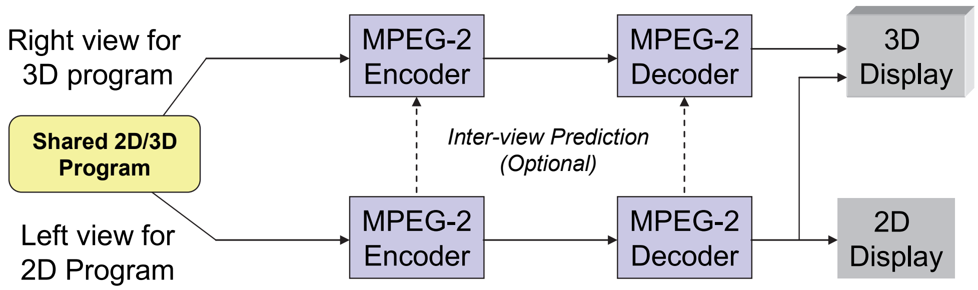

Figure 4.1 Shared MPEG-2 coding approach.

4.2 Architectures Based on Advanced Codecs (Scenario 2, Option B)

Rather than using MPEG-2 to encode the second view, it is also possible to consider the use of advanced codecs to encode the video required to form the 3D program. This is referred to as a hybrid solution since two different codecs (e.g., MPEG-2 and AVC) would be used to represent a single 3D program. (See Figure 4.2) In some configurations, a dependency may also exist.

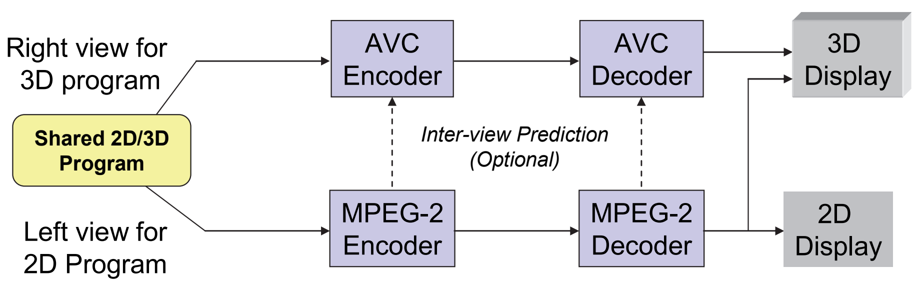

Figure 4.2 Hybrid architecture based on MPEG-2 and advanced codecs.

4.2.1 AVC for Second View (Scenario 2, Option B-1)

A simple option is to use the High profile of H.264/MPEG-4 AVC to independently encode the second view. Alternative codecs may also be used, but the advantage of AVC is that it is already well established and deployed. The subsequent analysis assumes AVC is used for this option.

4.2.2 AVC for Second View with Prediction (Scenario 2, Option B-2)

An alternative option is to consider a scheme in which the second view is adaptively predicted from pictures in the first view as well as temporal reference pictures to produce an enhancement layer that corresponds to residual data. Then, the resulting enhancement layer is encoded with an advanced codec such as H.264/MPEG-4 AVC or the emerging HEVC standard. Such an option would require the standardization of a new video codec.

5. DEPTH-BASED 3D FORMATS

5.1 Capabilities and Representations

Depth-based representations are another important class of 3D formats. Such formats enable the generation of virtual views through depth-based image rendering techniques, which may be required by auto-stereoscopic or multiview displays. Depth-based 3D formats can also allow for advanced stereoscopic processing, such as adjusting the level of depth perception with stereo displays according to viewing characteristics such as display size, viewing distance or user preference.

The depth information may be extracted from a stereo pair by solving for stereo correspondences or obtained directly through special range cameras; it may also be an inherent part of the content, such as with computer generated imagery. Depth estimation is currently a challenging problem and often leads to depth maps that are noisy and do not always correspond precisely to the scene. As a consequence, views generated by depth maps may suffer from visible artifacts.

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. The well-known 2D plus depth format (see Figure 5.1) is supported 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.

Figure 5.1 2D plus depth format.

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, a multiview video plus depth format with only 2 original input views and associated per pixel depth data can be considered. With two input views, high quality stereo video is provided and the depth information would enhance 3D rendering capabilities beyond 2D plus depth. However, for high-quality auto-stereoscopic displays, wide-baseline rendering with additional views beyond the stereo range may be required. For example, formats with 3 or 4 views with associated depth map data may be considered. Depth-based 3D formats beyond 2D plus depth are a current topic of study in MPEG.

5.2 Coding Architectures

Depth map information could be used to represent a 3D program or integrated into any of the emission formats described previously. For instance, in transmission scenario 1 (Section 3), the 3D program could be represented by a single view and its corresponding depth map, or the depth map may be supplemental to the stereo pair. On the other hand, in transmission scenario 2 (Section 4), the 3D program may be derived from the 2D program plus the depth map. It is also possible to encode the depth maps with either MPEG-2 or advanced video codecs such as AVC/MVC. In the following, we outline a select set of these possible combinations.

5.2.1 2D Plus Depth for Transmission Scenario 1

The 2D plus depth format is used to represent the 3D program in this option, as illustrated in Figure 5.2. Both channels of data are encoded with an advanced video codec to achieve high compression efficiency.

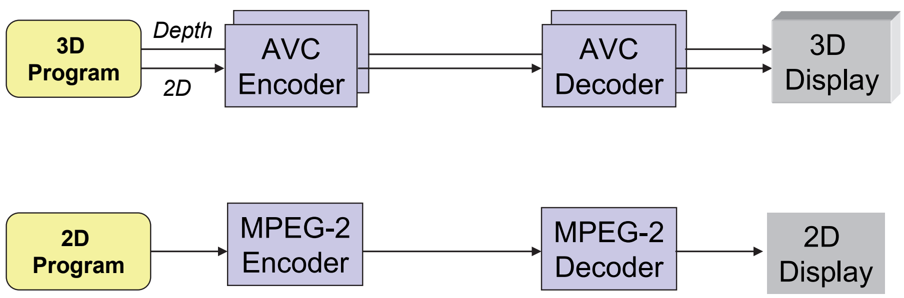

Figure 5.2 The 2D plus depth approach for Transmission Scenario 1.

5.2.2 2D Plus Depth for Transmission Scenario 2

In this option, the 3D program is dependent on the 2D program as with all architectures described under Transmission Scenario 2. The main difference with this option is that the additional channel is a depth map rather than one of the views of a stereo pair. (See Figure 5.3)

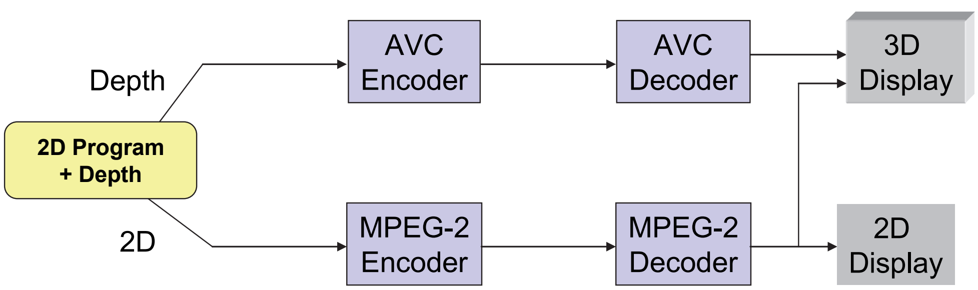

Figure 5.3 The 2D plus depth approach for Transmission Scenario 2.

5.2.3 Multiview Plus Depth for Transmission Scenario 1

In this option, the 3D program is represented by a stereo video and depth maps that correspond to each of the views. (See Figure 5.4) It is assumed that advanced video coders are used for compression to achieve the highest coding efficiency.

This option is an enhancement of the approach described in Section 3.2.2 (*Scenario 1, Option B-2*). As in Section 3.2.2, correlations between the corresponding left and right channels of data could be exploited with inter-view prediction.

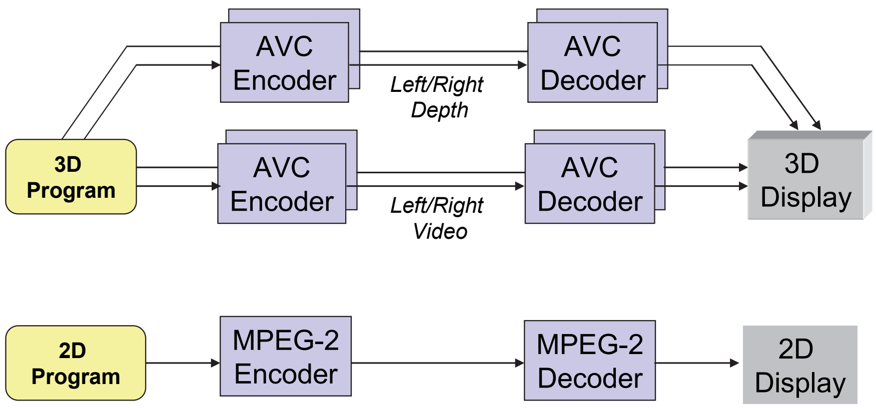

Figure 5.4 Multiview plus depth for Transmission Scenario 1.

5.2.4 Multiview Plus Depth for Transmission Scenario 2

This option is an enhancement of approach described in Section 4.2.1 (*Scenario 2, Option B-1*), where the second view of the stereo video is encoded with an advanced video codec. (See Figure 5.5) Additionally, the depth maps corresponding to the left and right views are encoded with an advanced video codec.

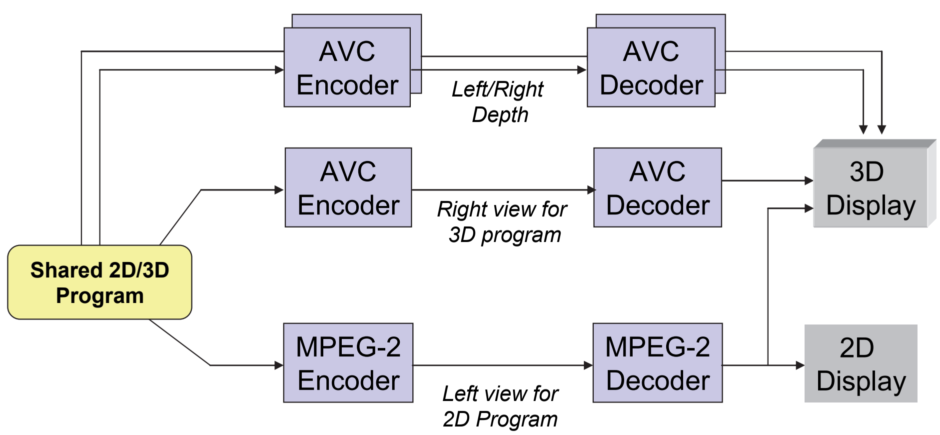

Figure 5.5 Multiview plus depth for Transmission Scenario 2.

6. ANALYSIS OF 3D FORMATS FOR REAL-TIME TRANSMISSION TO FIXED RECEIVERS

This section provides an analysis of the 3D formats described in Sections 3, 4, and 5 for real-time transmission to fixed receivers. It is assumed that the main service is HD video encoded with MPEG-2 and that the 3D program also targets HD resolution. Recommendations are also provided based on the analysis of the different formats for each transmission scenario.

6.1 Real-time Fixed for Transmission Scenario 1

Table 6.1 compares the two MPEG-2 architectures described in Section 3.1 (*Scenario 1, Option A*). While MPEG-2 receivers are widely deployed, it would be difficult, if not impossible, to satisfy the bandwidth requirements with either of these formats.

Table 6.1 Comparison of MPEG-2 Formats for Real-Time Fixed (Scenario 1)

	MPEG-2 Dual	**MPEG-2 Frame-Compatible**
2D/3D Program Dependency	The 3D program is not dependent on the 2D program.	
Bandwidth Requirements	Not possible to satisfy bandwidth requirements.	May be possible (barely) to satisfy bandwidth requirements, but the quality of either or both of the 2D program or the 3D program will likely not be sufficient.
Ease of Deployment	Minimal impact on existing emission infrastructure.	
	Impact to internal broadcast facility infrastructure—2x increase in baseband data. Requires dual encoding/decoding.	Impact to internal broadcast facility infrastructure—1x increase in baseband data. Requires single view encoding/decoding.
Standardization Status	Based on well-established and deployed standards. Would require appropriate system layer signaling.	
Picture Quality	Maintains full-resolution provided that there is sufficient bandwidth.	3D picture quality will be degraded for some scenes due to down-sampling
MVPD Compatibility	Not aware of any deployments or deployment plans for MPEG-2 dual,	Being deployed in initial phases of MVPD.
Multiview Output	No explicit support.	

Table 6.2 compares the four AVC-based architectures described in Section 3.2 (*Scenario 1, Option B*). Although the 3D picture quality may be degraded with the frame-compatible option, this format is the most favorable in terms of bandwidth, ease of deployment and MVPD compatibility. To achieve full-resolution picture quality, the MVC format, or a frame-compatible enhancement format, appear to be feasible options as well. However, these formats would require greater bandwidth and receivers with dual-decoding capability. While these formats are not compatible with current MVPD formats, there does appear to be interest in moving towards such full-resolution formats.

Table 6.2 Comparison of AVC-Based Formats for Real-Time Fixed (Scenario 1)

	Frame-Compatible	**MVC**	**Frame-Compatible Enhancement**	**Full-Resolution Frame Compatible**
2D/3D Program Dependency	The 3D program is not dependent on the 2D program.			
Bandwidth Requirements	It should be possible to satisfy bandwidth requirements with sufficient quality for both 2D program and 3D program.	Might be possible to satisfy bandwidth requirements, but the quality of either or both of 2D program or 3D program may not be sufficient.		Unlikely to satisfy bandwidth requirements with sufficient quality for both 2D program and 3D program
Ease of Deployment	New encoding format could be integrated into broadcast multiplex.			
	Minimal impact on existing emission infrastructure. Require AVC decoding in receiver.	Impact to internal broadcast facility infrastructure – 2x increase in baseband data.		
		Dual AVC decoding is required in the receiver.		Require a Level 5.0 codec for 3D program at HD resolution.

Standardization Status	Would require new signaling and codec for 3D program.			
	Based on well-established and deployed standards.		Possibility to use existing standards, but improved solutions may require new standardization.	Based on well-established and deployed standards.
Picture Quality	3D picture quality will be degraded for some scenes due to down-sampling.	Maintains full-resolution provided that there is sufficient bandwidth.		
MVPD Compatibility	Being deployed in initial phases of MVPD.	Being considered for future phases of MVPD.	Being considered for future phases of MVPD.	Not aware of any consideration by MVPD.
Multiview Output	No explicit support.			

Table 6.3 compares the two depth-based architectures described in Sections 5.2.1 and 5.2.3 for transmission scenario 1. While it may be possible to satisfy bandwidth requirements with the 2D plus depth format, ease of deployment and picture quality issues are less favorable. While the multiview plus depth format could likely overcome the picture quality issues, it requires significant bandwidth and also imposes significant decoding complexity.

Table 6.3 Comparison of Depth-Based Formats for Real-Time Fixed (Scenario 1)

	2D plus Depth	**Multiview plus Depth**
2D/3D Program Dependency	The 3D program is independent of the 2D program.	
Bandwidth Requirements	It may be possible to satisfy bandwidth requirements with sufficient quality for both data included in 2D program and 3D program.	Would be difficult to satisfy bandwidth requirements while maintaining the quality of both 2D and 3D programs (note that depth map bit rate could vary).
Ease of Deployment	New encoding format could be integrated into broadcast multiplex.	
	Impact to internal broadcast facility infrastructure—2x increase in baseband data. Dual AVC decoding would be required in the receiver. Require depth-image based rendering for stereo and multiview output.	Impact to internal broadcast facility infrastructure—4x increase in baseband data. Quad AVC decoding would be a significant hurdle in the near-term for receivers.
Standardization Status	Based on well-established and deployed standards. Would require new signaling and codec for 3D program.	
Picture Quality	Limited rendering capability and problem with occlusions.	Relative to 2D plus depth, the rendering capability of multiview plus depth could be improved; e.g., occlusions could be better handled.
MVPD Compatibility	Not aware of any deployments or deployment plans that utilize depth,	
Multiview Output	Facilitates generation of multiview output at the receiver.	

6.2 Real-time Fixed for Transmission Scenario 2

Table 6.4 compares the three architectures described in Section 4.1 (*Scenario 2, Option A*) and Section 4.2 (*Scenario 2, Option B*). While it should be possible to satisfy the bandwidth requirements by using MPEG-2 for the second view, greater efficiency could be achieved by using an advanced video codec such as AVC for the second view. One challenge with such a solution is the need for MPEG-2 and AVC codecs to operate simultaneously. Even higher coding efficiency could be expected if inter-view prediction is enabled, however this would require standardization of a new video codec.

Table 6.4 Comparison of Video Formats for Real-Time Fixed (Scenario 2)

	MPEG-2	**AVC**	**AVC with Prediction**
2D/3D Program Dependency	The 3D program is dependent on the 2D program; therefore it is not possible to transmit a 3D version of a program that is independent from the 2D program.		
Bandwidth Requirements	Should be possible to satisfy bandwidth requirements, but the quality of either or both of the 2D program or the 3D program may not be sufficient.	Bandwidth requirements could be satisfied with sufficient quality for both 2D program and 3D program.	Bandwidth requirements could be satisfied with sufficient quality for both 2D program and 3D program; this technique is expected to be more efficient than the AVC option that does not utilize inter-view prediction for the second view.
Ease of Deployment	Minimal impact on existing emission infrastructure with new signaling for 3D program. Impact to internal broadcast facility infrastructure—1x increase in baseband data.		
	Need to ensure that encoders and decoders are well synchronized.	Need MPEG-2 and AVC encoders and decoders to operate simultaneously; enhancement in the design of MPEG-2+AVC decoders would be required (prototypes are being evaluated and demonstrated in Korea).	
Standardization Status	Based on well-established and deployed standards.		Standardization of a new video codec would be required.
Picture Quality	Maintains full-resolution provided that there is sufficient bandwidth.		
MVPD Compatibility	Not aware of any deployments or deployment plans that utilize two full-resolution signals encoded with MPEG-2.	Not aware of any deployments or deployment plans that utilize a hybrid codec for reconstruction of 3D program. Transcoding would be required.	
Multiview Output	No explicit support.		

Table 6.5 compares the two depth-based architectures described in Sections 5.2.2 and 5.2.4 for transmission scenario 2. While it should be possible to satisfy bandwidth requirements with the 2D plus depth format, ease of deployment and picture quality issues are less favorable. While the multiview plus depth format could likely overcome the picture quality issues, it still requires significant bandwidth and also imposes significant decoding complexity. Both options also require a high degree of synchronization between MPEG-2 and AVC codecs.

Table 6.5 Comparison of Depth-based Formats for Real-Time Fixed (Scenario 2)

	2D plus Depth	**Multiview plus Depth**
2D/3D Program Dependency	The 3D program is dependent on the 2D program; therefore it is not possible to transmit a 3D version of a program that is independent from the 2D program.	
Bandwidth Requirements	It should be possible to satisfy bandwidth requirements with sufficient quality for both data included in 2D program and 3D program.	Might be difficult to satisfy bandwidth requirements while maintaining the quality of both 2D and 3D programs (note that depth map bit rate could vary).
Ease of Deployment	New encoding format could be integrated into broadcast multiplex.	
	Impact to internal broadcast facility infrastructure—1x increase in baseband data. Need MPEG-2 and AVC encoders and decoders to operate simultaneously; enhancement in the design of MPEG-2+AVC decoders would be required. Require depth-image based rendering for stereo and	Impact to internal broadcast facility infrastructure—3x increase in baseband data. Need MPEG-2 and AVC encoders and decoders to operate simultaneously; enhancement in the design of MPEG-2 and AVC decoders would be required. Additionally, dual AVC decoding is required for the depth

	multiview output.	channels.
Standardization Status	Based on well-established and deployed standards. Would require new signaling and codec for 3D program.	
Picture Quality	Limited rendering capability and problem with occlusions.	Relative to 2D plus depth, the rendering capability of multiview plus depth could be improved; e.g., occlusions could be better handled.
MVPD Compatibility	Not aware of any deployments or deployment plans that utilize depth,	
Multiview Output	Facilitates generation of multiview output at the receiver.	

6.3 Real-time fixed for Transmission Scenario 2 with Broadband Channel

In this delivery scenario, the 3D program is dependent on the 2D program that is transmitted in real-time. The view that is used for the 2D program is transmitted in real-time over the broadcast channel, while the second view is transmitted over the broadband channel using a streaming scheme. Table 6.6 presents an analysis of this case based on advanced codecs as described in Section 4.2.1 (*Scenario 2, Option B-1*) and the corresponding Figure 4.2 when extended to include a broadband channel.

Table 6.6 Analysis of Real-time Transmission with a Broadband Channel (Scenario 2)

	AVC for 2nd View
2D/3D Program Dependency	There is a dependency between the 2D real-time program and the 2nd view that is transmitted via a broadband channel to compose the 3D program. Frame-level synchronization between views via broadcast channel and broadband channel is required at the receiver.
Bandwidth Requirements	Possible to satisfy bandwidth requirements with sufficient quality for both 2D program and 3D program (depending on the broadband connection).
Ease of Deployment	Impact on existing emission infrastructure with new signaling and data on the broadband channel for the 3D program. Need to ensure that encoders are well synchronized.
Standardization Status	Codecs used for each view are based on well-established and deployed standards, but the combined use is not specified.
Picture Quality	Maintains full-resolution both on 2D program and 3D program.
MVPD Compatibility	Not aware of any deployments or deployment plans that utilize a hybrid codec for reconstruction of 3D program. Transcoding would be required.
Multiview Output	No explicit support.

6.4 Real-time fixed for Transmission Scenario 2 with Mobile Stream

In this scenario, the 3D program is the combination of the 2D program, which may be either HD or SD resolution, and a 2D program transmitted as part of an ATSC-M/H service (Table 6.7). One view of the 3D program is transmitted over the current 2D broadcast channel (the main 8-VSB channel) and the second view is transmitted over the mobile channel using ATSC M/H. This scenario is based on the architecture described in 4.2.1 (*Scenario 2, Option B-1*). The main difference is that the encoded stream for the second view would conform to the Baseline Profile of H.264/MPEG-4 AVC or the Scalable Baseline Profile of SVC, and is delivered through the mobile channel.

Table 6.7 Analysis of Real-time Transmission with a Mobile Stream (Scenario 2)

	AVC for 2nd View
2D/3D Program Dependency	There is a dependency between the 2D program and the 2nd view that is transmitted via a mobile channel to compose the 3D program. Frame-level synchronization between views via main and mobile channels is required at the receiver.
Bandwidth Requirements	Possible to satisfy bandwidth requirements with sufficient quality for both 2D program and 3D program.
Ease of Deployment	Impact on existing emission infrastructure with new signaling and data on the mobile channel for the 3D program. Need to ensure that encoders are well synchronized.
Standardization Status	Codecs used for each view are based on well-established and deployed standards, but the combined use is not specified.
Picture Quality	Asymmetrical left/right view video resolution may have some degradation in picture quality.
MVPD Compatibility	Not aware of any deployments or deployment plans that utilize a hybrid codec for reconstruction of 3D program. Transcoding would be required.
Multiview Output	No explicit support.

6.5 Recommendations

This section considers the benefits and limitations of various formats for real-time transmission of 3D content to fixed receivers. All of the options considered maintain backwards compatibility with existing 2D HD services delivered via the MPEG-2 Video format. From the analysis, a few general observations can be made:

- The use of advanced codecs for the additional data required by 3D services is highly beneficial to satisfy bandwidth requirements and maintain high picture quality.
- There appear to be viable options and benefits for 3D broadcast based on both transmission scenarios, including the case that the 3D program is independent with the 2D program, and the case that the 2D program can be extracted from the 3D program.
- There are several deployment strategies that would allow enhanced levels of 3D services to be introduced in a compatible manner. For example, one may consider an early introduction of frame-compatible formats for stereo services, then an enhancement of this service to full-resolution stereo, followed by the addition of depth data to support auto-stereoscopic or multiview displays.
- A broadband channel could enhance the bandwidth capability of the receiver such that 3D programs may be received with the same picture quality as current 2D programs received on legacy 2D receivers.

It is recommended that these observations be accounted for in the requirements of a New Work Item Proposal (NWIP) and in the subsequent development of 3D specifications for real-time transmission to fixed receivers. In particular, it is recommended that the specification for real-time transmission to fixed receivers be based on advanced codecs. Also, both transmission scenarios described in this report are candidates for standardization, and the benefits of a broadband channel should be considered. Finally, different phases of standardization could be considered as part of a long-term deployment strategy.

7. ANALYSIS OF 3D FORMATS BASED ON NRT DELIVERY

This section provides an analysis of 3D programs that make use of an NRT channel. The NRT content may be delivered over a terrestrial channel or the Internet. It is recognized that standardization of media formats for NRT is a current activity in ATSC; therefore the immediate subset of candidates for NRT focuses on a subset of options from Section 4 that are based on currently defined standards. Emphasis is also given to formats based on advanced video codecs.

Two distinct delivery scenarios involving an NRT channel are considered. In the first, 3D programs are delivered through the NRT channel. Several available formats are discussed and analyzed in this context. In the second scenario, a hybrid system that combines the real-time 2D transmission with NRT delivery of additional data that is used to form the 3D program is considered. The benefits and limitations of such a system are discussed and analyzed.

7.1 NRT Delivery of 3D Programs

In this delivery scenario, a 3D program is delivered through the NRT channel. It is possible to extract a 2D program from the 3D program, but the complexity of doing so is dependent on the type of encoding used. The following two cases are considered depending on whether the 3D content is appropriate for 2D extraction and viewing in 2D.

- **Incompatible content**: The 3D file is not expected to have a 2D extraction. However, the program could be watched on a 2D display if that display (or the receiver that it is connected to) is capable of performing the necessary extraction of a 2D program, and such an extraction is not prohibited by DRM rules and license. Alternatively, a separate 2D file can be delivered in addition to the 3D file to facilitate both 2D and 3D viewing.
- **Compatible content**: A common file intended for both 2D and 3D use is delivered. The 2D display (or the receiver that it is connected to) can extract a 2D program from the 3D file, and it would be a market requirement that all TVs/receivers have the necessary capabilities for extraction. A separate 2D file may also be delivered in addition to the 3D file to realize higher 2D picture quality than might be offered through the 2D extraction of the 3D file.

The focus of the analysis below (Table 7.1) is on the 3D formats for NRT delivery; i.e., 3D formats with supplemental 2D files are not explicitly considered. This subsection focuses on architectures based on advanced codecs as described in Section 3.2 (*Scenario 1, Option B*) and the corresponding Figures 3.3 to 3.6.

Table 7.1 Comparison of Currently Available 3D Formats for NRT Delivery

	Frame-Compatible	**MVC**	**Frame-Compatible Enhancement**	**Full-Resolution Frame Compatible**
2D/3D Program Dependency	The 3D NRT program may be independent of a real-time 2D program, in which case no synchronization between the two delivery streams is required.			
	Extraction of 2D program requires decoding and up-sampling of cropped region for each frame based on signaling information.	2D program directly decoded from 3D format.	Extraction of 2D program requires decoding and up-sampling of cropped region for each frame based on signaling information.	Extraction of 2D program requires decoding and cropping for each frame based on signaling information.
Bandwidth Requirements	Lowest download time and storage for 3D.	Moderate download time and storage for 3D.	Moderate download time and storage for 3D.	Higher download time and storage for 3D.
Ease of Deployment	New encoding format could be integrated into broadcast multiplex.			
	Minimal impact on existing emission infrastructure. Require AVC decoding	Impact to internal broadcast facility infrastructure—2x increase in baseband data.		
		Dual AVC decoding is required in the		Require a Level 5.0

	in receiver.	receiver.		codec for 3D program at HD resolution.
Standardization Status	Would require new signaling and codec for 3D program.			
	Based on well-established and deployed standards.		Possibility to use existing standards, but improved solutions may require new standardization.	Based on well-established and deployed standards.
Picture Quality	3D picture quality will be degraded for some scenes due to down-sampling.	Maintains full-resolution provided that there is sufficient bandwidth.		
MVPD Compatibility	Being deployed in initial phases of MVPD.	Being considered for future phases of MVPD.	Being considered for future phases of MVPD.	Not aware of any consideration by MVPD.
Multiview Output	No explicit support.			

7.2 Hybrid Real-Time and NRT Delivery of 3D Programs

In this delivery scenario, the 3D program is dependent on the 2D program that is transmitted in real-time (Table 7.2). Only the portion of the data that is transmitted to support 3D is delivered via the NRT channel. Specifically, the view that is used for the 2D program is transmitted in real-time, while the second view for the 3D program is transmitted at an earlier time via NRT and stored. This subsection focuses on architectures based on advanced codecs as described in Section 4.2 (*Scenario 2, Option B*) and the corresponding Figure 4.2. Furthermore, since there is no existing standard for *Scenario 2, Option B-2*, this subsection of the report focuses solely on *Scenario 2, Option B-1*.

Table 7.2 Analysis of Hybrid Real-Time/NRT System for 3D

	AVC for 2nd View
2D/3D Program Dependency	There is a dependency between the 2D real-time program and the 2nd view that is transmitted via NRT to compose the 3D program. As a result, frame-level synchronization between real-time and NRT streams at the receiver is required. Note that it is also necessary to coordinate storage between real-time and NRT streams. Also, it is not possible to transmit a 3D version of a program that is independent from the 2D program.
Bandwidth Requirements	Low download time and storage requirements for 3D delivery.
Ease of Deployment	Minimal impact on existing emission infrastructure with new signaling for 3D program. Impact to internal broadcast facility infrastructure—1x increase in baseband data. Need to ensure that encoders are well synchronized.
Standardization Status	Codecs used for each view are based on well-established and deployed standards, but the combined use is not specified.
Picture Quality	Maintains full-resolution provided that there is sufficient bandwidth.
MVPD Compatibility	Not aware of any deployments or deployment plans that utilize a hybrid codec for reconstruction of 3D program. Transcoding would be required.
Multiview Output	No explicit support.

7.3 Recommendations

This section has identified and evaluated several currently available 3D formats for NRT delivery. The set of formats includes frame-compatible, MVC, enhancement of frame-compatible, and full-resolution frame-compatible. All of these formats are based on the existing

H.264/MPEG-4 AVC video coding standard. The analysis has considered factors such as file size and ease of deployment, as well as requirements identified as part of the NRT 2.0 work such as 2D extraction capabilities. A summary of the analysis and recommendations are given below.

The frame-compatible format is favorable in terms of bandwidth, ease of deployment, and MVPD compatibility. 2D extraction can be supported provided that the receiver is capable of cropping a portion of the decoded video frames based on signaling information and up-sampling by a factor of 2. Picture quality may be compromised since only half the resolution for both 2D and 3D is provided. Given the ease of deployment, it is recommended that this format be specified in a standard by ATSC for NRT delivery of 3D content. The delivery mechanism would be the same as any AVC encoded video, but new signaling must be specified to indicate the properties of the frame packing arrangement. Extraction of 2D video from this format should also be documented.

The MVC format provides full-resolution picture quality for the 3D content with moderate ease of deployment. In the conventional configuration, one view serves as the base layer and is an AVC-compatible bitstream that could support 2D viewing. The second view is encoded as an enhancement to reconstruct the full-resolution 3D. Given the high picture quality and inherent support for 2D, it is recommended that this format be specified in a standard by ATSC for NRT delivery of 3D content. The specifications for delivery of AVC encoded video would need to be updated in accordance with multiview extensions to related transport and file format specifications. The extraction of 2D video from this format does not require any specification other than a conventional AVC decoder that is capable of decoding 2D.

In an alternative configuration of MVC, the base view is a frame-compatible video, and the decoded enhancement layer is used to reconstruct the full-resolution 3D. Post-processing to combine the base and enhancement layers would need to be specified. 2D video with half-resolution could be extracted in the same way as the frame-compatible video from the base layer only, or a full-resolution 2D video could be extracted after decoding and post-processing. Since this format could be used by receivers with varying capability to reconstruct both 2D and 3D video, it is recommended that this alternative configuration of MVC be considered for inclusion in standard by ATSC for NRT delivery of 3D content. The delivery would be the same as that for MVC, however new signaling to indicate the properties of the frame packing arrangement of the base layer would be needed. Methods for extraction of 2D video from this format should also be documented.

The full-resolution frame-compatible format offers full-resolution picture quality for the 3D content, but requires a decoder that is not currently being deployed in consumer markets. It is recommended that this format be considered in a future phase of NRT delivery of 3D content.

The hybrid real-time/NRT system should be considered for inclusion in a standard that utilizes an NRT channel for delivery of 3D content. It is noted that specific technology for implementation is required for synchronization of the NRT and broadcast content.

8. ANALYSIS OF 3D FORMATS FOR REAL-TIME TRANSMISSION TO MOBILE RECEIVERS

The ATSC-M/H standard specifies H.264/MPEG-4 AVC as the video compression format for transmission of video to mobile receivers. A base layer service is supported with a video resolution of 416x240 and maximum bit rate of 768 kbps. An enhancement layer may also be delivered to support resolutions up to 832x480 at higher bit rates; e.g., 2 to 3 Mbps[4]. Since ATSC-M/H receivers are not required to support legacy broadcast streams, the analysis focuses on the *Scenario 1* architectures described in Section 3. Furthermore, the analysis is limited to those options that are based on advanced codecs.

In addition to the main service, which delivers HD video encoded with MPEG-2, our analysis examines the transmission of 3D programs as part of the M/H service. Also, for the purpose of this analysis, it is assumed that the target resolution for each view is the maximum resolution that is supported by the base layer service of the existing ATSC-M/H standard.

8.1 Mobile Delivery of 3D Program

Table 8.1 compares the four AVC-based architectures described in Section 3.2 (*Scenario 1, Option B*) for real-time transmission to mobile receivers.

The frame-compatible formats would not have any dependency with the legacy M/H stream so these streams would be sent separate from the 2D mobile stream, whereas the 2D mobile stream could be used as a base view of the 3D stream if the 2D and 3D programs are compatible. At the target resolutions of interest for mobile receivers, all of the formats should be able to fit within the available bandwidth irrespective of whether the 2D and 3D programs are compatible or not.

Since existing mobile receivers support AVC decoding, the frame-compatible formats enable possible deployment of 3D to existing receivers with the addition of new signaling and corresponding processing of the decoded stereo signal at the receiver. Dual AVC decoding or a higher level AVC decoder is required for other formats.

As with other services types, the picture quality for some scenes may be degraded due to the down-sampling inherent in the frame compatible format. On the other hand, the target resolution is maintained with other formats including MVC, frame-compatible enhancement and full-resolution frame-compatible.

Table 8.1 Comparison of AVC-Based Formats for Real-Time Mobile

	Frame-Compatible	**MVC**	**Frame-Compatible Enhancement**	**Full-Resolution Frame Compatible**
2D/3D Program Dependency	The 3D program is not dependent on the 2D program in the main service.			
	No dependency on 2D mobile stream, if present.	If present, 2D mobile stream could be used as base view of 3D stream.	No dependency on 2D mobile stream, if present.	
Bandwidth Requirements	It should be possible to fit within the available bandwidth with sufficient quality for 2D stream in main service as well as 2D and 3D streams in M/H service.			
Ease of Deployment	Minimal impact on existing emission infrastructure. Existing AVC decoder in mobile receiver can be used.	New encoding format could be integrated into broadcast multiplex. Dual AVC decoding is required in the receiver.		Minimal impact on existing emission infrastructure. A higher level AVC codec is required.
Standardization Status	Would require new signaling for 3D program.	Would require new signaling and codec for 3D program.		

[4] These video bit rates will be multiplied by 2x to 4x to determine the effective bit rate in the broadcast channel due to channel coding.

	Based on well-established and deployed standards.		Possibility to use existing standards, but improved solutions may require new standardization.	Based on well-established and deployed standards.
Picture Quality	3D picture quality will be degraded for some scenes due to down-sampling.	Maintains full-resolution provided that there is sufficient bandwidth.		
MVPD Compatibility	Being deployed in initial phases of MVPD.	Being considered for future phases of MVPD.	Being considered for future phases of MVPD.	Not aware of any consideration by MVPD.
Multiview Output	No explicit support.			

8.2 Recommendations

This section considers the benefits and limitations of AVC-based formats for real-time transmission of 3D content to mobile receivers. All of the options considered maintain backwards compatibility with existing main and M/H services for 2D program delivery. From the analysis, a few general observations can be made:

- There do not appear to be any significant bandwidth limitation for delivery of 3D programs to mobile receivers for target video resolutions comparable to the current base layer service support by the existing ATSC-M/H standard.
- While bandwidth efficiency is important in the selection of 3D formats for mobile, ease of deployment and picture quality are also influential factors in determining a format for 3D delivery to mobile devices.

It is recommended that these observations be accounted for in the requirements of a New Work Item Proposal (NWIP[5]) and in the subsequent development of 3D specifications for real-time transmission to mobile receivers. In particular, it is recommended that the specification for real-time transmission to mobile receivers be based on advanced codecs. The need for higher video resolution for mobile services, including both 2D and 3D video, may also be considered.

9. 3D ACTIVITIES OF OTHER SDOS

A brief summary of related 3D activities of other SDOs is given below. The emphasis is on activity that is related to the specification of compression formats or delivery of 3D video content. Some select 3D services and trails are also highlighted.

MPEG: The MPEG committee has specified a number of 3D compression formats including frame-compatible stereoscopic formats based on MPEG-2 and AVC, as well as multiview extensions to AVC known as MVC. These formats have been discussed in Sections 3 and 4 of this report. MPEG has also initiated a new phase of standardization to define a new 3D data format and associated compression technology to facilitate the generation of multiview output. As discussed in Section 5 of this report, such a format will enable both advanced stereoscopic display processing and improved support for auto-stereoscopic displays. MPEG is also in the process of exploring technologies that could enable full-resolution enhancement of frame-compatible stereo. It is expected that stereo and multiview extensions of the next-generation HEVC standard will also be defined in the future. In addition to these various compression formats, MPEG has also specified transport and file delivery of 3D content,

[5] An NWIP on 3D-TV was approved by the ATSC Board of Directors in July 2011.

including stereo and multiview extensions to the MPEG-2 transport stream and ISO file format specifications.

DVB: In February 2011, the DVB published a specification on the delivery system for frame-compatible stereoscopic 3DTV services[6]. This specification defines methods to encode and deliver frame compatible video over existing broadcast infrastructures, and their decoding by a digital receiver. The selected frame-compatible formats include top-and-bottom and side-by-side formats. DVB also defines frame compatible signaling information and the handling of graphics and captions overlays in the receiver during the reception of a frame compatible signal. The DVB has recently shifted attention to development of a "Phase 2" 3DTV specification for the delivery of full-resolution 3D services that will require the use of a new set-top box.

In Europe, BSkyB launched 3D broadcasting in late 2010 based on the frame-compatible format, and are now considering full-resolution stereoscopic video delivery. In July 2011, Wimbledon was the first 3D broadcast of an event by the BBC, transmitted in a side-by-side frame-compatible format. The 2011 French Tennis Open was also broadcast live in 3D to more than 17 countries across Europe in cooperation with Eurosport and Orange TV. Additional 3D broadcast trials are also underway in other European countries including Italy's RAI.

SCTE: In May 2011, the SCTE announced approval of a project to define the development of a standard for stereoscopic 3D for cable. This new project has been designed as a two-stage effort. Phase 1 will define the video related formatting, signaling and encoding parameters for frame-compatible stereoscopic 3D, while Phase 2 will define requirements for full resolution stereoscopic 3D video systems. It is noted that CableLabs has published an updated specification on content encoding profiles that includes the specification of encoding formats for frame-compatible formats including side-by-side and top-and-bottom formats[7].

There are already 3D services available through cable in the US, including ESPN's dedicated 3D sports channel. A variety of 3D movies have also been made available by cable providers as part of their video-on-demand offerings. All of the content to date is based on frame-compatible encoding formats.

TTA: In April 2011, TTA completed a draft standard that defines the stereoscopic video format and transmission standard for terrestrial 3DTV broadcasting including the support of service-compatible hybrid-coded scheme (*Scenario 2, Option B-1*). They are planning to finalize the technical specifications by the end of 2011.

Organizations involved in the standardization of 3D video formats and delivery in Korea have also been conducting field trials using the newly standardized scheme, which has been publicly demonstrated at various trade shows and industry events.

SMPTE: In 2009, SMPTE completed a report on 3D that defined requirements for a stereoscopic 3D Home Master standard. The organization is now in the final stages of completing a series of standards that define an image format, metadata, as well as graphic overlays, subtitling and captioning for use in production environments to prepare content for

[6] http://www.dvb.org/technology/standards/a154_DVB-3DTV_Spec.pdf

[7] http://www.cablelabs.com/specifications/OC-SP-CEP3.0-I02-110131.pdf

downstream delivery. There is also ongoing work in SMPTE to define the signaling of frame-compatible formats as well as a dense disparity map format.

Part III: Content

1. SCOPE

There are essentially two types of content employed in terrestrial digital broadcasting. There is pre-produced content such as episodic television series and feature films. There are also real-time events such as news and sports which are typically broadcasted live. There is also a hybrid model such as reality type content that although pre-produced, is shot and recorded as live, and although the contributors to this report are not aware of any current reality content being produced in 3D it will eventually need to be considered as it is a staple of modern terrestrial broadcasting.

This report focusses more on the creation of live sporting events which is where the most television production has been done and therefore where the most experiential knowledge exists. There are of course home devices such as Blu-Ray players and broadband connected devices that are used to delivery 3D theatrical content but those delivery systems do not have the same bandwidth limitations that the terrestrial broadcast channel has; also, the target display for theatrical content is not the smaller home screen.

2. CURRENT 3D WORKFLOW

In our current environment, there are very few 3D enabled displays in homes so therefore the vast majority of the audience still experiences a program on a 2D display. (See Figure 2.1.) In this mass market advertising driven environment, the 2D production must not be compromised to create the 3D version. In live sports productions almost universally employ two completely separate productions with different cameras, mobile units, crews and producers. In many cases the commentators and even the commercials are different.

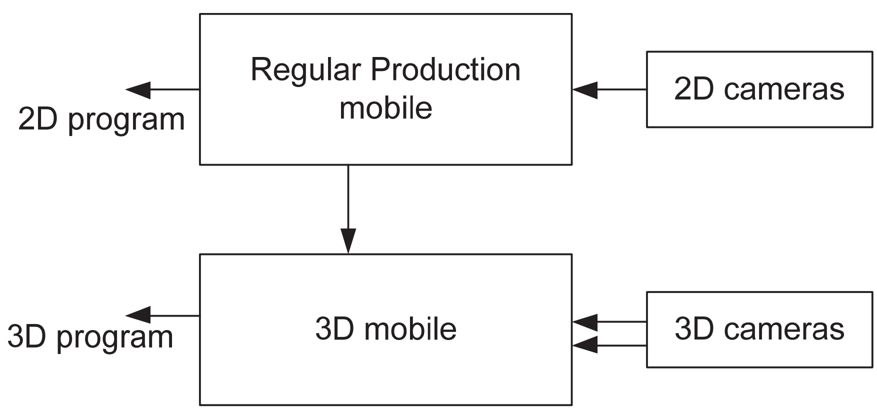

Figure 2.1 Simplified diagram of current 2D/3D live event workflow.

2.1 Camera

Three types of stereoscopic camera system are generally available for live 3D sports production. They are:

- Side by side rigs with two camera systems placed side by side. Because of their size and relative position between the two cameras, these rigs are typically used for long range shots.

- Beam-splitter rigs which consist of two cameras placed at an angle of 90 degree from each other with a 45° mirror used to create split video feeds for the two cameras. These rigs are still typically used for long to medium shots because of their size but the beam splitting technology allows for camera placement much closer together in the horizontal vision plane to more closely approximate human vision.
- Integrated 3D camera system with a single camera fitted with two lenses and sensor fitted with a fixed intra-ocular width approximately equal to human vision. These small units can be used as a portable hand held camera as well as fixed units.

Figure 2.2 shows examples of typical 3D camera rigs.

Figure 2.2 Examples of 3D camera rigs.

3D camera rigs are often significantly more expensive to acquire (either rent or purchase) than their 2D counterparts. For long and medium shots using the side-by-side or beam splitter rigs they are also significantly larger and require additional rigging and human resources to install. The added complexity and costs contribute to limiting the number and location of cameras for a 3D production. In the example cited by the Canadian Broadcasting Corporation (CBC) of their 2D/3D production of Hockey Night In Canada, the 2D production had access to some 20 to 30 different cameras while the 3D production was limited to 6 cameras. The result is a very different production. Similar observations were made by ESPN and CBS on their live 2D/3D sports productions.

Current 3D camera systems, especially the side-by-side and beam-splitter models, require significantly more attention than their 2D counterparts. Their size and weight make them much more cumbersome to manoeuvre. The two cameras must be precisely matched to maintain the proper imagery through the electronics. The lenses used on the camera must be precisely matched to maintain proper imagery through their focal and zoom ranges driven by their separate electromechanical servo systems. Any misalignment that results from vibrations, movement or environmental changes will lead to a degraded 3D image.

The weight and size means camera systems are more restricted on where they can safely be used. The fragility comes from the fact that in addition to the two cameras, extra gears and motors are fitted into the camera system for depth adjustments, these systems are precision systems as even micrometer maladjustment can create misaligned 3D. Figure 2.3 shows a side-by-side rig used by the CBC at the Calgary heritage game. Note that the rig had to be heated to stabilize the two camera and their shared control systems.

Figure 2.3 Comparison between 2D and 3D side-by-side camera rigs.

2.2 Framing and Motion

The goal of a television show is to provide the viewer with an esthetically pleasing experience containing content that compels the viewer to stay engaged with the program. The producer/director of the program will make camera shot choices and camera moves based on using the video to tell the story. There are a number of issues related how a scene is shot in 3D versus 2D that result in very different choices in how a scene is framed and how action is captured. The differences can be subtle to uncomfortable and frequently what works for 2D may not work in 3D.

In 2D content when a new character enters a scene a common method of bringing the character in while maintaining continuity is to center the shot on the character as they enter the scene and pan the camera until the character is fully established in the scene. Even a subtle move like this may be uncomfortable for the 3D viewer. In 3D the same shot would more likely be handled by a static camera shot with a slightly wider field of view allowing the new character to enter and establish position with no camera movement. In 2D this would make for a more passive and less engaging scene. (See Figure 2.4.)

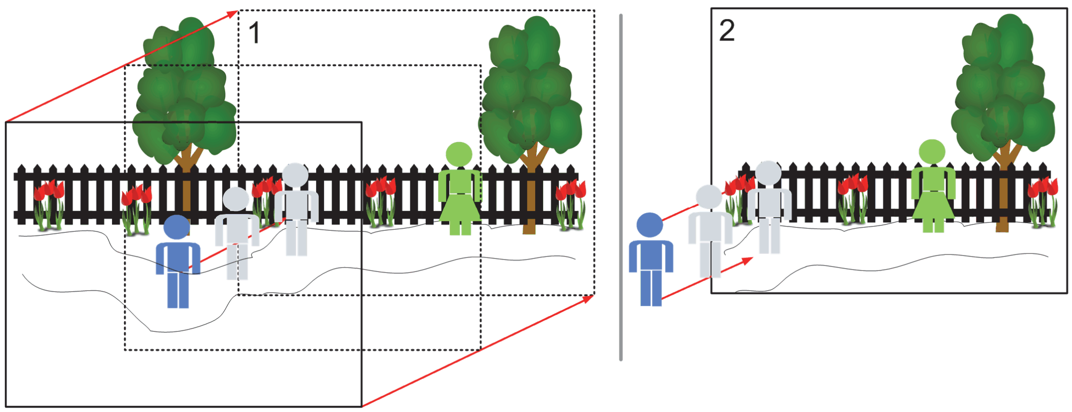

Figure 2.4 2D framing versus 3D framing.

Another example is the use of depth of focus which is a common method for manipulating the viewer attention within a 2D scene. To bring attention to an on screen element that is either in the background or foreground of the current focus point, the camera will change is focal point to bring the new element into focus. In 3D the director has available convergence/divergence in addition to depth of focus in order to manipulate the viewers' point of attention.

In the 3D production of the Masters Golf Tournament, CBS reports that rapid camera pans, following a tee shot down the fairway or an up and on wedge shot caused viewer discomfort such as headaches, nausea, eye-strain, and viewer fatigue. They also report that cuts between cameras with different depth settings (divergence, toe-in) also caused similar issues. The 3D and 2D productions are completely separate with the 3D production using more wide shots, fewer cuts between cameras, lower camera positions and a generally slower paced production while the director waits on the stereographer to set the shot depth. ESPN also reports that although they don't produce any event in 3D that is not produced in 2D, the productions are completely separate.

2.3 Lenses

Lens selections for 3D production add some additional challenges and limitations. As stated earlier, the lenses must be closely matched in their optical performance as well as the performance of their electromechanical servo systems. Differences between the optical performance of the two lenses leads to overall degradation of the combined 3D content. Issues with focus and zoom tracking performance between the two lenses will create additional varying degradation as the lenses zoom and focus.

Lens size is another significant limiting factor. In order to take advantage of the full range of a 3D rig, the lenses must either be compact units as used in a side-by-side rig or the rig becomes extremely large and unwieldy. It would be possible for example to construct a beam splitter rig using studio sized lenses but the rig would be enormous and impractical to use.

The vast majority of the 3D rigs available therefore use lenses with 20 times magnification factors which limit their effective range for placement in large venues and their size limit their placement in smaller venues. As noted earlier, there are smaller portable cameras that incorporate the two lenses in a single hand held package but they performance is limited by the size of the lenses and their magnification factor.

2.4 Graphics and Closed Captioning

Graphics and closed captioning also present a unique challenge in 3D versus 2D production. In 2D sports production we are used lower third supers and upper corner elements that provide statistics and details about the event we're watching, the players on the screen and other sporting events. These graphics along with the live action on the field of play coexist on the single focal plane of the display. In 3D production it is not possible to present these graphics on a single focal plane since the action on the screen is changing the viewers' focus point from in front of the screen to behind the screen. If the viewer is to be able to read the graphics and comprehend the images simultaneously without discomfort, then the graphics need to be adjusted relative to the depth of objects in the scene. The same will be required for closed captioning and may be even more critical since the viewer relying on closed captioning for the dialog will typically be more reliant on the text in the absence of sound.

2.5 Interstitials

All of the contributors to this document report that interstitial material such as commercials, promotional messages and underwriting credits are produced in 3D and are different than those used in the 2D production. In many cases, the cost of 3D production is underwritten or offset by special arrangements with companies that produce the products used to create the content and/or display the content at home. This is similar to the arrangements that Sony, Panasonic, and others had in the early roll out of high-definition production of sporting events.

3. LIMITATIONS IN CROSS UTILIZATION OF 3D AND 2D CONTENT

There is no doubt that a 2D program can be created from a 3D stereoscopic program. In many of the current trials, the left eye image is treated as the mono-scopic image for 2D display devices. While this works and in some cases produces acceptable results (e.g., when showing fish swimming in the ocean or a field of flower moving in a gentle breeze), action content such as a live sporting event generally does not lend itself to this solution. A new "production grammar" is evolving around 3D production, and it differs significantly from the production grammar employed for 2D. Viewing 3D requires continuous use of the viewer's eye convergence muscles, along with the attendant brain image processing to fuse the two images. In order to avoid eye fatigue and stress on the visual system it is important to avoid fast pans/zooms, and cuts between scenes with widely differing depths. Well-produced 3D is often more of a point-of-view experience. ESPN, the CBC, CBS, and others have all noted that 3D production moves slower, uses fewer cameras, wider shots, and less motion. It relies on the immersive nature of 3D to engage and hold the audience. Use of one view of the 3D content would be very dull and slow moving to the audience watching it on a 2D display.

In contrast, some anticipate that the distinctions between 3D and 2D will disappear, 3D cameras will replace 2D cameras, shooting will take place more or less the same as is now done for 2D, and the 2D feed will use one eye from the 3D. These expectations are associated with the assumptions that single productions are essential to make a business model work. It is probably too soon to conclude that a single production cannot satisfy both the 2D and 3D audiences. Depending on content type and economics we will see both single and dual productions employed.

4. CREATING 3D OUT OF 2D CONTENT

Another method of creating 3D programming is to acquire the content in 2D and use computer based technology to synthesize the 3D imagery. One method for doing this is to create a depth map from the 2D images. Although typically a post-production process, it is clearly within the capabilities of existing video processing engines to process 2D video, create credible depth maps and then synthesize a 3D image in real-time with reasonable latency. The results of a fully automated 2D-3D conversion process may be suitable for occasional camera shots. Achieve of uniformly high quality conversion does require manual effort to adjust the depth maps and deal with disocclusions, which make the process quite expensive and not suitable for live events.

In volume 57, issue 2 of *IEEE Transaction on Broadcasting* a paper entitled “3D-TV Production From Conventional Cameras for Sports Broadcast” presents the case for using conventional 2D cameras placed around a venue. The cameras capture baseline images of the field from known positions along with known pan, tilt, and zoom magnification data. During the event, the live action images along with the known data from the cameras are used to synthesize 3D images. The authors present a very sound proposal for using the existing 2D production cameras to create a 3D production without the need for a separate production crew and the expense of completely separate mobile production facility.

In both of these examples there is little doubt that the technology can be refined to allow for the creation of credible 3D imagery from 2D cameras. The issue here will be the inverse of the issue noted from creating 2D from 3D. In order to create a compelling 2D production, the producer/director will utilize camera views, movement and shot selection that will potentially generate a great deal of discomfort for 3D viewers.

5. OTHER CONSIDERATIONS IN CREATING STEREOSCPIC 3D

The CBC provided additional observations regarding their experience producing the program *Hockey Night in Canada*. Many of the creative and technical limitations they mention correspond to the issues raised by ESPN, CBS, and others. It is worth reviewing their list.

Lenses availability. 3D lenses are limited in size, weight and zoom ratio. High zoom ratio lenses are too big for your typical 3D rig. Close-up shots from these lenses are not available and change the resulting program intent.

Limited number of cameras. The number of cameras available on most 3D production is limited because of cost, size, and total production cameras positioning. Positioning is very limited and adding cameras for a dual production can be complicated. The result is the 3D production is working with a reduced number of views, and doesn’t really reflect the intent of today’s shows. Also, certain cameras are not available in 3D such as net cams, above the goal cams, slow motion cameras.

Reflections. The mirror found on a beam-splitter 3D ring can introduce unwanted reflection in the recorded video. While this defect is annoying on a 3D display, when viewed on a 2D display, the newly created artefact will be obvious on the display. This effect is shown in Figure 5.1.

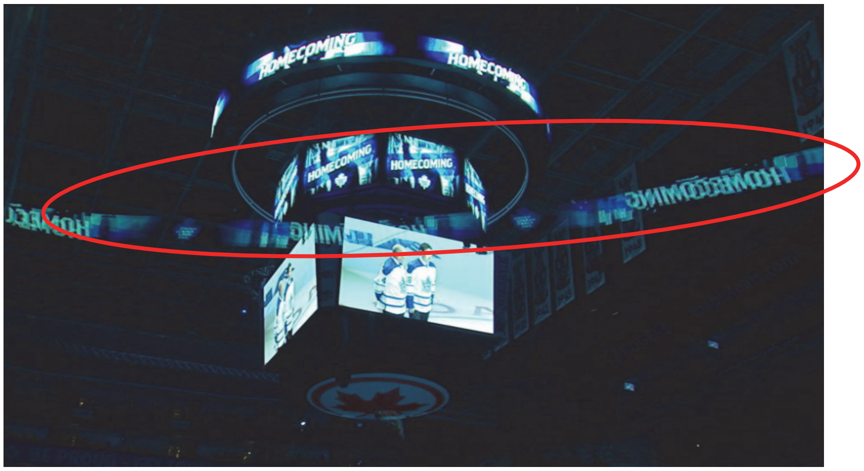

Figure 5.1 Reflection introduced by a mirror-rig.

Stereographer (depth perception). Current stereoscopic recording methods involve two cameras or sensors, each of them producing one eye of the stereoscopic 3D video feed. To adjust the perception of depth, most systems use the same methods: while one of the optical capture device positions is fixed, the other is mobile on two axes, horizontal for adjusting the intra ocular distance and rotational to adjust the convergence. Video material created by the mobile capture device can contain horizontal jitter or rotational variations as the stereographer adjusts the camera system to provide the optimal 3D effect. (See Figure 5.2.) Like out-of-focus or overexposed scenes, these adjustments are undesirable in both the stereoscopic and the regular program. To minimize the effect of the depth adjustment in a derivative 2D program, video material should be taken from the fixed capture device only.

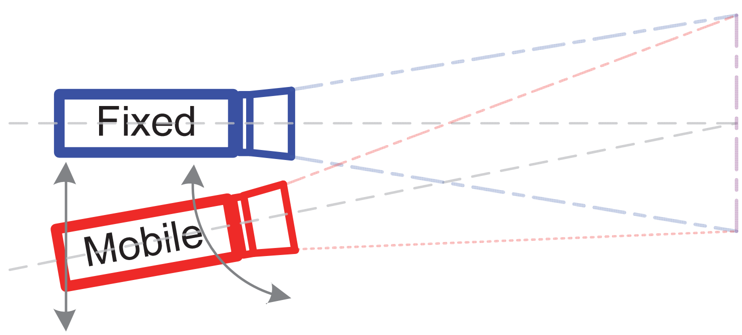

Figure 5.2 Diagram of 3D capture device.

Graphics and overlay. Stereoscopic 3D brings the chance to display a 3rd dimension, depth, which brings new creative possibility in terms of graphics and overlays above the produced video. Since these video sequences might be designed to take advantage of the depth produced by the stereoscopic feature of the display, the original meaning behind those graphics might be lost or greatly impaired if viewed in a 2D situation.

Loss of resolution. Processing by a stereoscopic image processor (SIP) to correct 3D problems may incur a reduction of the resolution of the produced video images, while this might not be easily apparent on a stereoscopic display, the loss of resolution will degrade the quality of a 2D video feed derivate from the 3D feed.

The SIP will correct distorted 3D video by cropping parts of the left and right feed to reduce or correct the problem, then resize the video back to the desired resolution.

Transport. Whether transported over fiber optics or copper, compressed or not, two methods for transport of the 3D feeds are currently readily available:

- The left and right images are transported simultaneously. While this ensures the maximum quality possible, this requires the double of the bandwidth usually required by standard production. This method was used by the CBC for local transport of the stereoscopic signals between the cameras and the production mobile. Over long distance using compression methods like MPEG-2 and H.264 synchronizing both signals can be hard.
- The second method involves resizing both image in a single video frame, either in a side-by-side, top-bottom or any other available methods. This method enables transport of stereoscopic 3D video signal using the same equipment and bandwidth than traditional 2D video. The main drawback is the reduced video resolution of each of the images (half vertical resolution for side-by-side images), as illustrated in Figure 5.3. Thus, deriving a 2D signal from a stereoscopic 3D signal that was fitted in a single video frame would severely impact on video quality. Resized material should not be used to create a 2D program.

Figure 5.3 Side-by-side video frame.

Contribution (portable camera). When shooting in stereoscopic 3D, video quality from portable camera transmitting their feed over wireless is limited due to the bandwidth limitation. Since the bandwidth is limited, the 3D feed is usually transmitted as side-by-side signal; a derivative 2D feed will be impaired.

6. CONCLUSION

Clearly there is great interest in the development and deployment of 3D technology. The market is currently being driven by the content creators and the manufacturers of the products that consumers will have to buy in order to experience 3D in their homes. There is no denying that 3D has been successful in the theaters. So far in 2011, there are at least 50 feature films released or scheduled for release in 3D. 3D versions of many feature films are available for sale, rent or download. ESPN, Discovery and others have launched or are planning the launch of cable/satellite delivered 3D networks. Consumer product manufacturers are delivering 3D ready displays for home viewers. There is a rapidly developing infrastructure for the creation and distribution of 3D content.

What is unclear at this time is whether 3D for the home is a sustainable business or a passing fad. The 3D experience in the controlled environment of a movie theater is quite different than what will be experienced in an average home. Will the experience of the big screen transfer to the small screen with enough quality and impact to compel consumers to invest more money in their home entertainment systems? How will the legacy content and 3D content coexist in the consumer's home?

For the broadcaster, will there be a compelling business model that makes investing in the technology to pass through or originate 3D content worthwhile? In the case of production and transmission of separate 2D and 3D content versions, a significant investment will need to be made within the broadcast environment to handle two independent programs. There are the additional uncertainties of spectrum availability, the fact that most broadcasters are fully utilizing their existing channels for HD and ancillary SD program multicasts, and the coming of mobile DTV. Against that backdrop, the case for speculative investment in 3D technology may be off the radar and outside of the current strategic plans of most local broadcast stations.