A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV

Lydia M. J. Meesters, Wijnand A. IJsselsteijn, and Pieter J. H. Seuntiëns

***Abstract*—A high-quality three-dimensional (3-D) broadcast service (3-D TV) is becoming increasingly feasible based on various recent technological developments combined with an enhanced understanding of 3-D perception and human factors issues surrounding 3-D TV. In this paper, 3-D technology and perceptually relevant issues, in particular 3-D image quality and visual comfort, in relation to 3-D TV systems are reviewed. The focus is on near-term displays for broadcast-style single- and multiple-viewer systems. We discuss how an image quality model for conventional two-dimensional images needs to be modified to be suitable for image quality research for 3-D TV. In this respect, studies are reviewed that have focused on the relationship between subjective attributes of 3-D image quality and physical system parameters that induce them (e.g., parameter choices in image acquisition, compression, and display). In particular, artifacts that may arise in 3-D TV systems are addressed, such as keystone distortion, depth-plane curvature, puppet theater effect, cross talk, cardboard effect, shear distortion, picket-fence effect, and image flipping. In conclusion, we summarize the perceptual requirements for 3-D TV that can be extracted from the literature and address issues that require further investigation in order for 3-D TV to be a success.**

***Index Terms*—Image quality, measurement paradigms, stereoscopic artifacts, three-dimensional television (3-D TV), visual comfort.**

I. INTRODUCTION

THREE-DIMENSIONAL television (3-D TV) could provide a dramatic enhancement in the television viewing experience comparable to the transition between black-and-white and color TV. Although the idea of stereoscopic television was already demonstrated in the 1920s by John Baird, it was not until the 1980s that experimental 3-D TV was presented to a large audience in Europe [1].

Despite the audience's enthusiasm, poor image quality and unwanted side effects of current 3-D TV systems as well as a lack of consensus on broadcast standards have hampered the successful introduction of 3-D TV to the mass consumer market. We strongly believe that the introduction of 3-D TV can only be a lasting success if the perceived image quality and the viewing comfort are at least comparable to conventional television. In addition, 3-D TV technology should be compatible with conventional television to ensure a gradual transition from one system to the other [2], [3]. This is becoming increasingly feasible because of recent technological advances in image processing, camera and display development, as well as an improved understanding of 3-D perception through human factor studies.

In Section II, we give a brief overview of the currently applied technology in 3-D image content generation [stereoscopic image acquisition, depth cameras and two-dimensional (2-D) to 3-D image conversion tools], 3-D compression methods, and near-term broadcast-style stereoscopic as well as autostereoscopic displays suited for single and multiple viewers. Next, we discuss perceptually relevant issues in relation to 3-D TV systems, with a particular emphasis on perceived image quality and visual comfort. With regard to this, we review in Section III measurement paradigms used to identify and measure perceived attributes underlying image quality in 3-D TV [e.g., focus groups or psychophysical scaling methods such as proposed by the International Telecommunications Union (ITU)]. In Section IV, we discuss how an image quality model for conventional 2-D images can be modified to suit image quality research for 3-D TV. Such an image quality model is intended as a tool to describe the relationship between physical 3-D TV system parameters and the most dominant perceptual factors contributing to 3-D image quality. Related to this model, we discuss in Section V typical 3-D TV artifacts (keystone distortion, depth-plane curvature, puppet theater effect, cross talk, cardboard effect, shear distortion, picket-fence effect and image flipping), and their effect on image quality and/or visual comfort. Finally, in Section VI, we summarize the perceptual requirements for 3-D TV that can be extracted from the literature and address the issues that require further investigation in order for 3-D TV to become a mature home-viewing system.

II. 3-D TV BROADCAST SYSTEM

A 3-D TV broadcast service comprises of the entire chain starting from content production and coding schemes for efficient transmission to adequate displays presenting a high-quality 3-D picture in the home (see Fig. 1). In this section, we give an overview of the most relevant technology to realize each of these components in a 3-D TV system.

Section II-A provides a description of different techniques to generate 3-D content including stereo cameras, depth cameras, and 2D-to-3D conversion tools. 3-D image compression is discussed in Section II-B. Stereoscopic and autostereoscopic displays for single or multiple viewers are covered in Section II-C.

Manuscript received January 10, 2003; revised October 2, 2003. This work was supported by the European IST-2001-34396 Project ATTEST. An earlier version of this paper was presented at the IS&T/SPIE Conference on Stereoscopic Displays and Virtual Reality Systems X 2003, January 21–24, 2003.

The authors are with the Human-Technology Interaction (HTI) Group, Eindhoven University of Technology, 5600MB Eindhoven, The Netherlands (e-mail: l.m.j.meesters@tue.nl; w.a.ijsselsteijn@tue.nl; p.j.h.seuntiens@tue.nl).

Digital Object Identifier 10.1109/TCSVT.2004.823398

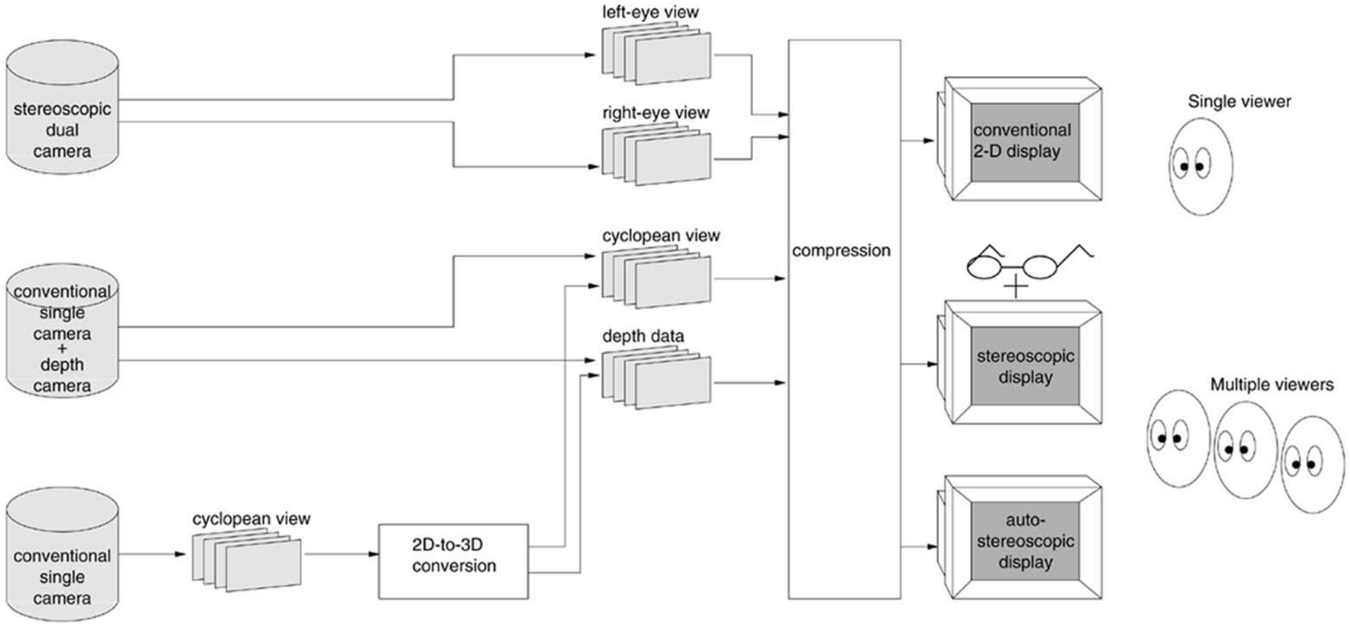

Fig. 1. Schematic representation of the complete 3-D TV broadcast chain including various options for content production and displays. Note that each of the displays can be viewed by either a single viewer or multiple viewers depending on the particular stereo output technologies used.

A. 3-D Content Generation

The majority of 3-D broadcast material available today has been produced using a dual-camera configuration giving a stereo pair where the left-eye and the right-eye views are separately recorded from slightly different perspectives. Filming parameters such as camera base distance (distance between the two cameras), convergence distance (distance from the cameras to the point where both optical axis intersect), and camera lens focal length can be used to scale the horizontal disparity and thus the degree of perceived depth. Two stereo camera configurations can be distinguished: 1) the parallel camera configuration and 2) the toed-in camera configuration, also called converging cameras. The geometry of such stereo camera configurations in relation to the display and the viewer is explained in Woods *et al.* [4]. These authors recommend a parallel camera configuration to avoid geometrical distortions like the keystone distortion and depth plane curvature. Stereoscopic film production requires that both director and camera operator are highly skilled regarding stereoscopic geometry and camera calibration [5]. Furthermore, the possibilities of adjusting the image pairs in postproduction are limited and time consuming. In terms of compatibility with a conventional 2-D broadcast system a double bandwidth is needed if the video streams would be transmitted uncompressed.

An approach to overcome some of these limitations is to use 3-D cameras that register one conventional RGB view and an accompanying depth map, containing the corresponding depth of each image point. This format allows for easier shooting and simplifies postproduction as only one RGB image needs to be manipulated. Examples of such systems that capture depth information in real time include the *AXI-Vision* camera developed by *NHK*, and the *Zcam* produced by *3DV Systems*.

The latest version of the *AXI-Vision* camera can simultaneously capture an HDTV color picture and a corresponding depth image. The depth image is recorded in real time and calculated from two measurements of reflected light intensities. The first measurement is obtained when an object is illuminated with continuously increasing light intensity and the second measurement when the intensity of the illumination is decreasing. A depth resolution of 1.7 cm is achieved when the distance from the camera to the object is 2 m and 5 cm for a distance of 10 m [6], [7].

The *Zcam* can be used as a camera add-on to an existing 2-D broadcast camera to record a depth map. The principle of the *Zcam* is that a succession of short light pulses (10 ns) transmitted in the direction of a real-world 3-D scene reflect back to the camera when the light collides with objects. Hence, for each pixel the measured depth corresponds inversely with the measured amount of energy. The *Zcam* operates in a range from 0.5 to 7 m with a depth resolution less than 0.5 cm for a distance of 1 m [8]–[10]. One of the main advantages is that a depth camera is easy to handle by producers since the depth range adapts automatically with varying shooting conditions, e.g., a transition from a close-up to a distance shot. At this point, depth cameras are mainly used in depth-keying to separate the foreground objects from the background. Their application in 3-D TV is a promising development although still in its infancy. One of the challenges of RGB-depth annotated video is to recover the left- and the right-eye views by image-based rendering (IBR) techniques [11]. Since two slightly different views are reconstructed from a single recorded view and one set of depth information, the main drawback is that occluded image regions cannot be retrieved correctly [3]. View synthesis techniques are needed that adequately handle image regions becoming visible in the reconstructed views. Fehn *et al.* [12] proposed to extrapolate background information where image information is missing for intermediate view reconstruction to support head-motion parallax on a 2-D display. At present, the perceptual effects of reconstructing occluded image regions are being investigated for 3-D TV applications. We are involved in the European ATTEST project and evaluated the perceived image quality of 3-D image sequences reconstructed from an RGB-depth annotated video format. The preliminary results are promising and indicate a good subjective image quality [13].

Three-dimensional content is sparse. Moreover, existing stereoscopic content (captured with a stereo camera) was mainly intended for viewing on a large screen. Hence, when this material would be viewed on a small 3-D television the parallax between the left- and right-eye images reduces such that the depth impression is lost [14]. Furthermore, the entertainment industry needs to make large investments in new 3-D cameras or add-on depth modules to compile the massive amount of stereoscopic video needed for broadcast purposes. It is probably not realistic that this will be accomplished in a short time frame. Therefore, 2-D-to-3-D conversion tools that transform 2-D video into 3-D video would increase the amount of video material suited for stereoscopic viewing considerably. Furthermore, existing 2-D movies can be viewed as a novel stereoscopic film. In principle, 2-D-to-3-D conversion algorithms derive a depth map from a 2-D still image or video sequence. Only a limited amount of the monoscopic video can be converted automatically into 3-D video if depth estimation techniques such as: 1) depth from motion and 2) structure from motion are used [15]. Dynamic Digital Depth Research Pty Ltd developed a semi-automatic algorithm to recover a depth map from a monoscopic video [15]. Within the European ATTEST project, effort has been spent in developing 2-D-to-3-D conversion algorithms that require minimal manual intervention [11].

B. 3-D Compression

Most 3-D compression schemes are developed for stereoscopic pictures or video that consist of two views taken from a slightly different perspective of a 3-D scene. The principle of compression schemes utilizing disparity estimation is to exploit the high correlation between the image information in both views by predicting one view (target) from the other view (reference). Hence, the reference view is in general coded with a traditional 2-D compression method whereas the target view can be represented by disparity vectors. Disparity estimation in coding schemes is analogous to the correspondence problem in theories of stereopsis, namely to cross correlate corresponding image points between the left– and right-eye views. Two basic categories of stereoscopic coding schemes using disparity estimation can be distinguished: 1) intensity-based methods and 2) feature-based methods. The former determine the disparity on the basis of corresponding image intensities while the latter methods use image features such as edges or objects [16]–[19]. Intensity-based methods and, in particular, block-based disparity estimation approaches are mostly used. The principle of block-based disparity estimation is to divide one of the views (target) into nonoverlapping blocks with a fixed size of $N \times N$ pixels (e.g., 8×8 pixels). For each $N \times N$ pixel block, the most similar block in the other view (reference) is determined by means of an error criterion (e.g., mean-squared error) that expresses the best corresponding match. A block-based coding approach that can handle geometric deformations, for instance, perspective distortions due to the camera configuration, is suggested by Seferidis and Papadimitriou [20]. The disparity-compensated transform-domain predictive coding (DCTDP) is a block-based method that takes into account the possible intensity differences between the left and right views which could impede the disparity estimation process [21]. One of the drawbacks of compression schemes using block-based disparity estimation is that occluded regions cannot be represented by the reference image and disparity vectors alone. No true matching blocks can be identified for areas (e.g., objects or image borders) in the target image that are occluded in the reference image. This obstacle can be overcome when the residual image (the difference between the reconstructed target image and the reference image) is coded and transmitted [22]. Instead of a fixed block-size, other effective 3-D compression methods are for instance based on quadtree decomposition and use an adaptive block-size to handle, amongst others, occluded image regions [20], [23], [24].

More recently, coding schemes evolved for 3-D picture formats containing dense depth images (resulting from, e.g., *AXI-Vision* camera or *Zcam*). Coding schemes are proposed that apply traditional 2-D compression techniques such as quantization, discrete pulse code modulation (DPCM), and motion compensation [25]. Krishnamurthy *et al.* [26] use region-of-interest (ROI) depth coding. These authors showed that visible artifacts in reconstructed views depend on the accuracy of the depth values at sharp discontinuities in the depth map. Hence, important regions can be detected by an edge operator and coded at a higher bit rate than less important image regions. Grau *et al.* [27] proposes to record metadata, e.g., the camera focus setting, which can be used in the coding process to assign higher bit rates to the objects in focus (e.g., persons).

Another stereoscopic coding approach is based on theories of binocular suppression. The assumption that the final percept is dominated by the high-quality component of a stereo pair is exploited to achieve compression. Thus, when one view of the stereo image is of high image quality, the other view can be degraded without introducing visible distortions in the binocular percept [21], [28]. However, the impact of coding distortions on the perceived stereoscopic image quality of such an asymmetric image pair depends on the visual appearance of the distortion, where blockiness appears to be much more disturbing than blur. The image quality of stereo images with a different degree of *blur* in the left- and the right-eye images seems to be dominated by the high-quality component. However, if both components are degraded by *blockiness*, the perceived stereoscopic image quality seems to be an average of the image quality of the left and right-eye views [29]. For instance, the perceived image quality of asymmetrically coded MPEG-2 sequences with a spatial resolution of 720×576 pixels remains good as long as one of the two views is not compressed at a lower bit rate than 3 Mb/s [30].

For 3-D multiview broadcast applications, compression schemes are proposed to exploit the inter-view redundancy. A major reduction of the data can be obtained by compressing the original multiviews into a smaller number of key views with sufficient disparity information. At the receiver side, intermediate views can be reconstructed from this sparse number of key views and the disparity information. The image quality of the reconstructed views depends amongst other things on whether occluded areas can be recovered from the coded key views and the disparity values. Tseng and Anastassiou [31] tackled the occlusion problem by coding the extreme left and right images as well as a third image representing all objects that are visible in intermediate frames but not in the extreme left and right images. In combination with dense disparity maps, intermediate views can be reconstructed. Siegel *et al.* [32] proposed a method to select key views with minimal occlusion and an interpolation method to generate intermediate views from a sparse disparity map. Different techniques to obtain a sparse depth or density map and interpolation methods to reconstruct an intermediate view are described in [33] and [34].

C. Stereoscopic and Autostereoscopic Displays

Since the introduction of Wheatstone's stereoscope in 1833, the concept of displaying two separate dissimilar images to each eye brought forth many techniques to realize a stereoscopic display. The distinguishing features characterizing stereoscopic displays are: 1) the method applied to separate the left- and right-eye views; 2) whether look-around capabilities are implemented (multiview); and 3) the number of viewers that can watch a stereoscopic sequence simultaneously. Usually, a distinction is made between stereoscopic and autostereoscopic displays. In the former case, the viewer wears an optical device to direct the left and right images to the appropriate eye (aided viewing) while in the latter case the technique to separate both views is integrated in the display (free viewing). Below we give an overview of different types of stereoscopic and autostereoscopic display techniques. Extensive reviews of these display techniques are given by Sexton and Surman [35] and Pastoor and Wöpking [36].

Stereoscopic displays with aided viewing (e.g., polarized glasses) are widely used and can be time-parallel or time-sequential. In time-parallel displays, the left- and right-eye views are displayed simultaneously on one or two screens depending on the image separation technique. Techniques used to direct the distinct views to the appropriate eye in a time-parallel display system are: 1) location multiplexing; 2) anaglyph or color-multiplexing; and 3) polarization multiplexing.

Location multiplexing is one of the oldest techniques and redirects the separately generated left and right views to the appropriate eye through separate channels. The Wheatstone and Brewster stereoscopes are early examples of this method. Stereoscopes were popular in Europe and the United States, and, from the mid-18th century to the 1930s, millions of stereoscopes were produced [37].

Anaglyph or color-multiplexing is also one of the primitive image separation techniques. The left and right views are filtered with near-complementary colors and viewed through respective color-filter glasses to direct both views to the appropriate eye. The technique is cheap and has been applied in experimental broadcast sessions, for instance, in Europe (two 45-min popular scientific series were broadcast in 1982 in several European countries), and the United States (the 3-D movie "Miss Sadie Thompson" broadcast in 1980 by the Los Angeles pay channel SelectTV). However, the color rivalry effect and poor image quality restrict its practical use in broadcast applications [1].

Polarization multiplexing uses polarized light for image separation. The hardware configuration may consist of two monitors or projectors covered with linear or circular polarizing filters and is viewed with polarized glasses to maintain separate left and right eye views.

The most frequently used stereo displays are based on the principle of time-sequential presentation, displaying the left- and right-eye views of a stereo image in rapid alternation. The stereo pairs are viewed using synchronized active shuttering glasses which open alternately for the appropriate eye while closing the other eye's view. This system exploits the human visual systems characteristic of integrating a stereo pair across a time-lag of up to 50 ms [36].

A different implementation of a time-sequential display is commercialized by, amongst others, StereoGraphics Corporation [38]. Their *Z-Screen* displays left- and right-eye views sequentially with opposite polarization and is viewed with passive polarized glasses with corresponding polarization directions. The main advantages of such time-sequential displays viewed with passive glasses are that they allow the use of lightweight, inexpensive polarized glasses and that less crosstalk is perceptible in comparison to active shutter glasses [36].

Autostereoscopic displays supporting free viewing are probably best suited for an application such as 3-D TV. In a home environment, the need for glasses can be unpractical and limits the viewers' freedom of movement. The main categories of autostereoscopic displays that can be distinguished are: 1) direction-multiplexed; 2) volumetric; and 3) holographic displays.

Direction-multiplexed displays can differ in the optical effect (e.g., diffraction, refraction, reflection, and occlusion) used to send the left- and right-eye views directly to the appropriate naked eye. Historically, parallax barriers and lenticular systems are primarily used stereoscopic techniques and examples of the occlusion and refraction approaches, respectively. An autostereoscopic display generates for each eye an image at a plane parallel to the display. If the observers' eyes are located within these two images or viewing "windows," the 3-D image is perceived correctly. In practice, only a limited horizontal viewing angle is supported such that the viewers' pupils stay in the corresponding viewing "window." A broader horizontal viewing angle can be achieved by direction-multiplexed multiview displays, where a number of discrete views are presented across the viewing area. Another option is to employ user-position tracking to adjust a set of steerable exit pupils [39].

Though volumetric and holographic displays could potentially realize the most optimal 3-D free-viewing experience, they suffer from various drawbacks that preclude them from being practically feasible as a 3-D TV application at this point in time.

Volumetric displays tend not to rely upon flat displays, but generate an image that occupies a limited volume in space. A 3-D image is divided into a very large number of planes (which is why these displays are also referred to as *multiplanar*) and image points of these planes can be plotted in rapid succession onto a volumetric medium. An example of such a system is the Texas Instruments "Omniview" system, which uses three laser beams (red, green, and blue) projected onto a rotating helix mirror. Commercial volumetric displays are the DepthCube from LightSpace Technologies [40], [41] and Perspecta developed by Actuality Systems [42], [43]. An important drawback in volumetric displays is that solid objects tend to look transparent and the size of the display volume is usually small.

Electroholographic displays have sometimes been presented as the ideal free-viewing 3-D technique and have received considerable attention over the past years from research groups in Japan, the United States, and the United Kingdom. For most practical purposes, an important drawback is the fact that coherent light is required during image recording, which means that holograms cannot be recorded with natural (incoherent) lighting. Also, the amount of data contained in a hologram is enormous, requiring specific data compression techniques. At present, only small displays are feasible for video-based holography [36]. For instance, at MIT's Media Lab, a holovideo display is developed that reproduces an image volume 150-mm wide, 75-mm high, and 150-mm deep [44], [45]. Technical implementations of these types of display are discussed comprehensively by Sexton and Surman [35] and Pastoor and Wöpking [36].

The acceptance and commercial success of any advanced technology aimed at the consumer market depends critically on the user's experiences with and responses toward the system. When evaluating 3-D technologies for a practical next-generation 3-D TV broadcast service, the human factors' perspective is of essential value. The authors believe that a viable 3-D TV system for home viewing needs to be autostereoscopic, enabling multiple viewers to freely move around while still maintaining adequate stereoscopic picture quality. Given individual differences and variable preferences between viewers, a level of control over the amount of depth is preferable. This is feasible using the novel 3-D representation format of RGB+depth. Moreover, to support the transition between traditional 2-D TV and 3-D TV, the 3-D TV system should be backward-compatible and enough interesting 3-D content should be available, motivating people to invest in it.

III. EVALUATION PARADIGMS

Subjective assessment methods for perceptual evaluation of monoscopic and stereoscopic television pictures are becoming widely accepted. Standards for subjective testing, defining adequate assessment methods and perceptual factors, such as recommendations of the ITU, are a necessity to compare competitive systems and monitor future developments. For instance, standardized measurement methods to quantify perceptual attributes such as perceived image quality, depth, and sharpness enable engineers to optimize and enhance their systems. Explorative studies, e.g., to explore viewers' unprimed reactions to conventional television or a novel 3-D television, are conducted to a lesser degree. In the next sections, we describe how explorative studies can be performed and how they can contribute to a better evaluation of new media like 3-D TV. Furthermore, a brief discussion of the ITU recommendation 500–10 psychophysical scaling paradigms is given, and we discuss the need for measurements in a home environment.

A. Explorative Studies

Evaluation criteria used to test conventional media like 2-D TV are not necessarily applicable to new media. New perceptual constructs could be needed to fully explore the benefits and drawbacks of a new technology. Therefore, explorative studies are better suited to get a deeper understanding of the added value and underlying perceptual constructs of a novel technology such as 3-D TV.

By means of explorative studies, no direct subjective ratings are acquired but instead the unprimed attitudes, feelings, and reactions toward 3-D TV are explored. An example are focus groups, extensively used in market research, where naive subjects participate in small discussion groups to express and share their experiences of the viewed stereoscopic image system. The experiment leader moderates the discussion and gives guidance, if necessary. Other methods such as the think-aloud procedure or co-discovery method, adopted from usability studies, are also suited to gather unprimed viewers' reactions to new technologies.

Freeman and Avons [46] used focus groups to explore viewers' reactions to conventional 2-D TV (a standard 20-in television) and novel 3-D TV (a 20-in stereoscopic time-parallel display with polarized light for image separation). The results showed that viewers report, with respect to stereoscopic sequences, a sense of "being there" before this concept was raised by the moderator. Furthermore, this feeling of "being there" was related to attributes such as realism, naturalness, and involvement. A second aim of the focus group was to identify program types suited for 3-D TV. In general, subjects preferred action movies and live events such as sports, theater, and concerts. Program types such as news, soap operas, documentaries, and talk shows were thought of as inappropriate for 3-D TV. Moreover, subjects indicated that they would like to decide on a program-by-program basis whether they wanted to watch it in 3-D or 2-D. However, these attitudes may be subject to change as the technology advances and people become more familiar with the 3-D experience.

Explorative studies can also be used to arrive at a better understanding of the attributes underlying a multidimensional concept like image quality, naturalness, or presence. Bech *et al.* [47] proposed the rapid perceptual image description (RaPID) method. In this approach, the descriptive analysis method is used and it is assumed that observers can describe the attributes underlying a multidimensional perceptual construct like image quality by means of standard words. As the authors pointed out, the attributes are defined by the standard observers and are not necessarily equal to those used by engineers or experts in the field. Moreover, the words used to describe an attribute may be specific to each language.

In summary, explorative studies can be used to: 1) collect unprimed viewers's descriptions of the sensations evoked by

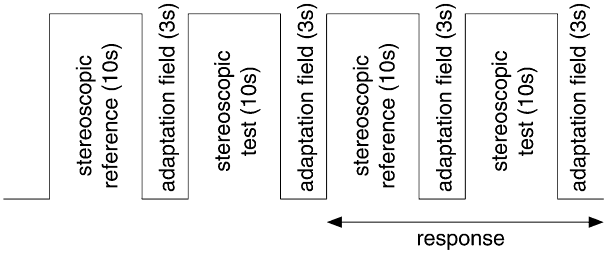

Fig. 2. ITU-R BT.500–10 recommendation: DSCQS method. Alternately, an unimpaired stereo image (reference) and an impaired stereo image (test) are shown. The reference and test image are presented twice. During the second presentation, the subjects are asked to judge the perceived image quality of the reference and the test image separately.

a stereoscopic image system; 2) investigate the added value of new image systems, e.g., 3-D TV, without imposing predefined appreciation criteria such as image quality; and 3) determine the attributes that may underlie concepts such as image quality, naturalness, and presence, without directed questions.

B. Psychophysical Scaling Paradigms

Psychophysical scaling paradigms can be used to measure and quantify perceptual concepts such as sharpness or image quality. The scaling concepts and perceptual attributes are known a priori and the stimulus set is manipulated such that variations in the sensations of the particular attribute are measurable. Two types of psychophysical assessment methods can be distinguished: 1) performance-oriented methods and 2) appreciation-oriented methods [48]. The former is applicable whenever the purpose of the application is to facilitate a certain task, for instance, a detection task. Appreciation-oriented assessment methods are applicable in appreciation-oriented applications, such as stereoscopic television, where the goal is to generate images that are as "pleasing" as possible. The emphasis is on visual comfort associated with the images. For instance, it is strenuous to watch a television program containing excessive binocular disparities. Watching programs that induce diplopia (i.e., double images) requires a great deal of effort and viewers experience this as unpleasant.

Appreciation-oriented subjective assessment of stereoscopic television pictures is described in the ITU-R BT.1438 recommendation [49]. The subjective assessment methods are adopted from the ITU-R BT.500–10 recommendation for conventional monoscopic television [50]. The proposed assessment methods are used to measure overall perceived impairment or image quality of degraded still images and image sequences. The same experimental paradigms can be applied to obtain ratings of the perceived strengths or sensation of attributes such as sharpness, depth, eye-strain, naturalness, or presence. In general, three different approaches are proposed: the double-stimulus-continuous-quality-scale method (DSCQS), single-stimulus methods, and stimulus-comparison methods.

In DSCQS, observers assess the overall image quality for a series of stereoscopic images presented separately in time (see Fig. 2). Alternately, an unimpaired stereo image (reference) and an impaired stereo image (test) are shown. For both stereo images (reference and test), observers assess the overall picture quality separately. Eventually the DSCQS assessment results are the difference in scores between the reference and test images. In single-stimulus scaling, the overall picture quality of each stereoscopic image in the stimulus set is assessed individually. In stimulus-comparison scaling, again a series of stereoscopic images are presented separately in time. In this procedure, observers assign a relation between two successive stereoscopic images. In order to restrict the number of observations, often just a sample of all possible combinations of stereo images in a stimulus set is used. The same single-stimulus and stimulus-comparison methods can be used to assess impairment. In the double-stimulus-impairment scale (DSIS) method, again a series of stereoscopic images are presented in time (alternately a stereoscopic reference and test image). However, the assessors are asked to judge only the test image, "keeping in mind the reference" [50].

TABLE I
ITU-R BT.500–10 RECOMMENDATION RATING SCALES

single stimulus quality scale		DSIS and single stimulus impairment scale	
5	excellent	5	imperceptible
4	good	4	perceptible but not annoying
3	fair	3	slightly annoying
2	poor	2	annoying
1	bad	1	very annoying

comparison scale	
-3	much worse
-2	worse
-1	slightly worse
0	the same
1	slightly better
2	better
3	much better

The scaling methods impose different grading scales to assess the perceived image quality. In DSCQS, a continuous graphical scale is used to avoid forcing subjects to answer within too coarse a category. The scale is often labeled with verbal terms such as *excellent*, *good*, *fair*, *poor*, and *bad* to guide the observer. For single-stimulus, stimulus-comparison, and DSIS, the rating scales usually applied such as verbal or numerical categories are given in Table I. The subjects express the perceived image quality, the impairment, or the relation between two stereoscopic images by placing the presented stimuli in one of these categories.

The average observer's quality judgements can be obtained by a number of different analysis methods. Methods such as averaging the judgements across observers and defining a confidence interval indicating the individual differences are specified by the ITU. More complex judgment models were proposed by Torgerson [51]. An application of such a model is described in [52].

The subjective assessment methods described above are used to obtain a single judgement of the overall image quality of still pictures or short video sequences of 10 s. An alternative assessment method, single-stimulus continuous-quality evaluation (SSCQE), was proposed to obtain continuous time-varying quality judgements of longer stereoscopic video sequences. Subjects continuously assess the picture quality by moving a hand-held slider. A subject indicates excellent image quality when the slider is positioned at the top of the grading scale, and bad image quality is indicated by moving the slider to the bottom. SSCQE is used to assess video that contains scene-dependent and time-varying impairment, for example, introduced by compression. Furthermore, television is usually watched for longer periods, so SSCQE is the most appropriate way to mimic home viewing conditions. This method of assessment has been applied in the context of 3-D TV by IJsselsteijn *et al.* [53] to continuously assess observers' sense of presence, depth, and naturalness, and more recently by Yano *et al.* [54]

C. The Need for Measurements in the Home

The previously described evaluation methods, and especially psychophysical scaling, are designed for subjective testing in a controlled testing environment where viewers are asked to carry out predefined tasks. Given the measurements and experiences acquired from such experiments, user requirements can be specified for 3-D TV. However, validation of some requirements and certainly the impact of a novel service such as 3-D TV on viewing behavior as well as the interaction among behavior, technology, and content can only be realistically addressed through studies performed in a home-environment situation.

For instance, conventional television, video cassette recorders (VCRs), and most recently interactive television (ITV) revealed that the type of medium and its technological implementation can lead to changes in the viewing behavior of people. The small screen size and fuzzy images inherent to the television sets of the 1950s were the predominant technological factors that imposed people to sit close to the television screen. This changed in the 1970s when affordable large television screens were introduced that allowed viewers to sit farther back from the screen. Another change in viewing behavior was observed when households were equipped with multiple television sets and television viewing became part of everyday life. Over the years, watching television shifted from a group activity, with focused attention to the broadcast program, toward a more individual and background activity. Nowadays people watch television while performing all kinds of activities such as cooking, dressing, or talking on the telephone [55]. The question is how 3-D TV will fit into existing viewing patterns or how viewing behavior will evolve. A number of studies showed that stereoscopic video adds "a feeling of being there" to the viewing experience. Judging from the content preference people expressed for 3-D TV, it appears that people may be aiming for a more intense TV experience rather then a background activity. Clearly, we need data from sustained field trials to be able to judge the impact of the introduction of 3-D TV on viewing behavior in the home.

IV. 3-D IMAGE QUALITY

Perceived 3-D image quality is one of the evaluation criteria to assess the performance of an image system such as 3-D TV. Hence, the 3-D TV systems' parameters can be tuned and optimized to the customers quality preference. However, subjective testing is time-consuming and needs to be repeated for each new parameter setting. Therefore, image quality models defining the relationship between physical system parameters and the perceived 3-D image quality could contribute to a lower cost 3-D TV design cycle. Moreover, competing systems can be compared by standardized assessment factors. For conventional television pictures, image quality models have been proposed to predict 2-D picture quality. Nevertheless, before the 3-D TV image quality can be modeled, a deeper understanding is needed of the relationship between the physical system parameters and the perceptual factors contributing to the perceived 3-D image quality. In Section IV-A, we discuss how the principles of a 2-D image-quality modeling approach can be used to gain insight into the relationship between 3-D TV parameters and 3-D image quality.

A. Image-Quality Model

The 2-D picture quality affected by imaging system parameters is considered to be a multidimensional attribute. The first step to model 2-D image quality is to define its most dominant perceptual factors and thus to obtain descriptions of these factors, for instance, blur, noise, or blockiness. Focus groups or the RaPID method are adequate assessment methods to define such subjective attributes (see Section III-A). Thereupon, psychophysical scaling methods are used to quantify the strengths of these attributes. Perceptual rules (e.g., Minkowski summation, which allows different weighting of separate attributes) are used to combine the measured attribute strengths into a prediction of the overall perceived image quality [56].

Engeldrum [57] applied this approach in the Image Quality Circle to relate the physical system parameters of a 2-D imaging system to the customer's quality preference.

Of course, 2-D picture quality models are not adequate to measure 3-D image quality since depth reproduction, the most important factor in a 3-D system, and typical stereoscopic distortions (for instance crosstalk) are not incorporated. We believe that a 3-D image quality model is required that is multidimensional, incorporating perceptual factors related to reproduced depth, 3-D image impairments, and visual comfort.

The added value of depth needs to be defined explicitly in a 3-D image quality model, especially when 2-D picture quality is used as reference. The positive contribution of depth to the perceived image quality was demonstrated for stereoscopic uncompressed and blur-degraded images [58], [59]. Also, Schertz [60] demonstrated that human observers seem to prefer DCT-coded stereoscopic images over the monoscopic originals, even though the perceived impairment was rated as perceptible and slightly annoying. Other research has shown, however, that, when observers were asked to rate the perceived image quality of images compressed with MPEG-2 and JPEG, the image quality results are mainly determined by the introduced impairments and not so much by depth [61], [62]. Seuntiëns *et al.* [62] also showed that JPEG distortions do not affect the depth appreciation as such, which also indicates that the added value of depth is not incorporated in the overall quality judgement for DCT-impaired images. One of the reasons for this might be the stimulus set or the scaling paradigm used in the reported experiments. For instance, when viewers judge unimpaired stereoscopic images against their monoscopic counterparts, the added value of depth is clearly recognized [58]. However, when image coding impairments are introduced, in particular DCT-coding distortions, observers may anchor their judgements on the most salient variations, e.g., blockiness level, in the stimulus set.

Increasing perceived depth implies increasing screen parallax, which can also negatively affect stereoscopic impairments such as crosstalk. For instance, the same level of crosstalk is more visible and more annoying in images with an increasing degree of depth [63], [13].

Studies giving evidence for a positive contribution of depth to the overall perceived image quality and those showing a negative effect of depth on the visibility of particular stereoscopic impairments indicate that perceptual summation rules (e.g., Minkowski summation) as applied in 2-D image-quality models are not sufficient as perceptual combination rules in a 3-D quality model.

Also, the experienced visual discomfort may degrade the perceived image quality. For instance, excessive disparities can cause eye strain and therefore degrade the perceived image quality [58].

Thus, to relate the technical parameters of a 3-D TV service to the customer's quality preference, perceptual factors such as depth reproduction, stereoscopic impairments, and visual comfort should be taken into account. Furthermore, perceptual rules to integrate the separate factors into an overall 3-D quality prediction need to consider the positive and negative contribution of depth.

In the next section, we discuss typical stereoscopic artifacts that degrade the perceived image quality.

V. STEREOSCOPIC IMPAIRMENTS

In the literature, several typical stereoscopic distortions have been described. These impairments would fit in an overall 3-D preference model, however, their relative perceptual importance will need to be determined by studying their mutual interactions in various combinations. Typical stereoscopic impairments induced by the camera configurations, compression, or (auto)stereoscopic displays include: keystone distortion, depth-plane curvature, puppet theater effect, crosstalk, cardboard effect, shear distortion, picket-fence effect, and image flipping. Below we discuss the possible causes of these impairments and their effects on perceived image quality and visual comfort.

A. Keystone Distortion

Woods *et al.* [4] reviewed several stereoscopic distortions including keystone distortion and depth-plane curvature. These are typically introduced in a stereoscopic image due to a converging camera configuration where the left and right cameras are positioned at an angle toward each other. In this case, the imaging sensors of the two cameras are directed toward slightly different image planes. This results in a trapezoidal picture shape in opposite directions for the left- and right-eye camera recordings. In a stereoscopic image, these oppositely oriented trapezoidal picture shapes may induce incorrect vertical and horizontal parallax. This incorrectly introduced vertical parallax is the source of keystone distortion. The keystone distortion is most noticeable in the image corners and increases with increasing camera base distance, decreasing convergence distance, and decreasing lens focal length. Incorrectly introduced horizontal parallax is the source of depth-plane curvature, whereby objects at the corner of the image appear further away from the observer compared to objects in the middle of the image. Perceptually, keystone distortion and depth-plane curvature may have a negative effect on appreciation-oriented assessments [58]. Furthermore, Woods *et al.* [4] reported that subjects experienced eye-strain at higher vertical parallax values. More recently, Stelmach *et al.* [64] demonstrated that a moderate vertical parallax, introduced with camera convergence distances in the range from 60 to 240 cm, hardly affects visual comfort ratings. Thus, although perceptual studies showed that eye strain is affected by keystone distortion, its effect is minimal, and vertical parallax seems to not be the most dominant factor attributed to visual discomfort.

B. Puppet Theater Effect

The puppet theater effect is an annoying miniaturization effect, making people look like animated puppets [65], [66]. Visual size distortions result if the angular retinal size of a displayed object and its perceived distance do not covary as in real-world conditions. In the real world, a change in angular size is corresponding to a change in distance. Yamanoue [66] showed that orthostereoscopic parallel shooting and display conditions (i.e., simulating human viewing angles, magnification, and convergence in the most natural way possible) do not cause the puppet theater effect. Hopf [67] describes a display technique which reduces the puppet theater effect. An autostereoscopic display with collimation optics enables a large volume of depth and thus allows a larger image to be presented at a greater distance behind the screen so that the puppet theater effect is not perceptible. Novel display techniques seem promising to avoid or reduce the puppet theater effect which contributes to an unnatural appearance of a 3-D image.

C. Crosstalk

Crosstalk or image ghosting in stereoscopic displays are caused primarily by: 1) phosphor persistence of a CRT-display and 2) imperfect image separation techniques by which the left-eye view leaks through to the right-eye view and vice versa. Crosstalk is perceived as ghost, shadow, or double contours and even a relative small amount of crosstalk can lead to headaches [65]. Crosstalk due to phosphor persistence may occur in time-sequential displays when the left and right views are displayed alternately and the image intensity leaks into the subsequent view. Thus, the right eye also receives a small proportion of the left-eye view. Additionally, for linear polarization techniques, an incorrect head position of an observer (e.g., tilted head) also causes annoying image ghosting. This is avoided in circular polarization techniques. Pastoor [65] demonstrated that the annoyance of crosstalk increases with increasing contrast and disparity values. The author suggests that the crosstalk of a display should not cross a threshold of 0.3%. Woods and Tan [68] introduced a crosstalk model which incorporates phosphor afterglow and shutter leakage. Moreover, an algorithm to compensate for crosstalk in stereoscopic images was proposed by Konrad *et al.* [69]. The authors showed that suppression of crosstalk enhances the visual comfort. The main problem of crosstalk is that it depends on the system (display and/or glasses) and thus an algorithm to compensate for it needs to be modified for each system. Autostereoscopic displays also suffer from crosstalk.

This is mainly due to the latency in the directional lenses to support motion parallax. Crosstalk is probably one of the main perceptual factors contributing to image quality and visual comfort. Moreover, the visibility of crosstalk increases with increasing screen parallax and thus limits the degree of depth that can be introduced while maintaining a high 3-D image quality.

D. Cardboard Effect

A typical stereoscopic distortion affecting perceived depth is the cardboard effect. The cardboard effect results in an unnatural depth percept whereby the objects appear flat as if the scene is divided into discrete depth planes. The effect can be compared to the scenery in a theater [60]. A cardboard effect can be caused by image acquisition parameters (e.g., lens focal length, camera base distance, and convergence distance) or compression parameters resulting in a coarse quantization of disparity or depth values [70], [60]. An explanation for this phenomenon is that the perceived size and perceived depth of an object do not correspond because an observer underestimates the objects' distance d on a stereoscopic display. The perceived depth obtained from disparity (inversely proportional to the square of the distance $1/d^2$) is too small compared with its perceived size (inversely proportional to the distance $1/d$) and as a result objects appear flattened [71]. The cardboard effect can appear even when the film parameters are set such that size and disparity of an object are reproduced correctly. To avoid or reduce the cardboard effect, camera parameters need to be tuned to one another such that the thickness of objects can be perceived [70]. In case the cardboard effect is due to the quantization of disparity values, it can manifest itself also as torn-up objects such that a coherent object is represented in several depth planes and perceived as a disjointed object. Temporal discontinuous-depth mismatches can occur if objects or parts of an object are assigned to different depth layers in time, which results in a flickering depth percept. Schertz [60] showed that scenes with torn objects were judged as more annoying than if the objects are reunited by reducing the depth resolution. Therefore, choosing a suitable number of depth layers is a subtle issue, where higher resolution does not always imply better quality. Apparently, in 3-D compression schemes, a tradeoff needs to be found between the needed accuracy of depth values to maintain high 3-D image quality and the gained compression ratio.

E. Shear Distortion

Shear distortion is typically experienced with stereoscopic displays that allow only one correct viewing position [4], [63]. For most stereoscopic displays, a stereoscopic image can only be viewed correctly from one particular viewpoint. If the observer changes his viewing position, the image seems to follow the observer and therefore appears perspectively distorted. Objects with uncrossed disparity move in the same direction of the observer and objects with crossed disparity seem to move in the opposite direction. This is experienced as unnatural since in real life an object remains stationary while the observer is allowed to move his head and look at it from slightly different positions. The shear distortion can be avoided by headtracking where the correct stereoscopic image is presented according to the viewpoint of the observer. For instance, in an autostereoscopic display, steerable exit pupils can be used to present either a single stereo pair correctly when viewed from different positions or a correct unique stereoscopic view onto the 3-D scene for different viewing positions. The latter approach requires a large data set containing multiple stereo image pairs of the same scene. A more efficient approach is based on IBR techniques, and the appropriate view can be reconstructed from an RGB-depth annotated 3-D image format. However, a drawback of this approach are the occlusion problems, as discussed in Section II

F. Picket-Fence Effect and Image Flipping

Typical multiview autostereoscopic display artifacts are the picket-fence effect and image flipping. Both artifacts are perceived if the observers move their head laterally in front of the display. The picket-fence effect is the appearance of vertical banding in an image due to the black mask between columns of pixels in the liquid crystal display (LCD). Image flipping indicates the noticeable transition between viewing zones which leads to discrete views and is experienced as unnatural compared to the continuous parallax experienced in the real world [35]. Display techniques can be improved such that both artifacts are less visible. For instance, in the work by Van Berkel *et al.* [72], a tilted lenticular sheet is put in front of the LCD as a result of which a constant amount of the black mask is always visible. Owing to habituation, an observer actually does not perceive the picket-fence effect anymore and image flipping is softened. Progress in autostereoscopic displays with user tracking is another promising approach to reduce image flipping.

VI. CONCLUSION: 3-D TV REQUIREMENTS AND FURTHER RESEARCH

The widespread introduction and acceptance of digital broadcasting makes the transmission of a stereoscopic signal increasingly feasible. Proponents of 3-D TV have argued that 3-D TV will bring the viewer a wholly new experience, a "fundamental change in the character of the image, not just an enhancement of quality" [73]. It is a widely held belief that 3-D television should be autostereoscopic. The main reason for this is that the need to wear glasses is unacceptable in a home situation where television is usually watched casually, i.e., with many interruptions for telephone calls, conversations, making a sandwich, or engaging in other activities with TV as a simultaneous background activity. Having to take a pair of spectacles on and off constantly is a nuisance.

Most current autostereoscopic displays, on the other hand, tend to restrict the viewer to a fairly rigid viewing position, in terms of the angle under which the stereoscopic imagery can be perceived without geometrical distortions or substantial artifacts (e.g., crosstalk or picket-fence effects). Stereoscopic TV should be able to provide good quality stereo pictures to multiple viewers who are free to move throughout the room. Current developments in multiview autostereoscopic displays provide hope that such a system may be feasible in the not-too-distant future [74], [39].

In addition to the requirements mentioned above, any stereoscopic system should also be able to display monoscopic images without problems, and with an image quality that is at least comparable but preferably superior to current TV sets. Other important considerations include cost and size[1] of the stereoscopic television system.

It is important to have a clear understanding of the potential added value and drawbacks of a 3-D TV broadcast service for the users. A stereoscopic image quality model could contribute to a lower cost design cycle for 3-D TV and the technological parameters can be optimized to the customer's quality preference. Finally, prototype systems need to be tested outside the controlled laboratory space, for it is the long-term, real-world use of 3-D TV that will prove its impact.

Despite significant advances over recent years, 3-D technology is not yet mature enough to introduce 3-D TV to the mass market. However, recent joint activities in industry and scientific research are promising. For instance, the European IST ATTEST project aims at a novel concept of 3-D TV which covers the entire broadcast chain, including 3-D recording, 3-D compression, and 3-D displays [11]. Furthermore, the goal of the MPEG-community working group 3DAV (3-D Audio Visual coding) is to standardize and define requirements for future 3-D audio-visual applications [75]. Moreover, in the spring of 2003, a 3-D consortium was formed to enhance the potential market for 3-D products and applications, including 3-D TV [76]. These efforts are a good basis for creating the necessary conditions to make 3-D TV a reality in the foreseeable future. Human factors research fulfills an essential role in this respect, addressing perceptual and usability issues, and optimizing the design of 3-D technologies from a user-centered perspective.

REFERENCES

[1] R. Sand, "3-DTV – A review of recent and current developments," *Proc. IEE Colloq. Stereoscopic Television*, pp. 1–4, 1992.

[2] P. Harman, "An architecture for digital 3D broadcasting," *Proc. SPIE*, vol. 3639, pp. 254–259, 1999.

[3] M. Op de Beeck and A. Redert, "Three dimensional video for the home," in *Proc. EUROIMAGE Int. Conf. Augmented, Virtual Environments and Three-Dimensional Imaging*, 2001, pp. 188–191.

[4] A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proc. SPIE*, vol. 1915, pp. 36–48, 1993.

[5] R. Spottiswoode, N. Spottiswoode, and C. Smith, "3-D photography – Basic principles of the three-dimensional film," *J. SMPTE*, vol. 59, pp. 249–286, 1952.

[6] M. Kawakita, K. Lizuka, T. Aida, H. Kikuchi, H. Fujikake, J. Yonai, and K. Takizawa, "AXI-vision camera (Real-time distance-mapping camera)," *Appl. Opt.*, vol. 39, pp. 3931–3939, 2000.

[7] M. Kawakita, T. Kurita, H. Kikuchi, and S. Inoue, "HDTV AXI-vision camera," in *Proc. IBC*, 2002, pp. 397–404.

[8] G. Iddan and G. Yahav, "3D imaging in the studio (and elsewhere …)," *Proc. SPIE*, vol. 4298, pp. 48–55, 2001.

[9] R. Gvili, A. Kaplan, E. Ofek, and G. Yahav, "Depth keying," *Proc. SPIE*, vol. 5006, pp. 554–563, 2003.

[10] 3DV Systems [Online]. Available: http://www.3dvsystems.com/news/press_releases.html

[11] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. van Gool, E. Ofek, and I. Sexton, "An evolutionary and optimized approach on 3D-TV," in *Proc. IBC*, 2002, pp. 357–365.

[12] C. Fehn, E. Cooke, O. Schreer, and P. Kauff, "3D analysis and image-based rendering for immersive TV applications," *Signal Processing: Image Commun.*, vol. 19, pp. 705–715, 2002.

[13] L. Meesters, B. Quante, P. Seuntiëns, and W. A. IJsselsteijn, "Report on Formative Evaluations of Software Prototype Chain and Display Systems,", ATTEST-IST-2001–34 396 rep. D6, 2003.

[14] P. Harman, "Home based 3D entertainment – An overview," in *Proc. ICIP*, vol. 1, 2000, pp. 1–4.

[15] P. Harman, J. Flack, S. Fox, and M. Dowley, "Rapid 2D to 3D conversion," *Proc. SPIE*, vol. 4660, pp. 78–86, 2002.

[16] M. Strintzis and S. Malassiotis, "Review of methods for object-based coding of stereoscopic and 3D image sequences," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 5, 1998, pp. 510–513.

[17] H. Aydinoglu and M. H. Hayes, "Stereo image coding: A projection approach," *IEEE Trans. Image Processing*, vol. 7, pp. 506–516, Apr. 1998.

[18] T. Naemura, M. Kaneko, and H. Harashima, "Compression and representation of 3-D images," *IEICE Trans. Inform. Syst.*, vol. E82-D, pp. 558–567, 1999.

[19] J. Jiang and E. A. Edirisinghe, "A hybrid scheme for low bit-rate coding of stereo images," *IEEE Trans. Image Processing*, vol. 11, pp. 123–134, Feb. 2002.

[20] V. Seferidis and D. Papadimitriou, "Improved disparity estimation in stereoscopic television," *Electron. Lett.*, vol. 29, pp. 782–783, Apr. 1993.

[21] M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. Commun.*, vol. 40, pp. 684–696, Apr. 1992.

[22] H. Yamaguchi, Y. Tatehira, K. Akiyama, and Y. Kobayashi, "Stereoscopic images disparity for predictive coding," in *Proc. ICASSP*, vol. 3, 1989, pp. 1976–1979.

[23] S. Sethuraman, M. W. Siegel, and A. G. Jordan, "A multiresolutional region-based segmentation scheme for stereoscopic image compression," *Proc. SPIE*, vol. 1429, pp. 265–274, 1995.

[24] H. Aydinoglu and M. H. Hayes, "Performance analysis of stereo coding algorithms," in *ICASSP*, vol. 4, 1996, pp. 2191–2195.

[25] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo images sequences using joint 3D motion/disparity compensation," *IEEE Trans. Video Technol.*, vol. VT-11, Mar. 1973.

[26] R. Krishnamurthy, B.-B. Chai, H. Tao, and S. Sethuraman, "Compression and transmission of depth maps for image-based rendering," in *Proc. ICIP*, Thessaloniki, Greece, 2001, pp. 828–831.

[27] O. Grau, S. Minelly, and G. Thomas, "Applications of depth metadata," in *Proc. IBC*, Sept. 14–17th, 2001.

[28] W. D. Reynolds and R. V. Kenyon, "The wavelet transform and the suppression theory of binocular vision for stereo image compression," in *Proc. ICIP*, vol. 1, 1996, pp. 557–560.

[29] D. V. Meegan, L. B. Stelmach, and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery," *J. Exp. Psychol.: Appl.*, vol. 7, pp. 143–153, 2001.

[30] L. B. Stelmach and W. J. Tam, "Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views," *Signal Processing: Image Commun.*, vol. 14, pp. 111–117, 1998.

[31] B. L. Tseng and D. Anastassiou, "A theoretical study on an accurate reconstruction of multiview images based on the viterbi algorithm," in *Proc. ICIP*, vol. 2, 1995, pp. 378–381.

[32] M. Siegel, S. Sethuraman, J. S. McVeigh, and A. Jordan, "Compression and interpolation of 3D-stereoscopic and multi-view video," *Proc. SPIE*, vol. 3012, pp. 227–238, 1997.

[33] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Disparity field and depth map coding for multiview 3D image generation," *Signal Processing: Image Commun.*, vol. 11, pp. 205–230, 1998.

[34] G.-C. Chang and W.-N. Lie, "Multi-view image compression and intermediate view synthesis for stereoscopic applications," in *Proc. ISCAS*, vol. 2, 2000, pp. 277–280.

[35] I. Sexton and P. Surman, "Stereoscopic and autostereoscopic display systems," *IEEE Signal Processing Mag.*, vol. 16, pp. 85–99, May 1999.

[36] S. Pastoor and M. Wöpking, "3-D displays: A review of current technologies," *Displays*, vol. 17, pp. 100–110, 1997.

[37] R. Zone, "The deep image: 3-D in art and science," *Proc. SPIE*, vol. 2653, pp. 4–8, 1996.

[38] StereoGraphics Corporation. [Online]. Available: http://www.stereographics.com

[39] P. Surman, I. Sexton, R. Bates, W. K. Lee, M. Craven, and K. C. Yow, "Beyond 3D television: The multi-modal, multi-viewer TV system of the future," in *Proc. SID/SPIE FLOWERS 2003*, 2003, pp. 208–210.

[40] A. Sullivan, "A solid-state multi-planar volumetric display," in *Proc. SID 2003*, 2003.

[41] LightSpace Technologies, Inc. [Online]. Available: http://lightspacetech.com

---

[1] Many of the systems proposed today require a housing size that is not acceptable for domestic settings.

[42] G. Favalora, R. K. Dorval, D. M. Hall, M. Giovinco, and J. Napoli, "Volumetric three-dimensional display system with rasterization hardware," *Proc. SPIE*, vol. 4297, pp. 227–235, 2001.

[43] Actuality Systems, Inc.. [Online]. Available: http://www.actuality-systems.com

[44] M. Lucente, S. A. Benton, and P. St.-Hilaire, "Electronic holography: The newest," in *Proc. Int. Symp. Three Dimensional Imaging and Holography*, Osaka, Japan, 1994.

[45] M. Lucente, "Interactive three-dimensional holographic displays: Seeing the future in depth," in *Proc. SIGGRAPH Compu. Graph.*, vol. 31, 1997, pp. 63–67.

[46] J. Freeman and S. Avons, "Focus group exploration of presence through advanced broadcast services," *Proc. SPIE*, vol. 3959, pp. 530–539, 2000.

[47] S. Bech, R. Hamberg, M. Nijenhuis, C. Teunissen, H. L. de Jong, P. Houben, and S. K. Pramanik, "The RaPID perceptual image description method (RaPID)," *Proc. SPIE*, vol. 2657, pp. 317–328, 1996.

[48] J. A. J. Roufs, "Perceptual image quality: Concept and measurement," *Philips J. Res.*, vol. 47, pp. 35–62, 1992.

[49] "Subjective Assessment of Stereoscopic Television Pictures," ITU, Recommendation BT.1438, 2000.

[50] "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU, Recommendation BT.500–10, 2000.

[51] W. S. Torgerson, *Theory and Methods of Scaling*. New York: Wiley, 1958.

[52] M. C. Boschman, "Difscal: A tool for analyzing difference ratings on an ordinal category scale," *Behavior Research Methods, Instruments, & Computers*, vol. 33, pp. 10–20, 2001.

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

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

[55] J. Carey, "The evolution of TV viewing," in *Proc. 4th Annu. TV Meets the Web Seminar*, Amsterdam, The Netherlands, 2002.

[56] H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments," *Proc. SPIE*, vol. 1666, pp. 16–26, 1992.

[57] P. Engeldrum, *Psychometric Scaling*. Winchester, MA: Imcotek Press, 2000.

[58] W. IJsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 225–233, Mar. 2000.

[59] A. Berthold, "The Influence of Blur on the Perceived Quality and Sensation of Depth of 2D and Stereo Images," ATR Human Information Processing Research Laboratories, 1997.

[60] A. Schertz, "Source coding of stereoscopic television pictures," in *Proc. IEE Inter. Conf. Image Processing and its Applications*, Maastricht, The Netherlands, Apr. 7–9, 1992, pp. 462–464.

[61] W. Tam, L. Stelmach, and P. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences," *Proc. SPIE*, vol. 3295, pp. 226–235, 1998.

[62] P. Seuntiëns, L. Meesters, and W. IJsselsteijn, "Perceptual evaluation of JPEG coded stereoscopic images," *Proc. SPIE*, vol. 5006, pp. 215–226, 2003.

[63] S. Pastoor, "Human factors of 3D displays in advanced image communications," *Displays*, vol. 14, pp. 150–157, 1993.

[64] L. Stelmach, W. Tam, F. Speranza, R. Renaud, and T. Martin, "Improving the visual comfort of stereoscopic images," *Proc. SPIE*, vol. 5006, pp. 269–282, 2003.

[65] S. Pastoor, "Human factors of 3D imaging: Results of recent research at Heinrich-Hertz-Institute Berlin," in *Proc. ASIA Display'95 Conf.*, 1995.

[66] H. Yamanoue, "The relation between size distortion and shooting conditions for stereoscopic images," *J. SMPTE*, pp. 225–232, April 1997.

[67] K. Hopf, "An autostereoscopic display providing comfortable viewing conditions and a high degree of telepresence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 359–365, Apr. 2000.

[68] A. J. Woods and S. Tan, "Characterising sources of ghosting in time-sequential stereoscopic video displays," *Proc. SPIE*, vol. 4660, pp. 66–77, Jan. 21–23, 2002.

[69] J. Konrad, B. Lacotte, and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," *IEEE Trans. Image Processing*, vol. 9, pp. 897–908, 2000.

[70] H. Yamanoue, M. Okui, and I. Yuyama, "A study on the relationship between shooting conditions and cardboard effect of stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 411–416, Apr. 2000.

[71] I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis*. New York: Oxford Univ. Press, 1995.

[72] C. van Berkel and J. Clarke, "Characterization and optimization of 3D-LCD module design," *Proc. SPIE*, vol. 3012, pp. 179–186, 1997.

[73] C. Smith and A. Dumbreck, "3-D TV : The practical requirements," *Television: J. Roy. Television Soc.*, pp. 9–15, Jan./Feb. 1988.

[74] G. J. Woodgate, D. Ezra, J. Harrold, N. S. Holliman, G. R. Jones, and R. R. Moseley, "Autostereoscopic 3D display systems with observer tracking," *Signal Processing: Image Commun.*, vol. 14, pp. 131–145, 1998.

[75] MPEG Ad-Hoc Group 3DAV [Online]. Available: http://www.chiariglione.org/mpeg/ working_documents.htm

[76] 3-D Consortium [Online]. Available: http://sharp-world.com/corporate/news/030 304.html

Lydia M. J. Meesters received the M.Sc. degree in computer science from the University of Nijmegen, Nijmegen, The Netherlands, and the Ph.D. degree from Eindhoven University of Technology, Eindhoven, The Netherlands.

During 1994–1996, she worked as a Software Engineer in the field of computer graphics and image processing. From 1996 to 2000, she was with the Vision Group, Institute for Perception Research, Eindhoven, The Netherlands, where she was involved in the European ACTS TAPESTRIES project. Her research focused on objective quality measures, subjective image quality models, and subjective scaling paradigms for compressed two-dimensional pictures. In 2002, she joined the Human-Technology Interaction (HTI) Group, Faculty of Technology Management, Eindhoven University of Technology. Currently she works on the European IST ATTEST project and she conducts research on three-dimensional (3-D) perception and human factors issues surrounding 3-D TV.

Wijnand A. IJsselsteijn was trained as a psychologist at Utrecht University, with a specialization in neuropsychology.

From 1996 to 2000, he was with the Institute of Perception Research (IPO), Eindhoven, The Netherlands, on the ACTS TAPESTRIES project, where his input was aimed at developing and applying psychological assessment methodologies that are suited to evaluate the subjective impact and potential side-effects of stereoscopic entertainment media. His research interests focus on the psychological evaluation of new display and communication media, in particular the sense of (social) presence these media may elicit. Since August 2000, he has been working as an Assistant Professor with the Human-Technolog Interaction (HTI) Group, Department of Technology Management, Eindhoven University of Technology, Eindhoven, where his research interests have broadened to include multilevel presence theory (ranging from neuroscientific to sociological perspectives), multisensory interactions, and haptic interfaces, studied from the perspective of applied experimental psychology. He was project coordinator for the IST PRESENCE project (2001–2002) and is currently workpackage leader for the perceptual evaluation work within the IST ATTEST project and project coordinator for the IST OMNIPRES project. He has published in various journals and conference proceedings and has co-edited a recent book on presence: *Being There—Concepts, Effects and Measurement of User Presence in Synthetic Environments* (Amsterdam, The Netherlands: IOS Press, 2003).

Pieter J. H. Seuntiëns was born in Eindhoven, The Netherlands, on February 6, 1976. He received the B.Sc. degree in mechanical engineering from Fontys University, Eindhoven, The Netherlands, and the M.Sc. degree in human technology interaction from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1999 and 2002, respectively, where he is currently working toward the Ph.D. degree.

Since April 2002, he has been working on the European ATTEST project. In close collaboration with the Video Processing and Visual Perception Group of Philips Research Laboratories, Eindhoven, his work focuses on perceptual issues and quality metrics regarding two- and three-dimensional imaging systems.