Abstract—Stereoscopic viewing of 3D content brings with it a variety of complex perceptual quality issues. For the percept of depth to be convincing, consistent, and comfortable, many parameters throughout the imaging pipeline must to be matched correctly. However, tradeoffs are inevitable for practical reasons, which may lead to various distortions. This paper reviews the issues from an engineering perspective, addressing in particular depth issues, multiview issues, display issues, viewer-specific issues, as well as possible measurements for the overall 3D quality of experience (QoE).

Index Terms—Perception, Quality of Experience (QoE), Human Visual System (HVS), 3D Television (3DTV)

I. INTRODUCTION

3D viewing is on everyone’s mind, including the TV industry’s, where it is currently making headlines. An overview of the technology is provided in [1]. Quality issues for images and 2D video have been studied quite extensively [2], and commercial quality assurance (QA) tools are already being deployed to monitor video quality in real time.

Stereoscopy adds another layer of complexity on top of the common 2D quality issues from video compression, network impairments, etc. [3]. Furthermore, stereoscopic content may even have potential physical effects: if 3D is not produced, processed and presented correctly, it can make viewers dizzy or nauseous. For example, Samsung has issued a health and safety information for its 3D displays, according to which possible side effects include altered vision, lightheadedness, dizziness, involuntary movements such as eye or muscle twitching, confusion, nausea, loss of awareness, convulsions, cramps, disorientation.

This underlines that 3D viewing comes with more severe concerns than 2D. Therefore, one of the primary practical goals must be to minimize or prevent possible discomfort caused by 3D content. Issues with stereoscopic viewing can be roughly classified as depth, multiview, subtitling, display, and user-specific issues, which are discussed in more detail in the following sections.

II. STEREOSCOPIC VIEWING BASICS

A. Depth Cues

3D is all about the perception of depth. There are actually a large number of depth cues that the human visual system uses when viewing a 3D scene [4]. 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 and binocular cues.

Oculomotor cues include accommodation and vergence. Accommodation 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 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, and shadows; motion parallax.

The most important binocular visual cue is the retinal disparity between points of the same object viewed from slightly different angles by the eyes. This effect is used in stereoscopic systems such as 3DTV.

B. Stereoscopy

The basics of stereoscopy can be briefly summarized as follows (see Figure 1). A point of a 3D object is projected onto a screen in two locations, representing the views of that 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 difference in apparent position of the object viewed along the two lines of sight, which is called parallax.

III. DEPTH ISSUES

A. Vergence-Accommodation Conflict

In normal 3D viewing conditions in the real world, both eyes would focus on the 3D object (using lens accommodation), and at the same time the eyes would converge on the object in question. In stereoscopic viewing, there is no real object; therefore, the eyes must focus on the objects projection on the screen, which is at a different distance from the virtual 3D object (vergence distance, see also Figure 1). This conflict between accommodation and vergence is one of the main reasons for discomfort [5]. It has been quantified by determining a comfort zone around normal vergence-accommodation matching conditions.

The depth range that can be comfortably presented to a viewer depends on the viewing distance (for example, the depth range is much larger in the cinema than for a TV at home). As a result, 3D content has to be adapted to each screen (because screen size largely determines viewing distance).
B. Depth Bracket

The depth bracket is the range of depth from closest to furthest object inside a scene. Because of the vergence-accommodation conflict described earlier, the depth bracket should not be too large.

C. Temporal Depth Discontinuities

Temporal depth discontinuities occur when the depth or depth distribution of a scene changes. Rapid depth variations can result in viewer discomfort, because the human visual system (HVS) is unable to follow the changes and to reconstruct depth properly. This is a common problem at transitions (e.g. scene cuts). In general, depth changes should happen slower and less frequently than in 2D.

An example of a tool for viewer-centric editing in 3D movie productions was introduced in [6]. One of the important functions in this editing tool is to blend the depth bracket during scene changes, so that the objects of interest maintain the same depth. The smooth transition of depth distributions at scene changes can mitigate the effect of temporal depth discontinuities.

D. Interaxial Distance

Typically, the two cameras for recording 3D content are situated at an inter-axial distance that is roughly equal to the distance between the average person’s eyes (inter-ocular distance, about 65 mm). If the inter-axial distance deviates from this, it can create unwanted effects, such as making close objects appear unnaturally large or too small [7]. Another issue is that not all people have the same inter-ocular distance. For example, children have a smaller inter-ocular distance, which is why 3D viewing can be more stressful for them.

E. Exaggerated Parallax

A parallax (i.e. the disparity of an objects projection on the screen) greater than the inter-ocular distance would force the eyes to diverge and place the object beyond infinity, which is impossible in nature and should be avoided.

F. Depth Mismatch

Stereoscopy alone recreates only one of the many visual cues the HVS uses to determine the 3D structure of a scene. As highlighted in Section II-A, there are many other depth cues. In the real world, all these cues match. In a stereoscopic 3D projection, they may be violated, and the mismatches can reduce or even destroy the 3D percept.

One example of mismatching cues 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. Because of the large number of depth cues and their complexity, depth mismatch can be particularly challenging to detect using automatic QA systems.

G. Depth Quality

One of the possible formats for 3D video storage and transmission is a color-plus-depth representation, where a monoscopic color image and a corresponding depth map are used to represent 3D video. Two (or more) slightly different views are generated by warping the color image with the depth information and addressing the disocclusion problem [8], [9]. Depth maps, whether they are estimated by stereo matching methods [10], structured light, or active depth cameras, are prone to errors due to the ill-posed problem of depth estimation and inherent physical limits of the sensor such as noise or interference. They may also be impaired during coding and transmission [11], [12]. The quality of the rendered 3D images is highly dependent on the accuracy of the estimated depth map, so that the effects of depth estimation errors and coding/transmission artifacts should be investigated [13]–[15]. An upper bound of the allowable depth error may provide a new insight into depth sensor development or coding/transmission system design [16].

IV. MULTIVIEW ISSUES

A. View Differences

Unwanted mismatches between corresponding left and right views may arise at various stages of the production and distribution chain, for example if any of the following are not identical [17]: Camera optics and sensors; White balance; Shutter speed; Aperture; Gamma; Geometry (camera angle and position; picture skew or cropping). Most of these mismatches can be corrected through careful calibration or during post-production. Compression can also lead to view differences, such as: Different artifact severity (blockiness, blur); Different time-varying quality (if multiple views are compressed separately), e.g. different GOP structure/length; Network impairments, error propagation (especially when the two views are contained in separate streams). Views may also get out of sync; a difference of just a few frames can be very annoying. Finally, view reversal may occur when the left view is presented to the right eye and vice versa.

If any of these differences become too severe, the HVS may be unable to fuse the two images into a consistent 3D percept and instead alternate between the two views. This is
also known as binocular rivalry. On the other hand, artifacts in one view but not the other may actually be masked (hidden) by the HVS, which is called binocular suppression. Both effects must be taken into account for accurate measurement of the perceptual impact of these differences.

The second effect has been exploited in asymmetric stereo video coding [18], where the two views are encoded at different quality (e.g., through spatial scaling or quantization). 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 [19]–[21]. 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 among spatial scaling, quality scaling, or their combination? Content- and display dependence also need to be investigated.

B. Monocular Occlusion

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. This is a particular problem for 3D content that has been poorly reconstructed from 2D content.

C. Aliasing

Aliasing happens due to the rendering of 3D content with high-frequency components on 3D displays. We can distinguish intra-perspective aliasing within each view due to the discrete 2D pixel of each view, and inter-perspective aliasing due to the discrete number of views [22], [23].

Various techniques have been proposed to alleviate 3D aliasing problems. Moller et al. [24] presented a spatially varying filter to reduce inter-perspective aliasing by leveraging the knowledge of per-pixel scene depth based on display bandwidth analysis. Konrad et al. [25] studied inter-perspective aliasing by analyzing a multiplexing process from a sampling perspective and then derived a filter to prevent the aliasing caused by nonorthogonal grid pattern of 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 prefocusing techniques [23]. This approach addresses aliasing within each view as well as inter-perspective aliasing [22]. Kim et al. [26] 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 scene is band-limited on 3D displays with the disparity-adaptive low-pass filtering for enhancing viewing comfort. This approach was extended into temporal aspects by considering disparity and motion together in 3D video [27]. Although the above-mentioned anti-aliasing filters reduce the aliasing artifacts, some viewers may prefer the aliased 3D video, which are generally sharper [23]. Further subjective evaluation is needed to determine the right balance between aliasing and blur [23].

V. Subtitles

Subtitling in 3D is a surprisingly complex process [28]. Subtitles, captions, logos, etc. pose a problem because they are often added onto existing content. Therefore, even if the underlying 3D content was produced and edited correctly, the addition of subtitles may introduce further issues.

A. Depth Conflicts

A major problem with subtitles or any other inserts in 3D is the danger of them appearing behind other objects in the scene, which can affect the 3D percept. At the same time, their depth should not be too different from the scenes depth bracket. Multiple inserts or subtitles in a scene can pose additional problems if they appear at different depths. Also, subtitles may be inserted into only one view by mistake.

Finally, subtitles can also cause unnatural depth perception when internal parameters of the stereo camera change at the production, whereas the subtitles remain fixed, for instance, when the camera focal length changes (zoom-in or out).

B. Geometric Misalignment

Subtitles are usually inserted into 3D video with a parallel configuration for simplicity. However, two (or more) views may be slightly mismatched geometrically due to some mistakes at the production. This geometric misalignment between subtitles and 3D content may cause discomfort.

VI. Display Issues

A. Crosstalk

Crosstalk happens when part of one view also appears in another [29]–[31]. 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 described in two different ways [32]:

- System crosstalk is defined as the leaking image from the other view (content-independent crosstalk).
- Viewer crosstalk is defined as the ratio of the luminance of the unwanted ghost image and the actual image (content-dependent crosstalk).

All stereoscopic 3D displays suffer from crosstalk [33]. In anaglyph displays, crosstalk may occur when the color filters of the glasses do not separate spectral components completely or do not match with the spectral emission of the display [34]. When active shutter glasses are used, the timing must be precisely synchronized with the display, otherwise crosstalk may occur. The rise and fall times of the display and the glasses are also an important parameter. For polarized viewing with passive glasses, crosstalk can occur when viewers tilt their head or lie down. Autostereoscopic displays are prone to crosstalk around the view boundaries due to their incomplete multiplexing [35].

A number of experiments demonstrate that ghosting from crosstalk causes discomfort and 3D quality degradation [29], [31], [36]. Recently, quality metrics for crosstalk have also been proposed [37]. While current display technologies do
not achieve perfect separation of the multiple views, some compensation is possible by digitally pre-processing the views before display [50], [58], [40]. In order to mitigate the effect of crosstalk (ghosting), the amount of intensity leakage from the unintended view is first estimated by modeling the inter-view dependency on the 3D display, and multiple views are then pre-distorted to compensate for the distortion from crosstalk. This cancellation procedure can be done in the illumination domain [30] or in 3D color space [40].

The inclined angle of slanted lenticular displays maps multiple views into the lenticular sheet with sub-pixel precision. However, the boundaries of subpixels cannot be covered exactly due to inherent physical limitation of lens elements, which causes crosstalk. Wang et al. [41] 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. [42], [43] exploited the geometric relationship between LCD subpixels and lenticular sheets with pattern images and then estimated a mapping matrix 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 by using stereo images and the corresponding depth maps [44].

B. 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. 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.

VII. USER ISSUES

Each viewer has different optimal viewing conditions due to individual differences in depth perception. This is affected by a combination of several factors such as age, gender, and degree of previous 3D viewing experience. These user issues must be taken into account for accurate evaluation of 3D quality.

A number of reports reveal that there is a difference in 3D depth perception over age groups [45], [46]. For instance, older people are less sensitive to perceiving depth and surface curvature [47]. Depth perception with age may vary according to the characteristics of 3D content such as disparity magnitude, disparity direction (crossed vs. uncrossed disparity), orientation difference of corresponding lines on each view [46]. Binocular rivalry and suppression are also dependent on viewers’ age [48].

Another age-dependent factor is the inter-ocular distance; for children, it is relatively smaller than for adults. It may hence cause discomfort when children watch 3D content and possibly affect the development of vision. Recently, Samsung and LG Electronics have warned that children, pregnant women, and elderly people should refrain from watching 3DTV in order to prevent potential risks that can be triggered by stereoscopic content. These potential risks should be evaluated carefully for all 3D applications such as 3DTV, 3D movies, 3D gaming, etc. [49].

Gender may be also important factor in perceiving 3D depth. It has been generally known that there is a difference between men and women in terms of visual perception abilities. The degree of previous 3D viewing experience may serve as an important factor as well.

VIII. 3D QoE Metrics

In addition to the individual quality issues and parameters identified above, it is useful to define metrics that quantify the combined impact of several such parameters or the overall viewing experience of a 3D presentation. The following QoE metrics for 3D have been proposed:

- **Discomfort.** Certain 3D content (e.g. extreme ranges in depth or disparity) may cause discomfort (eye strain, headache, fatigue) to viewers [50].
- **Naturalness.** The ease with which viewers can fuse L/R views into a natural-looking 3D percept with smooth depth representation [51]. A 3D scene that looks natural also enhances the viewers sense of presence [52], especially in interactive applications.
- **Value-add.** The perceived benefit (or detriment) of viewing a specific piece of content in 3D over viewing the same content in 2D [53].
- **3D Mean Opinion Score (MOS)** for overall 3D content quality. A separate 2D MOS could still be reported for each view.

IX. 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 is important in three different aspects:

1. 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.
2. 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 users become more experienced with 3D content and its distribution, these issues will likely diminish as well.
3. 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.
REFERENCES


