Livrable D1.2 of the PERSEE project : Perceptual-Modelling-Definition-of-the-Models

Junle Wang, Emilie Bosc, Jing Li, Vincent Ricordel

Projet PERSEE
SCHÉMAS PERCEPTUELS ET CODAGE VIDÉO 2D ET 3D
n° ANR-09-BLAN-0170

Livrable **D1.2** *15/10/2011*

Perceptual Modelling : Definition of the Models

Junle	WANG	IRCCyN
Emilie	BOSC	INSA
Jing	LI	IRCCYN
Vincent	RICORDEL	IRCCyN

Contents

1 Introduction

In this report we describe contributions for the task 1 within the PERSEE project.

The target of this task is to identify perceptual models for 2D and 3D contents, and to improve them. These models will be used by the other tasks of the project.

The document is organized as follows:

- Section 1 is about the representation models for the 3D video coding, and it details how the combination of blur and disparity affects the perceived depth;
- Section 2 presents 3D visual attention models and it details two studies. The first one is about the depth bias. The second is on the temporal effects, namely the effects of motion on visual disconfort.

The description of the experimental tests that have been done for these models, is in the deliverable D6.1 entitled "Perceptual Assessment: Definition of the scenarios".

2 Representation Models for the 3D Video Coding

2.1 Combination of blur and disparity, how it affects the perceived depth

2.1.1 Introduction

Recently, stereoscopic image and video production is gaining an increasing amount of attention. The stereoscopic productions are famous for the 3D viewing they can provide to the viewers. Because the displays nowadays used to show these 3D productions are usually planar, binocular disparity on the display plane is a pre-eminent depth cue that enable viewers to perceive depth.

However, several other depth cues besides binocular disparity affect also the apparent depth, e.g., perspective, blur, occlusion, motion parallax, and so on. Among these depth cues, blur is the one which might be affected in most of the steps of a broadcast chain of video. In video acquisition, blur can be created because of the optical feature of camera's lens; it can be caused by interpolation which comes from the coding and compression; displays can also generate blur to the viewers.

Since blur is a depth cue, the unintentional addition of blur may have unknow influence on perceived depth, which is the most important feature of stereoscopic production. On the other hand, the blur in the image or video can be taken advantage of to improve the efficiency of video coding. For instance, Budagavi et al[3] proposed a blur compensation algorithm that makes used of the blurring information to provide improved compression performance.

Therefore, it is significant to quantify how blur, interacting with disparity, affects the perceived depth in 3DTV. Some previous investigations have shown clear contributions of blur to depth perception[20][25][9][8], while others showed that blur has either no effect[16] or only some qualitative effects on perceived depth ordering[14][15][18].

2.1.2 Disparity and defocus blur

When people fixate an object in a three-dimensional scene, they can perceive defocus blur and binocular disparity simultaneously. Actually, the creations of defocus blur and binocular disparity in the retinal image have the same fundamental geometry. We first consider blur.

An ideal thin lens focuses parallel rays to a point on the opposite side of the lens. The distance between this point and the lens is the focal length, f. Light rays emanating from a distance d_1 in front of the lens will be focused to another point at distance s_1 at the opposite side of the lens. The relationship between these distances derives from the thin-lens equation:

$$\frac{1}{s_1} + \frac{1}{d_1} = \frac{1}{f}$$

When the lens focuses at an object at distance d_0, the image of this object will be formed at the image plane at distance s_0. This object is now in focus, while objects at other distance become out of focus and hence generate blurred images on the image plane (figure 1). We express the amount of blur in the image plane by the blur circle diameter c_1, which can be computed by the equation:

$$c_1 = \left| A \frac{s_0}{d_0} \left(1 - \frac{d_0}{d_1} \right) \right|$$

where A is the diameter of the lens (pupil) and s_0 is the posterior nodal distance (approximately the distance from the pupil to the retina). Note that the human eyes have imperfect optics, and this equation does not incorporate any of the eye's aberrations.

Nevertheless, previous researches showed that this equation can provide an accurate approximation of blur when the eye is defocused[5]. Based on this calculation, blur of the retinal image can be modeled as a two-dimensional Gaussian function. We consider the blur circle radius (c_1) as the radius of decay to exp(-0.5), i.e. the standard deviation sigma of the Gaussian function.

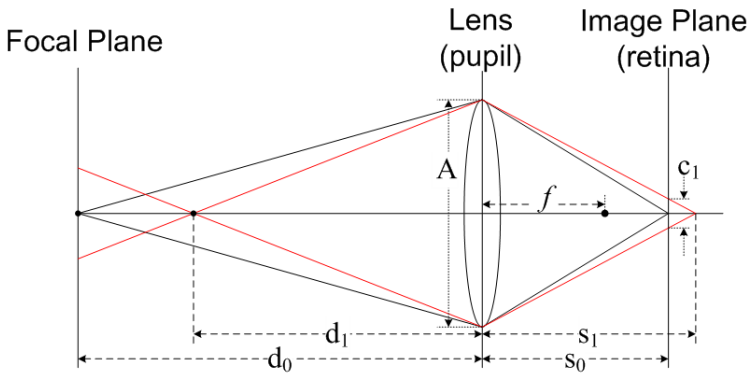


Figure 1: Schematic diagram of the generation of defocus blur.

Thus, given A and S_0, the magnitude of blur varies according to the focus distance and the distance between the object focused and the object defocused . The magnitude of blur increases as the distance from the fixated object increases. When an object is located outside of the depth of field, this object will be projected in the eye with a perceivable magnitude of blur. This relation enables blur as a cue of depth.

Based on the same geometry (figure 2), disparity provides also depth information to the human brain. When the two eyes separated by a distance p converge on an object at distance d_o, another object at distance d_1 creates images with an angular disparity δ which can be obtained by the following equation using small-angle approximation:

$$\delta = \Phi_L - \Phi_R = 2\left[tan^{-1}\left(\frac{P}{2d_1}\right) - tan^{-1}\left(\frac{P}{2d_0}\right)\right] \approx p\left(\frac{1}{d_0} - \frac{1}{d_1}\right)$$

We can consider that disparity is caused by differing two vantage points of two cameras (eyes), while blur is caused by differing the vantage point at two positions of one camera (eye). By converting the equation of blur circle radius into angular units, and rearranging both equations, we can find the relationship between the magnitudes of disparity and blur:

$$c_1 = \frac{A}{p}|\delta|$$

Therefore, the magnitudes of blur and disparity created by objects in a three-dimensional

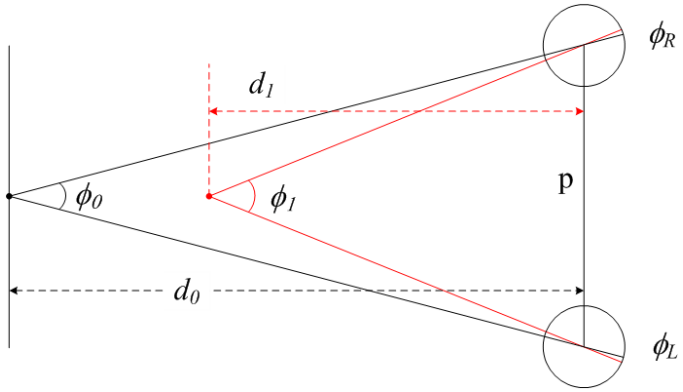


Figure 2: Schematic diagram of the generation of binocular disparity.

scene are proportional to each other. This relationship generally holds because accommodation and vergence are coupled in the real world, namely the eyes accommodate and converge to the same distance. It suggests that the human visual system might use information from both the two factors in correlating the two eyes' images. However, when people watch 3D images or videos displayed on a planar stereoscopic display, there exists a conflict between accommodation and vergence. No matter how the vergence changes, the image is always perceived sharpest when the eyes accommodate on the screen plane. This conflict suggests that a manual creation of blur on the focal plane (the screen) might inhibit the conflict and affect the perceived depth.

2.1.3 Experiment

The subjective experiment is briefly introduced in this section. The detail of this experiment will be presented in D6.1 of the report.

In our study, we conducted the subjective experiment using a state-of-art stereoscopic display system, the Samsung 22.5-inch 1680*1050@120Hz wide-screen LCD monitor working with active shutter glasses from NVidia. The stimuli used in the experiment contained a background plane and a single object in the foreground, both of which were chosen closer to natural content compared to the stimuli used in previous publications. The image of a butterfly was used as the foreground object, since it is spatially complex enough, containing regions with both low and high frequency.

In our experimental observations, a pair of stimuli were shown to the subjects in each trial. One stimulus contained a blurred background (BB-stimulus) and a sharp fore-

ground object, while the other stimulus contained a sharp background (SB-stimulus) and also a sharp foreground object.

Observers viewed stimuli in a two alternative forced choice (2AFC) task, being required to select the stimulus with larger depth interval between foreground and background. Two sources of perceived depth are used: disparity and blur. The perceived depth from disparity stems from the difference of disparity between the foreground object and the background. The perceived depth from blur stems from the amount of blur introduced to the background by convolution with a Gaussian kernel. Both the absolute position and relative distance between the foreground and background stay as a free parameter. This setup is able to evaluate how the combination of disparity and blur affects the perceived depth of objects located at different distance.

2.1.4 Result and analysis

For each condition, some observers considered the BB-stimulus as having a larger depth interval (between the foreground and the background), while the other observers chose the SB-stimulus. We measure the proportion of 'BB-stimulus contains a larger depth interval' responses, and plot the data as a function of the disparity difference between the relative depth in the BB-stimulus (D_{r_BB}) and the relative depth in the SB-stimulus (D_{r_SB}). The cumulative Weibull function was used as the psychometric function. The disparity difference corresponding to the 50% point can be considered as the Point of Subjective Equality (PSE). When measuring the disparity difference at that point, the increase of perceived depth is obtained. In total, by filtering out the data of 7 observers who made decisions in the test quite differently from other observers, 28 observations of each conditions were included in the computation. An example pattern of response and the fitted psychometric function is shown in Figure 3.

According to the setup of the experiment, blur was added to the backgrounds located at different absolute depth positions in different trials, while the depth interval between the background and the foreground object stayed also as a free parameter. We thus plot the curve of the PSE as a function of the background disparity representing the depth of the blurred background in Figure 4. Each of the five points on the curve is obtained by considering all the possible depth intervals ($D_{r_{BB}}$) which are with the same background depth (D_a). The steps are slightly different because the same distances equal to various disparity in angular units depending on the viewing distance.

As we can see in the Figure 4, the PSE curve shows a clear offset from the unit slope, indicating an increase of perceived depth caused by blur in binocular vision. Note that

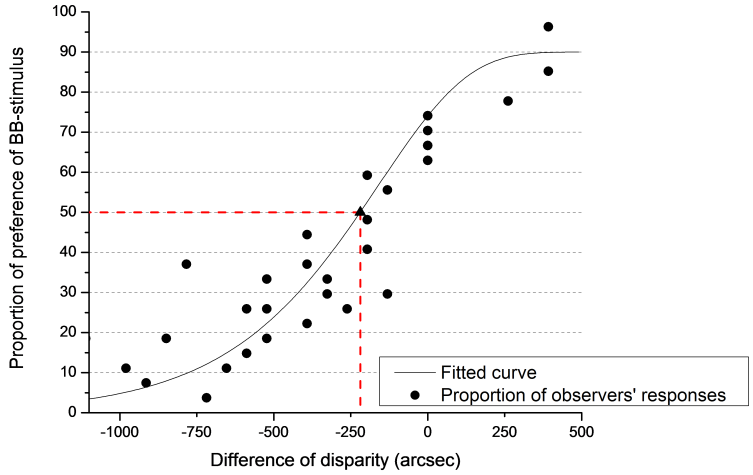


Figure 3: An example pattern of the proportion of observers' responses and the fitted psychometric function. In this trial, we consider D_{r_BB} = 6.6 cm and D_a = -19.7 cm, -13.2 cm, -6.6 cm, 0 cm, 6.6 cm. An equal apparent depth is reached at -220 arcsec

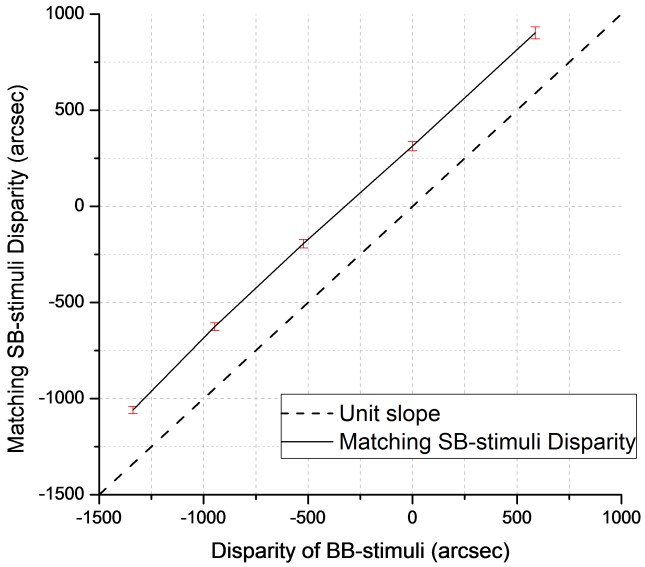


Figure 4: The PSE curve shows subjective matches between SB-stimuli and BB-stimuli. Negative values denote far disparities. The dashed line at unit slope indicates physically correct disparity matches.

the increase is almost constant (approximately 180 arcsec) regardless of the change of absolute depth of the background. This means that the increase of perceived depth caused by blur is insensitive to the disparity of the blurred background, e.g. the absolute position. This phenomenon makes sense because all the blur actually exists only on the screen plane where human eyes accommodate.

We plot also the curve of the PSE as a function of the depth interval (in length unit) between the foreground and background in figure 5. Each of the six points on the curve is obtained by considering all the possible background depth (D_a) which are with the same depth interval ($D_{r_{BB}}$). 0 cm means that the foreground and the background are at the same depth without any difference of disparity between them. According to the viewing geometry, the same depth intervals stand for slightly different disparity, therefore we plot the increase of perceived depth in disparity as a function of depth interval in length unit.

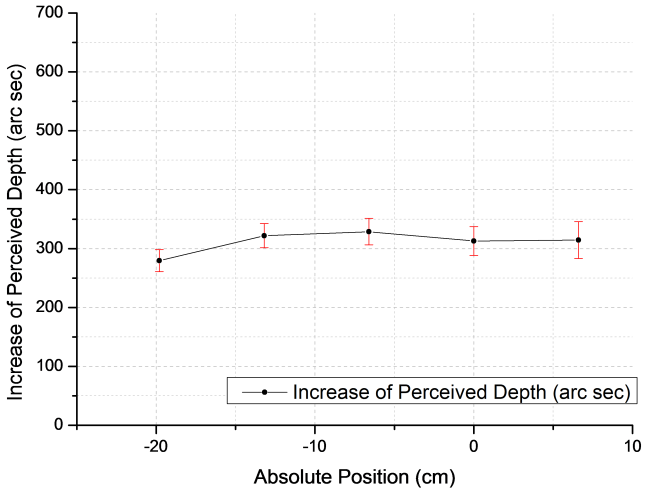


Figure 5: Increase of perceived depth as a function of absolute position.

Figure 6 shows also a clear increase of perceived depth created by the influence of blur. In the figure, we can find that the enhancement of perceived depth increases with increasing depth interval between the fixated sharp foreground object and the defocused blurred background. When the foreground object is close to the background (the relative distance is 0 cm or 6.6 cm), the influence of blur is relatively small. The 0 cm depth interval comes with a slightly larger increase of perceived depth than 6cm depth interval. The reason might be because when the depth interval is 0cm, it seems to the observers that they are looking at a planar image with a sharp foreground and blurred background. For this kind of image, intentional blur has been introduced by photog-

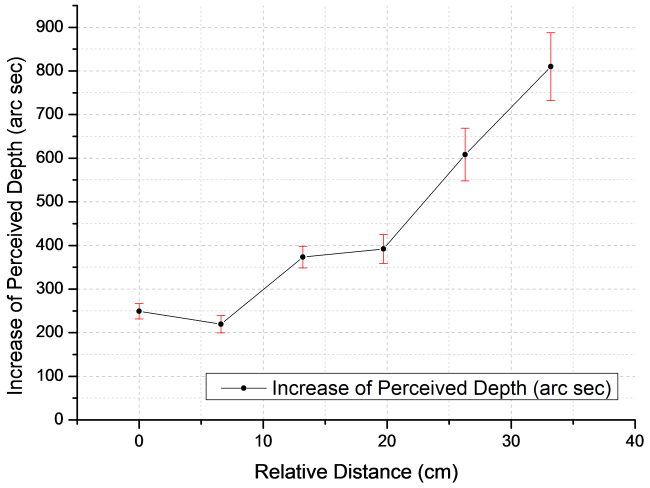


Figure 6: Increase of perceived depth as a function of relative distance between foreground and background.

raphers for a long time to induce some illusions of the existence of depth between the foreground and the background. When disparity differences appear in the image, observers start to notice that they are facing a stereoscopic image. This might be the reason why the 0cm depth interval shows a slightly larger increase of perceived depth than the 6cm depth interval. Starting from the second point (6.6 cm depth interval), the enhancement is monotonic.

As shown by the curves in figure 5 and figure 6, the overall tendency of how the depth interval ($D_{r_{BB}}$) and the absolute position of the background (D_a) affect the perceived depth seems easy to understand. However, we computed the PSEs of every combination of $D_{r_{BB}}$ and D_a, we then found that the variations of PSEs are with some uncertainties. In figure 7, we plot a surface (consisting of blue points) which is obtained from the curve in figure 6, while the red points come from the PSEs which are computed individually by each combination of $D_{r_{BB}}$ and D_a. The figure shows that the changes of perceived depth do not vary strictly according to the tendency we described previously. There exists a lot of variability at certain points. Namely, despite the overall tendency of how the blur affects perceived depth at different relative distance and absolute position is known, the predictions of perceived depth of objects at certain positions or the perceived depth of certain observers remain uncertain. Several types of analysis were performed in order to learn about the source of these aberrations. However, there was no indication found concerning a bi- or multimodal distribution of the observers. So the variance that is found in our experiment concerning the individual data points has to be considered as measurement noise for the moment.

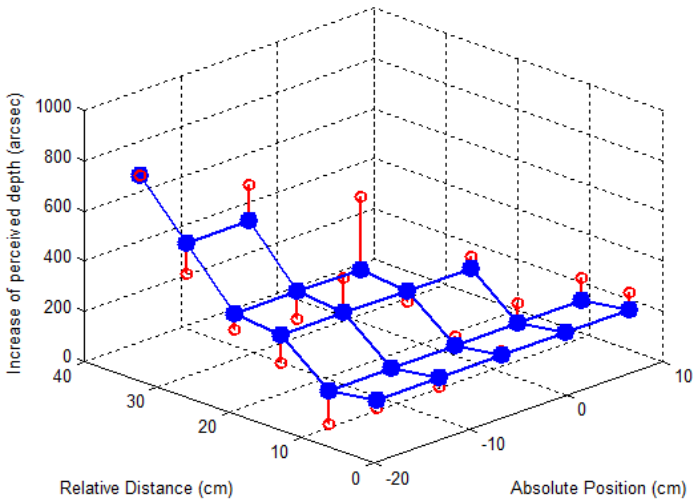


Figure 7: Increase of perceived depth as a function of relative distance between foreground and background.

2.1.5 Conclusions

The influence of defocus blur on the perceived depth in a stereoscopic scene and its relationship with binocular disparity were studied in this work. The experimental result indicates that image blur makes contributions to the impression of depth perceived in stereoscopic images, when measured against depth perceived in stereoscopic images without blur. The increase of depth can be considered as a function of the relative distance between the fixated foreground ground object and the blurred background, while this increase is insensitive to the distance between the viewer and the depth plane at which the blur is added. This phenomenon implies that blur should be well controlled during the every setp of video production chain, otherwise it could affect the depth perception of the final output.

On the other hand, the feasibility of enhancing the perceived depth by reinforcing a monocular cue, namely defocus blur, provides an interesting way to deal with the conflict between accommodation and vergence when 3D images are shown on a planar stereoscopic display. Generally, the foreground object popping out of the screen is the most important object in the scene, while the large disparity of the object may lead to visual discomfort when it is actually fixated by the observer. Our results show that it is feasible to decrease the disparity of this object without losing its pop-out effect by adding some blur on its background.

3 3D Visual Attention Models

3.1 Study of the depth bias

By comparing with the human eye-tracking data, recent studies of visual attention suggest that 2D visual attention computational models can provide a good estimation of where people look at. In these studies, human eye-tracking data shows a so-called "center-bias", which means that fixations of gaze are biased toward the center when observers are looking 2D stimuli. This bias can been taken advantage of by computational models as location prior to improve the performance. Sometimes a simple Gaussian blob locating at the center of the image can be considered as a model with good performance[10][22][24].

Nowadays, the research of 3D visual attention attracts more and more attention. However, in the 3D visual attention, depth is another feature having great influence on guiding eye movements. Relative little is known about the impact of depth. Several studies mentioned that people tend to look at the objects at certain depth planes. Therefore, it is reasonable to suppose the existence of a "depth-bias". But studies proving or quantifying this depth-bias are still limited.

Studies of proving the existence of depth-bias and quantifying this bias is not a trivial task, due to the fact that they can be beneficial for the development of 3D visual attention computational models. We conducted a binocular eye-tracking experiment by showing synthetic stimuli on a stereoscopic display. Observers were required to do a free-viewing through active shutter glasses. Gaze positions of both eyes were recorded for obtaining the depth of fixation. Stimuli were well designed in order to let the center-bias and depth-bias affect eye movements individually. Results showed that the number of fixations varies as a function of depth planes.

3.1.1 Fixation distribution in time

We first evaluated how the depth of fixations varied according to time. The relative depth position of each fixation in the depth range of a scene was computed by

$$D_{r_i} = (D_i - D_{min})/(D_{max} - D_{min})$$

where D_i is the absolute depth of the fixation, D_{min} and D_{max} are the minimum and maximum absolute depth of all the objects respectively. Depths of the first seven

fixations which located on objects were computed.

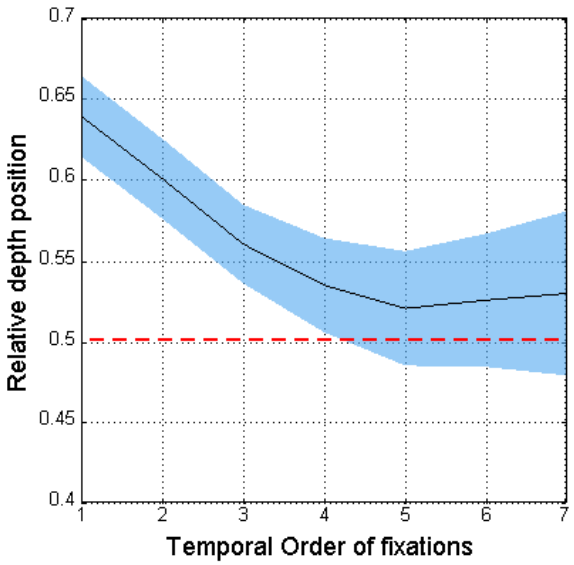


Figure 8: Relative depth position of fixations as a function of temporal order. The blue region is 95% confident interval.

An initial front response upon a new scene was observed in every participant and revealed in Figure 8. In the figure, the red dash line shows the average relative depth of all the objects displayed. If there was no depth-bias, the observers explored the scene equally in depth during the observation, each object in the scene had the same probability to be fixated. That means the average depth of all fixations should had varied little throughout fixations sequence, and stayed around 0.5. However, a one-way ANOVA showed an effect of fixation sequence on depth ($F(6,860) = 7.94$, $p < 0.01$). A post hoc paired t-test with Bonferroni correction showed that the depth of the first and second fixation were significantly higher than that of the following fixations ($p < 0.01$).

The curve of average depth as a function of temporally ordinal fixations shows a viewing strategy that observers tended to explore the scene from the objects closed to them. We can also find that the average depth of all the ordinal fixations is higher than the average depth of objects, which means observers pay more attention to the objects in the front part of the scene than the objects in the back part. An obvious depth-bias was showed.

3.1.2 Fixation distribution in depth

Besides the temporal analysis, another way to quantify depth-bias is to evaluate how the fixations distributed along the depth in every scene. For each scene, we considered the order of each object in depth instead of the absolute depth value of the object. We considered the object which was closest to the observer as the first object, the further one as the second object, and so on. The furthest object was considered as the N^{th} object, where N is the number of objects in the scene. All the 118 conditions were separated into 5 groups based on the number of objects contained in the scene.

Numbers of fixations located on each object were calculated for each observer and each scene, then they were transformed into a probability distribution function. The percentage values were draw as a function of the order of fixation's depth across participants and scenes. We considered the value $P_r = 1/N$ as the reference value for each type of scene, which was derived based on the assumption that each object attracts the same amount of fixations if there was no effect of depth-bias on the distribution of fixations.

Figure 9 shows how the first fixation of each observation distributed in the scene. As we can see in the figures, regardless the number of objects contained in the scene, the closest object to the observer attracted always most fixations (more than 30% of the total amount, about 2 times of the reference value). The percentage of fixations then decreases as the depth order increases in the front half part of the scene. The curves generally follow a very similar progression for all five conditions. The objects locating in a certain front range (about 30%) of the scene attracts more fixations than the reference value, while the others attract less.

Num. of objects	ANOVA result
5	$F(4,130) = 11.73$, $p < 0.05$
6	$F(5,156) = 12.42$, $p < 0.05$
7	$F(6,182) = 13.22$, $p < 0.05$
8	$F(7,208) = 13.8$, $p < 0.05$
9	$F(8,234) = 11.85$, $p < 0.05$

Table 1: The ANOVA result of the analysis of the fixation distributions in figure 9.

A one-way ANOVA was performed to check the significant difference among the values. The results (presented in table 1) confirmed that there existed an effect of fixation's depth order on fixation distribution (table X). A post hoc paired t-test with Bonferroni correction was then performed to check the significant difference between each pair of ordinal fixations in depth. For all the conditions, the percentage of fixations of the

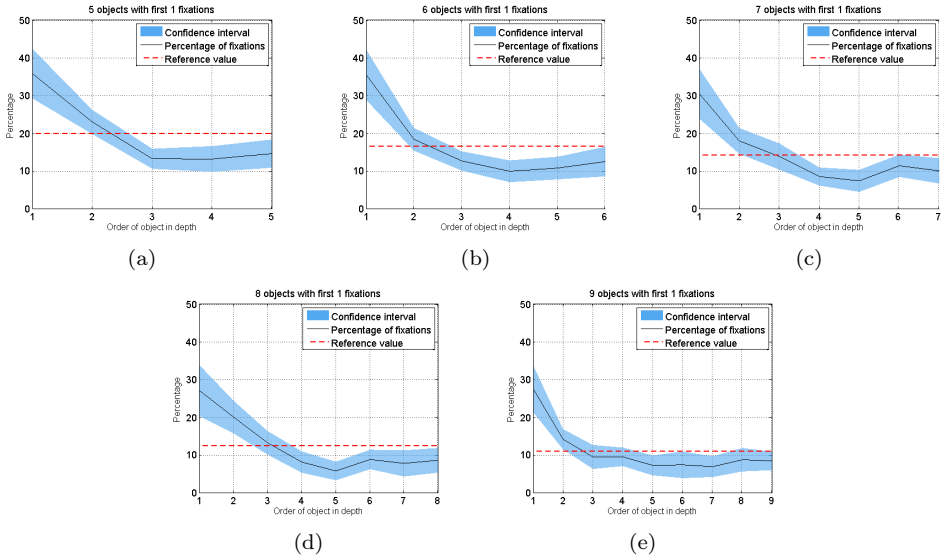


Figure 9: Fixation distribution (only the first fixation was considered) as a function of the order of object's depth for the scenes containing different number of objects ($N \in 5, 6, 7, 8, 9$). X axis is the order of objects; sub-figure (a) to (e) repesents the group of scenes that contain 5 to 9 objects, respectively. Y axis represents the percentage of number of fixations. The blue area represents the 95% confidence interval. The dash line represents the reference value of the percentage, which is obtained by 1/N

first object was significantly higher than the others, while the fixation percentage from the third to Nth ordinal fixations were not significantly different.

The curves in figure 10 shows the distribution when all the fixations during the whole observation were considered. We can find that the curves are flatter, but revealing a strong similarity of tendency when they are compared with the correspondent ones in the first column.

This phenomenon shows that, the depth bias still exists even with a longer observation time. The stronger depth-bias of the first fixation implies that depth-bias could be one of the bottom-up mechanisms. Shape of the curves also show that some visual attention models in the literature which considered the saliency distribution along the depth as a monotonic function[4] or step function were not appropriate enough.

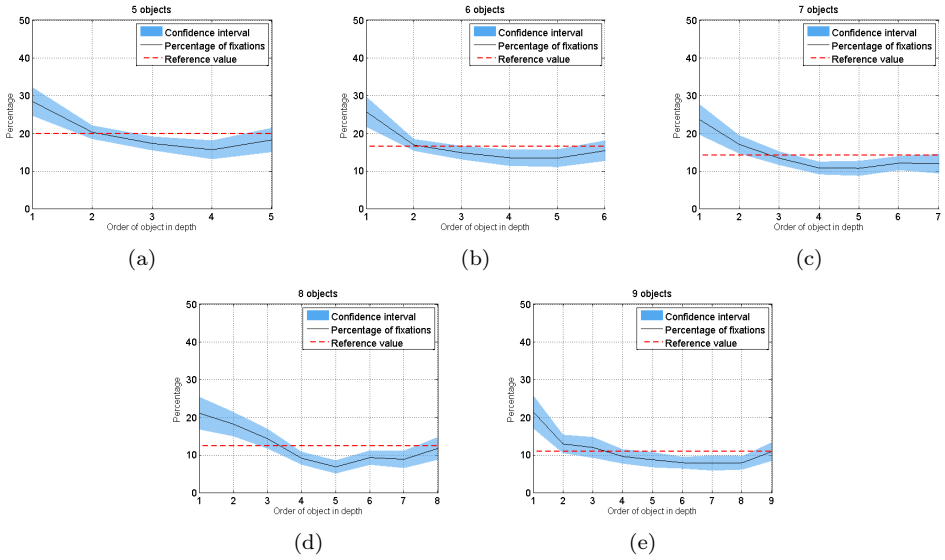


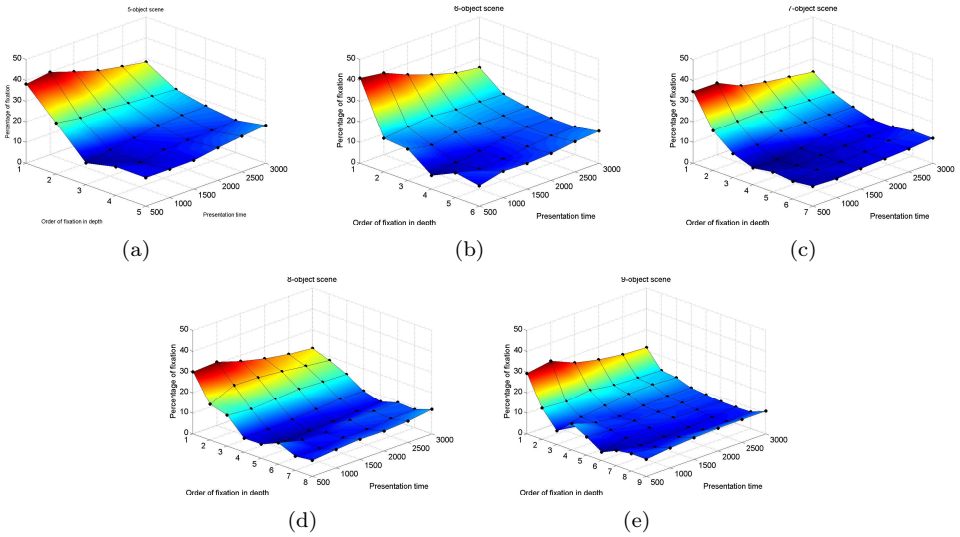
Figure 10: Fixation distribution (all the fixations were considered for all the conditions) as a function of the order of object's depth for the scenes containing different number of objects ($N \in 5, 6, 7, 8, 9$). The blue area represents the 95% confidence interval. The dash line represents the reference value of the percentage, which is obtained by 1/N

3.1.3 Convergent behavior

The analysis in the previous sections reveal also a variation of the level of depth-bias when different number of temporal ordinal fixations were considered. This variation implies that the level of depth-bias varies according to time. One might question that if the depth-bias still existes after ceratin observation time, and what is the final and stable situation of depth-bias.

In the condition of 2D content viewing, the distribution of fixations varies according to the presentation time, and becomes more stable for a increased presentation time [6]. In figure 8, we found that the average depth of fixation does not change significantly from the forth fixation. This increased stability suggests that the depth-bias of fixation distribution becomes more stable with longer presentation times.

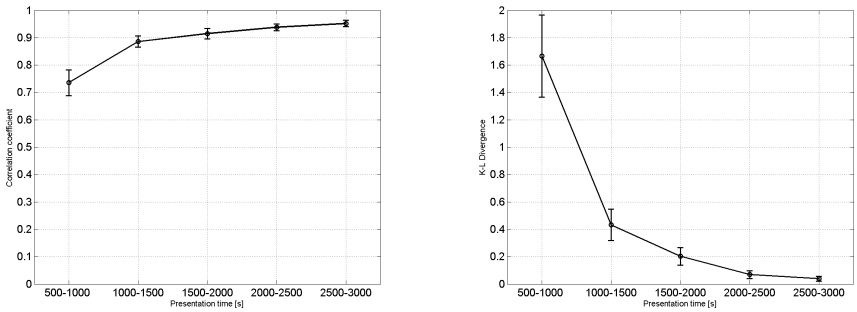
To verify this convergent behavior, we computed the fixation distributions $FD^{(t)}$ created from different presentation times t ($t \in 500, 1000, 1500, 2000, 2500, 3000ms$). All these FD for different types of conditions are shown in figure 11 as surfaces. We then



(a) (b) (c) (d) (e)

Figure 11:

analysis the Pearson linear correlation coefficient (PLCC) and Kullback-Leibler divergence (KLD) between FD created from two consecutive presentation time for all the conditions. These PLCC and KLD values are presented as a function of presentation time in figure 12.



(a) (b)

Figure 12:

For presentation time larger than t = 2.5 s, the average PLCC are above $\rho = 0.95$, and the average KLD are below $D = 0.05$. These two averages over all conditions illustrates that there is indeed a strong convergent behavior of the FD with observation

time. One can argue that after a presentation time of approximately 2.5 seconds, the FD along depth do not change drastically anymore.

This knowledge may aid in reducing experimental time, and show that the 3-second observation time in our experiment is long enough for evaluating the depth-bias. It also shows that depth-bias will not disapear as the observation time increases.

3.1.4 Conclusion and discussion

The experiment data illustrates the existence of depth-bias. Firstly, the variation of average depth of temporal ordinal fixations exhibit a viewing strategy that peoples prefer start the observation of a 3D scene from the objects close to them. Fixation distribution was biased to the objects locating in a certain front range (about 20%). The curves show that, the methods in the literature that considered the saliency distribution along the depth as a monotonic function or step function should be improved. Secondly, strong depth-bias of the first fixation implies that depth-bias might be a bottom-up process. Thirdly, the convergent behavior of fixation distribution shows that the 3-second observation time was long enough for the evaluation of depth-bias.

In this study, we considered only the relative depth. However, it has been shown the existence of disparity-selective neurons in Area V4. And we also know that the conflict between accommodation and vergence affects the viewing behavior of stereoscopic content. These evidences show that the absolute depth value (i.e. disparity) might affect the fixation distribution.

From the curves of fixation distribution, we can also find a similarity among the conditions that contained different number of objects. This similarity implies the possibility of a curve fitting for the development of a 3D visual attention model. A straight forward way is to analysis the scene by a 2D computational model, then combine the depth information as the complementary information.

3.2 Temporal effects

Motion in stereoscopic images can be classified into in-depth motion and planar motion. A lot of researches on the effects of motion component on visual discomfort and visual fatigue have been conducted. In 2002, Yano et al. found that a local minimum of visual comfort appeared for both high degree of parallax and large amount of motion [28]. Then in 2004, Yano et al. [27] pointed out that the visual fatigue occurred when the stereoscopic images involved in an in-depth motion component even if they were displayed within the range of depth of field. In 2006, Speranza et al. [21] concluded that motion in depth, i.e., the magnitude of binocular disparity varying over time, could play an important role in visual discomfort, and it might be more important in determining visual discomfort than the absolute magnitude of the binocular disparity. All of these researches showed that with the increase of motion velocity, the possibility that the in-depth motion component may induce visual discomfort would increase. Thus, it could be concluded that visual discomfort is negatively affected by the frequency of change of disparity.

The effects of planar motion on visual discomfort are learned less than the in-depth motion components. In fact, there are some conflicts about the effects of planar motion on visual discomfort. In [27], the authors concluded that no visual fatigue was found in lateral motion images. However, this conclusion was given under the condition that the velocity of lateral motion was varied over time. In [11], the authors showed that for motion component in vertical, horizontal and depth direction, visual discomfort may occur when the velocity exceeds the threshold, which means the planar motion also may induce visual discomfort. To make clear relationship between the visual discomfort and planar motion, we decided to design an experiment specific for this relationship. In our recent study [13][12], we found that planar motion indeed had influences on visual discomfort. With the increase of planar motion velocity, the observers are more likely to experience visual discomfort.

The following part will give a detailed introduction about our research on this topic.

3.2.1 Experimental setup

To avoid the influence of other factors on visual discomfort, we used computer-generated stereoscopic sequences for precise control. The stereoscopic sequences consisted of a left-view and a right-view image which were generated by the MATLAB psychtoolbox [2] [19]. There were 3 velocity levels and 5 angular disparity levels for the moving object. The velocity levels are 71.8, 179.5 and 287.2 degree/s and the disparity levels are 0, ± 0.65, and ± 1.3 degree. There were 15 stimuli for the experiment. An example of the stimuli is shown in Fig.13, in which the foreground object is placed in front of the screen with an angular disparity of 1.3 degree.

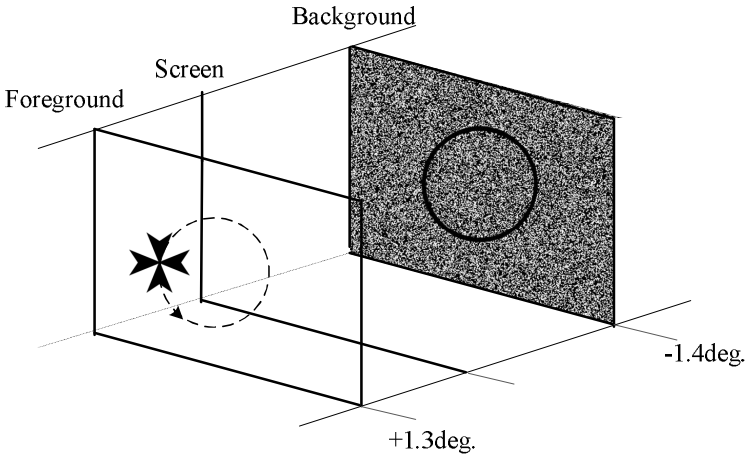


Figure 13: An example of a stereoscopic image in the experiment. The foreground object is moving at the depth plane with a disparity of 1.3 degree. The background is placed at the depth plane with a disparity of -1.4 degree. The motion direction of the Maltese cross is anti-clockwise.

In our study, the paired-comparison method was used. 10 experts observers and 45 naive observers participated in the experiment. They were asked to select the stimulus which was more uncomfortable for them in each trial.

3.2.2 Experimental results

The Thurstone-Mosteller [23] [17] model and Bradley-Terry model [1] are two widely used methods to analyze the paired comparison subjective experiment data, which can convert the paired comparison data to a scalable score for each stimulus. The Thurstone model was used to analyze the raw data owing to its origins in psychophysics. Bradley-Terry model is more developed mathematically [7]. It provides maximum-likelihood estimators for scales, confidence intervals and hypothesis tests for model fit, uniformity and differences among populations of judges. In this section we only check the consistency of the Bradley-Terry scores of all stimuli with the global and individual subjective experiment results. The program used in this study for Bradley-Terry model is available in [26]. In our experiment, both of the methods were used due to the different aims of the analysis. Bradley-Terry model was used to generate the visual discomfort scores for each stimulus, and the statistic analysis on the data. The Thurstone-Mosteller model was used to develop the objective visual discomfort models.

(1) Validation analysis on experts raw data

Before the data analysis, we use experts-only data to analyze if the presentation order has effects on the final results and if the observers gave random responses in the experiment. According to the experimental design of 3 velocity levels and 5 disparity levels, there were totally 15 stimuli. Each stimulus was compared with every other stimulus, and for each pair two trials were carried out because the sequence of the occurrence also matters. Therefore, there were 15×14=210 trials for the whole experts-only experiment. 210 comparison results and 15 stimuli which were not compared to themselves could be arranged in a 15×15 matrix with the diagonal value being 0.5. A three-dimensional matrix M(i,m,n) is used to express each individual subjective experimental results, as shown in Fig.14. The row represents the stimulus of the first presentation, and the column represents the stimulus of the second presentation. For the i_{th} individual subjective experiment, M(i,m,n) represents the number of times stimulus m is selected over stimulus n. For example, in the i_{th} individual subjective experiment, the pair of stimuli was stimuli m and n, and stimulus m is presented first. If the observer choose stimulus m as more uncomfortable, then, M(i,m,n)=1, otherwise M(i,m,n)=0. For each individual subjective experiment, the matrix M is binary without considering the diagonal elements. Before any further data analysis about the relationship of visual discomfort to binocular angular disparity and velocity, we checked the validity of the raw data regarding the following aspects:

1. Screening of the observers. To remove the data that stemmed from the observers who were inclined to give random answers in the experiment.

2. Verification of each pair condition. There might be a pair of stimuli in which most of the viewers always chose the one presented first. The influence of the presentation orders on visual discomfort, and the interaction effects of the pair of stimuli on visual discomfort can be verified in this process.

Ideally, for all pairs of stimuli, the observer's answers should not depend on the presentation orders. Thus, it can be expressed by the equation M(i,m,n)+M(i,n,m)=1. A statistical method was used to verify it. The Student's-t-Test was performed on each individual subjective experimental results, which correspond to the elements of the upper and lower triangular matrices of M for each subjective experiment, and expressed as M_{iu} and M_{il} respectively, as shown in Fig.15(1) . The question thus changed to if M_{iu} and 1- M_{il} were obtained from a Gaussian process with a common mean value. When the hypothesis was verified by the Student's-t-Test, it can be concluded that the observer gave consistent answers in this individual subjective experiment. Otherwise, the data from this individual experiment might not be valid and should be rejected. The whole process that utilizes Student's-t-Test will be referred to as "Consistency test" in this paper.

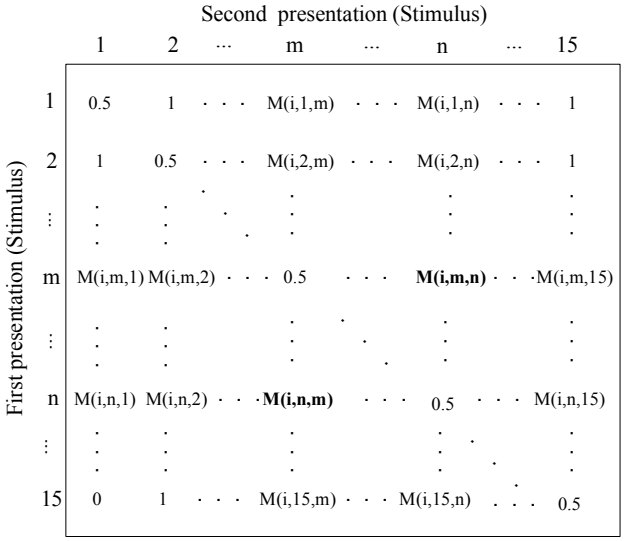


Figure 14: An example of the matrix M for the i_{th} individual subjective experimental results. The row represents the stimulus of the first presentation, and the column represents the stimulus of the second presentation. M(i,m,n) represents the number of times stimulus m is selected over stimulus n in the i_{th} individual subjective experiment and can thus be either 0 or 1.

Similarly, the correctness of each pair condition would be checked by the "Consistency test". It was performed on S_{mn} and 1- S_{nm}, where S_{mn} was a binary vector and expressed as [M(1,m,n), M(2,m,n),..., M(t,m,n)] assuming there were totally t individual subjective experiments, as shown in Fig.15(2). It represented all of the observers' selections one by one for the trial that stimulus m and n were compared, and stimulus m was presented first. In our experiment, the Student's-t-Test was performed at 5% significance level. Both, the screening of observers and the verification of each pair condition passed the "Consistency test".

According to the analysis above, we assume that the naive observers test results also obeys this conclusion. Besides, from the naive observer experiment results we found that it's possible that there are different classes of observers who give different opinions about the visual discomfort induced by the stimuli. Thus, we decided not to screen the observers before the data analysis.

(2) Analysis of experts and non-experts data on Bradley-Terry model

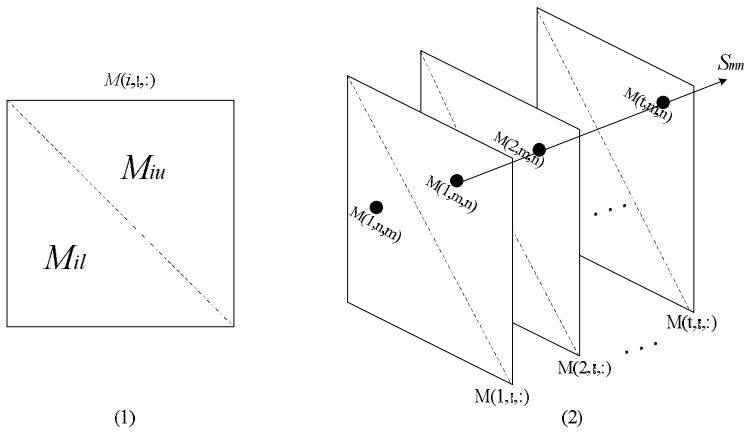


Figure 15: The matrices used in the raw data validation test. (1) The diagram of the upper and lower triangular matrix used for screening of the observers. (2) The diagram of the vector used for the verification of each pair condition.

As there were a foreground object and a background in the stimulus, the relative disparity between the foreground object and the background was used to analyze their effects on visual discomfort. The binocular angular disparity of the background was -1.4 degree, thus the 5 relative angular disparity levels of the foreground object were 0.1, 0.75, 1.4, 2.05, 2.7 degree. The Bradley-Terry scores for visual discomfort from experts and non-experts data are shown in Fig.16. Both the experts and non-experts Bradley-Terry scores for 15 stimuli give the same conclusion as what we have found in the previous experts-only study by utilizing Thurstone model. It shows that the visual discomfort increases with the relative angular disparity rather than the absolute angular disparity of the object. The influence of the vergence-accommodation conflict seems to be quite small under this experimental setup. It might be explained by the existence of the background and the moving foreground. There would be two vergence points in the stimulus for the viewers. When watching the stimulus, the viewers' attention may switch between the two objects. The larger of the depth distances between the visual attention points, the larger the abrupt change of the amount of vergence-accommodation mismatch when switch from one object to another, which might be seen as a reason that induces the visual discomfort.

It was clearly indicated that the perceived visual discomfort increases with velocity. This conclusion is in accordance with our previous study. And it is also consistent with the results that Lee et al. gave in [11] recently although the planar motion directions are different. They showed that the visual discomfort increased with the velocity of horizontal and vertical motion while a circular motion was used in our test.

Table 2: The significant difference test on experts data

Number	Stimuli Pairs(relative disparity, velocity)
1	(2.7, 71.8) & (2.05, 179.5)
2	(2.7, 71.8) & (1.4, 287.2)
3	(2.7, 71.8) & (0.75, 287.2)
4	(2.7, 179.5) & (2.05, 287.2)
5	(2.05, 71.8) & (0.1, 287.2)
6	(0.75, 71.8) & (0.1, 179.5)
7	(2.05, 179.5) & (0.75, 287.2)

In a practical application of our study, it may be concluded that for stereoscopic motion images, the depth budget for fast motion sequences should be significantly reduced and for slow motion sequences, the depth budget may be increased.

(3) Significant difference test on Bradley-Terry scores

From the Fig.16 we could find that some stimuli have a similar effect on visual discomfort. To find what kind of stimuli may have similar effects on visual discomfort, the statistic analysis is conducted on Bradley-Terry estimators. In order to compare these Bradley-Terry scores and test if they are significantly different, the confidence intervals for the comparison of the Bradley-Terry scores Vi, Vj of Stimulus i and j are calculated. If the value zero is located within the confidence interval, it can be concluded that the Bradley-Terry scores Vi and Vj are not significantly different. The stimuli which are not significantly different (significance level = 5 %) are shown in Table 2 and Table 3, where the stimulus pairs were expressed by '(relative disparity, velocity)'.

The results indicated that a stimulus which has large relative disparity and slow velocity might have a similar effect on visual discomfort compared to a stimulus which has small relative disparity but fast velocity.

(4) Evaluation of the Bradley-Terry model scores

For better illustration, some definitions of the matrix which will be used in the consistency test are given.B_{B-T}(m,n) is introduced to represent if the Bradley-Terry score of stimulus m is higher than that of stimulus n. Thus, B_{B-T} is a binary matrix without considering the diagonal elements. To compare the matrix B_{B-T} with the global subjective experiment results, a binary matrix which represents the global sub-

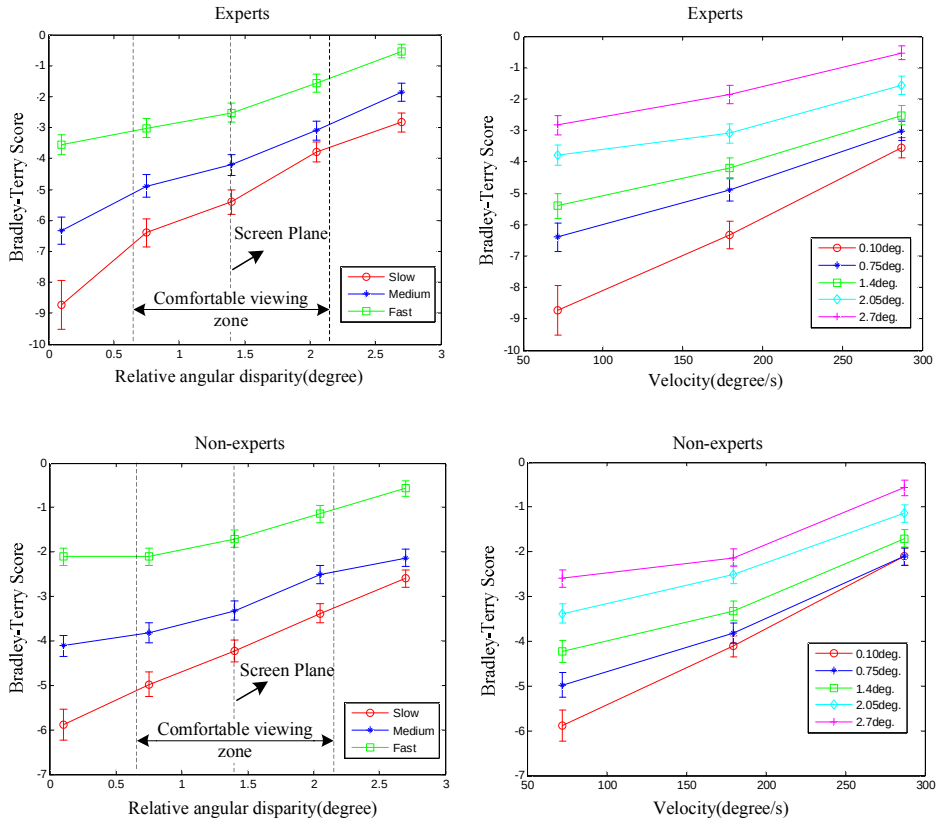


Figure 16: Bradley-Terry scores for visual discomfort. The top two figures are experts results. The bottom two figures are non-experts results. The different lines in the left figures represent the different velocity levels, where slow, medium and fast represent 71.8, 179.5 and 287.2 degree/s. The outer two dashed lines represent the upper and lower limits of the comfortable viewing zone, which are 0.66 and 2.14 degree. The dashed line in the middle represents the position of screen plane. The different lines in the right figures represent the different relative angular disparity levels.The error bars are the 95% confidence interval.

Table 3: The significant difference test on non-experts data

Number	Stimuli Pairs(relative disparity, velocity)
1	(2.7, 71.8) & (2.05, 179.5)
2	(2.7, 179.5) & (0.75, 287.2)
3	(2.7, 179.5) & (0.1, 287.2)
4	(2.05, 71.8) & (1.4, 179.5)
5	(1.4, 71.8) & (0.1, 179.5)
6	(0.75, 287.2) & (0.1, 287.2)

Table 4: The consistency and agreement test results

Observers	Consistency Test	Global Agreement	Individual Agreement	
			Mean	Std.
Experts	0(pass)	0.9619	0.7917	0.0832
Non-experts	0(pass)	0.9714	0.8142	0.0909

jective experiment results is needed and can be generated in the following way. Firstly, the probability matrix P is calculated where P(m,n) represents the probability that the stimulus m is selected over stimulus n. Then, the values in P which are below the threshold 0.5 are set to 0 and above the threshold to 1. This binary matrix is expressed by B_{obs}. Two evaluation methods were used to check the agreement of the Bradley-Terry scores with the subjective experiment results. The first one is the "Consistency test", which means using Student's-t-Test to check if B_{B-T}(m,n) and B_{obs}(m,n) with m<n were obtained from a Gaussian process with a common mean value. In our experiment, the Student's-t-Test was performed at 5% significance level. Secondly, an "Agreement test" was conducted both on the global and individual subjective experiment results, which means calculating the proportion that the value in each position of B_{B-T} was the same with the corresponding value in B_{obs} and M matrix of each observer.

The "Consistency test" and the "Agreement test" results for both experts and non-experts are shown in Table 4, each observer's agreement on the Bradley-Terry scores are shown in Fig.17. Generally speaking, the results indicated that the Bradley-Terry scores fit well with the subjective experiment results. However, it's easy to find that for some observers, their "Agreement test" results were lower, which means their opinions differ from the global observers' opinion. Based on this analysis, it's necessary to cluster the observers as several classes in which they have the similar opinions. This will be investigated in the section 3.2.3.

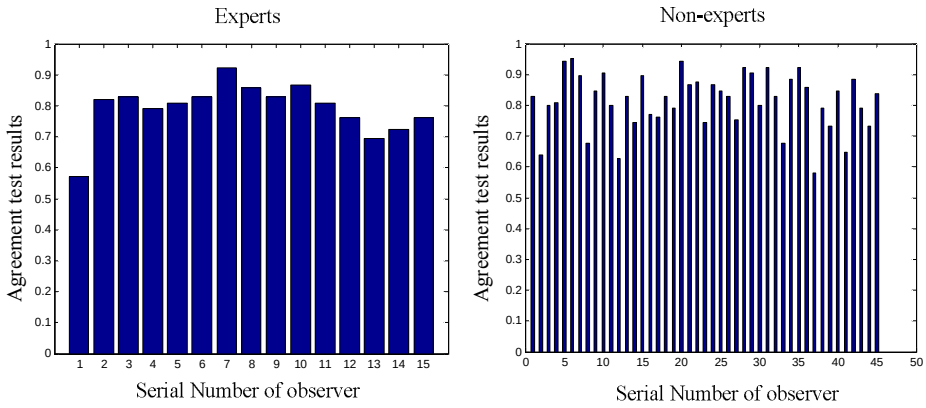


Figure 17: The individual “Agreement test" results for both experts and non-experts data with the Bradley-Terry scores.

3.2.3 Classification of observers

In section 3.2.2 it was already stated that there were some observers who had different opinions from the global subjective experiment results. Thus, it may be interesting to classify them as different groups and analyze the different influences of relative disparity and velocity on different observers. The relative disparity and velocity are two factors that may induce the visual discomfort in our study. Thus, the analysis of which factor is more dominant in determining the visual discomfort is conducted on each observer. There are two hypotheses in this analysis. One is “the relative disparity is predominant" and the other is “the velocity is predominant". Then, the proportion of each observer voting for the stimulus whose relative disparity is larger than the other one is calculated for hypothesis one, expressed as p1. And the proportion of voting for the stimulus whose velocity is faster than the other one is calculated is calculated for hypothesis two, expressed as p2. Each observer's opinion on these two hypotheses can be reflected by (p1, p2) which can be expressed by a point in a two-dimensional space. According to these points, the observers can be classified as different groups. In our study, the K-means clustering method was used. The clustering results are shown in Fig.18. The Bradley-Terry scores for all stimulus generated by each observer cluster are shown in Fig.19 and Fig.20, for experts and non-experts respectively.

There were two classes in the experts observers. 12 observers belonged to the first class, which gave a similar result as the global subjective results. The second class contained 3 observers, which gave the opinion that relative disparity was the predominant factor that had effect on visual discomfort. It's interesting to find that for small relative disparity, viewers perceived more visual discomfort when velocity was faster. However,

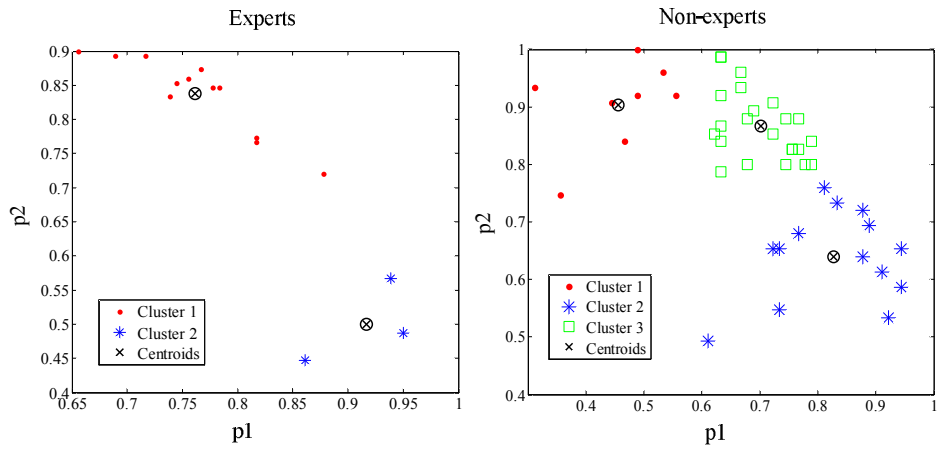


Figure 18: The clustering results for experts and non-experts observers. X-axis represents the agreement on "relative disparity is the predominant factor" and y-axis represents the agreement on "velocity is the predominant factor".

when the relative disparity changed largely, viewers felt more uncomfortable when velocity was slower. The experts who voted in this way gave the reasons that when the object's relative disparity was large but velocity was slow, it became difficult to fuse the foreground and the background at the same time, thus they would alternate the vergence between the two objects which made them more uncomfortable. But for fast velocity, they would not care about the other object as it appeared blurred due to its fast relative motion, which consequently reduced their visual discomfort.

There were 3 classes for non-expert observers. The first class with 8 observers considered that the velocity was the predominant factor in determining the visual discomfort. In fact, there often was no significant difference in visual discomfort between two stimuli with the same velocity but different relative disparities. For the second class with 14 observers, they gave the opinion that the relative disparity and fast velocity were more important in inducing visual discomfort. Slow and medium velocities had similar effects on visual discomfort. The third class with 23 observers gave the same opinion as characterized by the global results.

3.2.4 Objective visual discomfort model

(1) Construction of the objective visual discomfort model

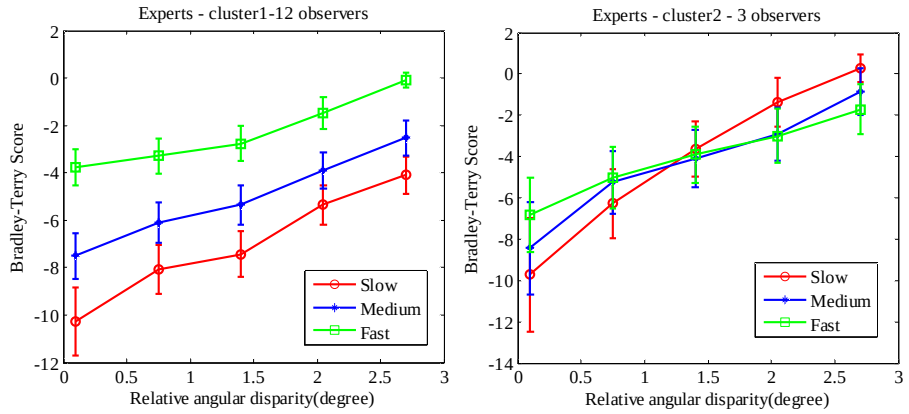


Figure 19: The Bradley-Terry scores of visual discomfort for experts clusters. The left figure is for cluster 1, and the right figure is for cluster 2.

From the Bradley-Terry score it could be found that the visual discomfort was affected by the relative disparity and velocity of the object in the stimuli. To better identify their relationship, a visual discomfort model should be constructed. As the relationship of visual discomfort to relative disparity and velocity was nearly linear, the simplest model would be the one in Equation 1, which will be noted as Model 1,where Q represents visual discomfort,a_1,a_2 and a_3 are coefficients, v is the velocity and d is the relative angular disparity.

$$Q = a_1 \cdot v + a_2 \cdot d + a_3 \tag{1}$$

However, it might be possible that the relative disparity and velocity have an interaction effect on visual discomfort. For each velocity level, the relationship between visual discomfort and relative disparity was nearly linear. Thus, the visual discomfort can be expressed as Equation2.

In our experiment, a linear function seems to be the most appropriate fitting method. So, f_1 and f_2 are modeled as linear functions of velocity. Finally, the visual discomfort can be described by Equation 3, where b_1,b_2,b_3 and b_4 are all constant and we define this as Model 2.

As the Bradley-Terry score is influenced by the stimulus number, and the score cannot be normalized by linear way, in this part, we use normalized Thurstone-Mosteller score to generate the objective visual discomfort model. A curve fitting was performed on the normalized Thurstone-Mosteller scores of experts-only subjective experimental data,

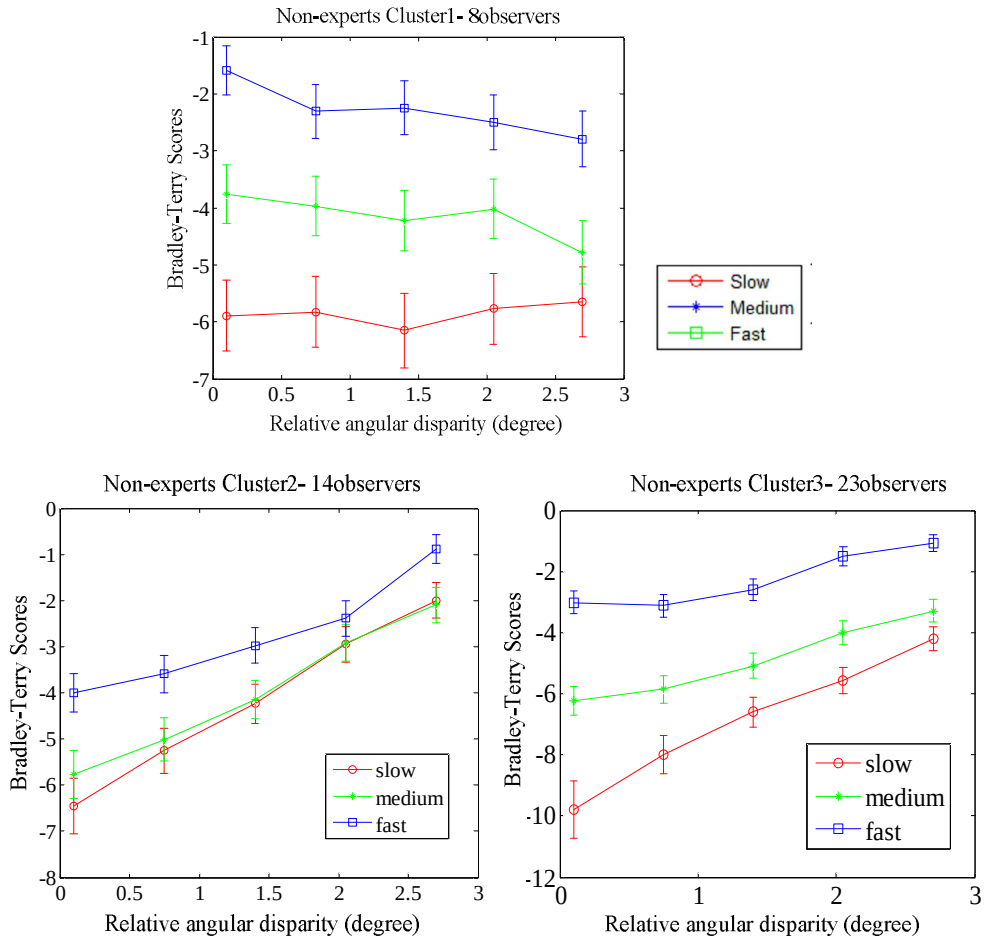


Figure 20: The Bradley-Terry scores of visual discomfort for non-expert clusters. The figures from left to right represent the cluster 1, 2 and 3.

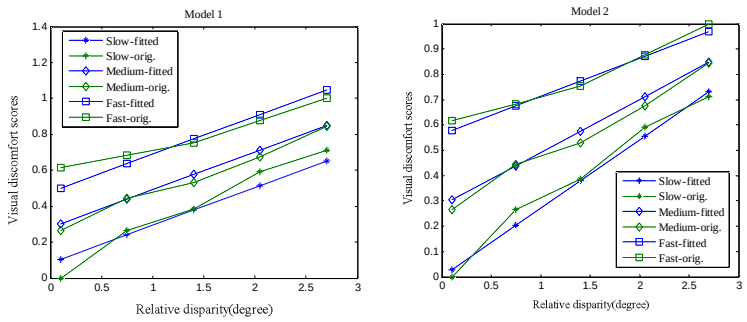


Figure 21: The predicted visual discomfort of Model 1 and 2 in function of disparity and velocity. Solid lines represent fitted scores. Dashed lines represent original normalized Thurstone-Mosteller scores.

the predicted coefficients for the two models were 0.0018,0.2102, -0.0477 for a_1,a_2 and a_3, and 0.3110, 0.0026, -0.0006, -0.1888 forb_1,b_2,b_3 and b_4, respectively. Fig.21 show the regression results of the two models for visual discomfort. The solid lines show the predicted visual discomfort scores for different relative disparity and velocity, the dashed lines denote the normalized Thurstone-Mosteller scores in this experiment.

$$Q = f_1(v) \cdot d + f_2(v) \tag{2}$$

$$Q = b_1 \cdot d + b_2 \cdot v + b_3 \cdot d \cdot v + b_4 \tag{3}$$

(2) Evaluation of the objective visual discomfort model

In this study, both of the objective models will be evaluated by comparing the predicted scores with the Bradley-Terry scores. Three metrics are used as the evaluation criterions: (1) Pearson linear correlation coefficient (CC), which provides an evaluation of prediction accuracy. (2) Spearman rank-order correlation coefficient (ROCC), which is considered as a measure of the prediction monotonicity. (3) Root mean squared error (RMSE), which reflects the validation of prediction. The evaluation results are shown in Table 5.

It can be concluded that the predicted visual discomfort scores from both of the models correlate quite well with the Bradley-Terry visual discomfort scores. Model 2 performs

Table 5: The performance of Model 1 and 2 on the experts and non-experts Bradley-Terry scores

Model	Experts			Non-experts		
	CC	ROCC	RMSE	CC	ROCC	RMSE
1	0.9876	0.9750	0.0406	0.9489	0.9000	0.0788
2	0.9949	0.9929	0.0257	0.9697	0.9286	0.0605

better than Model 1; however, when considering the complexity, Model 1 is slightly preferable for use as an index for the stereoscopic image related researches.

3.3 Conclusion

In this study, we investigated the effects of relative disparity and planar motion velocity on visual discomfort. The Bradley-Terry model was applied both on experts and non-experts subjective experiment data. The Bradley-Terry scores showed high agreement with our previous study. That is, the relative disparity between the foreground and background in the stimulus might be more significant in determining the visual discomfort than the binocular disparity of the foreground. Planar motion with faster velocity may result in more visual discomfort.

To quantify the effects of relative angular disparity and velocity on visual discomfort, two visual discomfort models were constructed. We also evaluated the objective visual discomfort models by the subjective data, the results showed that our models correlate quite well with the subjective perception.

As there were some observers who didn't agree with the global subjective experiment results, we classified these observers as different clusters according to which factor is predominant in determining their feeling of visual discomfort. The clustering results showed that most of the observers agreed with the global subjective experiment results. However, there were indeed some observers who considered either the relative disparity or velocity as the predominant factor in inducing visual discomfort while the other factor has small influence on their feelings.

In future work, some other factors which might also have influence on visual discomfort and consequently on visual attention will be studied.

References

[1] R.A. Bradley. 14 paired comparisons: Some basic procedures and examples. *Handbook of statistics*, 4:299–326, 1984.

[2] D.H. Brainard. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1997.

[3] M. Budagavi. Video compression using blur compensation. pages II–882. IEEE, 2005.

[4] C. Chamaret, S. Godeffroy, P. Lopez, and O. Le Meur. Adaptive 3D rendering based on region-of-interest. In *Proceedings of SPIE*, volume 7524, page 75240V, 2010.

[5] H. Cheng, J.K. Barnett, A.S. Vilupuru, J.D. Marsack, S. Kasthurirangan, R.A. Applegate, and A. Roorda. A population study on changes in wave aberrations with accomodation. *Journal of Vision*, 4(4), 2004.

[6] U. Engelke, H. Liu, H.J. Zepernick, I. Heynderickx, and A. Maeder. Comparing two Eye-Tracking databases: The effect of experimental setup and image presentation time on the creation of saliency maps. In *International Picture Coding Symposium*, 2011.

[7] John C. Handley. Comparative analysis of bradley-terry and thurstone-mosteller paired comparison models for image quality assessment. In *Proc. IS&Tís Image Processing, Image Quality, Image Capture, Systems Conference*, pages 108–112, 1081.

[8] R.T. Held, E.A. Cooper, J.F. O'Brien, and M.S. Banks. Using blur to affect perceived distance and size. *ACM Transactions on Graphics*, 29(2):1–16, 2010.

[9] D.M. Hoffman and M.S. Banks. Focus information is used to interpret binocular images. *Journal of vision*, 10(5), 2010.

[10] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 802–817, 2006.

[11] S. Lee, Y.J. Jung, H. Sohn, Y.M. Ro, and H.W. Park. Visual discomfort induced by fast salient object motion in stereoscopic video. In *Proceedings of SPIE*, volume 7863, 2011.

[12] J. Li, M. Barkowsky, and P. Le Callet. The influence of relative disparity and planar motion velocity on visual discomfort of stereoscopic videos. In *The third International Workshop on Quality of Multimedia Experience (QoMEX2011)*, pages 155–160, 2011.

[13] J. Li, M. Barkowsky, J. Wang, and P. Le Callet. Study on visual discomfort induced by stimulus movement at fixed depth on stereoscopic displays using shutter glasses. In *17th International Conference on Digital Signal Processing (DSP)*, pages 1–8, 2011.

[14] J.A. Marshall, C.A. Burbeck, D. Ariely, J.P. Rolland, and K.E. Martin. Occlusion edge blur: a cue to relative visual depth. *JOSA A*, 13(4):681–688, 1996.

[15] G. Mather. Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1367):169, 1996.

[16] G Mather and D R Smith. Depth cue integration: stereopsis and image blur. *Vision Research*, 40(25):3501–3506, 2000. PMID: 11115677.

[17] F. Mosteller. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9, 1951.

[18] S.E. Palmer and J.L. Brooks. Edge-region grouping in figure-ground organization and depth perception. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6):1353, 2008.

[19] D.G. Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4):437–442, 1997.

[20] Alex Paul Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4):523–531, July 1987.

[21] F. Speranza, W.J. Tam, R. Renaud, and N. Hur. Effect of disparity and motion on visual comfort of stereoscopic images. In *Proceedings of SPIE*, volume 6055, 2006.

[22] B.W. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 2007.

[23] L.L. Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.

[24] P.H. Tseng, R. Carmi, I.G.M. Cameron, D.P. Munoz, and L. Itti. Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of vision*, 9(7), 2009.

[25] Simon J. Watt, Kurt Akeley, Marc O. Ernst, and Martin S. Banks. Focus cues affect perceived depth. *Journal of Vision*, 5(10), December 2005.

[26] F. Wickelmaier and C. Schmid. A matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods*, 36(1):29–40, 2004.

[27] S. Yano, M. Emoto, and T. Mitsuhashi. Two factors in visual fatigue caused by stereoscopic hdtv images. *Displays*, 25(4):141–150, 2004.

[28] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites. A study of visual fatigue and visual comfort for 3d hdtv/hdtv images. *Displays*, 23(4):191–201, 2002.