Vis Comput (2015) 31:675–687
DOI 10.1007/s00371-014-0994-6

ORIGINAL ARTICLE

A response time model for abrupt changes in binocular disparity

Tai-Jiang Mu · Jia-Jia Sun · Ralph R. Martin · Shi-Min Hu

Published online: 19 June 2014

Abstract We propose a novel depth perception model to determine the time taken by the human visual system (HVS) to adapt to an abrupt change in stereoscopic disparity, such as can occur in a scene cut. A series of carefully designed perceptual experiments on successive disparity contrast were used to build our model. Factors such as disparity, changes in disparity, and the spatial frequency of luminance contrast were taken into account. We further give a computational method to predict the response time during scene cuts in stereoscopic cinematography, which has been validated in user studies. We also consider various applications of our model.

Keywords Stereoscopy · Perception · Response time · Visual comfort

1 Introduction

The goal of the stereoscopic film industry is to add realism to the screen by mimicking a sense of depth. Unlike in natural viewing, accommodation is separated from vergence—it is restricted to the display screen in stereoscopic three-dimensional (S3D) viewing [21,25]. This conflict [20] can be characterized by the binocular disparity, the difference of the object's retinal locations in left and right eyes [43]. Although a smoothly varying disparity range [33] providing stereo video can be achieved by carefully controlling the camera configuration [18,42], perceived depth jumps are inevitable and frequently occur between shots or during scene cuts in 3D film-making [33,38].

A depth jump in S3D movies is much more elaborate than in the real world for two reasons. First, the field of view in the real world is very large and the viewer controls where it is. Therefore he/she can make predictions about what will be seen next and thereby anticipate a change from far to near fixation. This is not easy with S3D movies where a scene cut may be completely unpredictable from the viewer's standpoint. Second, the real world maintains consistency between the stimulus for vergence and the stimulus for accommodation. So when the viewer changes fixation from a far object to a near one, vergence and accommodation remain consistent with one another. However, it is not so with S3D movies.

Vergence and accommodation are cross-coupled [47,48] in the human visual system (HVS) to maintain a clear single vision at the fixation. Vergence eye movements are driven by retinal disparity and accommodation is driven by blur. But vergence is also driven by accommodation (so-called accommodative vergence, AV) and accommodation is also driven by vergence (so-called vergence accommodation, VA). The cross-coupling has different temporal properties than the direct paths for vergence and accommodation control [47,48] and vergence would be faster when the accommodative stimulus changes in unison with it [9]. In S3D movies, the visual system has to fight against the cross-coupled responses to maintain accommodation at the screen and vergence on the stereo content that may be in front of or behind the screen. And that effort has been reported to be an important cause of visual discomfort [20]. Schor and Kotulak [49] have shown that cross-coupling pathways (AV and VA) are high-pass, i.e., they do not respond to slow changes in the vergence and accommodation stimuli. So, there is a strong likelihood

T.-J. Mu · J.-J. Sun · S.-M. Hu (✉)
Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
e-mail: shimin@tsinghua.edu.cn

R. R. Martin
School of Computer Science and Informatics, Cardiff University, 5 The Parade, Roath, Cardiff, Wales, CF24 3AA, UK
e-mail: ralph@cs.cf.ac.uk

Fig. 1 A scene cut (, The Smurfs 2 © 2013 Sony Pictures Animation Inc., left and middle), with disparity maps of the left view inset, and corresponding per-pixel response time map (right); brighter red indicates longer response time. We propose a novel response time model, based on perceptual experiments, to predict the time taken to adapt to the new binocular disparity during a scene cut

of visual fatigue when abrupt disparity changes occur frequently [15].

When adapting to a step change in disparity, the HVS takes time to accomplish stereo fusion starting from the previous incorrect vergence response [45,52], which in turn affects the 3D viewing experience [33]. However, little work has been conducted to investigate how the HVS responds to an abrupt change in binocular disparity.

In this paper, we explore the temporal performance of the HVS when an abrupt change in binocular disparity occurs. Binocular disparity is used to simulate the conflict between vergence and accommodation in S3D viewing. Human perception results have been increasingly important for computer graphics, such as computing, processing, and displaying [11,12,14,23,44,45]. We measure the time the HVS takes to fuse the new disparity, and hence obtain a clear vision of the new scene, through a series of *successive contrast* [21, Chapters 21.1] perceptual experiments. We take into account disparity, both crossed and uncrossed, as well as the disparity change, in both forward and backward directions. The influence of luminance contrast spatial frequency on the response time is also considered in our experiments, as it has previously been shown to have an impact on disparity sensitivity and discrimination [12,19,29]. We have statistically analyzed the data obtained and fitted a bilinear model which quantifies response time. Our results help to understand how the HVS adapts between different scene conditions, especially with respect to binocular disparity, and we provide a computational model to predict response time during a real scene cut (see Fig. 1). This model can be used to optimize the response time during a scene cut in S3D movies and used as a guideline when shooting or planning 3D film, or when arranging scene changes in post-production.

Contributions In summary, our work makes the following contributions:

(i) This is the first work to conduct systematic perceptual experiments on the temporal aspects of how the HVS adapts to abrupt changes in binocular disparities.

(ii) We show that the response time taken by the HVS to clearly fuse the new binocular disparity is significantly influenced by the change in disparity, no matter whether it is forward or backward, as well as the level of the new crossed or uncrossed disparity; it is also modulated by the spatial frequency of luminance contrast.

(iii) Following statistical analysis, we have built a bilinear model for response time, which takes into account the above factors as well as the new disparity and the change of disparity.

(iv) We provide a computational model to evaluate the expected response time during a scene cut and suggest several applications that could directly benefit from our model.

Limitations Though our experimental results and model reveal some temporal aspects of how the HVS responds to new S3D scenes, the mechanisms underlying how the HVS deals with S3D information are far from being elucidated. Further neurophysiological evidence is needed to provide understanding of the mechanisms underpinning how the HVS responds to disparity contrast [21, Chapters 21]. Additionally, we have only considered some of the parameters which could be relevant, and others could also have a significant impact, such as color contrast, luminance contrast, etc. Finally, our current model is an empirical one based on observations, and more sophisticated models based on an understanding of the brain would be preferable.

2 Related work

Various principles [33,38] have been proposed to reduce viewing discomfort in S3D cinematography [27]. Besides excessive disparity, discontinuous change in disparity is also a key contributor to visual fatigue when viewing S3D [15]. Repeated depth jumps force viewers to adjust vergence between shots frequently, which can inhibit the sense of

stereoscopic immersion [33], even when all depths lie in the zone of comfort [4,50].

There exists works on how disparity affects time to fusion. Akeley et al. [1] and Hoffman et al. [20] showed that the differences between vergence and focal distances affect the ability to fuse stereo stimuli—the stimulus can be fused quickly if the two distances change together, while more time is required if not.

Neri et al. [40] revealed that absolute and relative disparities are processed stereoscopically in different areas of the human visual cortex. Liu et al. [33] noticed that crossed and uncrossed depth jumps are not equal in effects, and the former are more problematic for some viewers in terms of the resting states of vergence and accommodation [50]. Yang et al. [54] conducted an identification task to discern differences and changes in object depths. Participants were asked to determine which of four circles was closest under various disparity conditions. Results showed that depth differences and motion in depth are more accurately and quickly discerned with a smaller baseline of crossed disparity for background objects and larger separations between objects.

A change in disparity results in convergence or divergence movements of the eye. Rashbass et al. [46] showed that eyes respond to an abruptly imposed disparity with an eye movement speed roughly proportional to the amount of abruptly imposed disparity. Erkelens et al. [16] found eye vergence tracking to be more accurate and less noisy when changes in object disparity are combined with changes in object size, compared to changing either alone, even though responses to size-change and disparity-change are independent. Later, vergence shifts between real objects, which provided all necessary natural cues for depth perception, were determined to be fast and accurate [17].

Most previous works on binocular disparity change either focus on disparity sensitivity or disparity discrimination threshold [11,12], or the eye movements induced by disparity change. Our work differs in that it provides a series of comprehensive and systematic perceptual experiments to study a more fundamental aspect of the HVS: how quickly the HVS adapts to changes in binocular disparity.

3 Methodology

We now present in detail our perceptual experiments used to measure the observers' response time to fuse a new disparity, after an abrupt change to an attentively perceived scene.

Parameters The first key factor to take into account is clearly the disparity d, as we wish to determine the response time the HVS takes to fuse the new disparity. Another factor of interest is the spatial frequency f of luminance contrast. The impact of luminance contrast spatial frequency on disparity discrimination has already been explored [19,29], and it has been exploited to develop a disparity discrimination threshold model [12], as well as a visual comfort metric for stereoscopic motion [14]. Though adaptation to luminance contrast is a well known phenomenon in human vision (dark adaptation and light adaptation), the influence of the frequency of luminance contrast on human vision, especially stereo adaptation, remains to be investigated.

As our experiments are dependent on changing quantities (see Sect. 3.1), two further derived parameters were also considered: the change in disparity Δd and the change in frequency δf. We measured d in terms of angular disparity (in degrees [14, Figure 2]) and f (measured in *cpd*) was the number of repeated luminance contrast patterns per viewing angle (in degrees).

Stimuli All stimuli were frontal-parallel disparity sinusoidal corrugations with luminance noise of various frequencies. Left and right views of the stereo stimulus were generated using image warping for the luminance pattern [12]. The mean disparity of the corrugation was set to d and the amplitude for all stimuli was fixed at 0.1°. We sampled the two explicit dimensions of our parameter space as follows:

- $d = \{-1.6°, -0.8°, 0°, 0.8°, 1.6°\}$;
- $f = \{1cpd, 4cpd, 16cpd\}$.

Taking all combinations produced a total of 15 different stimuli. Additionally, we produced four different corrugation orientations $\boldsymbol{o} = \{0°, 45°, 90°, 135°\}$ for each stimulus (see Sect. 3.1). The corrugation's spatial frequency was set to $0.3cpd$, which is within the peak sensitivity range of the HVS [11]. We set the maximum disparity as large as possible, from $-1.6°$, a crossed disparity corresponding to a position approximately 82 mm in front of the screen, to 1.6°, an uncrossed disparity indicating a position approximately 125 mm behind the screen. Stimuli exceeding this interval are difficult to fuse for some observers. Due to the angular disparity, it is convenient to extend to different viewing configurations, such as scenes in cinema or virtual reality. We set the sample step to be 0.8°, to restrict the number of stimuli, as well as being large enough to be well discerned between different disparities. The luminance contrast spatial frequency was sampled following Du et al. [14]. Example stimuli are shown in Fig. 2.

3.1 Procedure

We used a 23 inch interleaved *AOC* 3D display with a resolution of 1,920×1,080 pixels, viewed with passive polarized glasses. All subjects were seated 48 cm in front of the screen, with chins fixed so that the subjects' eyes looked at the center of the screen. We assumed a standard interpupillary distance

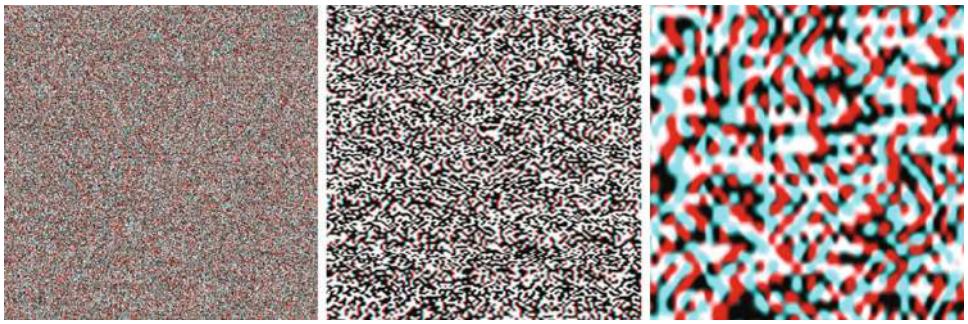


Fig. 2 Sampled stimuli, shown as *red-cyan* anaglyphs, with different combinations of corrugation mean disparity, corrugation orientation and spatial frequency of luminance contrast. From left to right: (0.8°, 135°, 16 *cpd*), (0.0°, 0°, 4 *cpd*), (−0.8°, 90°, 1 *cpd*)

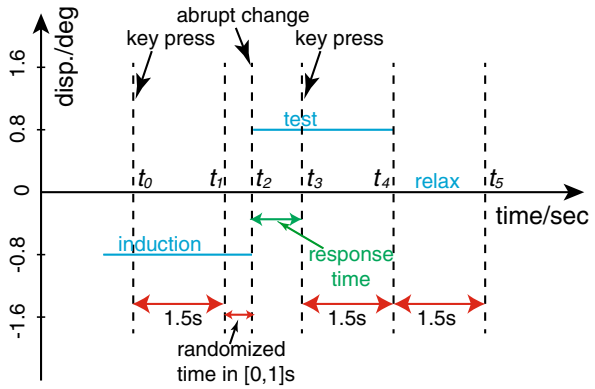


Fig. 3 A timeline for a typical trial considering disparity (ordinate). The subject presses a key when the corrugation's orientation is clearly noticed for the induction stimulus and test stimulus at t_0 and t_3, respectively, followed by a period of stereoscopic immersion until t_2 and t_4, respectively. The stimulus is abruptly changed to the test stimulus at t_2. The interval between t_2 and t_3 is the response time

of 65 mm. 15 subjects participated in our experiments (2 female, 13 male), all visually normal or corrected-to-normal, and having no difficulty in stereoscopic fusion. All participants were non-experts in stereoscopic research and were between 20 and 30 years of age, recruited from the campus.

Since the scenes before and after a cut are inherently successive in time, our experiments were conducted using a *successive contrast* approach [21, Chapters 21.1], unlike the disparity discrimination threshold experiments of Didyk et al. [11,12] which applied a *simultaneous contrast* approach [21, Chapters 21.1]. Specifically, for a successive contrast experiment, a first stimulus, the *induction* stimulus, is presented to the subject for a while, and then changed to the second stimulus, the *test* stimulus, with a different value of the feature being tested; the two stimuli are presented at the *same* time in a simultaneous contrast experiment.

The 15 stereo stimuli cases resulted in 210 trials, as we considered disparity changes in both directions i.e. both forward and backward jumps. Each trial used a pair of stimuli (induction, test). The timeline of a typical trial is illustrated in Fig. 3; the induction stimulus and test stimulus were presented sequentially. During the presentation of each stimulus, subjects were told to concentrate on fusing the corrugation's orientation, and to press a key as soon as they could clearly see the corrugation's orientation. After the key press at t_0, the induction stimulus, with orientation randomly picked from {45°, 135°}, was kept for 1.5 s plus a random additional time of up to 1 s to avoid predictive responses, until t_1. During this time subjects were told to continuously stare at the stimulus without loss of concentration. This was followed by an abrupt change at t_2 to the test stimulus, randomly selected from {0°, 90°}, lasting another 1.5 s until t_4 after the key press at t_3. The program recorded the interval between the abrupt change at t_2 and the key press at t_3, as the response time.

To encourage participants to participate diligently, warnings were given when an advanced key press (i.e. a second key press ahead of the abrupt change) or a late key press (i.e. a second key press after too long, saying 2 s) occurred, and such data were rejected as outliers. To avoid any learning effect as the experiment proceeded, a number of *irrelevant* white-to-black trials were randomly inserted. The subjects were asked to keep their fingers on the key to avoid unnecessary physical movement and to press the key with a steady speed during the whole experiment. To avoid adaptation to repeated disparity changes [15,52], the pairs of (induction, test) stimuli were randomly arranged for each subject. A single trial lasted between 3.5 and 5.5 s. Subjects took a 1.5 s break between consecutive trials during which a gray image was shown to relax the eye muscles and a longer, 120 s break was taken every 21 trials. It took a single subject 50–60 min to complete a whole batch of tests.

Prior to the actual trials, subjects were presented with the stereo stimuli and ten randomly picked trials to familiarize themselves with the experiment. From the measured response time, we also subtracted the keyboard response time and general human reaction time, measured using a red-to-green reaction test for each subject prior to the experiment.

4 Data analysis

We now present the statistical analysis of the collected data, discuss how the factors considered influence the measured response time, and present our model of time taken by the HVS to adapt to different stereoscopic content.

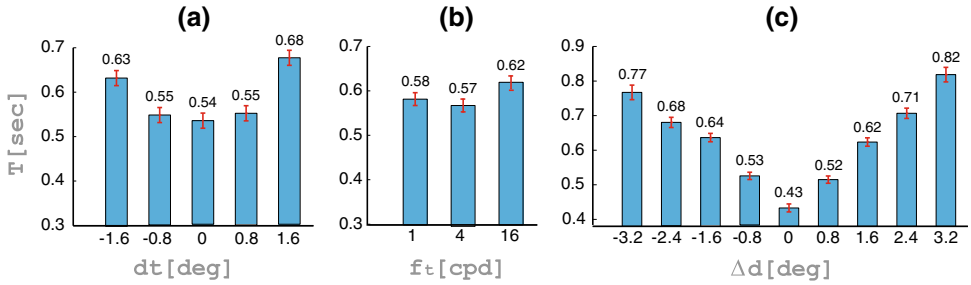


Fig. 4 Estimated marginal mean response time for five levels of target disparity (**a**), three levels of target luminance contrast spatial frequency (**b**) and nine levels of change of disparity (**c**). *Error bars* indicate the standard errors of the means: non-overlapping error bars indicate a significant difference under different conditions

The entire set of trials contained $210 \times 15 = 3{,}150$ records of response time, 35 of which were rejected as outliers (see Sect. 3.1). Each trial gives a parameter sample point $s_j = (d_i, f_i; d_t, f_t)$, where the subscript i and t indicate the *initial* content (induction stimulus) and *target* content (test stimulus), respectively. The response time T_j for each sample point s_j was averaged across all 15 participants, yielding 210 averages. To verify that the 210 average response times $\{T_j|_{j=1}^{210}\}$ were a suitable basis for further data manipulation, we performed a one-way analysis of variance (one-way ANOVA) on this data. It yielded an F value $F(209, 2{,}940) = 2.19$, which is larger than the F test critical value ($F_{\text{crit}} = 1.17$) for $p = .05$. This means inter point (intra participant) variances are larger than intra points (inter participant) variances, and so it is meaningful to use average response times.

To explore the impact of the level of depth jump on the response time T, $\Delta d = d_t - d_i$ and d_t were substituted for d_i. Similarly, $\delta f = f_t/f_i$ and f_t were substituted for f_i. Motivated by the work of Didyk et al. [11], the luminance contrast spatial frequency was manipulated in the logarithmic domain.

4.1 Results

Multi-variant ANOVA was used to evaluate the main effects of all factors considered in our experiments, as well as interaction effects, based on the 210 average response times.

Main effect

d_i and ***f_i*** ANOVA for the initial disparity d_i and spatial frequency f_i of luminance contrast yields an F value $F(4, 205) = 1.41$ with $p = .230$ and $F(2, 207) = .26$ with $p = .774$, respectively. This means neither the disparity nor the luminance contrast spatial frequency in the initial content significantly influences the response time T.

d_t and ***f_t*** The response time taken to fuse the target scene, on the other hand, *is* significantly influenced by the magnitude of target disparity d_t ($F(4, 205) = 13.39$, $p < .001$) and the level of target spatial frequency f_t of luminance contrast ($F(2, 207) = 3.47$, $p = .033$). Figure 4a gives the estimated marginal mean response time taken to clearly fuse the test stimuli for 5 different d_t. Figure 4b illustrates the estimated marginal mean response time to clearly fuse the test stimuli for 3 levels of f_t. Generally speaking, the response time significantly increases with increasing magnitude of target disparity d_t, whether crossed or uncrossed (see Fig. 4a). The mean response times ($\pm$ standard error) for target frequencies $1cpd$, $4cpd$ and $16cpd$ were 581 ± 14, 567 ± 14, and 619 ± 14 ms, respectively, with a *minimum* obtained around $4cpd$ (see Fig. 4b). This minimum agrees with previous findings on disparity discrimination thresholds [12] and effects of stereoscopic motion on visual comfort [14]; a shorter response time results in a more discriminatory and comfortable viewing experience. A further observation is that the mean response times for different f_t do not differ as much as for different d_t, as shown by the overlapping error bars in Fig. 4b.

Δd and ***δf*** A change in disparity Δd also significantly influences the response time ($F(8, 201) = 72.98$, $p < .001$). The change of luminance contrast spatial frequency δf, however, does not significantly affect the response time ($F(4, 205) = 1.90$, $p = .111$). Figure 4c indicates the mean response time for different magnitudes of Δd, regardless of d_t, f_t, and δf. They clearly show that the response time increases with increasing magnitude of delta disparity, no matter whether depth jumps are crossed or uncrossed. There is also a significant response time difference for different crossed or uncrossed Δd (as shown by non-overlapping error bars in Fig. 4c).

Interaction effect

As the response time is not significantly influenced by d_i, f_i, or δf, we only considered the pair-wise interaction impacts of Δd, d_t, and f_t on the response time.

Δd and ***d_t*** There is a strong interaction between Δd and d_t that influences the response time ($F(24, 185) = 35.78$, $p < .001$). The mean response times for different levels of target disparity with different delta disparities are presented in Fig. 5. Given a target disparity, whether crossed or uncrossed, the response time taken to clearly fuse images increases with magnitude of disparity change Δd for both crossed and uncrossed depth jumps, which is consistent with the findings

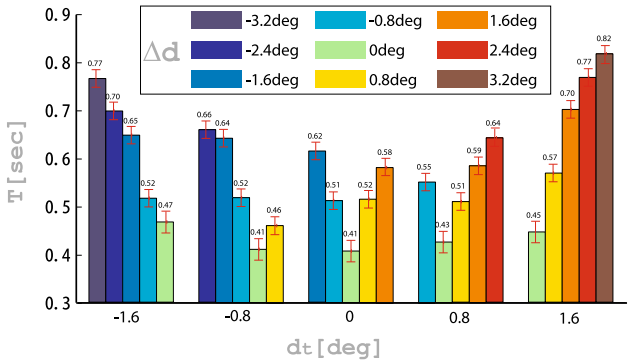


Fig. 5 Mean response time for clearly fusing the test stimuli for differing levels of target disparities and different delta disparities

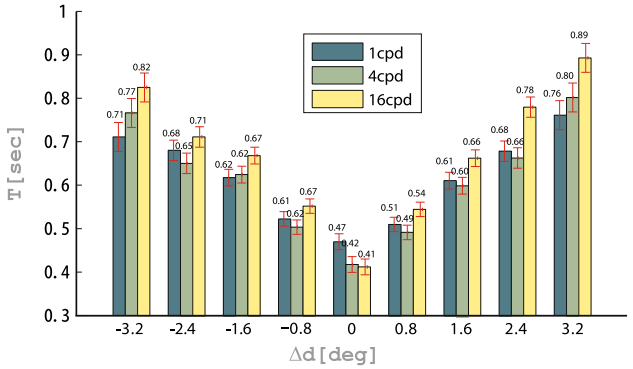


Fig. 6 Mean response time for clearly fusing the test stimuli for tested levels of delta disparities and target luminance contrast spatial frequencies

for the main effect of Δd. Moreover, the response time is greater for larger magnitudes of Δd and d_t. There is still a significant difference in response time with different delta disparities for all target disparities but $-0.8°$.

$\boldsymbol{\Delta d}$ and $\boldsymbol{f_t}$ The interaction between Δd and f_t also significantly affects the response time ($F(26, 183) = 29.16$, $p < .001$). Figure 6 illustrates the interaction effect of delta disparity and target luminance spatial frequency on the response time. The response time is significantly larger for a target luminance contrast spatial frequency of $16cpd$ than for $4cpd$, for all uncrossed delta disparities. However, the same is not true for all crossed delta disparities. Furthermore, no significant difference is found for luminance contrast spatial frequencies of $1cpd$ and $4cpd$ for either uncrossed or crossed depth jumps.

$\boldsymbol{d_t}$ and $\boldsymbol{f_t}$ The combination of d_t and f_t also has a significant influence on the response time ($F(14, 195) = 4.52$, $p < .001$). Figure 7 shows the interaction effect of target disparity and target luminance spatial frequency. The mean response time for target luminance contrast spatial frequency $4cpd$ is smaller than for either $1cpd$ or $16cpd$, for both uncrossed and crossed target disparities, in line with the main effect of target luminance contrast spatial frequency. However, no significant difference between different frequencies was found for a given target disparity, except for 1.6° (the response time for $16cpd$ is significantly larger than for both $4cpd$ and $1cpd$ at 1.6°).

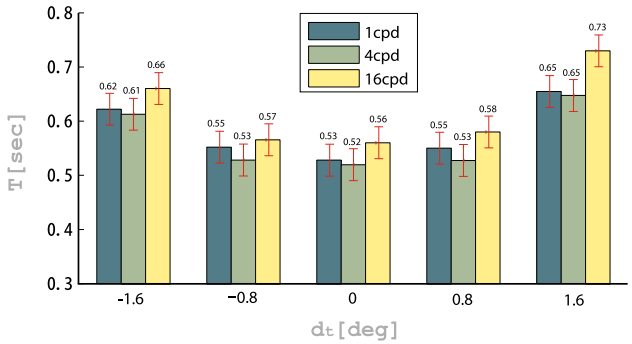


Fig. 7 Mean response time for clearly fusing the test stimuli for tested levels of target disparities and target luminance contrast spatial frequencies

4.2 Observations and discussions

These ANOVA analysis results allow us to make some basic observations about the time taken by the HVS to clearly perceive the target depth after an abrupt change in stereoscopic disparity.

- The response time is mainly affected by the change in disparity and increases with magnitude of the disparity change, both for forward and backward depth jumps. On the other hand, a change in luminance contrast spatial frequency has no significant influence on the response time.
- The target disparity and target luminance contrast spatial frequency, as well as their interaction, significantly influence the response time; the initial content has less significant effect on perception of the target depth. The response time increases with increasing magnitude of target crossed or uncrossed disparity, with a minimum reached at about $4cpd$ for target frequencies in $[1cpd, 16cpd]$.
- The response time is longer for greater target disparities and greater disparity changes, both crossed and uncrossed.
- Compared to target disparity and change in disparity, both the main and interaction effects of luminance contrast spatial frequency are much smaller in terms of response time.

We now further consider the effect of the luminance contrast spatial frequency on the response time. As for frequency, though a minimal response time occurs at $4cpd$ for the whole data (see Fig. 4b) or given a target disparity (see Fig. 7),

the differences between different frequencies are not significant as reported in Figs. 4b, 6 and 7: note the overlapping error bars. One possible explanation is that the time taken to process luminance information and disparity information is independent [36] and thus additive, while luminance information is processed much faster than disparity information, because processing disparity, especially disparity changes, involves an eye vergence movement [16,36]. Additionally, Fig. 4b shows that the difference between the maximal and the minimal mean response times across different frequencies is about 52 ms and is about 2× the standard error of the mean (14 ms), indicating that the time taken to process luminance information is easily influenced by random factors in the experiment.

4.3 Bilinear fitting

We now make a model of response time T: how long the HVS takes to clearly fuse the new stereoscopic content after a scene change, under different conditions, based on the data presented in Sect. 4.1 and observations in Sect. 4.2.

Previous perceptual experiments on disparity discrimination threshold [11,12] or visual comfort metric for stereoscopic motion [14] fitted functions for all the factors considered, using quadratic polynomials for all pair-wise combinations of factors as well as univariate terms, without considering the significance of each factor or their interactions (frequencies were represented in log-space). We also use a polynomial model for the function, but select terms to use based on the results in Sect. 4.1 and observations in Sect. 4.2.

First, we incorporate into the response time function $T(d_t, f_t, \Delta d)$ all terms previously found to be significant, i.e., d_t, f_t, Δd, $d_t \Delta d$, $\Delta d \lg(f_t)$, $d_t \lg(f_t)$. Furthermore, terms related to f_t are excluded from the current model, as the time taken to process luminance information is much less than for the other two factors. Additionally, evidence showed that eye movement speed was roughly proportional to the amount of abruptly imposed disparity [46], and the response time is approximately proportional to target disparity and/or disparity changes, whether crossed or uncrossed, as shown in Figs. 4a, c, 5, 6 and 7. We thus use a bilinear model for T with respect to d_t and Δd:

$$\begin{aligned} T &= T(d_t, f_t, \Delta d) \approx T(d_t, \Delta d) \\ &= p_1 d_t + p_2 \Delta d + p_3 d_t \cdot \Delta d + p_4, \end{aligned} \tag{1}$$

where the coefficient vector $\mathbf{P} = [p_1, p_2, p_3, p_4]$ is to be determined by fitting. Remember that our function measures the response time taken to fuse the target disparity d_t after a disparity change of Δd, and it is not equivalent to a differentiation of some response time function $T(d_t)$ on the target disparity d_t, which indicates the increased response time given an increasing in d_t.

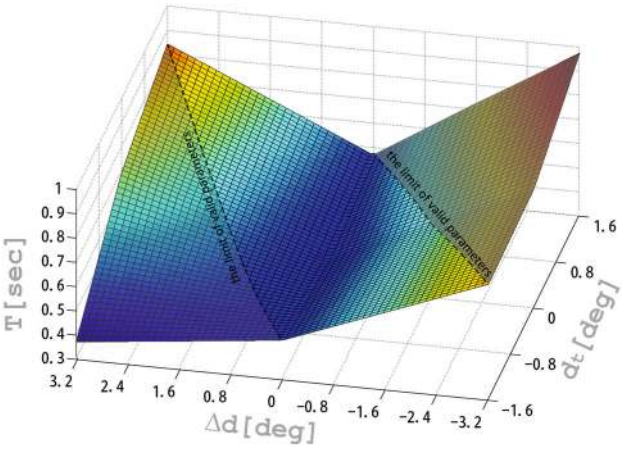


Fig. 8 The fitted response time function T. A *warmer color* indicates a longer response time. *Dashed lines* indicate parameter limits in our experiments

Furthermore, to explore how differently crossed and uncrossed disparity [14] affect the response time, as well as directions of depth jumps, we extend Eq. 1 to allow (p_1^+, p_2^+) and (p_1^-, p_2^-) for $d_t \geq 0, \Delta d \geq 0$ and $d_t < 0, \Delta d < 0$, respectively. Similarly, p_3 is extended to $(p_3^{++}, p_3^{+-}, p_3^{-+}, p_3^{--})$ for different sign combinations of d_t and Δd. The final coefficient vector $\mathbf{P} = [p_1^+, p_2^+, p_1^-, p_2^-, p_3^{++}, p_3^{+-}, p_3^{-+}, p_3^{--}, p_4]$ is obtained by optimizing

$$\arg\min_{\mathbf{P}\in\mathbb{R}^9} \sum_{j=1}^{210} \left(\frac{T(s_j) - T_j}{T_j}\right)^2.$$

The resulting coefficient vector is $\mathbf{P} = [.0294, .0849, -.0254, -.1165, .0203, -.0261, .0707, -.0112, .4186]$. Multivariate regression analysis gives an R^2 measure of goodness of fit $R^2 = .960$. The fitted function is visualized in Fig. 8.

This fitted function is consistent with the ANOVA results from Sect. 4.1. Other important quantitative conclusions can also be directly inferred from the function:

- The direction of a depth jump affects the response time. Response time changes more strongly after a crossed ($\Delta d < 0$) depth jump than it does after an uncrossed ($\Delta d \geq 0$) depth jump ($|p_2^-| = .117$, $|p_2^+| = .0849$): jumping from near to far is easier than the opposite.
- The sign of the target depth also affects the response time. The HVS behaves differently for crossed disparity ($|p_1^-| = .0254$) and uncrossed disparity ($|p_1^+| = .0294$), as noted in earlier works on visual discomfort [14,50].
- The combination of d_t and Δd influences the response time; moreover, this effect differs for different sign combinations of these quantities ($p_3^{++} = .020$, $p_3^{+-} = -.026$, $p_3^{-+} = .070$, $p_3^{--} = -.011$). In general, the response time is shorter when jumping to a crossed dis-

Fig. 9 Two scene cuts (disparity map inset, top row: Animal United © 2010 Constantin Film, Germany; bottom row: The Smurfs 2 © 2013 Sony Pictures Animation Inc.), with the corresponding per-pixel response time map. *Brighter red* indicates a longer response time. From *left* to *right* the scene (🚗) before the cut, the scene (🚗) after the cut, and the pixel-wise response time map

parity than to an uncrossed one, as $p_3 d_t \cdot \Delta d$ is negative when $d_t < 0$; see Figs. 4a and 7.

4.4 Real scene changes

So far, we have presented experiments and a model for response time given disparity changes in artificial stimuli. We now consider a practical computational approach to estimating the response time for changes in real stereoscopic content. In particular, we first consider a response time map for each pixel in the new scene, and then how to combine this information to determine the overall response time to a scene cut.

Given two scenes (I_i^l, I_i^r) and (I_t^l, I_t^r) with left and right views, we first compute the disparity maps D_i, D_t for the left view of each scene. For real-world scenes, pixel disparity maps may be estimated using the method of Smith et al. [51], and then converted to angular disparities.

Directly using the response time function presented in Sect. 4.3 allows us to determine a *pixel-wise* response time map for the target scene after a sudden scene change, indicating the time for a target position to be clearly fused after a scene change. Before the scene change, the viewer will typically focus on the most salient region of the image content [6]. We use a saliency map to characterize the corresponding effect on response time, as the most salient region will dominate the overall response: we expect the HVS to be most concerned with fusing the important content. We thus compute a region contrast-based saliency map $(S_i(\cdot), S_t(\cdot))$ for the left view using the method of Cheng et al. [6], giving saliency values for each region. A region in the initial scene contributes to a target pixel by averaging all its contained pixels. Finally, we compute the response time for each pixel in the target scene at location $\boldsymbol{x} \in \mathbb{R}^2$ as:

$$T_p(\boldsymbol{x}) = \sum_{\Omega \in I_i^l} S_i(\Omega) \cdot \frac{1}{|\Omega|} \sum_{\boldsymbol{y} \in \Omega} T(D_t(\boldsymbol{x}), D_t(\boldsymbol{x}) - D_i(\boldsymbol{y})), \tag{2}$$

where Ω represents a region in the initial scene.

Two observations may be made which allow this quantity to be quickly estimated. Firstly, in the initial scene, an average disparity can be used to represent a region's disparity based on the observation that disparities within a region vary little in general. Secondly, target disparities can be carefully quantized to allow use of a response time look-up table, as the same disparity leads to the same response time according to Eq. 2.

The pixel-wise response time is then spatially pooled to evaluate the *global* response time during a scene cut. As perceived depth at a point in the scene is influenced by its surroundings and scene arrangement [3,21], we incorporate contributions from all regions in the target scene into the final response time, instead of just using the area with longest response time [14,55]. Regions are weighted by their saliency values and the response time for a region is simply averaged across all contained pixels. Specifically, we estimate the overall response time for a scene cut as:

$$T_g = \sum_{\Omega \in I_t^l} S_t(\Omega) \cdot \frac{1}{|\Omega|} \sum_{\boldsymbol{x} \in \Omega} T_p(\boldsymbol{x}), \tag{3}$$

where Ω represents a region in the target scene. Section 5 presents tests performed to assess the validity of this approach.

Figures 1 and 9 show three scene cuts from two real 3D movies and the corresponding per-pixel response time maps; regions with different response times can be clearly seen. We show another example video clip, containing two consecutive scene cuts, in Fig. 10. Such maps could help producers to identify the potential problem regions which would result in a long response time if they were the main areas of interest. Alternatively, they could be used to help ensure that viewers' attention is drawn to regions requiring a short response time, by making them more salient.

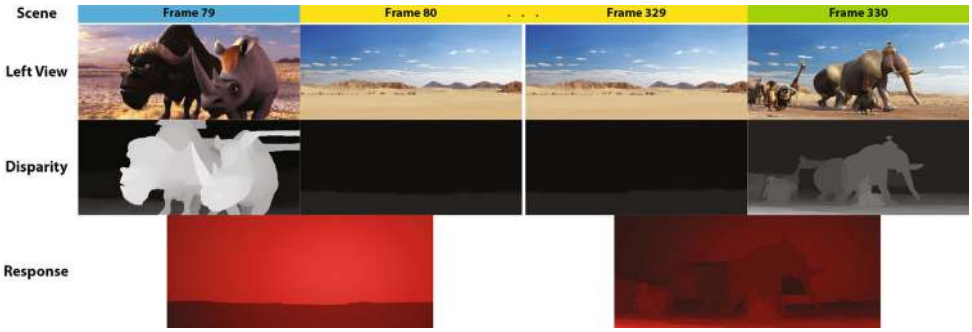


Fig. 10 An example of video clip (Animal United ©2010 Constantin Film, Germany) with two consecutive scene cuts. From *top* to *bottom* input frames, disparity maps, and per-pixel results (*brighter red* indicates a longer response time) of scent cuts

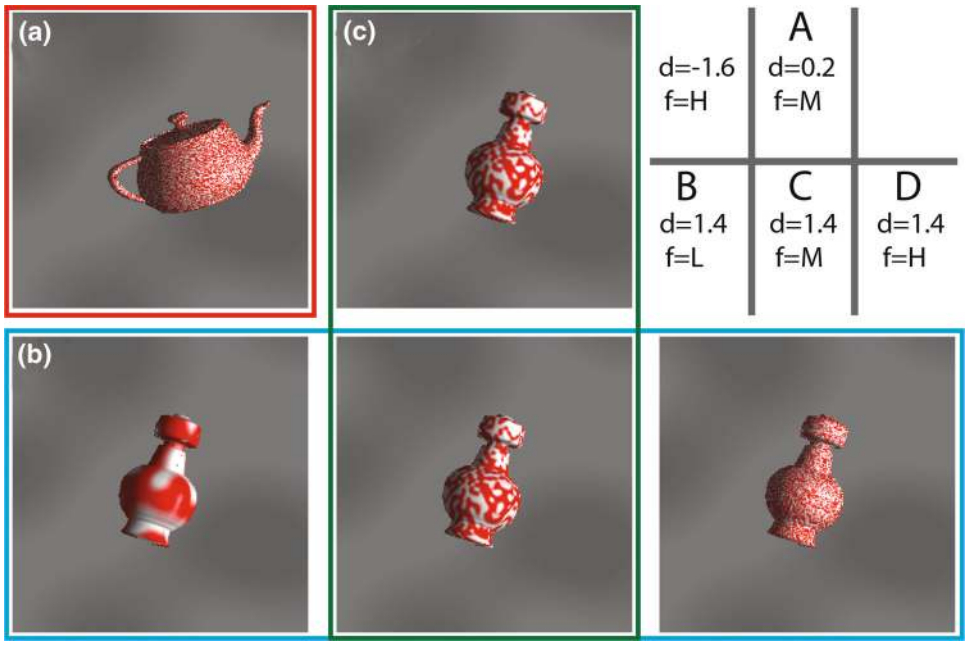


Fig. 11 Left view of controlled scene cuts used to validate our computational model. **a** Initial scene (*red*); **b** target scenes (*cyan*) with different luminance contrast spatial frequencies (low, medium, high); **c** target scenes (*green*) with salient object having different levels of disparity. The background disparity is 1.6° for all scenes. The target scenes are referred to as *A*, *B*, *C*, *D* in turn

5 Validation

In this section, we present a series of user studies used to validate our response time formula from Sect. 4.4, taking into account our assumptions on luminance contrast spatial frequency, and region saliency. These compared the response time predicted by our formula with subjects' actual response times to scene cuts. The relationship between viewing experience and response time during a scene cut was also investigated.

Model evaluation The first user study was performed in the same way as the one in Sect. 3.1: subjects were asked to press a key as soon as they had clearly accomplished stereo fusion. The initial scene (see Fig. 11a) was changed to scenes with different luminance contrast spatial frequencies (see Fig. 11b), and scenes with salient regions having different levels of disparity (see Fig. 11c). We further validated our computational model using real scene cuts which exhibited various depth jumps and target disparity ranges. Besides the three cuts, *SmurfsA, AU and SmurfsB*, shown in Figs. 1 and 9; another such scene cut, *Croods*, is presented in Fig. 12.

Ten subjects participated in this user study. Results are given in Fig. 13. For the controlled scenes, the response time differed significantly when the user focused on different salient regions, while no significant difference was found for different luminance contrast spatial frequencies, just as predicted by our model. Although the measured time is smaller than the one predicted by our formula, there is a significant correlation between them: Spearman's rank correlation coefficient is $\rho_S = .857$ ($p = .007 < .05$) and Pearson's linear correlation coefficient is $\rho_P = .993$ ($p < .01$).

Generally, the actual response time for both controlled and real scene cuts is smaller than the one predicted by our model. We believe this is mainly due to the availability of depth cues and it should be investigated further. Our model is merely based on the cue of binocular disparity (using random dot stereograms) to provide depth perception, while the scene cuts tested here provide other monocular cues, such as size, shade, perspective, occlusion etc., all of which can help to establish a more natural sense of stereopsis and facilitate a faster response [17]. Given the high linear correlation between the measured data and the predicted values,

Fig. 12 A further real scene cut (, The Croods ©2013 DreamWorks Animation Inc.) used to validate our computational model

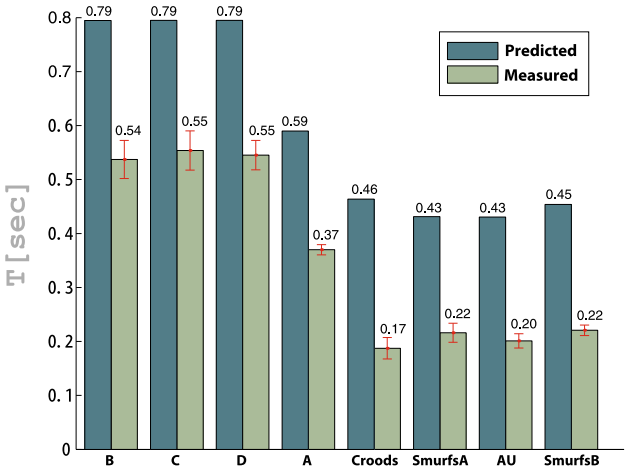


Fig. 13 Results of model validation. *B*, *C* and *D* show the response time for different luminance contrast spatial frequencies; *A* and *C* show the response time for salient objects having different disparities; the last four show response times for real scenes

we may apply a linear correction to the function in Eq. 3 to obtain a final model for response time, $T = .963T_g - .219$ ($R^2 = .987$).

Viewing experience We conducted a further user study to explore the relationship between viewing experience and the response time to a scene cut. Previous studies have shown that repeated or frequent depth jumps lead to deteriorating visual function [15] and interrupt the feeling of stereoscopic immersion [33].

We conducted trials with two consecutive controlled scene cuts (i.e. three consecutive scenes). The exposure duration to both the first and third scenes was 2 s, while for the second scene it was either 0.5× or 1.5× the response time for the first scene cut. We tested both forward jumps (the foreground object first changed from −1.4° to 1.4°, then to 0.0°) and backward jumps (the foreground object first changed from 1.4° to −1.4°, then to 0.0°). In each test, subjects were presented with both longer and shorter durations (in random order, with a one-minute break between trials), during which they were asked to attempt to fuse foreground objects in the second scene and indicate whether they had succeeded or not. After the test, subjects were asked to indicate which trial led to a better 3D viewing experience.

The same ten subjects from the previous user study took part in this experiment. All subjects reported that they could fuse the second scene in the longer trials and also had a better 3D viewing experience for the longer trials, for both forward and backward jumps. All subjects except one indicated that the second scene occurred too soon to fuse in the shorter trials, for both forward and backward jumps. The results confirm that response times during a stereo scene cut affect viewing experience.

6 Applications

In this section, we consider some applications that can benefit from our experiments and computational model derived in Sect. 4.4.

Optimizing for stereoscopic scene cuts Our computational model can be directly used to optimize the response time taken to fuse the new scene during a stereoscopic scene cut. The new response time can be calculated after linear or non-linear disparity mapping [28] for the initial scene and/or the new one. Viewers can even locally adjust the disparities in the regions of interest in an interactive manner until the proper response time is achieved.

Visual comfort assessment Many visual comfort metrics [7,8,24,26,55] have been proposed to evaluate long stereo videos. All have processed cuts and between-cuts in the same way. Recently, Du et al. [14] presented a visual comfort metric for stereoscopic motion for short videos (usually a single shot).

To evaluate the overall level of visual comfort for a long stereo video, the cuts and parts between cuts should be treated separately, as the HVS behaves differently in each circumstance. Our response time model can be incorporated into the visual comfort metric by first cutting the video into different scenes, and predicting the response time using our metric for scene cuts. By then comparing the time for which each scene lasts with the corresponding time needed to switch to the scene would indicate the comfort level related to the scene cut. Combining this approach with the metric from [14] metric for different scenes, an overall visual comfort score can be determined for a long stereoscopic video. This additional information is especially useful for videos with frequent scene switches, where target scenes appear fleetingly.

3D shooting 3D film shooting is still a tricky professional task which relies on experienced film-makers, although various principles have been suggested to guide shooting [33,38]. Methods have been introduced to plan a camera configura-

tion to achieve smooth depth jumps [18,42]. Our metric, on the other hand, can be exploited to quantitatively control the amount of depth jump which is perceptually acceptable.

The response time predicted by Eq. 3 is actually a threshold time for the target scene to be clearly fused, as revealed in the second validation: the target scene can be well fused if it lasts longer than this time, but fusing will not happen if it less than this time. Given the duration of the target scene, either the disparity change or the target disparity can be directly computed using our metric if the other quantity is constrained. This could help to guide the director in planning the layout of the scene to be filmed.

Stereoscopic video editing When applied to stereoscopic video, many stereoscopic editing tools, such as 2D-to-3D video conversion [32,56], disparity mapping [28,31], retargeting [2,5,10,30,37], warping [13,41], completion [39], scene arrangements [34,35,53], etc., could also benefit from our model by constraining the response time for scene cuts after editing. These methods usually rely on optimizing an energy function describing the operation, and our response time metric for scene cuts could be incorporated into the optimization process to ensure a sufficiently short response time after editing.

7 Conclusions

We have presented a novel model to describe the time taken by the HVS to clearly perceive a target scene following an abrupt depth change in S3D viewing. A bilinear response time model, involving the two most significant factors: the target disparity and the change of disparity, is a good fit to data collected from perceptual experiments. It was derived following a careful statistical analysis of the significance of both a single factor and interacting factors, taking into account initial and target disparity, and luminance contrast spatial frequency. Furthermore, a method has been given to estimate overall expected response time during a scene cut in stereoscopic videos. This has been validated in a further user study.

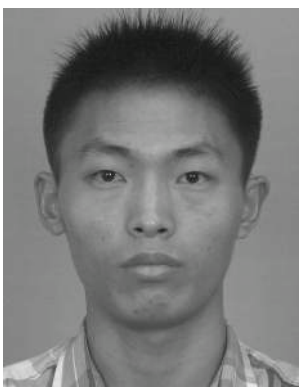
In future, other potentially important factors, such as luminance contrast and color contrast should also be considered and statistically analyzed in a similar way. It is clear that more sophisticated models could be proposed given the complexity of the HVS. Currently, we simply weight different scene regions by their saliency values, but a more sophisticated scheme is likely to provide an improvement, given that perceived depth is influenced by disparity gradient [3]. Overall, in combination with previous work on stereoscopic perception, such as disparity discrimination thresholds [11,12], and visual comfort with stereoscopic motion [14], we believe that our response time model can be used to help inform a better viewing experience for S3D content and applied to some interesting 3D versions of applications of visual media [22].

Acknowledgments We would like to thank the anonymous reviewers for their insightful comments. This work was supported by the National Basic Research Project of China (No. 2011CB302205), the Natural Science Foundation of China (No. 61272226 and 61120106007), the National High Technology Research and Development Program of China (No. 2013AA013903) and Tsinghua University Initiative Scientific Research Program.

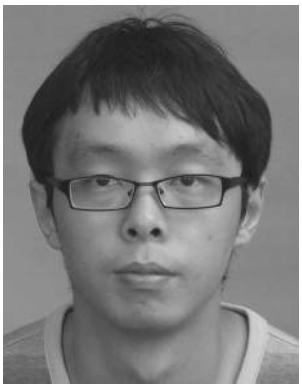
References

1. Akeley, K., Watt, S.J., Girshick, A.R., Banks, M.S.: A stereo display prototype with multiple focal distances. ACM Trans. Graph. **23**(3), 804–813 (2004)
2. Basha, T., Moses, Y., Avidan, S.: Geometrically Consistent Stereo Seam Carving. In: ICCV, pp. 1816–1823 (2011)
3. Burt, P., Julesz, B.: A disparity gradient limit for binocular fusion. Science **208**, 615–617 (1980)
4. Celikcan, U., Cimen, G., Kevinc, E.B., Capin, T.: Attention-aware disparity control in interactive environments. Vis. Comput. **29**(6–8), 685–694 (2013)
5. Chang, C.H., Liang, C.K., Chuang, Y.Y.: Content-aware display adaptation and interactive editing for stereoscopic images. IEEE Trans. Multimedia **13**(4), 589–601 (2011)
6. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global Contrast Based Salient Region Detection. In: IEEE CVPR, pp. 409–416 (2011)
7. Choi, J., Kim, D., Choi, S., Sohn, K.: Visual fatigue modeling and analysis for stereoscopic video. Opt. Eng. 51(1), 017206-1–017206–11 (2012)
8. Choi, J., Kim, D., Ham, B., Choi, S., Sohn, K.: Visual Fatigue Evaluation and Enhancement for 2d-plus-depth Video. In: IEEE ICIP, pp. 2981–2984 (2010)
9. Cumming, B.G., Judge, S.J.: Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey. J. Neurophysiol. **55**(5), 896–914 (1986)
10. Dahan, M.J., Chen, N., Shamir, A., Cohen-Or, D.: Combining color and depth for enhanced image segmentation and retargeting. Vis. Comput. **28**(12), 1181–1193 (2012)
11. Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., Seidel, H.P.: A perceptual model for disparity. ACM Trans. Graph. **30**(4), 96:1–96:10 (2011)
12. Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., Seidel, H.P., Matusik, W.: A luminance-contrast-aware disparity model and applications. ACM Trans. Graph. **31**(6), 184:1–184:10 (2012)
13. Du, S.P., Hu, S.M., Martin, R.: Changing perspective in stereoscopic images. IEEE Trans. Vis. Comput. Graph. **19**(8), 1288–1297 (2013)
14. Du, S.P., Masia, B., Hu, S.M., Gutierrez, D.: A metric of visual comfort for stereoscopic motion. ACM Trans. Graph. **32**(6), 222:1–222:9 (2013)
15. Emoto, M., Niida, T., Okano, F.: Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. J. Disp. Technol. **1**(2), 328–340 (2005)
16. Erkelens, C.J., Regan, D.: Human ocular vergence movements induced by changing size and disparity. J. Physiol. **379**, 145–169 (1986)
17. Erkelens, C.J., Steinman, R.M., Collewijn, H.: Ocular vergence under natural conditions. II. Gaze shifts between real targets differing in distance and direction. Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. R Soc (Great Britain) **236**(1285), 441–465 (1989)

18. Heinzle, S., Greisen, P., Gallup, D., Chen, C., Saner, D., Smolic, A., Burg, A., Matusik, W., Gross, M.: Computational stereo camera system with programmable control loop. ACM Trans. Graph. **30**(4), 94:1–94:10 (2011)
19. Hess, R.F., Kingdom, F.A., Ziegler, L.R.: On the relationship between the spatial channels for luminance and disparity processing. Vis. Res. **39**(3), 559–568 (1999)
20. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. J. Vis. **8**(3), 33:1–33:30 (2008)
21. Howard, I.P., Rogers, B.J.: Perceiving in Depth. Volume 2: Stereoscopic Vision. Oxford University Press, Oxford (2012)
22. Hu, S.M., Chen, T., Xu, K., Cheng, M.M., Martin, R.R.: Internet visual media processing: a survey with graphics and vision applications. Vis. Comput. **29**(5), 393–405 (2013)
23. Hulusic, V., Debattista, K., Chalmers, A.: Smoothness perception. Vis. Comput. **29**(11), 1159–1172 (2013)
24. Ide, S., Yamanoue, H., Okui, M., Okano, F., Bitou, M., Terashima, N.: Parallax distribution for ease of viewing in stereoscopic hdtv. SPIE **4660**, 38–45 (2002)
25. Inoue, T., Ohzu, H.: Accommodative responses to stereoscopic three-dimensional display. Appl. Opt. **36**(19), 4509–4515 (1997)
26. Kim, D., Sohn, K.: Visual fatigue prediction for stereoscopic image. IEEE Trans. Circ. Syst. Video Technol. **21**(2), 231–236 (2011)
27. Lambooij, M.T.M., IJsselsteijn, W.A., Heynderickx, I.: Visual discomfort and visual fatigue of stereoscopic displays: a review. J. Imaging Technol. Sci. **53**(3), 1–14 (2009)
28. Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., Gross, M.: Nonlinear disparity mapping for stereoscopic 3d. ACM Trans. Graph. **29**(4), 75:1–75:10 (2010)
29. Lee, B., Rogers, B.: Disparity modulation sensitivity for narrow-band-filtered stereograms. Vis. Res. **37**(13), 1769–1777 (1997)
30. Lee, K.Y., Chung, C.D., Chuang, Y.Y.: Scene warping: layer-based stereoscopic image resizing. In: IEEE CVPR, pp. 49–56 (2012)
31. Lee, S., Kim, Y., Lee, J., Kim, K., Lee, K., Noh, J.: Depth manipulation using disparity histogram analysis for stereoscopic 3d. Vis. Comput. **30**(4), 455–465 (2014)
32. Liao, M., Gao, J., Yang, R., Gong, M.: Video stereolization: combining motion analysis with user interaction. IEEE Trans. Vis. Comput. Graph. **18**(7), 1079–1088 (2012)
33. Liu, C.W., Huang, T.H., Chang, M.H., Lee, K.Y., Liang, C.K., Chuang, Y.Y.: 3d cinematography principles and their applications to stereoscopic media processing. In: ACM Multimedia, pp. 253–262 (2011)
34. Lo, W.Y., van Baar, J., Knaus, C., Zwicker, M., Gross, M.: Stereoscopic 3d copy and paste. ACM Trans. Graph. **29**(6), 147:1–147:10 (2010)
35. Luo, S.J., Shen, I.C., Chen, B.Y., Cheng, W.H., Chuang, Y.Y.: Perspective-aware warping for seamless stereoscopic image cloning. ACM Trans. Graph. **31**(6), 182:1–182:8 (2012)
36. Marr, D., Poggio, T.: A computational theory of human stereo vision. Proceedings of the Royal Society of London. Ser B Biol Sci **204**(1156), 301–328 (1979)
37. Masia, B., Wetzstein, G., Aliaga, C., Raskar, R., Gutierrez, D.: Display adaptive 3d content remapping. Comput. Graph. **37**(8), 983–996 (2013)
38. Mendiburu, B.: 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. Focal Press, England (2009)
39. Mu, T.J., Wang, J.H., Du, S.P., Hu, S.M.: Stereoscopic image completion and depth recovery. Vis. Comput. **30**(6–8), 833–843 (2014)
40. Neri, P., Bridge, H., Heeger, D.J.: Stereoscopic processing of absolute and relative disparity in human visual cortex. J. Neurophysiol. **92**(3), 1880–1891 (2004)
41. Niu, Y., Feng, W.C., Liu, F.: Enabling warping on stereoscopic images. ACM Trans. Graph. **31**(6), 183:1–183:7 (2012)
42. Oskam, T., Hornung, A., Bowles, H., Mitchell, K., Gross, M.: Oscam—optimized stereoscopic camera control for interactive 3d. ACM Trans. Graph. **30**(6), 189:1–189:8 (2011)
43. Palmer, S.E.: Vision Science: Photons to Phenomenology. The MIT Press, Cambridge (1999)
44. Pan, B., Zhao, Y., Guo, X., Chen, X., Chen, W., Peng, Q.: Perception-motivated visualization for 3d city scenes. Vis. Comput. **29**(4), 277–289 (2013)
45. Pollock, B., Burton, M., Kelly, J., Gilbert, S., Winer, E.: The right view from the wrong location: depth perception in stereoscopic multi-user virtual environments. IEEE Trans. Vis. Comput. Graph. **18**(4), 581–588 (2012)
46. Rashbass, C., Westheimer, G.: Disjunctive eye movements. J. Physiol. **159**(2), 339–360 (1961)
47. Schor, C.M.: A dynamic model of cross-coupling between accommodation and convergence: simulations of step and frequency responses. Optom. Vis. Sci. **69**(4), 258–269 (1992)
48. Schor, C.M.: The influence of interactions between accommodation and convergence on the lag of accommodation. Ophthal. Physiol. Opt. **19**(2), 134–150 (1999)
49. Schor, C.M., Kotulak, J.C.: Dynamic interactions between accommodation and convergence are velocity sensitive. Vis. Res. **26**(6), 927–942 (1986)
50. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: predicting visual discomfort with stereo displays. J. Vis. **11**(8), 11:1–11:29 (2011)
51. Smith, B.M., Li, Z., Jin, H.: Stereo Matching with Nonparametric Smoothness Priors in Feature Space. In: IEEE CVPR, pp. 485–492 (2009)
52. Takagi, M., Oyamada, H., Abe, H., Zee, D.S., Hasebe, H., Miki, A., Usui, T., Bando, T.: Adaptive changes in dynamic properties of human disparity-induced vergence. Invest. Ophthalmol. Vis. Sci. **42**(7), 1479–1486 (2001)
53. Tong, R.F., Zhang, Y., Cheng, K.L.: Stereopasting: interactive composition in stereoscopic images. IEEE Trans. Vis. Comput. Graph. **19**(8), 1375–1385 (2013)
54. Yang, S., Gowrisankaran, S., Younkin, A.C., Corriveau, P.J., Sheedy, J.E., Hayes, J.R.: Discernible difference and change in object depth afforded by stereoscopic three-dimensional content. In: SPIE, vol. 8648, pp. 86481C–86481C–11 (2013)
55. Yong, J.J., Lee, S.I., Sohn, H., Hyun, W.P., Yong, M.R.: Visual comfort assessment metric based on salient object motion information in stereoscopic video. J. Electron. Imaging **21**(1), 011008-1–011008-16 (2012)
56. Zhang, G., Hua, W., Qin, X., Wong, T.T., Bao, H.: Stereoscopic video synthesis from a monocular video. IEEE Trans. Vis. Comput. Graph. **13**(4), 686–696 (2007)



Tai-Jiang Mu received his BS degree in computer science from Tsinghua University in 2011. He is currently a PhD candidate in the Department of Computer Science and Technology, Tsinghua University. His research interests include computer graphics, stereoscopic image/video processing, stereoscopic perception.



Jia-Jia Sun received his BS degree in computer science from Xi'dian University in 2012. He is currently a PhD candidate in the Institute for Interdisciplinary Information Sciences, Tsinghua University. His research interests include computer graphics, stereoscopic perception and image rendering.

Ralph R. Martin is currently a Professor at Cardiff University. He obtained his PhD degree in 1983 from Cambridge University. He has published more than 200 papers and 12 books, covering such topics as solid and surface modeling, intelligent sketch input, ge- ometric reasoning, reverse engineering, and various aspects of computer graphics. He is a Fellow of: the Learned Society of Wales, the Institute of Mathematics and its Applications, and the British Computer Society. He is on the editorial boards of Computer Aided Design, Computer Aided Geometric Design, Computer & Graphics, Graphical Models and the International Journal of Shape Modeling.

Shi-Min Hu received the PhD degree from Zhejiang University in 1996. He is currently a professor in the Department of Computer Science and Technology at Tsinghua University, Beijing. His research interests include digital geometry processing, video processing, rendering, computer animation, and computer-aided geometric design. He is associate Editor-in-Chief of The Visual Computer (Springer), and on the editorial boards of IEEE Transactions on Visualization and Computer Graphics, Computer-Aided Design (Elsevier) and Computer & Graphics (Elsevier). He is a member of the IEEE, ACM and CCF.