Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation

PIETER SEUNTIENS
and
LYDIA MEESTERS
and
WIJNAND IJSSELSTEIJN
Eindhoven University of Technology

JPEG compression of the left and right components of a stereo image pair is a way to save valuable bandwidth when transmitting stereoscopic images. This paper presents results on the effects of camera-base distance (B) and JPEG-coding on overall image quality, perceived depth, perceived sharpness and perceived eye-strain. In the experiment, two stereoscopic still scenes were used, varying in depth (three different camera-base distances: 0, 8 and 12 cm) and compression ratio (4 levels: original, 1:30, 1:40 and 1:60). All levels of compression were applied to both the left and right stereo image, resulting in a 4x4 matrix of all possible symmetric and asymmetric coding combinations. The observers were asked to assess image quality, sharpness, depth and eye-strain. Results showed that an increase in JPEG coding had a negative effect on image quality, sharpness and eye-strain but had no effect on perceived depth. An increase in camera-base distance increased perceived depth and reported eye-strain but had no effect on perceived sharpness. Results on asymmetric and symmetric coding showed that the relationship between perceived image quality and average bit-rate is not straightforward. In some cases, image quality ratings of a symmetric coded pair can be higher than for an asymmetric coded pair, even if the averaged bit rate for the symmetric pair is lower, than for the asymmetric pair. Furthermore, sharpness and eye-strain correlated highly and medium, respectively, with perceived image quality.

Categories and Subject Descriptors: ... [**...**]: ...

General Terms: Human factors
Additional Key Words and Phrases: asymmetric/symmetric JPEG coding, depth, eye-strain, image quality, sharpness, stereoscopic images

1. INTRODUCTION

Complex, natural scenes contain a wide variety of visual cues to depth. Our visual system utilises monocularly available information such as accommodation, occlu-

Author's address: P.J.H. Seuntiëns, Eindhoven University of Technology, Faculty of Technology Management, Den Dolech 2, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands. E-mail: p.j.h.seuntiens@tue.nl

sion, linear and aerial perspective, relative size, relative density, and motion parallax to construct a perception of depth. The effectiveness of monocular cues is illustrated by the fact that we can close one eye and still have a considerable appreciation of depth. These cues are already available in monoscopic displays, such as conventional 2-D television. The binocular cues, stereopsis and vergence, require both eyes to work in concert. Random-dot stereograms or Julesz patterns [Julesz 1971] demonstrate that in the absence of consistent monocular information, stereopsis alone provides the visual system with enough information to extract a fairly accurate estimate of depth and shape.

Binocular disparity is available because the human eyes are horizontally separated - on average approximately 6.3 cm [Dodgson 2004]. This horizontal separation causes an interocular difference in the relative projections of monocular images onto the left and right retina. When points from one eye's view are matched to corresponding points in the other eye's view, the retinal point to point disparity variation across the image provides information about the relative distances of objects to the observer and the depth structure of objects and environments.

Stereoscopic display techniques utilise this principle by taking two images and displaying them in such a way that the left view is only seen by the left eye, and the right view only seen by the right eye. There are a number of ways of achieving this [Okoshi 1980; Pastoor and Wöpking 1997; Sexton and Surman 1999]. Today, stereoscopic displays have application utility in many areas such as simulation systems (e.g., flight simulators), medical systems (e.g., endoscopy), telerobotics (e.g., remote inspection of hazardous environments), computer-aided design (e.g., car interior design), data visualisation (e.g., molecular or chemical modelling, oil and gas exploration, weather forecasting), telecommunication (e.g., videoconferencing) and entertainment (e.g., TV, gaming). Stereoscopic systems reinforce the perception of physical space in both camera-recorded and computer-generated environments. The capacity to provide an accurate representation of structured layout, distance and shape can be utilised for precise perception and manipulation of objects, and has added value for communication and entertainment purpose.

The storage and transmission of stereoscopic image material involves a large amount of data. Therefore, a considerable research effort is focused on realizing digital image compression (such as JPEG or MPEG coding) to obtain savings in bandwidth and storage capacity. This is of particular relevance in the case of stereoscopic HDTV, where a single uncompressed HDTV channel may cost up to one Gbit/s transmission bandwidth, or in the case of stereoscopic video transmission over low-bandwidth transmission channels, such as the Internet [Johanson 2001].

The same compression techniques used in two-dimensional image material can also be applied independently on the left and right view of a stereoscopic image pair. Image compression may compromise perceived image quality however, through loss of detail and the introduction of compression artefacts such as blockiness, blur, or ringing. In order to ensure that the applied compression algorithms and levels still yield perceptually acceptable results, subjective testing using human viewers has been the only accurate method to date for assessing compressed stereoscopic video systems.

In this paper we evaluate a number of relevant perceptual attributes regarding symmetric and asymmetric JPEG coding of stereoscopic images. First, we will discuss previous research on perceived image quality, perceived depth, perceived sharpness and perceived eye-strain in relation to 3D television. Subsequently, we will describe an experiment where we have manipulated camera-base distance (B) and JPEG coding level and have measured the effects on perceived quality, sharpness, depth and eye-strain.

1.1 Asymmetric coding and perceived image quality

Based on theories of binocular suppression, it is assumed that the binocular percept of a stereo image pair is dominated by the high quality component [Levelt 1965]. Thus, theoretically, when one image of the stereo pair is compressed such that a high quality is maintained, the other view can be coded more heavily without introducing visible artefacts in the binocular percept. The mixed resolution concept was introduced by Perkins [1992]. Mixed resolution coding assumes that the binocular percept is not affected when one view is of high quality and the other view of lower quality. Perkins [1992] applied low-pass filtering (introducing blur) as compression algorithm resulting in a high-resolution and a low-resolution image for each view of a stereo image pair. The author concludes that mixed-resolution coding is easy to implement, and the reduction of the bit rate is significant with respect to a system that employs no coding. Stelmach and Tam [1998] and Tam et al. [1998] applied a different compression ratio on the left- and right-eye views of a stereoscopic sequence using MPEG-2 (introducing blockiness) and low-pass filtering (introducing blur). The results showed that the subjective image quality of a stereo sequence was approximately the average of the monoscopic quality of the left- and right-eye images when MPEG-2 coding was used. Subjective image quality of an asymmetric low-pass filtered stereo sequence was dominated by the high quality component. Meegan et al. [2001] studied the binocular combination of asymmetric blur and blockiness impairment images. In the case of asymmetric blur-impaired images the binocular percept was dominated by the high quality component. The binocular percept of the asymmetric MPEG-2 impaired images was approximately the average of the two monoscopic components. From this research it can be concluded that the success of asymmetric coding depends on the type of coding artefacts.

1.2 Perceived depth

The use of disparity information produces a compelling sense of depth, which defines the added value of stereoscopic TV. IJsselsteijn et al. [1998] investigated the perception of depth and the naturalness of depth when viewing stereoscopic image material. As soon as binocular disparity was introduced, the ratings of perceived depth and naturalness of depth increased. Research of Westheimer and McKee [1980] documented a decrease in stereopsis threshold with asymmetric blur that is larger than with symmetric blur. Stelmach et al. [2000] investigated the effect of spatial and temporal low-pass filtering on perceived depth for symmetric and asymmetric stereo pairs. The results indicate that spatial low-pass filtering has no effect on perceived depth. Temporal low-pass filtering produced poor image quality but the sensation of depth was relatively unaffected. An explanation is that low-pass filtering leaves the low spatial frequencies, that are sufficient to carry the disparity signal, unaffected. In their studies, depth shows a weak correlation with image quality and sharpness. These results suggest that depth is a dimension of perceptual experience that is largely independent of sharpness and overall image quality. This appears to be at variance with the results of IJsselsteijn et al. [2000] where perceived image quality could be expressed as a function of perceived depth and experienced eye-strain. These results were obtained using uncompressed images that varied in terms of camera-base distance, convergence distance, and focal length. A number of stimuli contained excessive disparities, thus making it likely for subjects to base their quality judgements on different image attributes than with the Stelmach et al. [2000] study.

1.3 Perceived sharpness

Perceived sharpness in stereoscopic images can be affected by several parameters, e.g., camera defocus, coding, or binocular disparity. Berthold [1997] reported that stereo images with different degrees of Gaussian blur were perceived sharper than non-stereo images. Tam et al. [1998] on the other hand found that the subjects rated the symmetric MPEG-2 coded stereo and non-stereo images equally sharp or the stereo images even slightly less sharp. A high correlation was found between sharpness and image quality in both studies. Stelmach et al. [2000] investigated the effect of mixed-resolution on perceived sharpness and concluded that spatial low-pass filtering gives an acceptable sharpness. Sharpness was biased towards the view with the greater spatial resolution. On the other hand, temporal low-pass filtering produced very poor images with blurred edges. Meegan et al. [2001] confirmed these findings in an experiment measuring the visibility of blur in asymmetric processed stereo images. When the lower quality view contained blur artefacts, the higher quality view was over-weighted by the visual system.

1.4 Perceived eye-strain

Many studies report a clear preference for stereoscopic images over non-stereoscopic ones [IJsselsteijn et al. 1998; Freeman and Avons 2000; Yano and Yuyama 1991]. However, viewing stereo images can be more fatiguing than viewing conventional two-dimensional images. Because eye-strain can be extremely annoying in stereoscopic displays, it is important to have an understanding of its subjective magnitude and impact on the user. IJsselsteijn et al. [2000] investigated the effect of stereoscopic filming parameters and display duration on the subjective assessment of eye-strain. The averaged results of the eye-strain ratings show a clear linear increase with increasing disparities. There was no significant effect of display duration on the eye-strain scores, but the display durations were relatively short (1-15 seconds). Mitsuhashi [1996] found that observers experienced more eye-strain for binocular vision than with the conventional television picture, using an objective measure known as the critical flicker frequency (CFF). The critical flicker frequency is the highest frequency at which a particular person still sees flicker. At any higher frequency, the subject sees a steady light source. Watching stereoscopic television caused a significant CFF decrease within 30 minutes. It was also found that the CFF decreases are related to a subjective feeling of eye-strain. Okuyama [1999] evaluated visual fatigue with visual function testing (objective measure) and interviews (subjective measure). Visual function testing showed a mismatch between convergence and accommodation. The interviews reported more eye pain, an 'alien feeling' in the eyes and eyes filled with tears. Both evaluations show an increase in visual fatigue. Kooi and Toet [2004] concluded that disparity, crosstalk and blur are the most important parameters that cause eye-strain.

1.5 Experiment

In sum, it seems that the mixed resolution concept is appropriate for stereoscopic transmission although the quality of the binocular percept will depend on the type of distortion. Previous experiments on asymmetric coding, however, did not manipulate camera-base distance. Since it is possible that different camera separation settings influence the visibility of coding artefacts, we decided to perform an experiment that would be aimed at investigating the effects of asymmetric/symmetric coding while using different levels of camera separation. The reason for varying camera separation is that larger separations lead to larger differences between the left and right eye view in terms of view perspective and image content at the borders of the image (image information appears and disappears with varying camera separation). In JPEG coding of stereoscopic image pairs, each image is encoded separately. Thus, the coding artefacts (e.g. blockiness) for each image will be unique and the differences between the coding artefacts in the left and right eye image will become more pronounced as the inter-camera separation increases. Thus, image coders (such as JPEG or MPEG) produce different artefacts in the left and right eye view in terms of intensity, position and shape when varying camera separation. So, matching the left and right eye view (based on corresponding points in the images) may result in a different perception of the artefacts. Thus, we predict that image quality would be negatively affected. Secondary, we also wanted to control for a possible effect of eye dominance, where images presented to the dominant eye would potentially contribute more to the overall stereoscopic percept than images presented to the non-dominant eye. Further, we want to extend the study of the effects of camera-base distance and compression (JPEG coding) on perceived depth, sharpness, image quality and eye-strain, as results from experiments reported in the literature appear inconclusive and somewhat contradictory.

2. METHOD

2.1 Observers

Forty non-expert observers were paid to participate in this experiment. They were divided in 4 groups assessing 1 attribute per group, i.e., a between-subjects design. The attributes to be assessed in this experiment were perceived sharpness, perceived depth, perceived image quality and perceived eye-strain. The observers, mostly students, came from the same age group (18-27 years old). All participants had a visual acuity of ≥ 1, good stereo vision <30 seconds of arc (as tested with the Randot stereo test) and good colour vision (as tested with the Ishihara test). Eye dominance and inter-pupillary distance were also measured. The Finger-Point method was used to determine eye dominance. Subjects pointed naturally at an object with both eyes open and the face square to the object. The eyes were closed alternately. When the dominant eye is closed the finger appears to jump away from the original location. Thirty-one subjects were right eye dominant and the average inter-pupillary distance was 6,2 cm (slightly below the population average of 6,3

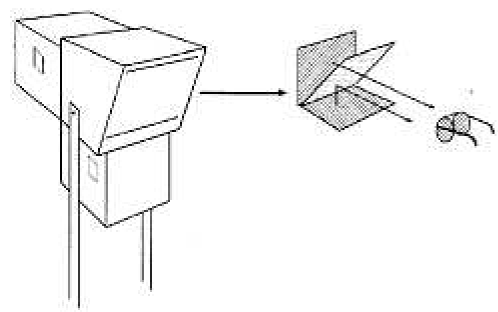

Fig. 1. The AEAT system consisting of two Barco CRT monitors displaying the left and right eye images at the same time. The polarized glasses are used to separate the left and right views.

cm [Dodgson 2004]). The viewing distance was 80 cm.

2.2 Materials

2.2.1 *Equipment.* An AEA-Technology stereoscopic display was used in this experiment to display the stereoscopic material. The AEAT system consists of two Barco CPM 2053FS CRT colour monitors mounted perpendicular to each other (see Figure 1). The dual monitor system displayed the right and left image at the same time using a half see-through mirror and a polarization filter in front of each screen. The observers wore polarized glasses in order to provide left-right separation with very little crosstalk in the stereo pair (less than 0.1% with linear polarized filters, [Pastoor and Wöpking 1997]). A SUN ISP system provided the CRT monitors with a video signal. Custom built software was used to control the display duration and synchronized the output of the 2 codecs transferring the images.

2.2.2 *Stimuli.* The image material used in this experiment consisted of two still scenes, *Playmobiles* and *Bureau*, that varied in camera-base distance (B) and compression ratio. The scene *Playmobiles* consists of a colourful toy landscape with mountains in the background and numerous *Playmobiles* in the foreground. The scene *Bureau* consists of a tailor's dummy sitting behind a desk on which some office equipment is located. The original scenes are shown in Figure 2.

The scenes were recorded in a studio set-up, using a professional stereoscopic studio camera in a toed-in configuration. For each scene, lens focal length and convergence distance of the cameras to the scene were fixed to 20 mm and 1.30 m, respectively. Each scene was recorded at three different camera-base distances, namely 0 cm (i.e. monoscopic), 8 cm and 12 cm. The scenes originated in the European DISTIMA project were kindly provided to us by CCETT in France.

An increase in camera-base distance results in an increase in disparity values and thus perceived depth, while the size of the objects and the field of view remains constant. A camera-base distance of 0 cm introduces no disparity between the left and the right image and thus no perceptible stereoscopic depth, while depth is highly noticable with a camera-base distance of 8 cm and 12 cm.

The stimulus set contained the original, uncompressed version of each scene and

Fig. 2. The left panel shows the original of the scene *Playmobiles* and the right panel shows the original of the scene *Bureau*.

Table I. Bytes per pixel (bpp) and the compression ratio for JPEG coded images.

Q-parameter	bpp	Compression Ratio	Stereoscopic image bpp		left Eye			
					Org	Q30	Q20	Q10
Org	3.00	1:1	Right Eye	Org	3.00	1.55	1.54	1.53
Q30	0.11	1:30		Q30	1.55	0.11	0.10	0.08
Q20	0.08	1:40		Q20	1.54	0.10	0.08	0.07
Q10	0.05	1:60		Q10	1.53	0.08	0.07	0.05

three JPEG coded versions. The Baseline Sequential JPEG compression software package of the Independent JPEG Software Group[1] with default quantization table was used to generate for each scene different versions at various compression rates. The compression rate was determined by the 'Q-parameter'. Images with a high compression ratio were obtained by low 'Q-values', and therefore contained the most conspicuous distortions. The JPEG 'Q-parameters' used in this experiment were Q30, Q20 and Q10 for the scenes *Playmobiles* and *Bureau*. The coding levels were chosen carefully based on the visibility of artefacts. Stereo image pairs were formed by symmetric or asymmetric coding of the left and right eye images. In a symmetric stereoscopic image pair the same compression ratio is applied to the left and right components. In asymmetric coding the compression ratio of the two components are different. Table I gives the bytes per pixel (bpp) of the separate component of a stereo pair and the bpp of the symmetric and asymmetric image pairs, which is the averaged bpp of the left and right eye components.

In order to test a potential effect of eye dominance, each combination of bit-rates was presented to both eyes. Each JPEG coding level was displayed an equal number of times to each eye (left and right) and all combinations of the coding levels were presented once. Thus in total, 2 scenes, 3 camera-base distances and 4x4 coding levels were used. This resulted in a stimulus set of 2x3x16 = 96 images.

[1]http://www.ijg.org

2.3 Procedure

A set of 96 stereoscopic images was randomized and presented sequentially. Observers were asked to rate according to the single stimulus scaling method. Each attribute was rated by a different group of 10 observers and each observer rated only one attribute. The perceived overall image quality was rated on a quality category scale from 1 up to 5 corresponding with 1 for bad image quality and 5 for excellent image quality. The scale was labeled with the adjective terms [bad]-[poor]-[fair]-[good]-[excellent] according to the International Telecommunication Union [2000] recommendation on subjective quality assessment. Perceived sharpness was rated on an impairment scale from 1 up to 5. The least sharp image corresponded with 1 and the sharpest image with 5. Experienced eye-strain was rated on an impairment scale from 1 up to 5 with the adjective terms [very annoying]-[annoying]-[slightly annoying]-[perceptible, but not annoying]-[imperceptible] according to the International Telecommunication Union [2000] recommendation. Perceived depth was rated on a numerical scale from 1 up to 5. The image with no perceived depth was to be rated 1 and the image containing most perceived depth was to be rated 5. No adjectives were used on the depth and sharpness scale.

The stimulus set of 96 images was judged for each attribute in two subsessions, containing 48 stimuli each, with a small break in between. Each subsession of 48 images took approximately 20 minutes. Before the experiment started the observers were asked to read the instructions explaining the task and attribute they had to rate. After that the observers participated in a trial of 16 stimuli to get acquainted with the stimulus set and the range of variations in image parameters (camera-base distance, compression ratio).

3. RESULTS

A way to analyze data obtained by numerical category scaling experiments is to transform the data into an interval scale assuming a psychologically linear continuum. Thurstone's law of categorical judgement can be used for such transformations [Thurstone 1927]. The Thurstone model assumes that the attribute strength is measured on an internal psychological scale, i.e., an interval scale with Gaussian noise distribution. For all subjects, the raw category scaling data obtained in the experiment were transformed to a Thurstone scale using the software package ThurcatD [Boschman 2000]. Equal distances on the scale correspond with equal differences in the judged percept because the Thurstone scale is a true interval scale. The calculated Thurstone scale values are scaled back to the original scale using a linear transformation.

3.1 Eye dominance

The effect of eye dominance was tested for the asymmetric combinations between ratings of left-eye dominant (averaged over 9) and right-eye dominant (averaged over 31) subjects for each attribute. As tested with a Wilcoxon Signed Rank Test, which is a conservative test, there were no significant differences between the ratings of left-eye dominant and right-eye dominant subjects for all asymmetric combinations and attributes. Therefore we pooled the data of the left- and right-eye dominant subjects. Next, we tested the effect between the ratings of the asymmetric combina-

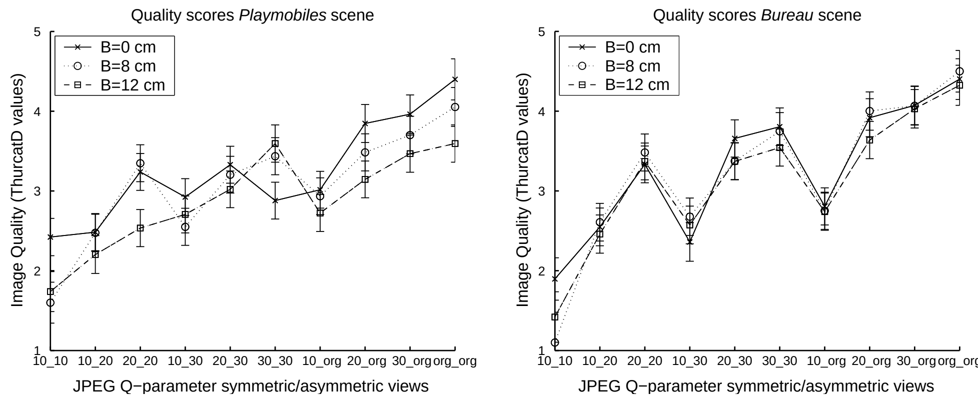

Fig. 3. Thurstone values and error bars for image quality for the scenes *Playmobiles* and *Bureau*. The x-axis represents the JPEG Q-parameter (increasing bit-rate to the right) for the symmetric and asymmetric image pairs and the three lines in the figure represent the 3 camera-base distances.

tions and their reverse version per attribute and also found no significant difference between them.

3.2 Perceived Image Quality

Figure 3 shows the Thurstone values for image quality and the standard errors for the scenes *Playmobiles* and *Bureau*. On the x-axis the symmetric and asymmetric coding combinations of the stereoscopic images are presented in increasing bit-rate. The y-axis represents the Thurstone scale values from 1 (bad) to 5 (excellent). The three lines in the figure represent the 3 camera-base distances 0, 8 and 12 cm.

The quality scores show an increasing trend with increasing bit-rate for both scenes. Remarkable are the two quality dips in the *Bureau* scene. At these points one of the two views of the stereoscopic image is coded with a JPEG compression factor of Q10. This image contains a lot of annoying artefacts (mostly blockiness) which can hardly be reduced by the high quality component of the stereoscopic image pair. The *Bureau* scene contains some homogeneous areas were the blockiness artefact is more visible. The quality dip in the *Playmobiles* scene, containing less homogeneous areas, is smaller but also at Q10. These quality dips indicate that the relationship between perceived image quality and average bit-rate is not straightforward, at least for stereoscopic images. For example, image quality ratings of a symmetric coded pair (Q30_Q30) can be higher than for an asymmetric coded pair (Q10_org), even if the averaged bit rate for symmetric pair (Q30_Q30) is lower, than for the asymmetric pair (Q10_org). An increase in camera-base distance (B) has almost no effect on perceived image quality. The *Bureau* scene shows almost no differences between the three camera-base distances and the quality judgements of the *Playmobiles* scene differs only slightly.

The weighting of bi-ocular inputs (left and right view) in binocular combination was also investigated. When the left and right view of an asymmetric coding pair are viewed separately, the less compressed view would have a high subjective

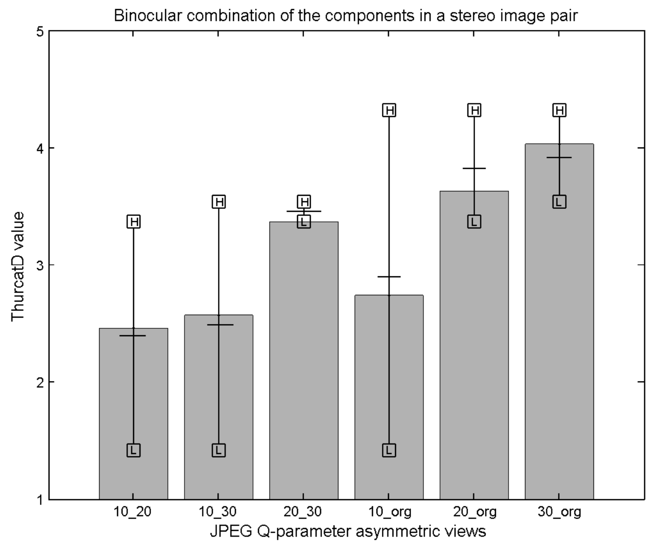

Fig. 4. Binocular weighting of image quality for the *Bureau* scene with a camera-base distance of 12 cm. In the first case, labeled as 10_20 on the x-axis, the judged image quality of the symmetric image pairs are labeled as 'L' and 'H' corresponding to JPEG coding Q10_Q10 and Q20_Q20. The judged image quality of the asymmetric image pair is the height of the bar (Q10_Q20).

image quality. On the other hand, the highly compressed view would have a low subjective image quality, and JPEG coding artefacts are visible. In Figure 4, the image quality of the symmetric inputs (high 'H' quality e.g. Q20_Q20, and low 'L' quality e.g. Q10_Q10 of the views) are presented by the whiskers (endpoints) and the asymmetric combination (stereo view of the high and low quality inputs, e.g. Q10_Q20) is presented by the height of the bar. The image quality of the symmetric inputs is the bi-ocular combination of two images with the same compression ratio. Figure 4 presents the symmetric and asymmetric stereoscopic image pairs of the *Bureau* scene with a camera-base distance of 12 cm.

The results of Figure 4 show that the perceived image quality of the binocular combination was approximately the average of the bi-ocular high quality and low quality input. The results show a tendency towards the low quality input. This may be due to the fact that in the lower compression images the blockiness artefact is highly visible. There are some differences between the scenes but these are small. The tendency towards the low quality input is a bit stronger in the *Bureau* scene than in the *Playmobiles* scene when Q10 is in the asymmetric image pair. This can be explained by the fact that the blockiness artefact is more visible in the *Bureau* scene at high compression rates because there are more homogeneous areas.

3.3 Perceived Depth

The Thurstone values for perceived depth and the error bars for the *Playmobiles* and *Bureau* scene are presented in Figure 5. The x-axis represents the symmetric and asymmetric coding combinations in increasing bit-rate. The y-axis represents

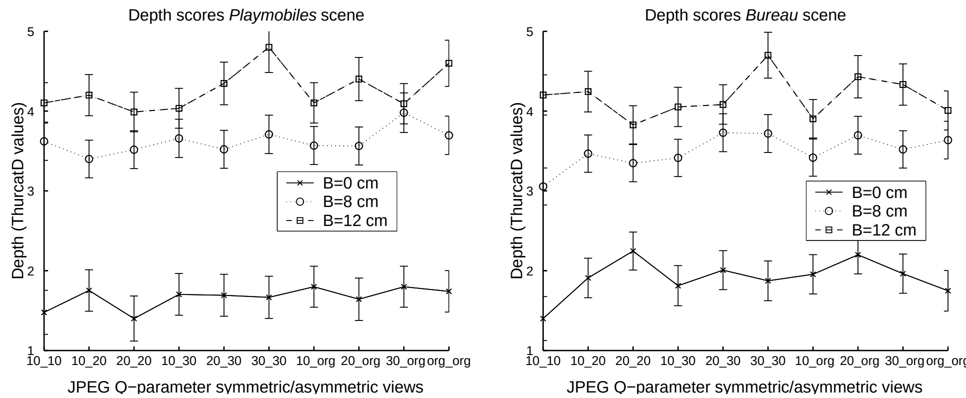

Fig. 5. Thurstone values and error bars for perceived depth for the *Playmobiles* and *Bureau* scenes. The x-axis represents the JPEG Q-parameter (increasing bit-rate to the right) for the symmetric and asymmetric image pairs and the three lines in the figure represent the 3 camera-base distances. The y-axis shows the Thurstone values of the judged depth.

the Thurstone values from 1 (no perceived depth) to 5 (maximum perceived depth).

As expected, the perceived depth scores increased when the camera-base distance increased. The perceived depth between camera-base distance 8 and 12 cm increased less than between 0 and 8 cm. Furthermore, JPEG coding had no effect on perceived depth. For all JPEG compression levels and combinations, the perceived depth remains nearly the same for each camera-base distance.

3.4 Perceived Sharpness

In Figure 6 the Thurstone values for sharpness of the *Playmobiles* scene and the *Bureau* scene are given. The x-axis represents the symmetric and asymmetric coding combinations in increasing bit-rate. In this case, the y-axis represents the Thurstone values from 1 (blurred) to 5 (sharp).

The results for perceived sharpness show great similarity with the perceived image quality results. Perceived sharpness increased when the bit rate increased. Also in these figures the perceived sharpness scores dropped dramatically as soon as JPEG compression level Q10 was presented in one of the two views of the stereoscopic pair. The sharpness scores in the *Bureau* scene were approximately the same for the three camera-base distances. There were little differences visible in the *Playmobiles* scene between the three camera-base distances. So, introducing image disparity appears to have no effect on perceived sharpness in our stimulus set.

3.5 Perceived Eye-strain

Figure 7 represents the Thurstone scale values for eye-strain scores of the subjects. The 5-grade impairment scale was used running from 1 (very annoying) to 5 (imperceptible).

The results show less annoyance with increasing bit-rate (less compression) and more annoyance with increasing camera-base distance. As in the quality and sharp-

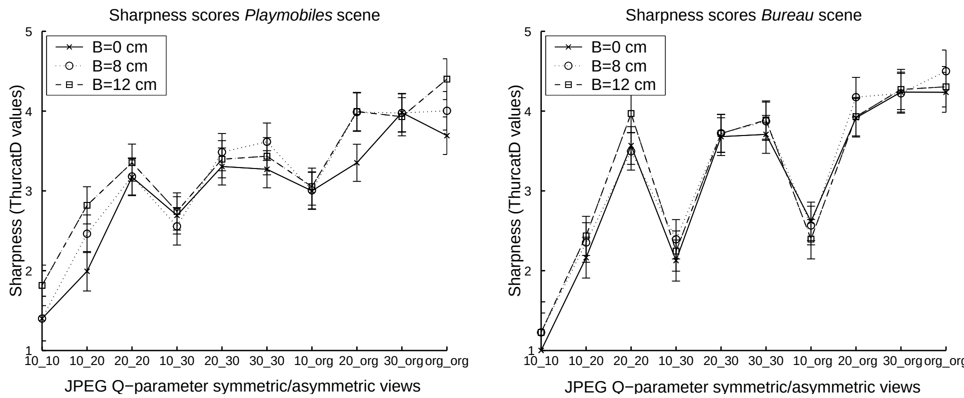

Fig. 6. Thurstone values and error bars for sharpness for the *Playmobiles* and *Bureau* scenes. The x-axis represents the JPEG Q-parameter (increasing bit-rate to the right) for the symmetric and asymmetric image pairs and the three lines in the figure represent the 3 camera-base distances. The y-axis shows the Thurstone values for judged sharpness.

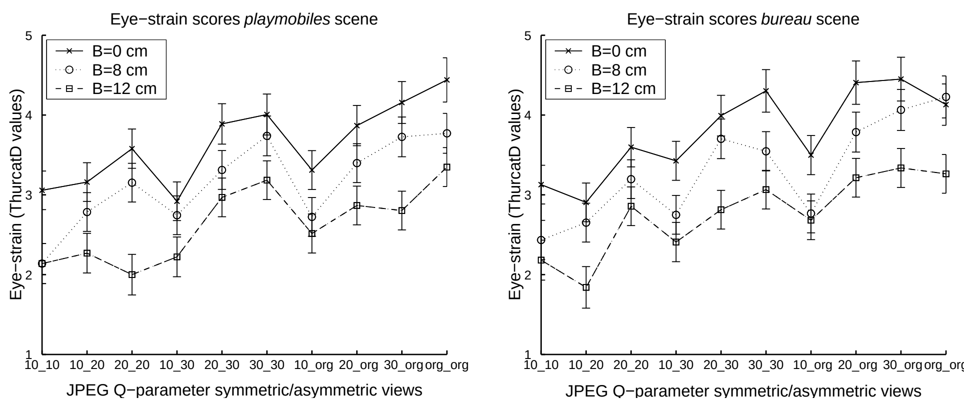

Fig. 7. Thurstone values and error bars for eye-strain for the scenes *Playmobiles* and *Bureau*. The x-axis represents the JPEG Q-parameter (increasing bit-rate to the right) for the symmetric and asymmetric image pairs and the three lines in the figure represent the 3 camera-base distances. The y-axis shows the Thurstone values for eye-strain.

ness figures, there is an increase in reported eye-strain as soon as JPEG coding level Q10 is presented in one of the two views of a stereoscopic image pair.

3.6 Correlation between attributes

In this experiment we found a high correlation coefficient of determination between image quality and sharpness (R^2=0.86) and a medium correlation coefficient of determination between image quality and eye-strain (R^2=0.58). No significant correlation was found between image quality and depth as can be deduced when comparing Figures 3 and 5. In Figure 8 the scatter plots of the average scores of the attributes are shown for each combination of JPEG coded stereo images and

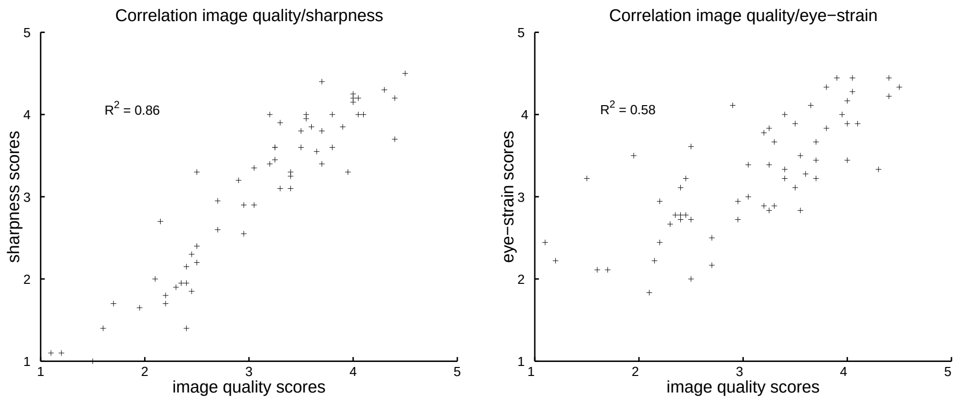

Fig. 8. Scatter plots of the attributes with a high correlation.

for all three camera-base distances.

3.7 Symptom Checklist

To check whether there were any negative side effects as a result of watching 45 minutes of stereoscopic images, we asked subjects to fill out a symptom checklist before and after the complete experiment. The symptom checklist consisted of 6 items which subjects had to rate namely (1) General discomfort, (2) Fatigue, (3) Headache, (4) Eye-strain, (5) Difficulty focussing and (6) Blurred vision.

Figure 9 shows the averaged results of the symptom checklist. The categories correspond with the labels on the x-axis. The y-axis shows the averaged scores (0 = none, 1 = slight, 2= moderate, 3 = severe) over 10 subjects for each attribute. The figures are shown separately for each of the subjective ratings the observers were asked to perform. This is done so as to visualise potential priming or sensitisation processes in relation to negative side effects as a consequence of having provided a certain attribute rating (e.g., eye-strain). In all figures a slight increase in symptoms can be observed. A Wilcoxon Matched-Pairs Signed Ranks test was used to reveal statistical significant differences (criterium $p < 0.05$) between the averaged checklist results before and after the experiments. The symptom checklist of the attribute quality shows a significant difference for the items headache ($p=0.034$) and eye-strain ($p=0.034$). No significant differences were found for the attributes depth and sharpness. The attribute eye-strain reveals significant differences for the items general discomfort ($p=0.014$), fatigue ($p=0.034$) and eye-strain ($p=0.014$). The items headache and difficulty focussing almost reached significance. The averaged ratings of the attribute eye-strain differed remarkably from the averaged ratings of the other attributes, although all subjects saw the same picture set. A possible explanation maybe the direct association of the attribute eye-strain with the items on the symptom checklist, sensitising observers to potential negative effects associated with viewing a stereo display.

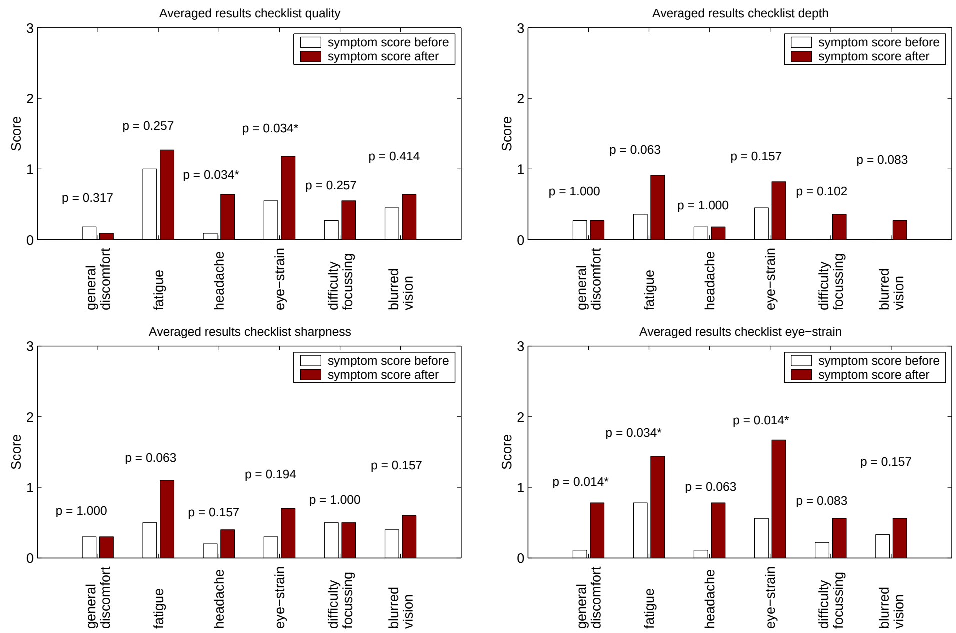

Fig. 9. Averaged symptom checklists for all attributes. The x-axis shows the items (1) general discomfort, (2) fatigue, (3) headache, (4) eye-strain, (5) difficulty focussing and (6) blurred vision. The y-axis shows the averaged scores (0 = none, 1 = slight, 2= moderate and 3 = severe).

4. DISCUSSION

The perceived image quality of monoscopic images (B=0 cm) was rated about the same as the stereoscopic images (B=8 and 12 cm). The reason for this finding may be that the added value of perceived depth was not taken into account when judging image quality. This may be due to the used stimulus set and experimental paradigm. Stelmach et al. [2000] also found a low correlation between image quality and perceived depth using low-pass filtering. On the other hand, IJsselsteijn et al. [2000] described an empirical relation between perceived depth, eye-strain and image quality for uncompressed stereoscopic images. The authors showed that an increase in image quality ratings could be attributed to an increase in perceived depth (when kept within natural bounds). However, quality judgements were attenuated by the eye-strain ratings, thus arriving at a simple stereoscopic image quality model for uncompressed images, describing quality as the difference between the added value of depth diminished by experienced eye-strain. In the current experiment, image artefacts were introduced which may have dominated the image quality assessment, making participants focus less on the depth dimension.

The results of this experiment show that an increase in JPEG coding decreases perceived image quality and sharpness and slightly increases perceived eye-strain. No effect of JPEG coding was found on perceived depth. Results of Stelmach et al. [2000] showed that spatial low-pass filtering had no effect on perceived depth. In an earlier study, Stelmach and Tam [1998] showed that subjective image quality of a stereo sequence was approximately the average of the monoscopic quality of the left- and right-eye images. In this experiment, the stereoscopic image quality of the binocular combination was approximately the average but tended a bit towards the low-quality component of the stereoscopic picture for all disparities (B= 0, 8 and 12 cm). From the image quality figures, it can be concluded that there is almost no decrease in image quality when coding at org_org, org_q30 and org_q20. The quality drops heavily when one of the views is coded with a JPEG compression of Q10. This can be explained because the blockiness artefact is highly visible in Q10. These results are in line with Meegan et al. [2001] who also found under-weighting of the high quality component when the blockiness artefact was present in the image. So, asymmetric coding is a valuable way to save bandwidth but one view must be of high quality (preferable the original) and the compression level of the coded view must by within acceptable range (in this experiment Q20 is allowed). Also an interesting conclusion of this experiment is, that the relation between perceived image quality and average bit-rate is not always straightforward, at least for stereoscopic images. Thus, an increase in average bit-rate does not always result in an increase in perceived image quality. The image quality of Q30_Q30 is higher than the image quality of Q10_org, whereas the average bit-rate of Q30_Q30 is lower than for Q10_org.

Camera-base distance had no significant effect on perceived image quality. In the *Playmobiles* scene there were slight differences and in the *Bureau* scene there were no differences. As expected, perceived depth increased with increasing camera-base distance. An increase in camera-base distance from 0 cm to 8 cm significantly increased perceived depth. The increase from 8 cm to 12 cm resulted in a smaller increase in perceived depth. Westheimer and McKee [1980] documented a decrease in stereopsis threshold with asymmetric blur that is larger than with symmetric blur. Results of current research showed no difference in perceived depth between symmetric and asymmetric image pairs above stereopsis threshold (B = 8 and 12 cm). As in the image quality figures, an increase in camera-base distance had no effect on the perceived sharpness. This is slightly surprising because introducing disparity information in a stereo image would sharpen the edges by reducing positional uncertainty. The last attribute, eye-strain, also showed an increase when camera-base distance increased. This was also found by IJsselsteijn et al. [2000] where averaged results of the eye-strain ratings showed a clear linear increase with increasing disparities. Also the results of Kooi and Toet [2004] showed that increasing disparity is one of the most important parameters that cause eye fatigue.

Looking towards the future, we want to investigate the image quality percept more precisely and find out what concepts determine the perceived stereoscopic image quality. In other studies it was reported that the added value of 3D, namely the depth dimension, may be related to concepts of naturalness or presence, as has been argued by IJsselsteijn [1998; 2000; 2004]. As well, the research of Yamanoue et al. [2002] conducted at NHK showed that viewers experience a sense of presence (related to naturalness, power) when watching 3D images. In the present study we showed that the added value of depth is not always or consistently taken into account when judging image quality (no increase in image quality when depth increases). This appears to depend rather heavily on the type of image impairments present in the stereoscopic pair. Introducing compression artefacts (2D) or crosstalk artefacts (3D) may change the weighting of perceptual attributes, making stereoscopic image quality predictions more complicated. In the case of crosstalk, the perceptual benefits of increased depth can be neutralized by the perceptual costs of increased crosstalk. First steps to arrive at a more complete understanding of stereoscopic image quality are being taken by the authors. So, for the evaluation of 3D TV other concepts than image quality may be needed.

REFERENCES

BERTHOLD, A. 1997. The influence of blur on the perceived quality and sensation of depth of 2D and stereo images. Tech. rep., ATR Human Information Processing Research Laboratories.

BOSCHMAN, M. 2000. ThurcatD: A tool for analyzing ratings on an ordinal category scale. *Behavior Research Methods, Instruments, and Computers 32*, 379–388.

DODGSON, N. 2004. Variation and extrema of human interpupillary distance. *Proceedings of the SPIE 5291*, 36–46.

FREEMAN, J. AND AVONS, S. 2000. Focus group exploration of presence through advanced broadcast services. *Proceedings of the SPIE 3959*, 530–539.

IJSSELSTEIJN, W. 2004. *Presence in Depth*. Eindhoven University of Technology, The Netherlands.

IJSSELSTEIJN, W., DE RIDDER, H., FREEMAN, J., AND AVONS, S. 2000. Presence: Concept, determinants and measurement. *Proceedings of the SPIE*, 252–255.

IJSSELSTEIJN, W., DE RIDDER, H., HAMBERG, R., BOUWHUIS, D., AND FREEMAN, J. 1998. Perceived depth and the feeling of presence in 3DTV. *Displays 18*, 207–214.

IJSSELSTEIJN, W., DE RIDDER, H., AND VLIEGEN, J. 2000. Effects of stereoscopic filming parameters and display duration on the subjective assessment of eye strain. *Proceedings of the SPIE 3957*, 12–22.

INTERNATIONAL TELECOMMUNICATION UNION. 2000. Methodology for the subjective assessment of the quality of television pictures. *Recommendation BT.500-10*.

JOHANSON, M. 2001. Stereoscopic video transmission over the internet. *Proceedings of the IEEE Workshop on Internet Applications*, 12–19.

JULESZ, B. 1971. *Foundations of Cyclopean Perception*. University of Chicago Press, Chicago, IL.

KOOI, F. AND TOET, A. 2004. Visual comfort of binocular and 3D displays. *Displays 25*, 99–108.

LEVELT, W. 1965. *On Binocular Rivalry*. Royal VanGorcum, Assen, The Netherlands.

MEEGAN, D. V., STELMACH, L. B., AND TAM, W. J. 2001. Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery. *Journal of Experimental Psychology: Applied 7*, 143–153.

MITSUHASHI, T. 1996. Evaluation of stereoscopic picture quality with CFF. *Ergonomics 39*, 1344–1356.

OKOSHI, T. 1980. Three-dimensional displays. *Proceedings of the IEEE 68*, 548–564.

OKUYAMA, F. 1999. Evaluation of stereoscopic display with visual function and interview. *Proceedings of the SPIE 3639*, 28–35.

PASTOOR, S. AND WÖPKING, M. 1997. 3-d displays: A review of current technologies. *Displays 17*, 100–110.

PERKINS, M. G. 1992. Data compression of stereopairs. *IEEE Transactions on Communications 40*, 684–696.

SEXTON, I. AND SURMAN, P. 1999. Stereoscopic and autostereoscopic display systems. *IEEE Signal Processing Magazine*, 85–99.

STELMACH, L., TAM, W. J., MEEGAN, D., AND VINCENT, A. 2000. Stereo image quality: Effects of mixed spatio-temporal resolution. *IEEE Transactions on Circuits and Systems for Video Technology 10*, 188–193.

Stelmach, L. B. and Tam, W. J. 1998. Stereoscopic image coding: effect of disparate image-quality in left- and right-eye views. *Signal Processing: Image Communications 14*, 111–117.

Tam, W., Stelmach, L., and Corriveau, P. 1998. Psychovisual aspects of viewing stereoscopic video sequences. *Proceedings of the SPIE 3295*, 226–235.

Thurstone, L. 1927. A law of comparative judgment. *Psychological Review 34*, 273–286.

Westheimer, G. and McKee, S. 1980. Stereoscopic acuity with defocused and spatially filtered retinal images. *Journal of the Optical Society of America 7*, 772–778.

Yamanoue, H., Ide, S., Okui, M., Okano, F., Bitou, M., and Terashima, N. 2002. Parallax distribution for ease of viewing in stereoscopic hdtv. Tech. rep., NHK Laboratories Note No. 477.

Yano, S. and Yuyama, I. 1991. Stereoscopic HDTV: Experimental system and psychological effects. *Journal of the SMPTE 100*, 14–18.