

# Stereo/Multiview Picture Quality: Overview and Recent Advances

Stefan Winkler, Dongbo Min \*

*Advanced Digital Sciences Center (ADSC),  
University of Illinois at Urbana-Champaign (UIUC),  
Singapore 138632*

---

## Abstract

Stereoscopic 3D content brings with it a variety of complex technological and perceptual issues. For the percept of depth to be convincing, consistent, and comfortable, a large number of parameters throughout the imaging and processing pipeline need to be matched correctly. In practice, tradeoffs are inevitable, which may then affect the quality or comfort of the 3D viewing experience.

This paper reviews stereo/multiview picture quality from an engineering perspective. With a focus on recent or emerging approaches and technologies used in 3D systems, it addresses in particular depth issues, multiview issues, display issues, viewer issues, as well as possible measurements and standards for 3D quality.

*Key words:* Depth Perception, Quality of Experience (QoE), Human Visual System (HVS), 3D Television (3DTV)

---

## 1 Introduction

Hardware upgrading costs, lack of 3D content, and the need for glasses have long been limiting factors for the widespread acceptance of stereo/3D technology outside of special settings such as virtual reality or gaming. With the arrival of more affordable equipment, a surge in 3D content production, and autostereoscopic displays, 3D viewing may receive a significant boost. As the

---

\* Corresponding author

*Email addresses:* stefan.winkler@adsc.com.sg (Stefan Winkler),  
dongbo@adsc.com.sg (Dongbo Min).

<sup>1</sup> This work is supported by the research grant for ADSC's Human Sixth Sense Programme from Singapore's Agency for Science, Technology and Research (A\*STAR).

technology becomes more widely adopted and mature, quality issues rise to the forefront of concerns. For 3D to become a widespread success, high quality of the presentation and comfortable 3D percepts are essential.

Quality issues for images and 2D video have been studied quite extensively [1], and commercial quality assurance (QA) tools are already being deployed to monitor video quality in real time. Most of these tools are designed to pick out common spatial and temporal distortions of the video resulting from compression and transmission.

Stereoscopic 3D adds another layer of complexity on top of the common 2D impairments from video compression and transmission issues [2]. Furthermore, stereoscopic content may have potential psychological and physiological effects [3, 4], especially when 3D is not produced, processed, and presented correctly. Symptoms such as eye strain, headache, or dizziness are commonly reported by viewers. This underlines that 3D viewing comes with more severe concerns than 2D. These effects need to be better understood, and one of the primary practical goals must be to minimize or prevent possible discomfort caused by 3D content.

There are some excellent reviews of the perception of stereoscopic pictures [5, 6] as well as a 3-volume treatise on the subject [7]. A special issue [8] delves into many other 3DTV-related topics. Building on an earlier short compendium [9], our aim here is to provide a coherent overview of 3D picture quality and in particular its inter-relationship with recent algorithms and emerging techniques for 3D production, processing, and display.

The paper is organized as follows. We first give a brief overview of stereoscopic viewing basics as well as various technologies used in the 3D imaging pipeline. We then discuss specific issues with 3D viewing, in particular depth, multiview, display, and viewer-specific issues. Finally, we identify a number of possible Quality of Experience (QoE) indicators for 3D and review current standardization efforts for subjective and objective quality measurement.

## 2 Stereoscopic Viewing Basics

### 2.1 Depth Cues

3D is all about the perception of depth. There is a multitude of depth cues that the human visual system (HVS) uses to understand a 3D scene [10]. These can be classified into oculomotor cues coming from the eye muscles, and visual cues from the scene content itself. They can also be classified into monocular

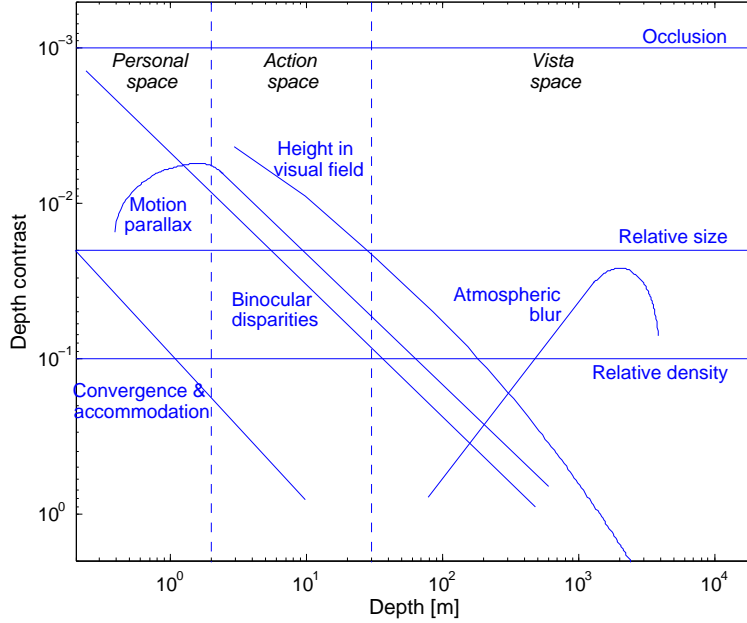


Fig. 1. Depth discriminability thresholds for different depth cues as a function of distance [12, 13].

and binocular cues [11].

Oculomotor cues include accommodation and vergence. Accommodation, a monocular cue, refers to the variation of the lens shape and thickness (and thus its focal length), which allows the eye to focus on an object at a certain distance. Vergence, a binocular cue, refers to the muscular rotation of the eyeballs, which is used to converge both eyes on the same object.

Just like the oculomotor cues, the visual cues consist of monocular and binocular cues. There are many monocular visual cues, such as relative size, familiar size, texture gradients, perspective, occlusion, atmospheric blur, lighting, shading, shadows, and motion parallax. Retinal disparity is the main visual binocular cue, and also the primary additional cue offered by stereoscopic systems such as 3DTV.

Not all cues have the same importance in the visual system, and their relative importance depends on distance, among other factors. Depth discrimination functions for different depth cues from [12] are shown in Figure 1. Some depth cues are independent of distance, such as occlusion or relative size, whereas others are distance-dependent, such as disparity or vergence. At the same time, binocular disparity is one of the most accurate and efficient depth cues [14]. It works best near fixation [15]. Elsewhere, vision relies on other depth cues such as perspective or familiar size, but those depend on the scene content and are therefore not always reliable. Depth from defocus blur relies on fewer assumptions and can fill in the parts of visual space where disparity is imprecise [16].

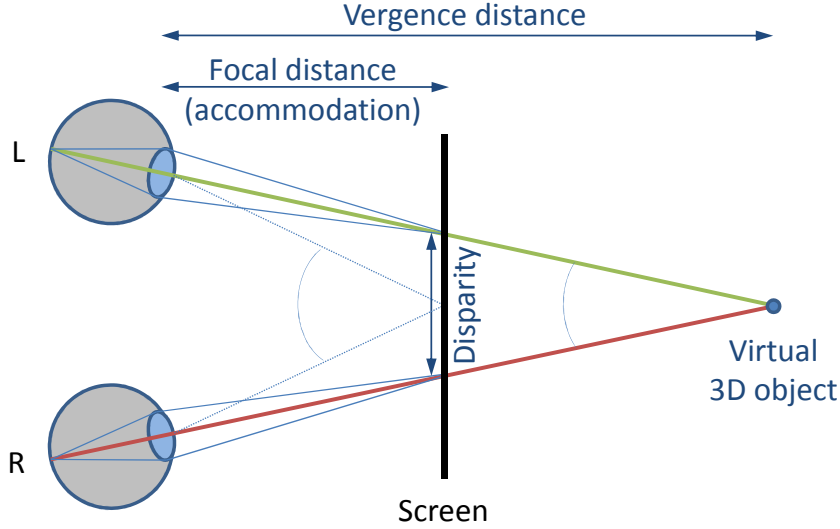


Fig. 2. Basics of stereoscopy.

However, it is still unclear how the different depth cues are combined to form an overall depth impression.

## 2.2 Stereoscopy

The basics of stereoscopy can be briefly summarized as follows (see Figure 2). A point of a 3D object is projected onto a screen in two locations, representing the views of the 3D object from the left and right eye, respectively. The left view is visible only to the left eye, and the right view only to the right eye. The disparity between the left and right views translates into a relative displacement of an object viewed along the two lines of sight caused by viewpoint change, which is called parallax.

## 3 Technical Issues

In this section, we describe various technical issues with regard to 3D formats, 3D production, 3D video coding, and 3D display. Table 1 provides a summary.

### 3.1 3D Formats

3D content can be represented either as two or more individual views from slightly different viewpoints, or as a combination of such views and depth maps (a.k.a. color-plus-depth). Since capture, processing, and display may work with

Table 1  
Summary of 3D Technologies.

Formats	Production
<b>Stereo/Multiview representation</b> - Frame/service compatible formats  <b>Color+Depth representation</b> - Multiview color-plus-depth video  <b>Conversions</b> - Depth image based rendering - View interpolation - Stereo matching	<b>Stereo/Multiview cameras</b> - Camera calibration  <b>Depth sensing</b> - Time of flight - Structured light - Depth video enhancement  <b>2D-to-3D conversion</b> - Using depth cues - Semi-automatic approaches
Coding	Displays
<b>Stereo/Multiview coding</b> - H.264/MPEG-4 AVC extension - View scalability - Backward compatibility  <b>Depth coding</b> - Non-standard transforms - H.264/MPEG-4 AVC encoding	<b>Anaglyph</b>  <b>Polarized</b> - Passive glasses  <b>Time-sequential</b> - Active shutter glasses  <b>Autostereoscopic</b> - Parallax barrier or lenticular

different representations, methods for high-quality conversions between them are also essential.

Storing the individual views separately is perhaps the most obvious approach. Alternatively, the views may be combined in various ways, e.g. for compatibility with existing formats. In many of today's distribution systems for stereoscopic 3D, two views (left/right) are packed into a single frame [17]. This frame-compatible format allows the direct use of existing 2D video system components for 3D content. Different types of packing exist; examples include side-by-side, top-and-bottom, various interleaving options, or time-sequential arrangement of frames.

The packing of two views into a single frame of the same size inevitably leads to a loss of spatial resolution, which can be further aggravated by multiple repacking and reconstruction steps.<sup>2</sup> However, the concept of resolution in 3D

<sup>2</sup> Newer standards such as the High-Definition Multimedia Interface (HDMI) 1.4 also specify a Full HD 3D (FHD3D) mode, where the two views are stacked verti-

is not sufficiently described by the 2D resolution of the individual views [18]. Also, compression for frame-compatible systems is often done with standard encoders, which may not distinguish between 2D content and 3D packed content. The resulting video compression artifacts (macroblocks, ringing, etc.) may cross view boundaries and lead to 3D artifacts.

Color-plus-depth is another common representation, where one (or more) color images/views and associated depth maps are used to represent 3D content [19, 20]. Unless the depth information is captured directly using depth sensors (see the following section on 3D production), disparity maps can be estimated using stereo matching methods. A number of approaches have been proposed, based on local cost aggregation [21, 22] using high-quality nonlinear smoothing algorithms, or global labeling optimization approaches using belief propagation or graph cut [23], but still exhibit various problems related to illumination and occlusion.

Conversely, missing views can be generated from color-plus-depth by warping a color image with the depth information, which is known as depth image based rendering (DIBR) [24]. Disocclusion problems can be addressed with simple interpolation or image inpainting techniques [25, 26]. Certain holes that are complementary among the views can be compensated for by a weighted summation of rendered (warped) images. In cases when color cameras are set with a non-parallel configuration, or when a virtual view is beyond the field of view of the color cameras, the holes on the synthesized view should be filled by image inpainting, which can be guided by color images and/or depth maps. Color values corresponding to the holes are inferred by propagating inherent structure and/or texture information in a semantic manner [25, 26].

This is also important if more views are required by the display than are available in the video stream. Virtual views can be synthesized using correspondence information, which again requires reliable disparity maps. For instance, in order to generate  $M > N$  views from  $N$  input views, two neighboring color images are first warped using the associated depth maps. Virtual views are then generated by interpolating the warped images [27].

Layered depth image (LDI) [28] is another popular format widely adopted in DIBR techniques. LDI is very efficient for rendering 3D scenes of complex geometry [29–31]. Unlike a 2D depth map, each pixel of the LDI contains multiple depth values measured based on a reference camera viewpoint. These multiple depths are then sorted according to their value. Such multiple depth layers enable rendering arbitrary views of scenes. The LDI can be directly generated with a set of multiple depth images, or it can be estimated from input color images without explicit depth maps, e.g., using the voxel coloring

---

cally on top of one another, with a corresponding doubling of lines in the resulting “packed” frame and no loss of resolution.

algorithm [32].

Alternatives for image interpolation without dense correspondence maps have been proposed to reduce the effect of matching errors, applications of which are 3D depth percept adjustment of stereoscopic video and frame-rate up-conversion of monoscopic video [33, 34].

### 3.2 3D Production

3D content production methods<sup>3</sup> can be classified into three categories, namely direct acquisition by stereo or multi-view cameras, active depth sensing, and 2D-to-3D conversion.

#### 3.2.1 Stereo/Multiview Cameras

Stereo or multiview cameras can be used to produce stereoscopic videos. For instance, Matusik [35] designed a distributed and scalable 3DTV system including real-time acquisition, transmission and 3D display of dynamic scenes in order to manage the high computation and bandwidth demands. Zitnick et al. [29] used eight color cameras to build a multiview capturing system and proposed a novel stereo matching algorithm for generating high-quality free-view video. The Stanford University multiview camera array [36] was built to capture 4-D light fields at video rate and render an arbitrary view. With the increasing demand for 3D multimedia content, commercial 3D cameras are being produced by a number of manufacturers. 3D smartphones, equipped with a stereo color camera and a 3D auto-stereoscopic display, are also becoming available.

Precise calibration and temporal synchronization of the cameras is very important for capturing high-quality multiview video. Temporally-synchronized multiple color videos are usually acquired using pre-calibrated cameras and signals triggered from multiple computers (or dedicated hardware). The color cameras are calibrated using specific calibration devices, e.g. checker board, sphere, or LED light. While popular camera calibration toolboxes [37–39] are generally designed for a single camera, they can be directly used for calibrating each of the multiview cameras.

Multiview videos have to be captured at the same moment using hardware- or software-based synchronization approaches. Hardware-based approaches use a control unit that triggers the multiview cameras via synchronization

---

<sup>3</sup> We consider mainly natural stereoscopic pictures, not computer graphics content.

signals [40]. Software-based approaches synchronize the cameras by networking host computers e.g. through the Ethernet [41]. Some recent works have attempted to address the lack of temporal synchronization in consumer camcorders by using subframe warping along estimated optical flow vectors [42] or trajectory filtering and matching schemes [43].

### 3.2.2 *Active Depth Sensing*

Active depth sensing technologies comprise time-of-flight (ToF) sensors such as those manufactured by Mesa Imaging and PMD Technologies, and methods based on structured light such as Microsoft’s Kinect. ToF sensors are based on a special complementary metal-oxide semiconductor pixel structure. They estimate a distance between the sensor and an object by extracting phase information from received light pulses. The structured-light approach usually recovers 3D shape from monocular images using a projector to illuminate objects with special patterns. Given a hybrid system consisting of color cameras and depth sensors, noisy and of low-resolution depth maps can be enhanced using the corresponding color images [44, 45].

Similar to multiview color cameras, in the multiple color and depth sensor setup, a fast, accurate, and practical calibration method is needed to produce geometrically aligned (single or multiview) color-plus-depth video. Given the popularity of consumer depth sensors (e.g. Kinect), many calibration methods have been proposed recently [46–48]. Unlike the existing algorithms [37–39] typically designed to calibrate a color camera, these methods often take into account both color and depth information. This helps improve calibration accuracy of the color-plus-depth camera pair and makes the algorithms more robust against matching outliers. A comparative study of calibration algorithms (especially for Kinect-style cameras) was presented in [49].

### 3.2.3 *2D-to-3D Conversion*

A number of methods have been proposed to create stereoscopic 3D video from conventional 2D content, called 2D-to-3D conversion [50]. Existing 2D content can be converted to 3D video by considering several depth cues such as motion parallax [51, 52], vanishing points/lines [53], or camera motion in a structure-from-motion framework [54, 55]. For instance, motion parallax is employed to extract approximated depth maps, assuming that objects with different motions are likely to have different depth values, i.e. near objects move faster than far objects [51, 52]. For static scenes with camera motion only, structure-from-motion techniques can produce depth maps from a series of video sequences using estimated camera calibration parameters [55].

In general, recovering the stereoscopic image from a single image is an ill-



posed problem, and fully automatic methods work well only on very limited scenes. Recently, most research has focused on semi-automatic conversion with minimal user input such as scribbling in key frames [56, 57].

### 3.3 3D Video Coding

For efficient storage and transmission of 3D video, multiview video coding (MVC) is gaining more attention from both academia and industry. The Joint Video Team (JVT) from the Moving Picture Experts Group (MPEG) of ISO/IEC and the Video Coding Experts Group (VCEG) of ITU-T recently defined requirements of test data and evaluation conditions for MVC [58] after including it as an extension of the H.264/MPEG-4 Advanced Video Coding (AVC) standard. These test data have been widely used to evaluate a variety of MVC algorithms. MVC provides a compact representation of 3D video with strong inter-view dependencies, enabling efficient compression of stereo and multiview video signals.

There are a few specific requirements for MVC [59]: view scalability, inter-view random access, and backward compatibility. View scalability means that a subset of the MVC bitstream can be accessed to decode a specific view from the multiview video data. The multiview video data should be accessed and decoded from a small portion of the MVC bitstream at any random access point, which is referred to as inter-view random access. Also, MVC should be backward-compatible with existing 2D H.264 decoders, e.g. the base view is standard format compatible, and the auxiliary views can be generated by MVC decoders, if needed.

Given multiple views of the same scene, inter-view dependencies as well as temporal dependencies are extensively exploited for efficient view prediction. A number of algorithms have been proposed to exploit view dependencies, e.g. hierarchical B-pictures in temporal and inter-view dimensions [60], illumination compensation [61], adaptive reference filtering to handle focus mismatch [62], view synthesis prediction using depth maps [63], etc.

Recognizing the high compression performance and backward compatibility of MVC, the Blu-ray Disc Association (BDA) recently selected the stereo high profile of the MVC extension as the coding format for stereoscopic HD video. BDA finalized the specifications of “Blu-ray 3D<sup>TM</sup>”, which are compatible with various stereoscopic 3D video formats. The specification is display agnostic, so Blu-ray 3D products can deliver 3D content to any compatible 3D display, regardless of the display type. The base stream is regular 2D content, compressed in standard H.264/MPEG-4 AVC format, so that 2D Blu-ray Disc players can read the base view as a traditional 2D video stream. The auxiliary

stream is compressed using the MVC extension, where the base stream video may be utilized for inter-view prediction, resulting in considerable bitrate savings.

Color-plus-depth video coding methods have received less attention so far. Earlier studies focused on developing new transform-based coding algorithms that are optimal with respect to the characteristics of the depth map, e.g. based on platelets [64] or shape-adaptive wavelets [65]. These transform-based approaches have shown to achieve better performance than existing compression algorithms [64, 65], but they are not compatible with current standards such as H.264. Thus, most studies have focused on reducing compression artifacts in the depth video when it is encoded with standard H.264, by utilizing post-processing algorithms as an in-loop filter [66, 67].

### 3.4 3D Display

We briefly describe stereoscopic and auto-stereoscopic 3D displays; volumetric and holographic displays are beyond the scope of this paper. For a more detailed review, please refer to [68].

Stereoscopic displays achieve the separation of the views for the two eyes by various means of view multiplexing, including anaglyph, polarized, and time-sequential 3D displays.

Anaglyphs rely on a pair of complementary color filters, most commonly red and cyan. They are popular in print and web content because glasses are cheap and can be used in combination with any traditional display. However, images are not of very high quality and usually monochrome; unmatched color filters often lead to crosstalk. The idea has been extended to the super-anaglyph technique [69], where narrow spectral bands in the red, green, and blue spectral regions with slightly different center frequencies are used for the separation of the two views. The color filters are integrated into both the glasses and the display. This technique is able to generate full-color 3D images.

For polarized 3D displays, the light is subjected to either linear or circular polarization with different orientations. The (passive) glasses also contain a pair of polarizing filters that match the display polarization. Polarization techniques are well suited for video projection, where both views are superimposed on the same screen; they are the method of choice for 3D cinema because of the inexpensive and maintenance-free glasses. In TV screens, polarization typically alternates at the scan-line level, resulting in a loss of spatial resolution.

Time-sequential 3D displays multiplex light by switching between the two views, i.e. both views are reproduced at full spatial resolution by a single

screen. Viewers wear active shutter glasses that switch on and off for the left and right eyes in sync with the display.

Some 3D displays combine the above two techniques: they actively switch between polarizations in a time-sequential manner. There is no loss of spatial resolution, and the presentation can be viewed with passive polarized glasses.

Auto-stereoscopic 3D displays, which allow a viewer to watch 3D video without glasses, typically use either parallax barrier or lenticular methods. The former employs an array of strips in order to multiplex light to the appropriate viewing regions. Light from the display screen such as liquid crystal display (LCD) or plasma display plate (PDP) is split with a parallax barrier, so that alternate columns of pixels on the screen reach the two eyes separately. For the latter, the display screen is located in the focal plane of lenticular lenses, which refract the incident light from the display screen and direct stereoscopic images onto the left and right eyes. Early vertical lenticular displays suffered from vertical banding and flipping effects between viewing zones. Slanted lenticular displays have been widely utilized to address these problems [70].

## 4 Depth Issues

### 4.1 *Vergence-Accommodation Conflict*

In normal 3D viewing conditions in the real world, when the eyes fixate on something, that point is brought into focus using accommodation of the lens, while the eyes converge on the same point. The fixation point falls on the horopter, a set of locations in space that project onto corresponding retinal points. Its shape is discussed in detail in [71]. Points located in front of or behind the horopter create negative or positive disparities on the retina, respectively, which is used as a binocular depth cue by the human visual system.

A disparity of zero means the object is perceived on the screen plane. Objects with positive disparity appear behind the screen. The inter-ocular distance places an upper limit on disparity; beyond this limit, the eyes are forced to diverge, placing the object beyond infinity, which is impossible in nature and should be avoided [18, 72].

With negative disparity (object appearing in front of the screen), the viewer's eyes cross; this can be strenuous, which is why filmmakers are cautious about using excessive negative disparities for extended periods of time [73]. Another issue with objects appearing in front of the screen is a possible border or window violation. This happens when a (virtual) object with negative disparity

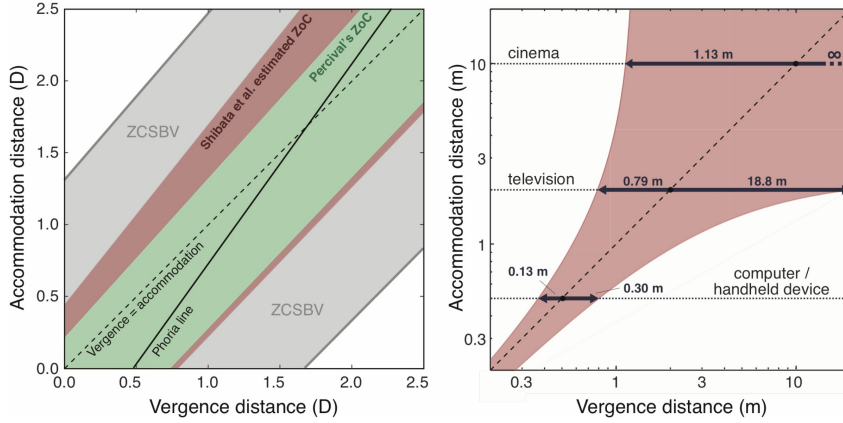


Fig. 3. Zone of comfort (ZoC) [6, 78]. Accommodation distance as a function of vergence distance in diopters (left) and meters (right). Gray: zone of clear single binocular vision (ZCSBV); green: Percival’s zone of comfort; red: estimate from [78].

is cut off at some point by the physical border of the screen. This violates our intrinsic assumption that the screen is a window through which we view the scene and destroys the 3D percept. This is particularly problematic at the left and right sides of the screen, allowing one eye to see what the other does not.

In stereoscopic viewing, there is no real object; therefore, the eyes must focus (via accommodation) on the object’s projection on the screen, which is often at a different distance from the virtual 3D object on which the two eyes converge (vergence distance, see Figure 2). This conflict between accommodation and vergence is one of the main reasons for discomfort [6, 74, 75].

The HVS is tolerant of a certain amount of blur, such that a small region around the accommodated point is still perceived to be in focus. This is also known as the depth of focus. Its size varies with pupil diameter, which in turn varies with light level. In general, visual comfort decreases with increased depth of focus, with 0.2 diopters representing an acceptable range [76].

There is a certain range of both vergence and accommodation around the empirical horopter where the images from the two eyes can be comfortably fused into a single percept by the HVS [77]. This range is also known as Panum’s fusion area; points outside of it result in double vision. The size of Panum’s fusion area determines the range of depth in a scene (also called “depth bracket”) that can be comfortably presented to a viewer.

Ideally, stereoscopic presentations should be displayed such that they fall entirely within this range [79, 80]. There are many approaches to characterize the size of the comfortable viewing zone; a thorough comparison of the various limits is given in [81]. Figure 3 shows comfort zones determined by various methods and experiments, including the zone of clear single binocular vision, which describes the maximum attainable decoupling of accommodation and

vergence, Perceival’s zone of comfort, and the zone of comfort according to recent measurements from [78].

The plots also highlight the impact of viewing distance on the size of the comfort zone; for example, the depth range is much larger in the cinema than for a TV at home, and even smaller on a mobile device. Since viewing distance and the amount of ambient light are major factors, 3D content should be adapted to individual viewing conditions (as screen size largely determines viewing distance) – see more on this below. Furthermore, the spatial and temporal properties of the content also affect the size of the comfort zone [82].

Limitations on disparity and depth range as discussed above apply mainly to static scenes. However, dynamic changes of apparent object depth or depth bracket are common in 3D video, sometimes leading to depth discontinuities. Rapid depth variations can result in viewer discomfort, because the human visual system is unable to follow the changes and to reconstruct depth properly [73,79]. This is a common problem at transitions (e.g. scene cuts). Rapid depth variations can be even more detrimental to viewing comfort than a large depth bracket [80]. In general, depth changes should be slower and less frequent than in 2D. One of the important functions of 3D movie editing tools is to adjust the depth bracket across scene changes, so that the objects of interest maintain the same depth [83]. The smooth transition of depth distributions at scene changes can mitigate the effect of temporal depth discontinuities.

## *4.2 Depth Adjustments*

In general, 3D depth perception is influenced by various factors, which can be classified into content characteristics (e.g. depth variation and scene complexity) and viewing environments (e.g. display size and viewing distance) [5,84]. These internal and external factors should be considered carefully in producing stereoscopic 3D content. In many circumstances, the content has to be post-processed to adjust production settings. In particular, the stereoscopic depth has to be matched with the viewing environment and conditions [72]; otherwise, the depth effect would not be as desired, for example if 3D content produced for the cinema is viewed on a TV screen or on a mobile device without corresponding depth adaptation. In the following, we briefly describe techniques for adjusting the 3D viewing experience, namely depth grading and depth adaptation.

### *4.2.1 Depth Grading*

When stereo images captured in a parallel camera configuration are fused by viewers, the two (left and right) views always produce a negative horizontal

disparity  $d_h(x_l, y_l) < 0$ , where  $d_h(x_l, y_l) = x_r - x_l$  for two matched pixels  $(x_l, y_l)$  and  $(x_r, y_r)$  on the left and right images. Note that  $d_h = 0$  for a 3D point at infinite distance.

In order to adjust the zero parallax setting (ZPS) in the parallel camera structure, Fehn and Pastoor [85] proposed to horizontally translate color sensors of stereoscopic cameras in the parallel structure by a small shift relative to the position of the lenses, allowing one to choose the convergence distance in 3D scene. Such a sensor shift method is often used as an alternative to the ‘toed-in’ approach, because it does not cause keystone distortions and depth-plane curvature in the stereo image.

In post-production, depth grading can be achieved by a similar process of horizontal image translation (HIT) [86]. Because of the potential need to enlarge and to crop the image during this process, filmmakers usually shoot in overscan mode, leaving a buffer zone around the edges of their frame.

#### 4.2.2 Depth Adaptation

For various reasons, stereoscopic video often needs to be adapted or re-targeted to specific viewing conditions [87–89]. The stereoscopic video is rescaled e.g. by considering depth range/order, display size, and viewing distance. Such operations are, however, not a simple linear scaling, which may lead to serious depth distortion during the stereo video manipulation.

Recently, several approaches have been proposed using an energy minimization formulation [88, 89], which is often called ‘stereo retargeting’. Their ideas are usually extended from ‘seam carving’ [90], which was originally proposed to perform content-aware image resizing while preserving important visual saliency present in a 2D image. Instead of removing a seam from 2D video, these 3D stereo retargeting algorithms attempt to remove a pair of seams from the stereoscopic video during the rescaling procedure. Thus, the retargeted video contains different depth variation, depth range, and image size, yet is geometrically consistent with the original input video, i.e., the original depth order is preserved.

Jung and Ko [91] presented a depth sensation enhancement method considering the HVS for adjusting depth range. Depth values are modified based on the just noticeable depth difference (JNDD) [92], which is the minimum threshold for perceiving depth variations between two neighboring objects. The spatial resolution of the depth map is maintained. This approach is also based on an energy optimization framework consisting of depth fidelity, depth order preservation, and depth difference expansion [91]. Another recent automatic depth grading tool adapts stereoscopic content to digital cinema or TV screens by positioning the objects of interest in the same plane using a sparse

disparity map [93].

### 4.3 *Depth Mismatch*

As highlighted in Section 2.1, stereoscopy recreates only one of many visual depth cues the HVS makes use of. In the real world, all the different cues match. In a stereoscopic 3D projection, they may be violated, and the mismatches can reduce or even destroy the 3D percept. In fact, stereoscopic 3D viewing almost always leads to some cue conflicts due to the nature of the medium and the technology; yet most viewers find 3D content compelling despite these violations [6].

The mismatch between oculomotor cues and monocular cues caused by the vergence-accommodation conflict is one such example (see Section 4.1). Another is a conflict between visually induced motion, which is naturally stronger for a 3D presentation, and vestibular signals in the brain, which provide information about our own movement and spatial orientation. Stereoscopic images are more effective in visually inducing motion sickness than 2D images [94,95].

How the brain deals with cue conflicts is still an open question [13]. In a study on conflicts between depth cues of perspective and binocular disparity for example, observers were found to use a weighted linear combination of cues for small conflicts, and cue dominance for large conflicts [96].

Because of the large number of depth cues and their visual complexity, depth mismatch can be particularly challenging to detect using objective quality measurement methods [13,97].

Post-production insertions (subtitles, captions, logos, etc.) into 3D content can also represent a major problem in terms of possible depth mismatches [98]. Even if the underlying 3D content was produced and edited correctly, these inserts may introduce further issues. For example, it is important to check the depth range of the scene before insertion, otherwise subtitles may be visible when they should be occluded by other objects in the scene closer to the viewer, which can violate the consistent 3D percept. At the same time, their depth should not be too different from the overall depth bracket of the scene. Multiple inserts or subtitles in a scene can pose additional problems if they appear at different depths. Subtitles can also cause unnatural depth perception when internal parameters (e.g. focal length) of the stereo camera change, whereas the subtitles remain fixed. All these different requirements make subtitling in 3D a surprisingly complex process [99].

#### 4.4 Depth Quality

In depth estimation (or acquisition) and compression/transmission stages, the performance of algorithms has usually been measured by the accuracy of the depth maps [100]. In many 3D applications, however, depth maps are used to generate synthesized views corresponding to virtual viewpoints. For instance, each viewer may have a different optimal viewing condition due to individually different depth perception. Thus, by synthesizing the virtual view, the baseline distance between two images in a stereoscopic (glasses-based) 3D display can be adjusted according to the optimal viewing angle of each viewer [101]. Interpolated views are even more important for auto-stereoscopic 3D displays. For wide viewing of aspect coverage and distances, these displays require a large number of views (e.g. 18). Given the practical limitations in the number of cameras for capturing content, interpolated views can be generated more efficiently from depth maps [102]. Therefore, quality metrics should be defined based on their application (e.g. rendered view quality), not the depth map itself [67].

However, providing an accurate depth map at high resolution is challenging. Depth maps, whether they are estimated by stereo matching methods, structured light, or active depth sensors, are prone to errors because of the ill-posed problem of depth estimation and inherent physical limits of the sensor such as noise or interference. They may be further impaired during coding and transmission [19, 64]. The quality of a virtual view is highly dependent on the accuracy of the estimated depth map, especially at object boundaries, so the effects of depth estimation errors and related coding/transmission artifacts have to be investigated [103, 104]. More specifically, depth discontinuities should be well-preserved for providing high-quality rendered images. An upper bound on the allowable depth error could provide new insights into depth sensor development and coding/transmission system design [105].

### 5 Dual- and Multiview Issues

Good correspondences between views are essential to creating a high-quality 3D viewing experience. Various possibilities for mismatch are presented in this section. Table 2 summarizes challenges and possible solutions of these issues, which apply to both dual- and multiview 3D presentations.



Table 2

Summary of Dual- and Multiview Issues.

Issues	Approaches	Challenges
<b>Color/Focus Mismatch</b>	<ul style="list-style-type: none"> <li>– Correspondence based color correction [61, 106]</li> <li>– Key-frame based handling [107]</li> <li>– Depth-adaptive filtering [62]</li> </ul>	<ul style="list-style-type: none"> <li>– 3D quality of color/focus mismatch correction</li> <li>– Content/display dependency</li> </ul>
<b>Aliasing</b>	<ul style="list-style-type: none"> <li>– Display bandwidth analysis &amp; pre-filtering [108–111]</li> <li>– Disparity adaptive anti-aliasing filtering [112, 113]</li> </ul>	<ul style="list-style-type: none"> <li>– Blur at discontinuities</li> <li>– Quality of anti-aliasing filter</li> </ul>
<b>Compression</b>	<ul style="list-style-type: none"> <li>– Asymmetric stereo coding (binocular suppression) [114–118]</li> <li>– Inter-view quality consistency</li> </ul>	<ul style="list-style-type: none"> <li>– Upper bound of asymmetry</li> <li>– Content/display dependency</li> </ul>
<b>Monocular Occlusion</b>	<ul style="list-style-type: none"> <li>– Disocclusion handling with interpolation/inpainting [25, 26]</li> </ul>	<ul style="list-style-type: none"> <li>– Temporal consistency</li> </ul>

### 5.1 View Mismatches

Capturing stereo and multiview video is challenging due to the difficulty of controlling multiple cameras (see Section 3.2). Unwanted mismatches between corresponding views may arise at various stages of the production and distribution chain, if any of the following are not matched [119, 120]: camera optics and sensors; white balance; shutter speed; aperture; gamma; geometry (camera angle and position; picture skew or cropping). Small geometrical distortions that are limited to the image edges generally do not impact viewing experience; however, vertical offsets or different blur levels are noticeable even in small amounts [121]. An example of color differences between multiple views is shown in Figure 4. Most of these mismatches can be corrected through careful calibration or during post-production. A special (extreme) case is view reversal, which occurs when the left view is presented to the right eye and vice versa.

Compression can also lead to view differences in terms of artifact severity (blockiness, blur), time-varying quality (e.g. different coding parameters), network impairments, and error propagation (especially when the views are contained in separate streams). Views that are out-of-sync even by only 20-30 ms (e.g. in field-sequential displays) can cause depth errors [122, 123].

Monocular occlusion refers to regions of a scene that can only be seen by one eye. These may be inadvertently added, distorted, or simulated improperly.

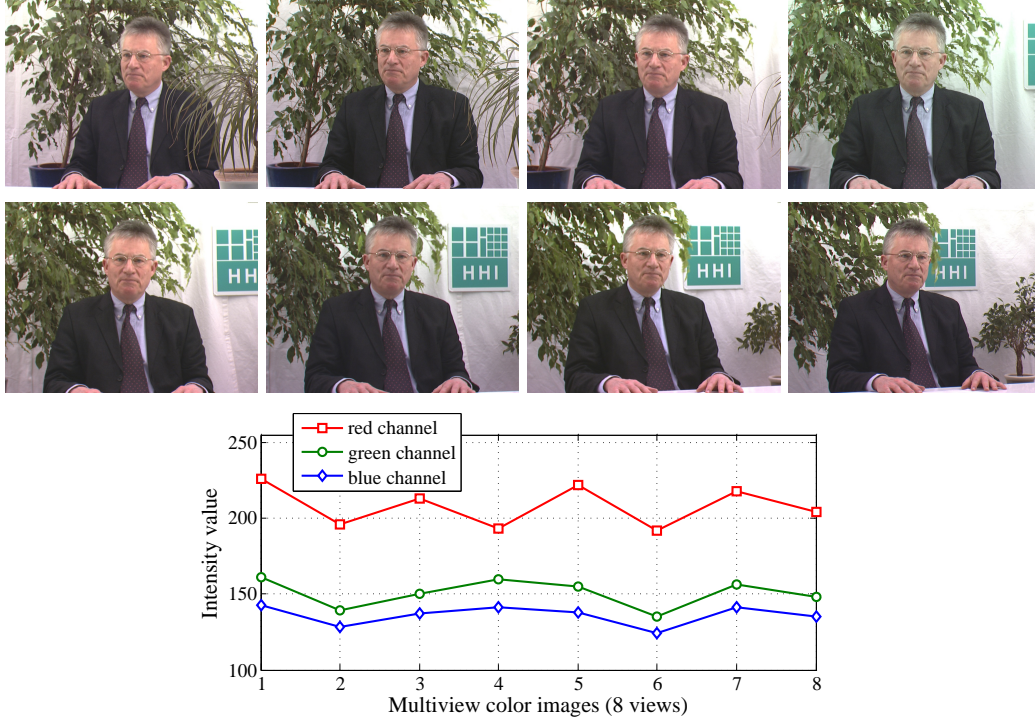


Fig. 4. Color and brightness differences between multiple views in the ‘Uli’ sequence. The plot shows the RGB intensities of a single matched pixel in the face across all views.

This is a particular problem for 3D content that has been poorly reconstructed from 2D content (cf. Section 3.2.3). Since occlusion is one of the most significant depth cues, this can lead to erroneous percepts quite easily.

If any of these view differences become too severe, the HVS may be unable to fuse the two images into a consistent 3D percept; in other words they will be in competition with each other. This is known as binocular rivalry. At first, the HVS may alternate between the two views. However, the HVS normally cannot tolerate rivalry for long and attempts to reconcile the conflicting inputs by suppressing (masking) one of them. This can affect the entire view of one eye, or just certain parts of the visual field from either eye [7], a phenomenon also known as binocular suppression. Both effects must be taken into account for accurate measurement of the perceptual impact of these differences.

## 5.2 Coding-Related Issues

Binocular suppression has been exploited in asymmetric stereo video coding [114], where the two views are encoded at different quality (e.g. through spatial scaling or quantization). It has been shown that the HVS is capable of compensating for the lack of high frequency in the auxiliary view if the refer-

ence view is of sufficiently high quality. Given the same bitrate for stereo video, asymmetric coding provides better depth perception than symmetric coding when the quality of the auxiliary view is above a certain threshold [115–117]. Zhang and Li [118] evaluated the 3D depth percept of various indoor and outdoor sequences with different depth variation, scene complexity, and motion of visual content. They controlled the quality of two views independently by compressing them at different bit rates, and showed that 3D depth percept may vary dramatically according to visual content as well as compression quality. However, there are still a lot of open questions, such as: Where is the upper bound of this asymmetry? Which method provides the best 3D perception between spatial scaling, quality scaling, or their combination? Content and display dependence also need to be investigated.

As mentioned in the previous section, color differences and focus mismatch may be present due to incomplete camera calibration, different shooting position, illumination change, and so on. Various methods have been proposed to address these problems in the generic stereo/multiview coding concept [61, 62, 106, 107]. They argue that mismatch correction can improve compression efficiency significantly, allowing one to employ predictive coding of the DC offset components across the different views. It also mitigates visual artifacts that may occur in virtual view synthesis. For instance, Hur *et al.* presented an adaptive illumination compensation method for each macroblock (MB) [61]. This alleviates the performance degradation due to illumination changes in MVC, assuming that the DC component in the MB is influenced by local illumination changes. In [106], a color correction method based on a least-squares regression was proposed as a pre-processing step for MVC. A similar color correction method was extended to the temporal domain for improving efficiency [107]. Kim *et al.* proposed a new method that handles illumination and focus mismatches between different views in MVC [62]. Disparity vector and illumination change are jointly estimated for each block.

Most of the methods have focused on improving compression efficiency by addressing view mismatches in MVC. However, the effects on quality and depth perception as well as content and display dependency have not yet been investigated in detail.

### 5.3 Aliasing

Aliasing happens due to the rendering of 3D video with high-frequency components on 3D displays. We can distinguish the well-known intra-perspective aliasing within a single view due to the discrete 2D pixels of each view, and inter-perspective aliasing due to the discrete number of views [108, 109], which is unique to stereo/multiview content. Various techniques have been proposed

to alleviate 3D aliasing problems.

Moller *et al.* [110] presented a spatially varying filter to reduce inter-perspective aliasing by leveraging the knowledge of per-pixel scene depth based on display bandwidth analysis. The modulation transfer function, which is a function of depth and spatial frequency, was used to filter 3D video by finding the minimum sampling frequency for a given maximum depth. Konrad *et al.* [111] studied inter-perspective aliasing by analyzing the multiplexing process from a sampling perspective, based on which they derived a filter to prevent the aliasing caused by the nonorthogonal grid pattern of the 3D display.

A unified approach based on the frequency analysis of light fields was also proposed by combining re-sampling of light fields and display prefiltering techniques [109]. Light field video obtained from multiple camera devices is reparameterized to the coordinates of the auto-stereoscopic display, and the light field signals are then prefiltered to match the Nyquist bandwidth of the display pixel grid. This approach addresses aliasing within each view as well as inter-perspective aliasing [108].

Kim *et al.* [112] proposed a disparity-adaptive anti-aliasing filter, based on a frequency analysis of the 3D image from a geometry model of depth perception. The depth distribution of the scene is band-limited by disparity-adaptive low-pass filtering for enhancing viewing comfort. This approach was extended temporally by jointly considering disparity and motion in 3D video [113].

Although the above-mentioned anti-aliasing filters reduce aliasing artifacts, some viewers may actually prefer the aliased 3D video, which is generally sharper [109]. Further subjective evaluation is needed to determine the right balance between aliasing and blur.

## 6 Display Issues

Presenting two or more views on a display usually takes its toll, leading to reduced brightness and reduced spatial or temporal resolution, compared to showing a single 2D view on the same display. Two other important issues that affect all stereoscopic display technologies are discussed in the following.

### 6.1 Geometric Distortions

Many displays have one or more “sweet spots” for the best viewing experience. In particular, off-center oblique viewing angles lead to geometric distortions of

objects/angles, sometimes referred to as “lopsided keystone”. The HVS cannot compensate for the oblique viewing of stereoscopic 3D images in the same way it does for 2D images [6]. The reason is that the binocular disparities specify not only orientation and distance of the picture surface, but also the layout of the picture content [124]. It was found that sitting in an oblique position to the screen attenuates perceived immersion, but also motion symptoms [95]. Software is available that allows the simulation of the impact of various parameters on perception.<sup>4</sup>

## 6.2 Crosstalk

Crosstalk happens when part of one view also appears in another [125, 126]. This is mainly a display issue, although other sources are possible (e.g. compression artifacts or transmission errors, especially in frame-compatible systems).

Crosstalk can be classified into two categories according to how it is measured [127]: System crosstalk is defined as the leaking image from another view (content-independent crosstalk), whereas viewer crosstalk is defined as the ratio of the luminance of the unwanted ghost image vs. the desired image (content-dependent crosstalk).

Most stereoscopic 3D displays (except the stereoscope) suffer from crosstalk [127]. In anaglyph displays, crosstalk may occur when the color filters of the anaglyph glasses do not separate spectral components well enough, or when they do not match with the spectral emission of the display [128].

For polarized viewing with passive glasses, crosstalk can occur when viewers tilt their head or lie down. Linear polarization is particularly susceptible to this problem, whereas circular polarization is less sensitive (although any mismatch between the eye plane and the original camera plane will still affect the depth percept). Crosstalk can also happen for large disparities and imperfect polarization.

When active shutter glasses are used, the timing must be precisely synchronized with the display, otherwise crosstalk will occur. Furthermore, the response times (rise and fall times) of the display must be fast enough to complete the switch between the two views while the glasses switch. This is particularly challenging in areas with large brightness or color differences between views, or at scene cuts.

Auto-stereoscopic displays are prone to crosstalk around the view boundaries

---

<sup>4</sup> <http://www.macbarnacle.com/robin/Software.html>

due to their imperfect multiplexing [129]. An inclined angle of a slanted lenticular display maps multiple views into the lenticular sheet with sub-pixel precision. However, the boundaries of subpixels cannot be covered exactly due to inherent physical limitations of the lens elements, which causes crosstalk. Wang et al. [130] formulated the relationship between crosstalk coefficients of each image on a slanted lenticular 3D display and then proposed a method to reduce the crosstalk by correcting the luminance values of each image displayed on screen. Lee *et al.* [131] exploited the geometric relationship between LCD subpixels and lenticular sheets with pattern images and then estimated a mapping between LCD subpixels and multiview images. This approach can adjust the viewing angle of the lenticular display by simply changing the mapping matrix. In order to address misalignment due to the subpixel mapping, a floating viewpoint image is synthesized using stereo images and corresponding depth maps [132].

While current display technologies do not achieve perfect separation of multiple views, some compensation is possible by digitally pre-processing the views before display [133]. The amount of intensity leakage from the unintended view is first estimated by modeling the inter-view dependency on the 3D display; multiple views are then pre-distorted to compensate for the distortion from crosstalk. This procedure can be carried out in the illumination domain [125] or in 3D color space [134]. Cancellation can also be achieved by locally modulating brightness levels [135, 136]. In [137], a crosstalk-free image obtained in the previous frame was used to remove crosstalk from the current frame.

A number of experiments demonstrate that ghosting from crosstalk causes discomfort and 3D quality degradation [126, 138]. Recent trials attempt to quantify the impact of crosstalk on quality [139, 140]. Interestingly, it was reported in [141] that in auto-stereoscopic displays, some crosstalk helps mitigate dizziness that often appears due to rapid view switches when an observer moves viewing spots, resulting in better 3D depth perception.

## 7 Viewer Issues

About 5% of viewers have defective stereo vision [142]. Furthermore, each viewer has different optimal viewing conditions due to individually different (depth) perception. Aside from differences in visual acuity or optical corrections between left and right eyes [95], this is determined by several other factors such as age, gender, and degree of previous 3D viewing experience [143–145]. Therefore, viewer idiosyncracies must be taken into account for an accurate evaluation of 3D quality [146]; unfortunately our understanding of the distribution of the visual sensitivity of the population for 3D viewing is far from complete.

The degree of previous 3D viewing experience can be an important factor. It is expected that as viewers become more experienced with 3D viewing, visual discomfort is gradually reduced, but at the same time expectations of 3D content quality may also rise, similar to the transition from standard-definition to high-definition television, or from VHS tape to DVDs.

Gender may also be a factor in 3D perception. There are differences between men and women in terms of visual perception abilities due to different inter-ocular distances [147] and different response times in 3D perception [143].

An important parameter that varies between viewers is the inter-ocular distance, or more specifically inter-pupillary distance. For adults, the average is about 63 mm, but it varies from 45 to 80 mm. For children, it can be as low as 40 mm [147]. For the same disparity stimulus, the perception of depth increases for smaller inter-pupillary distance, and vice-versa. Therefore, especially for children, watching 3D content that has been produced for the average adult may cause discomfort. Although there is no evidence of permanent damage to the visual system due to 3D viewing, there is also no evidence to the contrary. These risks should be evaluated carefully for all 3D applications such as 3DTV, 3D movies, 3D gaming, etc. [148].

A number of reports reveal that there is a difference in stereoscopic depth perception across age groups. Visual abilities in general deteriorate with age, in particular the ability for lens accommodation and the accommodation-convergence relationship [144]. For instance, older people are less sensitive to perceiving depth and surface curvature [149]. Depth perception with age may vary according to the characteristics of 3D content such as disparity magnitude, disparity direction (crossed vs. uncrossed disparity), and orientation difference of corresponding lines in each view [145]. Some user studies showed that the discrimination performance of an observer was reduced for high disparity magnitudes, and objects with crossed disparity produced better discrimination performance [145]. In [150], it was argued that binocular suppression of older observers is greater than that of younger observers.

## 8 3D Quality Measurement and Standardization

Many of the issues discussed above can affect the 3D viewing experience. Both physiological and psychological effects can be observed:

- *Physiological effects* can be measured objectively. Visual fatigue<sup>5</sup> is the

---

<sup>5</sup> In the literature, visual fatigue and discomfort are often used interchangeably; we follow the terminology suggested by [151].

most common physiological symptom associated with stereoscopic 3D [75]. It comes with a measurable decrease in the performance of the visual system, such as changes in eye movement characteristics, eye blink rate, or pupil diameter [152, 153]. Other physiological effects include changes in heart rate or brain activity [3].

- *Psychological effects* are often linked to physiological symptoms, but are measured using viewer surveys and subjective ratings. Eye strain, headache, dizziness, or nausea belong to this category. In a recent study of 3D cinema, two thirds of viewers reported one or more symptoms during the movie, and one quarter still two hours afterwards [4].

In addition to the individual quality issues, parameters, and symptoms identified in this paper, it is useful to define descriptors that quantify the overall viewing experience of a 3D presentation. The following quality descriptors for 3D content have been proposed:

- *Depth Quality*. The depth characteristics of 3D content need to be analyzed to validate whether the content is suitable for viewing [13].
- *Naturalness*. The ease with which viewers can fuse left and right views into a natural-looking 3D percept with smooth depth representation [154].
- *Immersion* or *Presence*. A 3D scene that looks natural enhances the viewers' sense of presence [155], especially in interactive applications.
- *Value-add*. The perceived benefit (or detriment) of viewing a piece of content in 3D over viewing the same content in 2D [156].
- *Discomfort*. The combined subjective sensation resulting from physiological and/or psychological effects of 3D viewing content [151].
- *Overall 3D Quality of Experience (QoE)* as typically measured by subjective Mean Opinion Score (MOS).

Unfortunately, there are no commonly accepted methods for measuring any of the above quality descriptors yet, whether subjectively or objectively, although standards have recently been emerging to address this.

Several interesting challenges for subjective quality assessment are highlighted in [18], such as selecting the right video content, finding the best physiological and psychological indicators of QoE, determining and testing for individual variations in depth perception, and analysis methods for multi-dimensional quality indicators. To address the last issue for example, Strohmeier et al. [157] extended the Open Profiling of Quality (OPQ) approach, which combines a conventional quantitative perceptual evaluation with a qualitative descriptive evaluation based on naive participants' individual vocabulary, and successfully applied it to Mobile 3DTV.

In parallel to the subjective quality assessment issues that still need to be worked out, there are also ongoing efforts in objective quality assessment of



3D content [158]. This research is still in the early stages. As opposed to traditional 2D content, where a direct analysis of the signal can produce accurate quality estimates, in 3D the rendered version of the signal has to be analyzed, which requires new approaches. Another challenge is designing and thoroughly testing models for all possible combinations of distortions, because the number of variables is so much larger than for 2D.

Many standard organizations have also started addressing these issues from both subjective and objective measurement angles:

- ITU-R has issued a new Recommendation on Subjective Methods for the Assessment of Stereoscopic 3DTV systems [159], which focuses on the three primary dimensions of picture quality, depth quality, and visual comfort. Similar work is on-going in ITU-T, including visual fatigue and safety assessment guidelines for 3D video as well as 3D display requirements.
- The Video Quality Experts Group (VQEG) is addressing three main areas, namely establishing ground truth data for subjective assessment methodology validation, determining the dependency of 3D quality assessment on the viewing environment, and validating objective 3D video quality measurement methods.
- IEEE launched work on a standard for the Quality Assessment of 3D Displays, 3D Contents and 3D Devices based on Human Factors (P3333); it will look at viewer, content, environment, display, and device characteristics.
- The 3D@Home Consortium, which recently merged with the International 3D Society (I3DS), has a steering team for human factors; together with the MPEG Industry Forum (MPEG IF), it published a glossary of terms for providing a common language for discussing, identifying and improving the subjective quality of stereoscopic video [160].
- The International Committee for Display Metrology (ICDM) published the Information Display Measurement Standard (IDMS1), a document for standard measurement procedures to quantify electronic display characteristics and qualities. It includes a section specifically on 3D and stereoscopic displays.

## 9 Conclusions

We discussed various quality issues of stereoscopy that need to be quantified and monitored. In many cases, this still requires determining the appropriate parameter ranges and acceptable thresholds for a comfortable viewing experience. Quality assurance (QA) is important in three different aspects:

- Technical issues, such as idiosyncrasies of the various display types. QA for these technical issues is generally done in the lab, when a technology is

evaluated. As technologies become more mature, we expect these issues to become less prevalent.

- Practical issues. These include all glitches, errors, mistakes, shortcuts, etc. that might happen when working with a complex system such as 3D video production and distribution. Here the role of QA is primarily to identify issues as they occur and alert operators accordingly. As viewers become more experienced with 3D content and its distribution, these issues will likely diminish as well.
- Intrinsic physical or physiological issues with a stereoscopic 3D presentation. While these cannot be overcome, they can be controlled and mitigated. The role of QA here is to minimize their impact on viewers.

Although there have been many advances in 3D production, 3D coding, transmission, and 3D display technologies, the quality of 3D content has mostly been investigated independently and separately for each stage. To achieve better 3D quality and comfort, the dependencies between content, display, and the viewer need to be considered in joint quality assessment and optimization. This is particularly true for objective 3D picture quality assessment. Such research can be conducted by leveraging multi-disciplinary expertise in image processing, video compression/transmission, 3D imaging/display, and vision science. Furthermore, individual user constraints, preferences, and perception characteristics must be considered, by giving users more control over some parts of 3D rendering, in order to make the 3D viewing experience as pleasant as possible.

## References

- [1] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008, invited paper.
- [2] L. M. J. Meesters, W. A. IJsselstein, and P. J. H. Seuntiëns, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 14, no. 3, March 2004.
- [3] S. Kim, J. Yoshitake, H. Morikawa, T. Kawai, O. Yamada, and A. Iguchi, "Psycho-physiological effects of visual artifacts by stereoscopic display systems," in *Proc. SPIE*, vol. 7863, San Jose, CA, Jan. 23–27, 2011.
- [4] A. G. Solimini, A. Mannocci, D. Di Thiene, and G. La Torre, "A survey of visually induced symptoms and associated factors in spectators of three dimensional stereoscopic movies," *BMC Public Health*, vol. 12, no. 779, 2012.
- [5] S. J. Daly, R. T. Held, and D. M. Hoffman, "Perceptual issues in stereoscopic signal processing," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 347–361, 2011.

- [6] M. S. Banks, J. C. A. Read, R. S. Allison, and S. J. Watt, "Stereoscopy and the human visual system," *SMPTE Motion Imaging J.*, vol. 121, no. 4, pp. 24–43, May 2012.
- [7] I. P. Howard and B. J. Rogers, *Perceiving in Depth*. Oxford University Press, 2012.
- [8] L. Zhang, S. Daly, O. Grau, N. Hur, and W. J. Tam, "Special issue on 3D-TV horizon: Contents, systems, and visual perception," *IEEE Trans. Broadcasting*, vol. 57, no. 2, June 2011.
- [9] S. Winkler and D. Min, "Stereoscopic image quality compendium," in *Proc. ICICS*, Singapore, Dec. 13–16, 2011.
- [10] R. Blake and R. Sekuler, *Perception*, 5<sup>th</sup> ed. McGraw-Hill, 2005.
- [11] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," in *Proc. SPIE*, vol. 7690, Orlando, FL, April 6–8, 2010.
- [12] J. Cutting and P. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Perception of Space and Motion*, W. Epstein and S. Rogers, Eds. Academic Press, 1995, pp. 69–117.
- [13] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet, "Evaluating depth perception of 3D stereoscopic videos," *IEEE J. Sel. Topics Sig. Proc.*, vol. 6, no. 6, pp. 710–720, Oct. 2012.
- [14] M. Mikkola, S. Jumisko-Pyykkö, D. Strohmeier, A. Boev, and A. Gotchev, "Stereoscopic depth cues outperform monocular ones on autostereoscopic display," *IEEE J. Sel. Topics Sig. Proc.*, vol. 6, no. 6, pp. 698–709, Oct. 2012.
- [15] R. Held, E. Cooper, and M. Banks, "Blur and disparity are complementary cues to depth," *Current Biology*, vol. 22, pp. 426–431, March 2012.
- [16] D. Hoffman and M. Banks, "Focus information is used to interpret binocular images," *J. Vision*, vol. 10, no. 5, 2010.
- [17] A. Vetro, "Frame-compatible formats for 3D video distribution," in *Proc. ICIP*, Hong Kong, Sept. 2010.
- [18] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New requirements of subjective video quality assessment methodologies for 3DTV," in *Proc. VPQM*, Scottsdale, AZ, Jan. 13–15, 2010.
- [19] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. ICIP*, San Antonio, TX, Oct. 2007.
- [20] P. Kauff *et al.*, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Sig. Proc.: Image Comm.*, vol. 22, no. 2, pp. 217–234, 2007.

- [21] K.-J. Yoon and I. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. PAMI*, vol. 28, no. 4, pp. 650–656, 2006.
- [22] D. Min and K. Sohn, "Cost aggregation and occlusion handling with WLS in stereo matching," *IEEE Trans. Image Processing*, vol. 17, no. 8, pp. 1431–1442, 2008.
- [23] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo using identical MRF parameters," in *Proc. ICCV*, Nice, France, Oct. 2003, pp. 900–906.
- [24] C. Fehn, "Depth image based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. SPIE*, vol. 5291, San Jose, CA, Jan. 2004, pp. 93–104.
- [25] K.-J. Oh, S. Yea, and Y.-S. Ho, "Hole-filling method using depth based in-painting for view synthesis in free viewpoint television (FTV) and 3D video," in *Proc. PCS*, Chicago, IL, May 2009.
- [26] S. Choi, B. Ham, and K. Sohn, "Space-time hole filling with random walks in view extrapolation for 3-D video," *IEEE Trans. Image Processing*, vol. 22, no. 6, pp. 2429–2441, 2013.
- [27] D. Min, D. Kim, S. Yun, and K. Sohn, "2D/3D freeview video generation for 3DTV system," *Sig. Proc.: Image Comm.*, vol. 24, no. 1-2, pp. 31–48, 2009.
- [28] J. Shade, S. J. Gortler, L.-W. He, and R. Szeliski, "Layered depth images," in *Proc. ACM SIGGRAPH*, 1998, pp. 231–242.
- [29] L. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 600–608, 2004.
- [30] S.-U. Yoon and Y.-S. Ho, "Multiple color and depth video coding using a hierarchical representation," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 17, no. 11, pp. 1450–1460, 2007.
- [31] K. Müller, A. Smolic, K. Dix, P. Kauff, and T. Wiegand, "Reliability-based generation and view synthesis in layered depth video," in *Proc. MMSP*, 2008, pp. 34–39.
- [32] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," *Int. J. Comp. Vis.*, vol. 35, no. 2, pp. 151–173, 1999.
- [33] D. Mahajan, F.-C. Huang, W. Matusik, R. Ramamoorthi, and P. N. Belhumeur, "Moving gradients: A path-based method for plausible image interpolation," *ACM Trans. Graphics*, vol. 28, no. 3, 2009.
- [34] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. H. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graphics*, vol. 29, no. 4, 2010.

- [35] W. Matusik and H. Pfister, “3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes,” *ACM Trans. Graphics*, vol. 23, no. 3, pp. 814–824, 2004.
- [36] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. R. Antúnez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” *ACM Trans. Graphics*, vol. 24, no. 3, pp. 765–776, 2005.
- [37] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. PAMI*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [38] J. Y. Bouguet, “Camera calibration toolbox for matlab,” [http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc).
- [39] T. Svoboda, D. Martinec, and T. Pajdla, “A convenient multicamera self-calibration for virtual environments,” *Presence*, vol. 14, no. 4, pp. 407–422, 2005.
- [40] Point Grey Research, <http://www.ptgrey.com/products/multisync/>.
- [41] G. Litos, X. Zabulis, and G. Triantafyllidis, “Synchronous image acquisition based on network synchronization,” in *Proc. Computer Vision and Pattern Recognition Workshop (CVPRW)*, 2006, p. 167.
- [42] D. Bradley, B. Atcheson, I. Ihrke, and W. Heidrich, “Synchronization and rolling shutter compensation for consumer video camera arrays,” in *Proc. PROCAMS*, 2009, pp. 1–8.
- [43] A. Elhayek, C. Stoll, K. I. Kim, H.-P. Seidel, and C. Theobalt, “Feature-based multi-video synchronization with subframe accuracy,” in *Proc. DAGM/OAGM Symp.*, 2012, pp. 266–275.
- [44] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” *ACM Trans. Graphics*, vol. 26, no. 3, 2007.
- [45] D. Min, J. Lu, and M. N. Do, “Depth video enhancement based on weighted mode filtering,” *IEEE Trans. Image Processing*, vol. 21, no. 3, 2012.
- [46] C. Zhang and Z. Zhang, “Calibration between depth and color sensors for commodity depth cameras,” in *Proc. ICME*, 2011, pp. 1–6.
- [47] D. Herrera C., J. Kannala, and J. Heikkila, “Joint depth and color camera calibration with distortion correction,” *IEEE Trans. PAMI*, vol. 34, no. 10, pp. 2058–2064, 2012.
- [48] I. V. Mikhelson, P. G. Lee, A. V. Sahakian, Y. Wu, and A. K. Katsaggelos, “Automatic, fast, online calibration between depth and color cameras,” *Vis. Comm. Image Repres.*, (in press).
- [49] A. Staranowicz and G. L. Mariottini, “A comparative study of calibration methods for Kinect-style cameras,” in *Proc. Int. Conf. on Pervasive Technologies Related to Assistive Environments (PETRA)*, 2012, p. 49.

- [50] L. Zhang, C. Vazquez, and S. Knorr, "3D-TV content creation: Automatic 2D-to-3D video conversion," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 372–383, 2011.
- [51] D. Kim, D. Min, and K. Sohn, "A stereoscopic video generation method using stereoscopic display characterization and motion analysis," *IEEE Trans. Broadcasting*, vol. 54, no. 2, pp. 188–197, 2008.
- [52] M. T. Pourazad, P. Nasiopoulos, and R. K. Ward, "Generating the depth map from the motion information of H.264-encoded 2D video sequence," *EURASIP J. Image Video Proc.*, vol. 2010, 2010, Article ID 108584.
- [53] S. Battiato, A. Capra, S. Curti, and M. L. Cascia, "3D stereoscopic image pairs by depth-map generation," in *Proc. IEEE 3DPVT*, Thessaloniki, Greece, Sept. 2004, pp. 124–131.
- [54] K. Moustakas, D. Tzovaras, and M. G. Strintzis, "Stereoscopic video generation based on efficient layered structure and motion estimation from a monoscopic image sequence," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 15, no. 8, pp. 1065–1073, 2005.
- [55] G. Zhang, W. Hua, X. Qin, T.-T. Wong, and H. Bao, "Stereoscopic video synthesis from a monocular video," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 4, 2007.
- [56] M. Guttman, L. Wolf, and D. Cohen-or, "Semi-automatic stereo extraction from video footage," in *Proc. ICCV*, Kyoto, Japan, Oct. 2009, pp. 136–142.
- [57] M. Liao, J. Gao, R. Yang, and M. Gong, "Video stereolization: Combining motion analysis with user interaction," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 7, pp. 1079–1088, July 2012.
- [58] ISO/IEC JTC1/SC29/WG11, "Call for proposals on 3D video coding technology," MPEG2011/N12036, 2011.
- [59] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, 2011.
- [60] P. Merkle, K. Müller, A. Smolic, and T. Wiegand, "Efficient compression of multi-view video exploiting inter-view dependencies based on H.264/MPEG4-AVC," in *Proc. ICME*, 2006, pp. 1717–1720.
- [61] J.-H. Hur, S. Cho, and Y.-L. Lee, "Adaptive local illumination change compensation method for H.264/AVC-based multiview video coding," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 17, no. 11, pp. 1496–1505, 2007.
- [62] J. H. Kim *et al.*, "New coding tools for illumination and focus mismatch compensation in multiview video coding," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 17, no. 11, pp. 1519–1535, 2007.
- [63] S. Yea and A. Vetro, "View synthesis prediction for multiview video coding," *Sig. Proc.: Image Comm.*, vol. 24, no. 1-2, pp. 89–100, 2009.

- [64] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P. H. N. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Sig. Proc.: Image Comm.*, vol. 24, no. 1-2, pp. 73–88, 2009.
- [65] M. Maitre and M. N. Do, "Joint encoding of the depth image based representation using shape-adaptive wavelets," in *Proc. ICIP*, San Diego, CA, Oct. 2008.
- [66] K.-J. Oh, A. Vetro, and Y.-S. Ho, "Depth coding using a boundary reconstruction filter for 3-D video systems," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 21, no. 3, pp. 350–359, 2011.
- [67] V.-A. Nguyen, D. Min, and M. N. Do, "Efficient techniques for depth video compression using weighted mode filtering," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 23, no. 2, pp. 189–202, Feb. 2013.
- [68] P. Benzie *et al.*, "A survey of 3DTV displays: Techniques and technologies," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 17, no. 11, pp. 1647–1658, 2007.
- [69] H. Jorke and M. Fritz, "Advanced stereo projection using interference filters," *J. Society for Information Display*, vol. 17, no. 5, pp. 407–410, May 2009.
- [70] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: A review and applications analysis," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 263–371, June 2011.
- [71] K. M. Schreiber, J. M. Hillis, H. R. Filippini, C. M. Schor, and M. S. Banks, "The surface of the empirical horopter," *J. Vision*, vol. 8, no. 3, pp. 1–20, March 2008.
- [72] ANSI/SCTE 187-3, "Informative guidance for stereoscopic video," American National Standards Institute and Society of Cable Telecommunications Engineers, 2012.
- [73] Y. Nojiri, H. Yamanoue, S. Ide, S. Yano, and F. Okano, "Parallax distribution and visual comfort on stereoscopic HDTV," in *Proc. IBC*, Amsterdam, 2006.
- [74] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vision*, vol. 8, no. 3, March 2008.
- [75] K. Ukai and P. Howarth, "Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations," *Displays*, vol. 29, no. 2, pp. 106–116, Feb. 2008.
- [76] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone," in *Proc. SPIE*, vol. 7863, San Jose, CA, Jan. 23–27, 2011.
- [77] J. L. Semmlow and D. Heerema, "The role of accommodative convergence at the limits of fusional vergence," *Inv. Ophthalmol. Vis. Sci.*, vol. 18, no. 9, pp. 970–976, Sept. 1979.

- [78] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vision*, vol. 11, no. 8, July 2011.
- [79] 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, no. 4, pp. 191–201, April 2002.
- [80] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, no. 9, pp. 141–150, Sept. 2004.
- [81] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual comfort," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 335–346, June 2011.
- [82] C. Schor and C. Tyler, "Spatio-temporal properties of Panum's fusional area," *Vision Res.*, vol. 21, no. 5, pp. 683–692, 1981.
- [83] S. J. Koppal *et al.*, "A viewer-centric editor for 3D movies," *IEEE Computer Graphics and Applications*, vol. 31, no. 1, pp. 20–35, 2011.
- [84] D. Kim and K. Sohn, "Visual fatigue prediction for stereoscopic image," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 21, no. 2, pp. 231–236, Feb. 2011.
- [85] C. Fehn and R. Pastoor, "Interactive 3-D TV – concepts and key technologies," *Proc. IEEE*, vol. 94, no. 3, pp. 524–538, 2006.
- [86] D. K. Broberg, "Guidance for horizontal image translation (HIT) of high definition stereoscopic video production," in *Proc. SPIE*, vol. 7863, San Jose, CA, Jan. 23–27, 2011.
- [87] F. Devernay, S. Duchene, and A. Ramos-Peon, "Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis," in *Proc. SPIE*, vol. 7863, 2011.
- [88] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 589–601, 2011.
- [89] T. Basha, Y. Moses, and S. Avidan, "Stereo seam carving a geometrically consistent approach," *IEEE Trans. PAMI*, vol. 35, no. 10, pp. 2513–2525, Oct. 2013.
- [90] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, July 2007.
- [91] S.-W. Jung and S.-J. Ko, "Depth sensation enhancement using the just noticeable depth difference," *IEEE Trans. Image Processing*, vol. 21, no. 8, pp. 3624–3637, 2012.
- [92] V. De Silva, A. Fernando, S. Worrall, H. K. Arachchi, and A. M. Kondoz, "Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-D displays," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 498–506, 2011.



- [93] C. Thébaud, D. Doyen, P. Routhier, and T. Borel, “Automatic depth grading tool to successfully adapt stereoscopic 3D content to digital cinema and home viewing environments,” in *Proc. SPIE*, vol. 8648, San Jose, CA, 2013.
- [94] H. Ujike and H. Watanabe, “Effects of stereoscopic presentation on visually induced motion sickness,” in *Proc. SPIE*, vol. 7863, San Jose, CA, Jan. 23–27, 2011.
- [95] S.-N. Yang, T. Schlieski, B. Selmins, S. Cooper, R. Doherty, P. J. Corriveau, and J. E. Sheedy, “Stereoscopic viewing and reported perceived immersion and symptoms,” *Optometry & Vision Science*, vol. 89, no. 7, pp. 1068–1080, July 2012.
- [96] D. A. Wismeijer, C. J. Erkelens, R. van Ee, and M. Wexler, “Depth cue combination in spontaneous eye movements,” *J. Vision*, vol. 10, no. 6, June 2010.
- [97] A. Saxena, S. H. Chung, and A. Y. Ng, “3-D depth reconstruction from a single still image,” *Int. J. Comp. Vis.*, vol. 76, no. 1, pp. 53–69, Jan. 2008.
- [98] J. Schild and M. Masuch, “Designing stereoscopic information visualization for 3D-TV: What can we learn from S3D gaming?” in *Proc. SPIE*, vol. 8288, San Jose, CA, Jan. 2012.
- [99] M. Lambooi, M. J. Murdoch, W. A. IJsselstein, and I. Heynderickx, “The impact of video characteristics and subtitles on visual comfort of 3D TV,” *Displays*, vol. 34, no. 1, Jan. 2013.
- [100] <http://vision.middlebury.edu/stereo>.
- [101] Z. Arican, S. Yea, A. Sullivan, and A. Vetro, “Intermediate view generation for perceived depth adjustment of stereo video,” in *Proc. SPIE*, vol. 7443, San Diego, CA, Aug. 2009.
- [102] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “View synthesis for advanced 3D video systems,” *EURASIP J. Image Video Proc.*, vol. 2008, 2008, Article ID 438148.
- [103] C. Hewage, S. Worrall, S. Dogan, S. Villette, and A. Kondoz, “Quality evaluation of color plus depth map-based stereoscopic video,” *IEEE J. Sel. Topics Sig. Proc.*, vol. 3, no. 2, pp. 304–318, 2009.
- [104] V. De Silva, W. Fernando, G. Nur, E. Ekmekcioglu, and S. Worrall, “3D video assessment with just noticeable difference in depth evaluation,” in *Proc. ICIP*, Hong Kong, Sept. 2010.
- [105] H. T. Nguyen and M. N. Do, “Error analysis for image-based rendering with depth information,” *IEEE Trans. Image Processing*, vol. 18, no. 4, pp. 703–716, April 2009.
- [106] C. Doutre and P. Nasiopoulos, “Color correction preprocessing for multiview video coding,” *IEEE Trans. Circ. Syst. Video Tech.*, vol. 19, no. 9, pp. 1400–1405, 2009.

- [107] F. Shao, G.-Y. Jiang, M. Yu, and Y.-S. Ho, “Fast color correction for multi-view video by modeling spatio-temporal variation,” *Vis. Comm. Image Repres.*, vol. 21, no. 5-6, pp. 392–403, 2010.
- [108] A. Vetro *et al.*, “Overview of multiview video coding and anti-aliasing for 3D displays,” in *Proc. ICIP*, San Antonio, TX, Sept. 2007.
- [109] M. Zwicker, W. Matusik, F. Durand, and H. Pfister, “Antialiasing for automultiscopic 3D displays,” in *Proc. Eurographics Rendering Symposium*, Nicosia, Cyprus, June 2006.
- [110] C. Moller and A. Travis, “Correcting interperspective aliasing in autostereoscopic displays,” *IEEE Trans. Vis. Comput. Graph.*, vol. 11, no. 2, pp. 228–236, 2005.
- [111] J. Konrad and P. Agniel, “Subsampling models and anti-alias filters for 3-D automultiscopic displays,” *IEEE Trans. Image Processing*, vol. 15, no. 1, pp. 128–140, Jan. 2006.
- [112] W.-J. Kim, S.-D. Kim, and J. Kim, “Analysis on the spectrum of a stereoscopic 3-D image and disparity-adaptive anti-aliasing filter,” *IEEE Trans. Circ. Syst. Video Tech.*, vol. 19, no. 10, pp. 1561–1565, Oct. 2009.
- [113] W.-J. Kim, S.-D. Kim, N. Hur, and J. Kim, “Temporal anti-aliasing of a stereoscopic 3D video,” *ETRI Journal*, vol. 31, no. 1, pp. 1–9, Feb. 2009.
- [114] P. Seuntiëns, L. Meesters, and W. IJsselsteijn, “Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation,” *ACM Trans. Applied Perception*, vol. 3, no. 2, pp. 95–109, April 2006.
- [115] G. Saygili, C. Gurler, and A. Tekalp, “Quality assessment of asymmetric stereo video coding,” in *Proc. ICIP*, Hong Kong, Sept. 2010.
- [116] P. Aflaki, M. Hannuksela, J. Hakkinen, P. Lindroos, and M. Gabbouj, “Subjective study on compressed asymmetric stereoscopic video,” in *Proc. ICIP*, Hong Kong, Sept. 2010.
- [117] G. Saygili, C. G. Gürler, and A. M. Tekalp, “Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming,” *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 593–601, 2011.
- [118] X. Zhang and B. Li, “Impact of disparate video quality in stereo channels on 3D perception,” in *Proc. VPQM*, 2010.
- [119] A. J. Woods, T. Docherty, and R. Koch, “Image distortions in stereoscopic video systems,” in *Proc. SPIE*, vol. 1915, San Jose, CA, Feb. 1993.
- [120] L. Goldmann, F. De Simone, and T. Ebrahimi, “Impact of acquisition distortion on the quality of stereoscopic images,” in *Proc. VPQM*, Scottsdale, AZ, Jan. 13-15, 2010.

- [121] F. L. Kooi and A. Toet, “Visual comfort of binocular and 3D displays,” *Displays*, vol. 25, no. 2, pp. 99–108, Aug. 2004.
- [122] D. C. Burr and J. Ross, “How does binocular delay give information about depth?” *Vision Res.*, vol. 19, pp. 523–532, 1980.
- [123] L. Goldmann, J.-S. Lee, and T. Ebrahimi, “Temporal synchronization in stereoscopic video: Influence on quality of experience and automatic asynchrony detection,” in *Proc. ICIP*, Hong Kong, Sept. 26–29, 2010.
- [124] R. T. Held and M. S. Banks, “Misperceptions in stereoscopic displays: A vision science perspective,” in *Proc. APGV*, New York, NY, Aug. 2008.
- [125] M. Barkowsky, P. Campisi, P. L. Callet, and V. Rizzo, “Crosstalk measurement and mitigation for autostereoscopic displays,” in *Proc. SPIE*, vol. 7526, Jan. 2010.
- [126] P. J. H. Seuntjens, L. M. J. Meesters, and W. A. IJsselstein, “Perceptual attributes of crosstalk in 3D images,” *Displays*, vol. 26, no. 4–5, pp. 177–183, 2005.
- [127] A. J. Woods, “Crosstalk in stereoscopic displays: A review,” *J. Electronic Imaging*, vol. 21, no. 4, pp. 040 902–1–040 902–21, 2012.
- [128] A. J. Woods and C. R. Harris, “Comparing levels of crosstalk with red/cyan, blue/yellow, and green/magenta anaglyph 3D glasses,” in *Proc. SPIE*, vol. 7524, San Jose, CA, Jan. 2010.
- [129] M. Salmimaa and T. Jarvenpaa, “3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays,” *J. Soc. Info. Display*, vol. 16, no. 10, pp. 1033–1040, 2008.
- [130] Q. Wang, X. Li, L. Zhou, A. Wang, and D. Li, “Cross-talk reduction by correcting the subpixel position in a multiview autostereoscopic three-dimensional display based on a lenticular sheet,” *Applied Optics*, vol. 50, no. 7, pp. B1–B5, 2011.
- [131] Y. G. Lee and J. B. Ra, “New image multiplexing scheme for compensating lens mismatch and viewing zone shifts in three-dimensional lenticular displays,” *Opt. Eng.*, vol. 48, no. 4, 2009.
- [132] H. Lim *et al.*, “A simultaneous intermediate view interpolation and multiplexing algorithm for a fast lenticular display,” *Opt. Eng.*, vol. 46, no. 11, pp. 114 003 1–8, 2007.
- [133] J. Konrad, B. Lacotte, and E. Dubois, “Cancellation of image crosstalk in time-sequential displays of stereoscopic video,” *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 897–908, 2000.
- [134] F. A. Smit, R. van Liere, and B. Fröhlich, “Three extensions to subtractive crosstalk reduction,” in *Proc. IPT-EGVE Symp.*, Weimar, Germany, July 2007.

- [135] W. Tan, S. Z. Zhou, and S. C. Read, “Methods and systems for reducing or eliminating perceived ghosting in displayed stereoscopic images,” U.S. Patent No. US20110080401.
- [136] D. J. McKnight, “Enhanced ghost compensation for stereoscopic imagery,” U.S. Patent No. US20100040280.
- [137] G. A. Slavenburg, T. F. Fox, and D. R. Cook, “System, method, and computer program product for compensating for crosstalk during the display of stereo content,” U.S. Patent No. US20080036696.
- [138] I. Tsirlin, L. M. Wilcox, and R. S. Allison, “The effect of crosstalk on the perceived depth from disparity and monocular occlusions,” *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 445–453, 2011.
- [139] L. Xing, T. Ebrahimi, and A. Perkis, “Subjective evaluation of stereoscopic crosstalk perception,” in *Proc. SPIE VCIP*, vol. 7744, Huang Shan, China, July 2010.
- [140] L. Xing, J. You, T. Ebrahimi, and A. Perkis, “A perceptual quality metric for stereoscopic crosstalk perception,” in *Proc. ICIP*, Hong Kong, Sept. 2010.
- [141] A. Jain and J. Konrad, “Crosstalk in automultiscopic 3-D displays: Blessing in disguise?” in *Proc. SPIE*, vol. 6490, San Jose, CA, Jan. 2007.
- [142] L. B. Stelmach and W. J. Tam, “Stereo-anomalous vision in a sample of young adults,” in *Proc. SPIE*, vol. 2657, San Jose, CA, Jan. 28–Feb. 2, 1996.
- [143] C. Akai, “Depth perception in real and virtual environments an exploration of individual differences,” Master’s thesis, University of British Columbia, 1998.
- [144] A. Bruce, D. Atchison, and H. Bhoola, “Accommodation-convergence relationships and age,” *Inv. Ophthalmol. Vis. Sci.*, vol. 36, no. 2, pp. 406–413, Feb. 1995.
- [145] J. F. Norman *et al.*, “Stereopsis and aging,” *Vision Res.*, vol. 48, no. 23-24, pp. 2456–2465, 2008.
- [146] J.-H. Lee and J.-K. Song, “Individual variation in 3D visual fatigue caused by stereoscopic images,” *IEEE Trans. Circ. Syst. Video Tech.*, vol. 58, no. 2, pp. 500–504, May 2012.
- [147] N. A. Dodgson, “Variation and extrema of human inter-pupillary distance,” in *Proc. SPIE*, vol. 5291, San Jose, CA, Jan. 2004.
- [148] P. A. Howarth, “Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: a review,” *Ophthalmic and Physiological Optics*, vol. 31, no. 2, pp. 111–122, March 2011.
- [149] J. F. Norman, T. Dawson, and A. Butler, “The effects of age upon the perception of depth and 3-D shape from differential motion and binocular disparity,” *Perception*, vol. 29, no. 11, pp. 1335–1359, Nov. 2000.

- [150] J. F. Norman *et al.*, “Aging and the depth of binocular rivalry suppression,” *Psychology and Aging*, vol. 22, no. 3, pp. 625–631, 2007.
- [151] M. Lambooij and W. IJsselsteijn, “Visual discomfort and visual fatigue of stereoscopic displays: A review,” *J. Imaging Sci. Tech.*, vol. 53, no. 3, pp. 1–14, May 2009.
- [152] T. Bando, A. Iijima, and S. Yano, “Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review,” *Displays*, vol. 33, pp. 76–83, 2012.
- [153] P. Neveu, M. Philippe, A. E. Priot, P. Fuchs, and C. Roumes, “Vergence tracking: A tool to assess oculomotor performance in stereoscopic displays,” *J. Eye Movement Res.*, vol. 5, no. 2, pp. 1–8, 2012.
- [154] P. Seuntiëns, I. Heynderickx, and W. IJsselsteijn, “Viewing experience and naturalness of 3D images,” in *Proc. SPIE*, vol. 6016, Boston, MA, Oct. 2005.
- [155] W. IJsselsteijn, “Presence in depth,” Ph.D. dissertation, Eindhoven University of Technology, Netherlands, 2004.
- [156] J. Hakala, “The added value of stereoscopy in still images,” Master’s thesis, Aalto University, Finland, 2010.
- [157] D. Strohmeier, S. Jumisko-Pyykko, K. Kunze, and M. O. Bici, “The extended-OPQ method for user-centered quality of experience evaluation: A study for mobile 3D video broadcasting over DVB-H,” *EURASIP J. Image Video Proc.*, no. 538294, 2011.
- [158] Q. Huynh-Thu, M. Barkowsky, and P. Le Callet, “Video quality assessment: From 2D to 3D – challenges and future trends,” in *Proc. ICIP*, Hong Kong, Sept. 26–29, 2010.
- [159] ITU-R Recommendation BT.2021, “Subjective methods for the assessment of stereoscopic 3DTV systems,” International Telecommunication Union, Geneva, Switzerland, 2012.
- [160] S. McCarthy, “Glossary for video & perceptual quality of stereoscopic video,” 3D@Home Consortium and MPEG Industry Forum, Aug. 2010.