



## Technical Section

## A perceptual approach for stereoscopic rendering optimization

Abdullah Bulbul\*, Zeynep Cipiloglu, Tolga Capin

Department of Computer Engineering, Bilkent University, 06800 Ankara, Turkey

## ARTICLE INFO

## Article history:

Received 27 May 2009

Received in revised form

30 September 2009

Accepted 13 November 2009

## Keywords:

Stereoscopic rendering

Binocular vision

Binocular suppression

Perception

## ABSTRACT

The traditional way of stereoscopic rendering requires rendering the scene for left and right eyes separately; which doubles the rendering complexity. In this study, we propose a *perceptually-based approach* for accelerating stereoscopic rendering. This optimization approach is based on the Binocular Suppression Theory, which claims that the overall percept of a stereo pair in a region is determined by the dominant image on the corresponding region. We investigate how binocular suppression mechanism of human visual system can be utilized for rendering optimization. Our aim is to identify the graphics rendering and modeling features that do not affect the overall quality of a stereo pair when simplified in one view. By combining the results of this investigation with the principles of visual attention, we infer that this optimization approach is feasible if the high quality view has more intensity contrast. For this reason, we performed a subjective experiment, in which various representative graphical methods were analyzed. The experimental results verified our hypothesis that a modification, applied on a single view, is not perceptible if it decreases the intensity contrast, and thus can be used for stereoscopic rendering.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Technologies underlying 3D autostereoscopic displays have matured to the point that several commercial products are now available in the mass market. These displays are an extension of the conventional 2D displays, by their ability to emit a different image for each eye. Binocular head-mounted displays have also matured to the level that they are widely used in a number of applications. The main difficulty of these stereoscopic and autostereoscopic displays is that they require a rendering phase for each view, which multiplies the rendering time by the number of views. Consequently, there is a need to optimize solutions for stereoscopic rendering.

The traditional way of stereoscopic rendering is to handle the left and the right eye views separately, which is still the model in use in graphics APIs such as OpenGL [1]. A number of stereoscopic and multi-view rendering techniques have recently been proposed. These approaches can be categorized as *pipeline-based solutions*, which aim to optimize the rendering on the rasterization stage of the rendering pipeline [2–4]; and *image-based solutions*, where one view is rendered using the graphics rendering pipeline, and the other view is generated from this image, using the correspondences of the two views [5–7].

In this paper, we propose a new *perceptually-based solution* for optimization, by utilizing the suppression theory of binocular vision. According to the Binocular Suppression Theory, the less dominant view will be suppressed by the dominant one; and when the images from corresponding regions differ in an appropriate way, they fuse but the disparities are registered and used for impression of depth. It has been shown that the result of which view suppresses the other depends on the visual properties of the two images [8,9]. Section 2.2 overviews the Binocular Suppression Theory that our solution is based on.

We investigate how binocular suppression mechanism can be utilized for optimization, by comparing the effects of different graphics rendering and modeling methods. Our aim is to identify the rendering and modeling features that do not affect the overall quality of a stereo image pair when simplified in one view. We applied our approach to a number of representative and commonly used methods used in rendering, including framebuffer upsampling, mixed-level antialiasing, specular highlight, mixed shading, mesh simplification, texture resampling, and mixed shadowing. We performed an experimental study in order to evaluate each method's perceptual effect on the overall perceived 3D image, in terms of quality, sharpness, depth, and comfort. The experimental results show that the overall perceived stereo image quality is not affected when one of the views is modified by a technique that decreases the *intensity contrast*. On the other hand, when a modification that increases the intensity contrast is applied on a single view; it will be visible and the overall perceived stereo image quality will be affected towards the modified image.

\* Corresponding author. Tel.: +90 312 2902163; fax: +90 312 2664047.

E-mail addresses: [bulbul@cs.bilkent.edu.tr](mailto:bulbul@cs.bilkent.edu.tr) (A. Bulbul).[zeynep@cs.bilkent.edu.tr](mailto:zeynep@cs.bilkent.edu.tr) (Z. Cipiloglu), [tcapin@cs.bilkent.edu.tr](mailto:tcapin@cs.bilkent.edu.tr) (T. Capin).

The main contributions of this study are as follows:

- a new approach for stereo rendering optimization, based on the usage of different stimuli, each with different quality, for each eye,
- a content creation guideline, describing when it is appropriate to use each optimization,
- a formal experimental study to verify the proposed hypothesis.

The rest of the paper is organized as follows. First, we survey previous work on stereoscopic rendering methods and Binocular Suppression Theory. Then, we explain our perceptually-based approach and the graphical methods we have used. Lastly, we provide the experiment design and our analysis of the results.

## 2. Previous work

### 2.1. Stereoscopic rendering optimization

A number of techniques have been proposed to optimize stereoscopic rendering. The first group of solutions follows a *graphics pipeline-based approach*, by utilizing the coherence between neighboring views. Adelson and Hodges [10] simultaneously render a triangle to both images by using the x-axis coherence in device coordinates to accelerate the stereoscopic rendering process. Kalaiah and Capin [4] propose a GPU-based solution that reduces the number of vertex shader computations needed for rendering multiple views, the vertex shader is split into two parts – view-independent and view-dependent. Performing vertex shader computations once for the view-independent part, instead of per-view calculation, reduces the rendering complexity. Hasselgren and Akenine-Moller [3] propose a multi-view pipeline-based method, called approximate rendering, where fragment colors in all neighboring views can be approximated from a central view when possible. As a result of approximate rendering, many per-pixel shader instructions are avoided.

Another group of solutions uses an *image-based approach*. In these solutions, one view is reconstructed from the other, 3D rendered view, by exploiting the similarity between the two views. In these techniques, the rendering time of the second image depends on only the image resolution, instead of the scene complexity, therefore saving rendering computations for one view. Fu et al. [5] compute the right image by warping the left image; however the resulting image contains holes, which require to be filled by interpolation. Wan et al. [6] fill these holes by raycasting. Similarly, Fehn [7] uses a depth buffer to generate multiple views from a single image. Blurring the depth buffer by a Gaussian filter is used for handling the hole-filling problem. Zhang et al. also use depth images to generate the second view [11]. In this method, the image for one view is used to construct a depth image, and then the second view is constructed using this depth image. Lastly the holes that occur in the previous step are filled by averaging the textures from neighboring pixels. Halle uses epipolar images that contain the rendered primitives interpolated between the two most extreme camera viewpoints for extracting the in-between views [12]. Stereo images produced with these techniques are generally an approximation to the original stereo image rendering result.

Finally, a third group of solutions has been proposed for stereoscopic rendering optimization targeted for ray tracing and volume rendering. Adelson and Hodges [10] propose a solution to stereoscopic ray tracing, where a ray-traced left image is used to construct part of the right image and the rest of the right image is

calculated by ray tracing. He and Kaufman [13] speed up stereoscopic volume rendering by re-projecting the samples for the left view to the right image plane and compositing several samples simultaneously while raycasting.

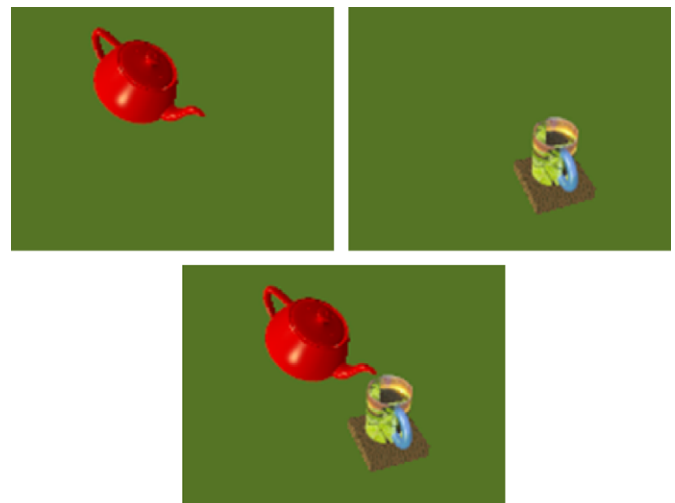
### 2.2. Binocular suppression

Our approach uses the *Binocular Suppression Theory* for binocular vision. According to this theory, when dissimilar images are shown to each eye, one of the views suppresses the other at any one time, and the dominating view alternates over time. But when similar images (e.g. in a stereo pair) are shown to each eye, similar images falling on corresponding retinal regions form a unitary visual impression, while each region in the visual field contains input from a single eye at any one time [8].

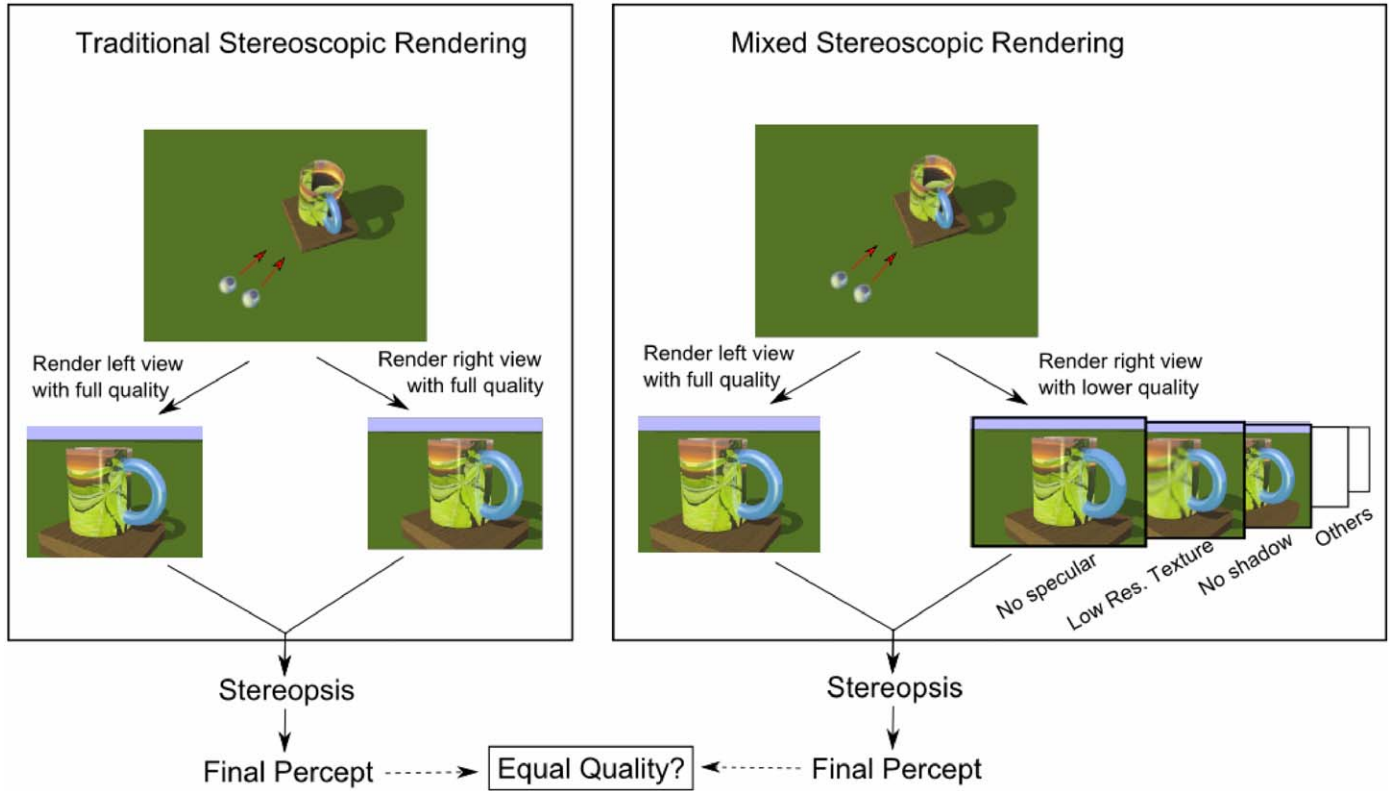
Even though the actual process of the binocular vision is not fully identified, there are cases which support the Binocular Suppression Theory in perception research. For instance, in an experimental study, subjects were asked to wear a lens for myopia for one eye, and hyperopia for the second eye, and were observed to see all distances in sharp focus, because the focused image suppresses the unfocused eye. This further supports the Binocular Suppression Theory that one view is suppressed by the other, with no effect on the final percept [8].

According to the Binocular Suppression Theory, when one view is suppressed by the other, a perceptual competition occurs between the two views. This is known as binocular rivalry and this property has been studied extensively. Asher states that rivalry occurs in local regions of the visual field, and only one eye's view is dominant within these regions [14]. Fig. 1 illustrates this mechanism. In the combined view, the teapot and the glass completely suppress the corresponding portion of the green ground seen by the other eye. Blake et al. also examined the principles of the binocular vision and claimed that stronger competitors have larger dominance [13]. For instance, a high-contrast figure will dominate over a low-contrast one, or a brighter stimulus has an advantage over a dimmer one from the perspective of predominance.

Once the binocular rivalry mechanism is confirmed, the next question becomes: what are the factors that affect the *strength* of a region for rivalry? Yang et al. state that a pattern with higher spatial frequency in one eye suppresses a pattern with lower spatial frequency in the other eye; therefore it is stronger [8].



**Fig. 1.** Binocular rivalry mechanism. When the left-eye (top-left) and right-eye (top-right) views are shown, the combined view (bottom) merges the dominant regions from the two views.



**Fig. 2.** Left: traditional stereoscopic rendering approach, Right: our rendering approach for optimization. On the upper images, the object space is illustrated along with the viewpoint. The images below show the corresponding left and right eye views of the above scene. The right view on the mixed stereoscopic rendering approach is generated with lower quality caused by simplifying a set of modeling and rendering features.

Similarly, a region becomes stronger when the contrast [15] or the number of contours increase, which in turn cause a higher spatial frequency. Color variance also has a positive effect on stimulus strength [16]. One other factor that causes a stimulus to be stronger is motion [8]. According to Breese, a moving grating has an advantage over a stationary one and the strength increases as the speed of motion increases [8].

Binocular Suppression Theory has recently gained interest in the image processing and compression fields. Perkins [17] studied mixed-resolution stereo image compression where one view is low-pass filtered and has lower resolution, and demonstrated that the resultant 3D percept is of adequate image quality, when compared to the reference content. In a related work, Berthold showed that apparent depth is relatively unaffected by spatially filtering both channels of a stereo image [18]. Therefore, the image processing research to-date suggests that it is possible to low-pass filter one or both views of a stereo pair without affecting the subjective impression of sharpness, depth, and quality of the image sequence. Stelmach et al. [18] has built on these results, and presented a solution for mixed-resolution stereo image compression, and provided favorable experimental results.

### 3. Approach

In this paper, we present a Binocular Suppression Theory based approach to stereoscopic graphics rendering. The proposed method exploits the fact that the overall perception of the stereo pair in a region will be determined by the dominant image on the corresponding region, instead of summation of the effect of two images. Our goal is to explore how the rendering quality can be

reduced in the suppressed view, without reducing the overall perceived quality of the rendered 3D image. If such features can be detected, rendering computations of those features for one eye can be reduced, thus increasing the overall speed of rendering.

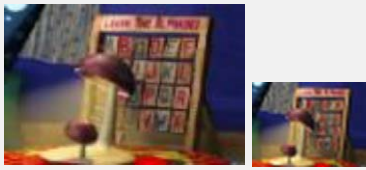
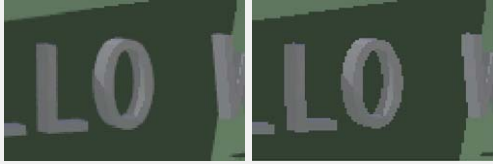
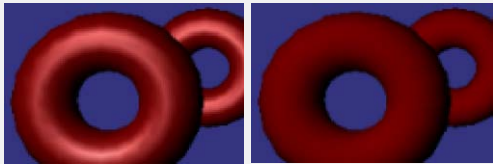
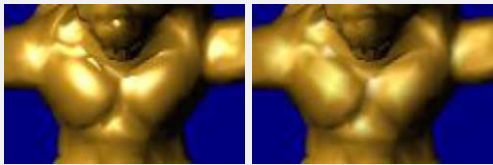
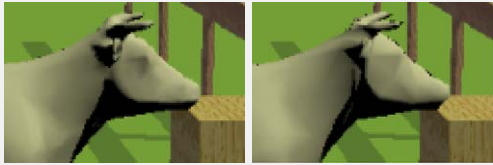
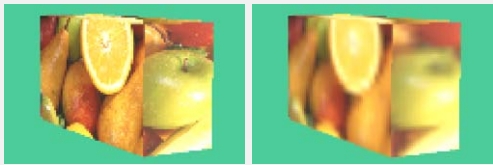

Fig. 2 illustrates the traditional and proposed mixed stereoscopic rendering approaches. In the traditional approach, the left and right views are generated with the same rendering technique and the same quality. On the other hand, in the proposed approach, two views can have different parameters of rendering – one of the views is generated with the original quality and the other view with lower quality, thus decreasing the overall rendering cost. However, this approach is feasible when the overall quality of the final 3D percept is determined by the high quality image.

#### 3.1. Mixed stereo methods

The proposed mixed stereoscopic rendering method manipulates various graphics rendering and modeling conditions, in different levels of intensity. In this work, a number of representative and commonly used methods, which are employed in virtual environments, have been investigated. These methods include framebuffer upsampling, mixed-level antialiasing, specular highlight removal, mixed shading, mesh simplification, texture resampling, and mixed shadowing. Table 1 and the rest of this section illustrate the used methods in detail.

**Framebuffer upsampling:** In framebuffer upsampling, the 3D scene is rendered to a smaller framebuffer in one view, and then this buffer is upsampled to match the framebuffer resolution for the high-quality view. In this paper, 4 different sizes of framebuffers have been used, each level halving the width and height of

**Table 1**  
Methods used for mixed stereoscopic rendering.

Method	Levels	Image A – image B
Framebuffer upsampling	Level 1: original ( <b>image A</b> ) Level 2: $\frac{1}{2}$ size ( <b>image B</b> ) Level 3: $\frac{1}{4}$ size Level 4: $\frac{1}{8}$ size	
Mixed-level antialiasing	On: antialiased ( <b>image A</b> ) Off: not-antialiased ( <b>image B</b> )	
Specular highlight	On: specular highlight is used ( <b>image A</b> ) Off: specular highlight is not used ( <b>image B</b> )	
Mixed shading	Level 1: phong shaded ( <b>image A</b> ) Level 2: gouraud shaded ( <b>image B</b> )	
Mesh simplification	Level 1: original mesh ( <b>image A</b> ) Level 2: # faces = $\frac{1}{2}$ of original Level 3: # faces = $\frac{1}{4}$ of original ( <b>image B</b> ) Level 4: # faces = $\frac{1}{8}$ of original	
Texture resampling	Level 1: texture size $512 \times 512$ ( <b>image A</b> ) Level 2: texture size $256 \times 256$ Level 3: texture size $128 \times 128$ Level 4: texture size $64 \times 64$ ( <b>image B</b> )	
Mixed shadowing	On: shadows are used ( <b>image A</b> ) Off: shadows are not used ( <b>image B</b> )	

the previous level ( $1/4$  area of the previous size). For upsampling, the Lanczos resampling algorithm has been used [19].

**Mixed-level antialiasing:** Antialiasing, based on sampling more than one sample per pixel, is used widely in graphics applications. Different sampling patterns have been proposed, including Grid, Checker, Quincunx sampling schemes [20]. Although hardware-based antialiasing solutions are fast, processing of more than one sample per pixel is still required. Therefore, in the mixed antialiasing method, one of the views is rendered using antialiasing, and the other view is rendered with antialiasing turned off. Although different antialiasing schemes could also be used for the

two views, one of the views is rendered with no antialiasing, to better illustrate its effect. In this work, a  $3 \times 3$  grid super-sampling is used as the antialiasing method [20].

**Specular highlight removal:** Specular highlight is the bright spot on a reflector surface caused by the reflection of the light, which depends on the viewing angle. In computer graphics, specular highlight is simulated in various specular reflectance models such as Phong, Cook-Torrance, etc. [21]. In this work, the Phong model is used to exhibit the specular highlight for one view, and the specular component of the material for the other view is ignored, resulting in a pure Lambertian reflectance model [22].

**Mixed shading:** In order to investigate the effects of illuminating two views with different interpolation methods, we implemented Phong and Gouraud shading, which are widely used in computer graphics. In Phong shading, normals are interpolated when calculating the color values inside a polygon in order to obtain a smooth appearance [23]; whereas in Gouraud shading only the colors are interpolated with lower cost [24]. Although a wide variety of advanced shading techniques, such as the use of BRDF, have been recently used, we have chosen two widely-used solutions for illustrating the effect of mixed shading.

**Mesh simplification:** In order to determine the effects of object-based techniques, mesh simplification is employed for all objects in one view. The simplification is done using the Quadric Edge Collapse method [25]. The number of faces of a mesh in a level is approximately half of the number of faces in the previous level (Table 1).

**Texture resampling:** To verify the effect of mixed-resolution textures in a scene, a texture resampling method is employed. In this method, the textures in the scene are rendered with lower resolution in one view. Various levels of texture resampling have been used: the size of the texture map used for a level is half of the previous level in terms of both width and height (Table 1). Linear filtering is used for resampling of the texture images [26]. In this work, further methods for antialiasing of textures, such as mipmapping or anisotropic filtering, are not tested, to verify only the resampling effect.

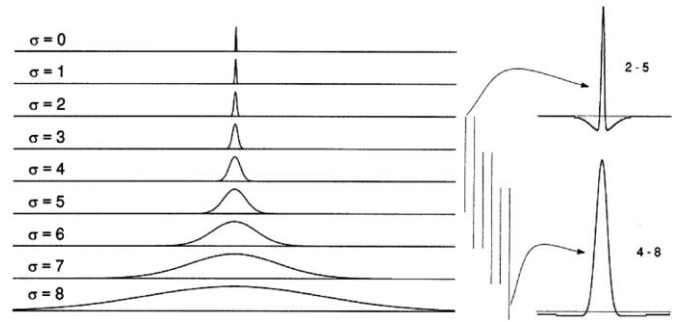
**Mixed shadowing:** As adding shadows to a 3D scene requires expensive calculations, avoiding these calculations for one view without affecting the final 3D percept would be an appropriate optimization. In the Mixed Shadowing method, no shadow is used for one view; and point light sources with hard shadows are used for the second view [27].

### 3.2. Intensity contrast

According to the Binocular Suppression Theory, the methods described above are effective when the regions in the simplified view are suppressed by the high-quality view. Therefore, the properties of an image, which allow a view to be suppressed, and which therefore keep the modification unnoticed, should be characterized.

Previous studies suggest that stronger competitors (e.g.: higher spatial frequency, more color variance or a faster motion) are more likely to suppress [9,8]. According to our observations, the properties that make the competitors of the binocular rivalry stronger are also the features that attract visual attention: The regions which attract more attention are likely to be the candidates for being strong competitors in a stereo pair. Itti and Koch [28] state that the visual attention is selective, and eye gaze is oriented towards regions that show large contrast, and these regions can be defined as *salient*. According to Itti, a region is more salient – thus attracts more attention – when it differs from its surroundings regarding a number of properties, such as intensity, color opponency, motion, and orientation [28]. These properties are consistent with the properties that increase the strength of the competitor.

To measure the strength of a view, we use a heuristic, *intensity contrast*, and obtain the change in the image intensity contrast caused by a modification, to decide whether it is sufficient to apply the modification to only a single view. For this purpose, we have followed the saliency calculation method [29], in which a center-surround mechanism is used to compute three separate saliency maps for three channels: intensity, color and orientation. In our work, only intensity maps are needed, since the methods used for modification do not have a significant



**Fig. 3.** Gaussian pixel widths for the nine scales used in the intensity contrast calculation. Scale  $\sigma=0$  corresponds to the original image, and each subsequent scale is coarser by a factor 2. On the right, two examples of the six center-surround DoG filters are shown, for scale pairs 2–5 and 4–8. [29].

effect on the color and orientation attributes of the stereoscopic image.

The first step in calculating the intensity contrast map of an image is extracting the intensity map of the image. The average of the RGB values in a pixel gives the intensity value:

$$I = \frac{R+G+B}{3}$$

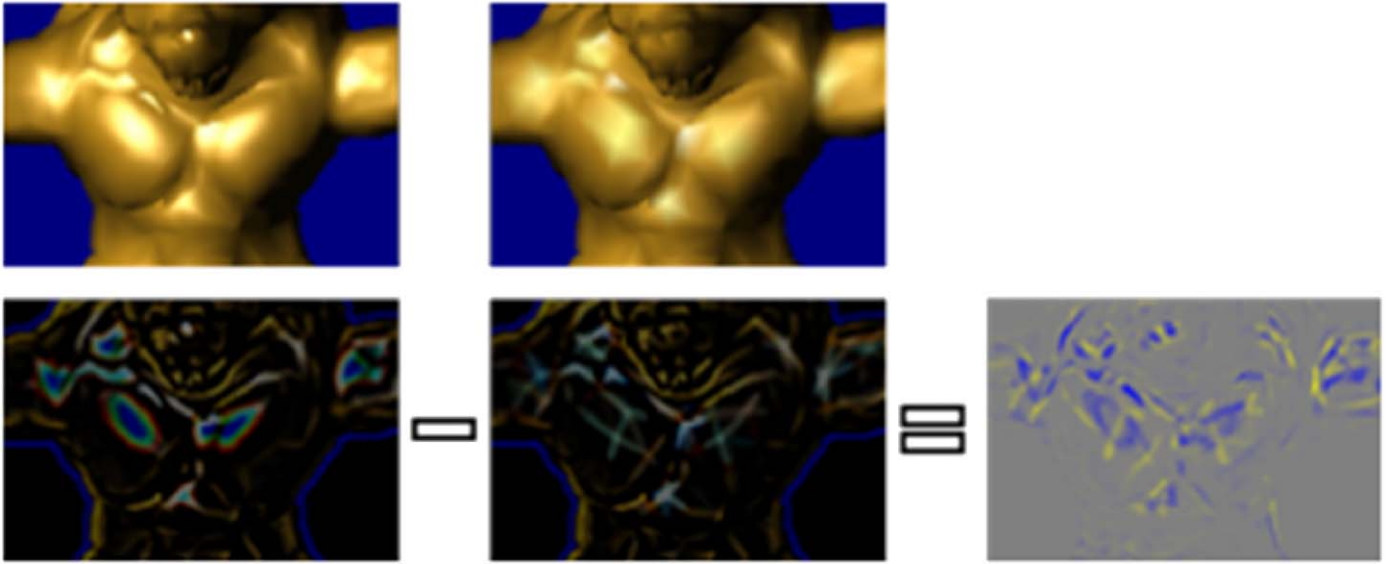
Then, the intensity contrast map is generated from the intensity image using a center-surround operator [29]. In this method, DoG (Difference of Gaussian) filters are calculated as the difference of Gaussian filters in fine (center) and coarse (surround) scales. The center consists of the pixels with a closer distance than QUOTE and the surround consists of the pixels with a closer distance than  $s=c+\delta$ , where  $c \in \{2,3,4\}$  and  $\delta \in \{3,4\}$ . Thus, six DoG filters are calculated using fine and coarse scales as  $\{2-5, 2-6, 3-6, 3-7, 4-7, 4-8\}$  (Fig. 3) and each of them is used to generate an intensity contrast map. These six maps are added pixel by pixel to construct a final intensity contrast map.

Fig. 4 shows a pair of intensity contrast maps and their difference. In this figure, brighter regions in the intensity contrast maps show the parts with greater intensity contrast. To calculate the effect of a modification on the intensity contrast; intensity contrast map of the modified image (rendered with low quality) is subtracted from the intensity contrast map of the original image. In Fig. 4, the positive values are colored as blue and the negative values are colored as yellow. The negative values on the result (yellow regions in the figure) indicate an increase in the intensity contrast due to the modification.

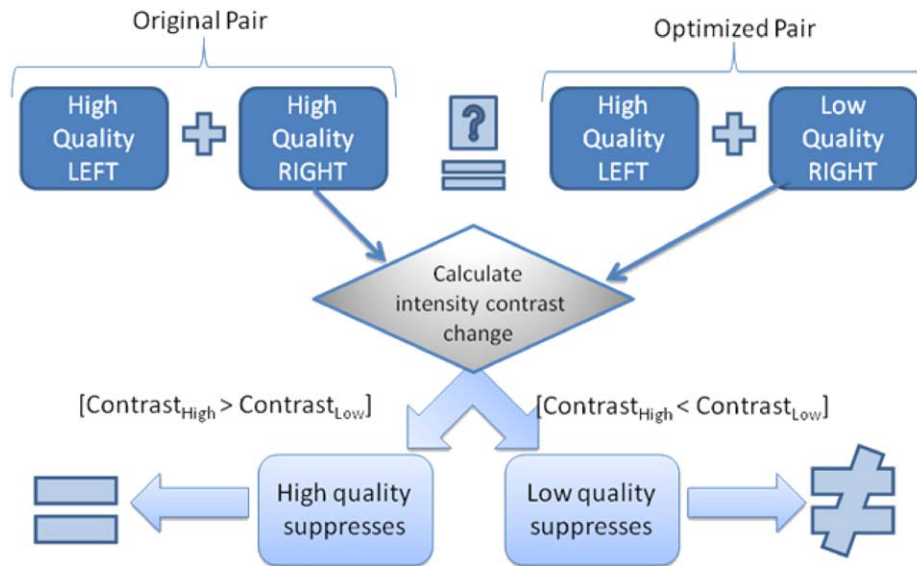
We are proposing that if the intensity contrast of a view is greater than the other view, it has the privilege of being dominant. Hence, we are hypothesizing that if the intensity contrast of the modified view is lower than the original image, then the optimized pair provides the same percept as the result of the traditional rendering. On the other hand, if a modification raises the intensity contrast, its effect will be perceived when it is applied on one of the views. Therefore, if this kind of modification reduces the quality, it cannot be used as an optimization since the quality of the overall percept is determined by the low quality image. Fig. 5 summarizes our hypothesis.

Note that it is sufficient to use the intensity contrast change map only at the beginning of the optimization process, in order to estimate whether a method is suitable for our optimization approach. If a method is decided to be suitable for optimization, it can be applied without recalculating the intensity contrast change map at each frame.

Fig. 6 contains sample intensity contrast change maps of the applied methods. According to our hypothesis, a modification is



**Fig. 4.** Top left: original image, top right: modified image, bottom left: intensity contrast map of the original image, bottom middle: intensity contrast map of the modified image (Brighter regions are the parts with higher intensity contrast in the intensity contrast maps.), bottom right: calculation of intensity contrast change (Difference of the two intensity contrast maps results in the right-most image, where the blue (dark) regions are the parts that the intensity contrast is greater in the original image and yellow (bright) regions are the opposite.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).



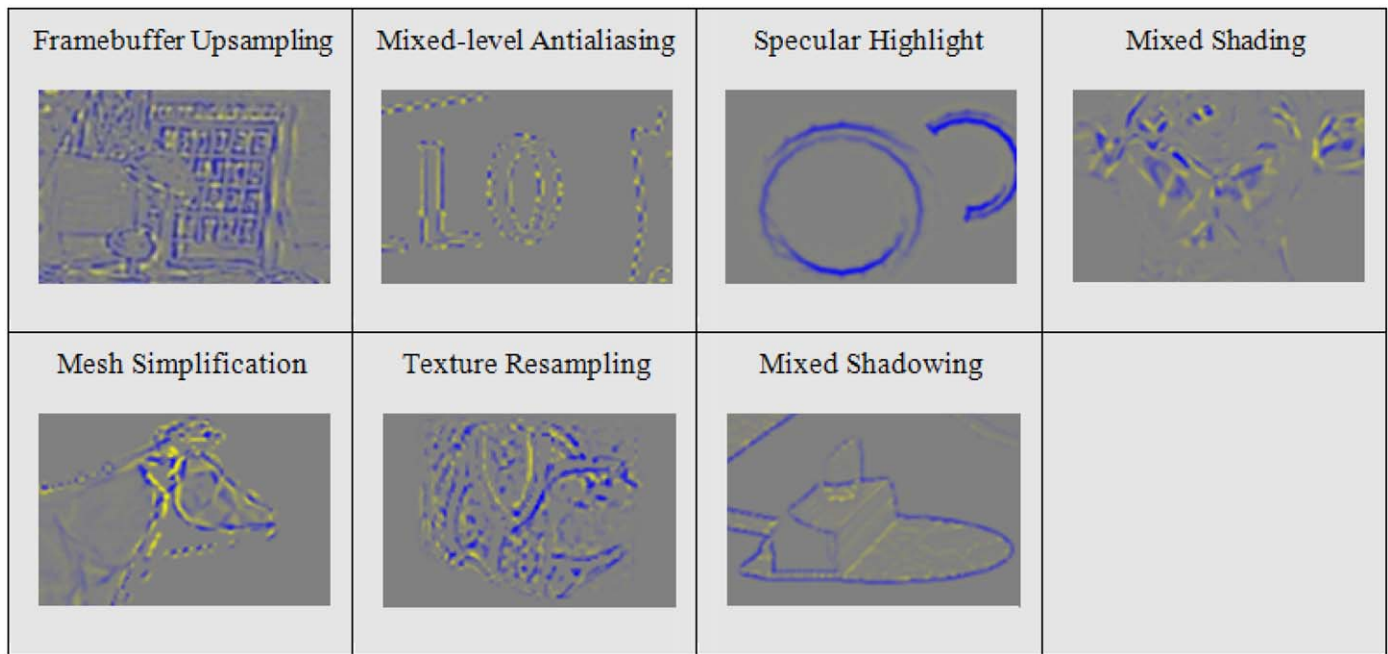
**Fig. 5.** Summary of the hypothesis. Optimized pair provides the same percept if the intensity contrast of the modified view is lower than the original ( $\text{Contrast}_{\text{High}}$ : intensity contrast of high quality image). Note that, the low quality image does not have to be the right view; it is also possible to decrease the quality of the left view.

not recognizable in blue regions of the intensity contrast change map. For instance, the intensity contrast change map of the specular highlight method contains only blue; therefore we expect that the image with the specular highlight will suppress the other image in the final percept, and this method is suitable for our optimization approach. In framebuffer upsampling and texture resampling methods; blue regions are considerably more than the yellow regions which also lead us to expect that original images are dominant in general. Thus, these methods are expected to be appropriate for stereoscopic rendering optimization. The yellow regions cover a large area in the figures of antialiasing and mesh simplification methods, therefore these methods are not suitable for our optimization approach according to our hypothesis. For mixed shadowing, although the shadowed image has apparently more intensity

contrast in the borders of the shadow; the opposite holds for the interior parts of the shadow. Hence, shadow is expected to be a strong factor for suppression, therefore it should be applied on both of the views. For mixed shading, even though there are regions in which the Gouraud shaded image is dominant, Phong shaded image is stronger in general.

#### 4. Experiment

We have implemented the proposed methods, and performed a formal experiment to observe whether the use of each method is perceptible. We have decided to base our work on users' subjective ratings, instead of objective evaluation in which users perform a task (such as measuring the time and error when



**Fig. 6.** Intensity contrast change due to selected methods (level 1 and level 3 are used for the upsampling, mesh simplification and texture resampling methods).

**Table 2**

Test cases for scalable methods.

Ref vs. test	Ref vs. test	Ref vs. test
1–1 vs. 1–2	2–2 vs. 1–2	3–3 vs. 1–2
1–1 vs. 1–3	2–2 vs. 1–3	3–3 vs. 1–3
1–1 vs. 1–4	2–2 vs. 1–4	3–3 vs. 1–4

placing an object at a depth with respect to another object). Our subjective evaluation may cause rater bias in the results due to the individual characteristics of the subjects; however, our within-subject experiment design and statistical analysis decrease the effect of such bias.

Our main approach is to compare differences of two cases: one of the cases is altering both left and right views in the same scale, and the other case is altering only one of the views with a different scale, using the methods in Table 1.

For the methods that can be applied in different scales, the first level is not modifying, and the fourth level is applying the method with the greatest strength. The actual correspondence of the scales for each method is shown in Table 1. For these methods, comparison cases are shown in Table 2. In the table, each cell corresponds to a comparison case. For instance, 1–1 vs. 1–2 stands for the comparison of reference content in which the method is applied with level 1 for both views, with the test content in which the method is applied with level 1 for one view and with level 2 for the other view. Thus, there are 9 cases in total for each scalable method.

For the methods with two levels, there are two options: applying the method or not. The details of this type of methods are also shown in Table 1, and the comparison cases for these methods are shown in Table 3. In this table, ‘on’ stands for applying the method and ‘off’ stands for not applying the method. On the right side of the table, levels for the mixed shading method are shown.

**Table 3**

Test cases for non-scalable methods.

Ref vs. test	Ref vs. test
on–on vs. on–off	phong–phong vs. phong–gouraud
off–off vs. on–off	gouraud–gouraud vs. phong–gouraud

#### 4.1. Subjects

We have recruited 61 subjects: 47 males and 14 females with a mean age of 24.6. The subjects were among voluntary undergraduate and graduate students with computer science background; and most of them do not have previous experience on rendering on stereoscopic displays. The subjects were not informed about the purpose of the experiment. All have self-reported normal or corrected vision.

#### 4.2. Display

We have used a Sharp Actius AL3DU stereoscopic laptop, which has an NVIDIA GeForce Go 6600 graphics processor and a 15-inch XGA (1024 × 768) TFT 3D LCD display. In this display, the 3D effect is provided by the parallax barrier technology. 3D perception is available for a single viewer, in a limited view angle and in a limited distance.

#### 4.3. Procedure

For the experiment design, we have followed the double-stimulus continuous-quality scale (DSCQS) method [30]. According to this procedure, subjects were shown a content, either test or reference, for about 10 s; after a 3 s break, they were shown the other content. Then, both contents were shown for the second time, to obtain the subjective evaluations. The order of the reference and the test contents is determined randomly and subjects do not know whether they see the reference or the test content first. This process is illustrated in Fig. 7.

Our rating scale was also consistent with the scale that is recommended in ITU-R 500 [30]. A screenshot from our experiment system is shown in Fig. 8. The rating scale is continuous and red pixels show the subject's rating for the corresponding field. In order to guide the users and prevent them from grading

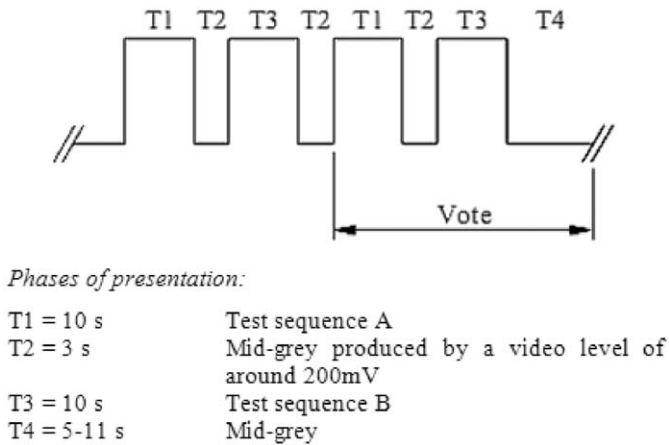


Fig. 7. Presentation of test material [30].



Fig. 8. Rating scales used for subjective assessments.

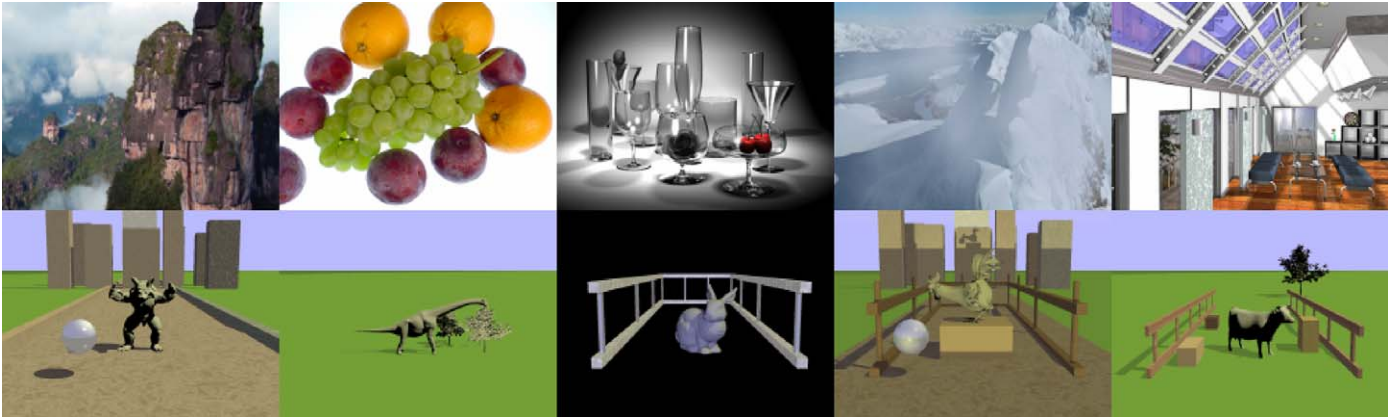


Fig. 9. Sample contents that are shown to the subjects (Top: samples from test images, bottom: samples from test videos.).

inconsistently, the rating scale provides equally-sized major intervals, which are labeled as “bad”, “poor”, “fair”, “good”, and “excellent”, and minor intervals inbetween.

In the experiment, there are 3 methods with 9 comparison cases and 4 methods with 2 comparison cases (Tables 1–3). Thus, in total, there are 35 ( $=9 \times 3 + 2 \times 4$ ) different cases to be evaluated. Either 10 different videos or 25 different images were used per case on average (A sample of the test images are shown in Fig. 9.). Each case is tested 15–20 times in total. For this purpose, each subject performed the experiment for about 30 min, including a training case at the beginning and a 5 min break in the middle of the experiment.

4.4. Assessment of contents

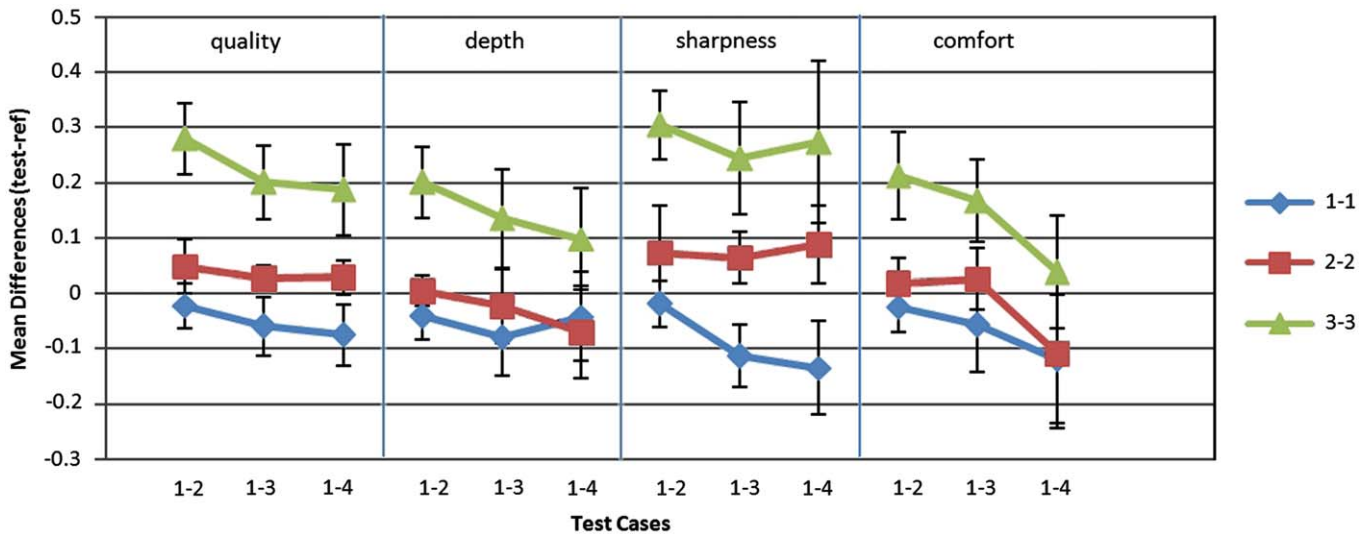
Subjects evaluated both test and reference contents of all the cases separately, with respect to four criteria, as shown in Fig. 8. These four criteria are commonly used in the perceptual evaluation of stereoscopic contents [18,31,32]. The meaning of each criterion was explained to the viewers before the experiment begins. The motivation behind selecting these grading criteria is as follows:

**Quality:** The primary goal of the experiment is to compare the quality of test and reference pairs. Quality denotes the perceived overall visual quality of the shown content.

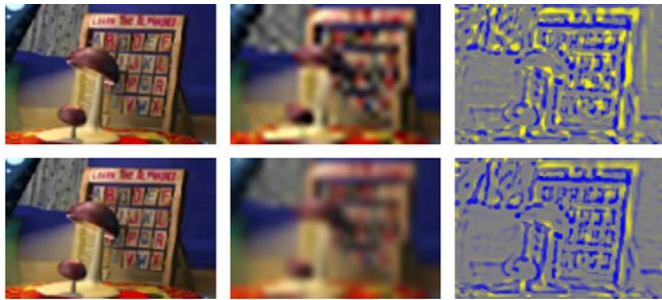
**Depth:** This criterion measures the apparent depth as reported by the user. Since stereoscopic displays are most beneficial for providing a better sense of depth, the effect of various modifications on apparent depth should be taken into account. For instance; if any modification that seems to retain quality causes a considerable amount of decrease in depth perception, the proposed optimization approach would be ineffective.

**Sharpness:** This criterion is the subjective clarity of the details in an image, which is an important factor while evaluating graphical or image-based contents [18]. Sharpness has also been reported to be well correlated with the quality of the contents[32,33].

**Comfort:** This criterion measures how distracting the scene is to the users. The visual comfort of a stereoscopic content may be affected by a number of factors such as left/right image misalignment, bad content creation, convergence–accommodation conflict and difficulty of getting the correct viewing position [34]. Therefore, the perceived comfort of a stereoscopic content is also reported as important and widely-used as a criterion in subjective experiments for stereoscopic displays [34,35]. Similar to depth criterion, the resulting comfort of an applied method should be taken into consideration to find out whether it is affected by our optimization approach.



**Fig. 10.** Experimental results for framebuffer upsampling method. Levels: 1—original, 2— $\frac{1}{2}$  size, 3— $\frac{1}{4}$  size, 4— $\frac{1}{8}$  size (Error bars show the 95% confidence interval of the mean.).



**Fig. 11.** Comparison of upsampling algorithms. Top-left & bottom-left: original image, top-middle: level 2 image upsampled with Lanczos filter, top-right: intensity contrast change map of top-middle image (compared to the original image), bottom-middle: level 2 image upsampled with B-Spline filter, bottom-right: intensity contrast change map of bottom-middle image (compared to the original image).

## 5. Experimental results and discussion

To determine the difference between the reference and test content, the *Test minus Reference* score was used. A score of zero means that the test sequence was rated equivalently to the reference sequence, and a negative score means that the test content was rated lower than the reference content. Error bars in the figures below show the 95% confidence interval of the mean, which corresponds to the range within which the mean is expected to fall with 95% certainty. Data points in the non-overlapping error bars indicate statistical difference at the  $p < 0.05$  level. For each of the 35 test cases, a *paired samples t-test* was also applied; with  $p < 0.05$  level to represent statistically significant difference between the pairs.

The parallax-barrier display that was used in the experiment has a narrow range of correct viewing position. Occasionally, this may cause a difficulty in getting the proper viewing position and leads to abnormal ratings for the contents that are seen from the wrong viewpoint. Therefore, it is likely to have outliers among the experimental results. The outliers were detected as the ratings that lie outside the region of  $mean \pm 2 \times std\ dev$  [36], and the cases reported by subjects.

### 5.1. Framebuffer upsampling

Fig. 10 shows the experimental results of framebuffer upsampling method. The quality and sharpness results show that all mixed pairs (1–2, 1–3, and 1–4) are perceived better than 2–2 pair, without any loss of apparent depth and comfort. The differences in sharpness are even larger for 3–3 pair, which is consistent with the hypothesis. Furthermore, the 1–2 pair is close to 1–1 pair in all rating criteria. On the other hand, the 1–1 pair statistically differs from 1–3 and 1–4 pairs, which may be the result of the Lanczos algorithm that was used for upsampling. Therefore, other upsampling algorithms, such as B-Spline based methods, may fit better to higher level of simplification; since they result in upsampled images that have lower intensity contrast (Fig. 11).

This pattern of results suggests that memory and computation savings can be achieved with application of the framebuffer upsampling method in Level 2 to one view, with equivalent perceived quality and sharpness. Moreover, the cost of advanced rendering techniques, such as real-time ray tracing, can be decreased for one view by using framebuffer upsampling.

### 5.2. Mixed-level antialiasing

Fig. 12 shows that the mixed pair for antialiasing was rated close in overall quality to the non-antialiased pair, whereas antialiasing both views show higher quality than the mixed pair. The results for perceived sharpness and comfort show a similar pattern to results of perceived quality. On the other hand, the results show that antialiasing has no effect on the apparent depth.

These results indicate that the antialiased image is suppressed by the non-antialiased image. This outcome is in accordance with our expectations, since antialiasing decreases the intensity contrast. Thus, turning off antialiasing in only one view is not appropriate, and it should only be applied or disabled on both of the views.

### 5.3. Specular highlight

The results for the specular highlight method are shown in Fig. 13. It is apparent that the perceived quality, depth, sharpness, and comfort of the mixed pair were rated similarly to the reference pair, while the quality of the pair with specular

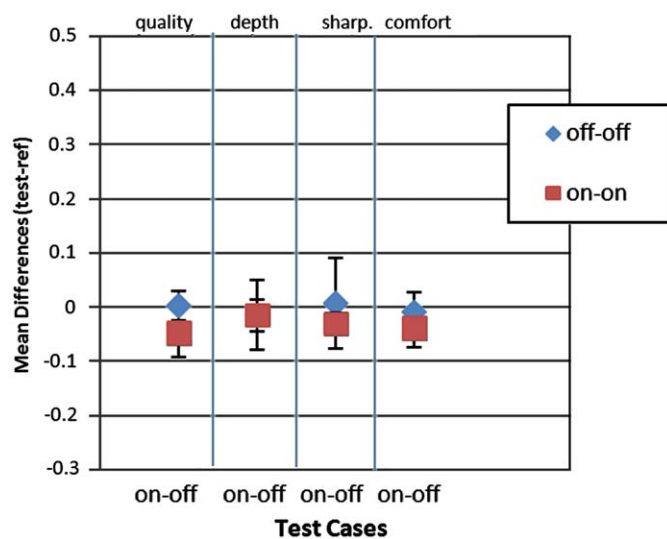


Fig. 12. Experimental results for mixed-level antialiasing method. Levels: on—antialiased, off—not antialiased (Error bars show the 95% confidence interval of the mean.).

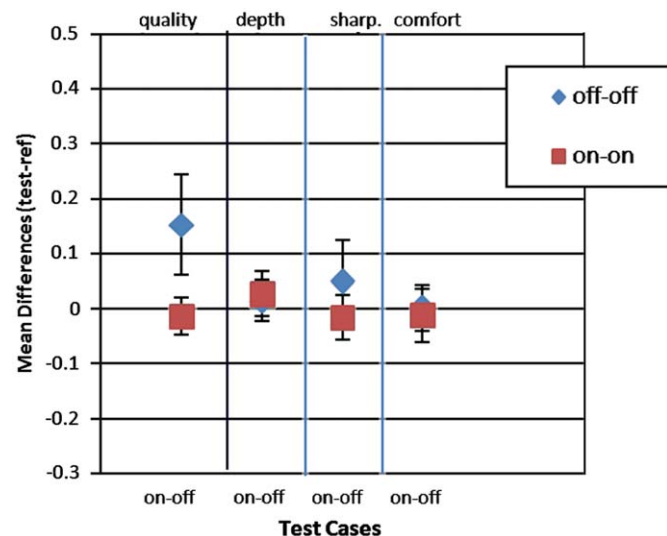


Fig. 13. Experimental results for specular highlight method. Levels: on—specular highlight used, off—specular highlight not used (Error bars show the 95% confidence interval of the mean.).

highlight removed in both views was rated lower than the mixed pair. Therefore, removing specular highlight in one view in a stereo pair has a small effect on the overall stereo image and thus can be used for mixed stereo rendering.

#### 5.4. Mixed shading

Fig. 14 shows that the mixed shading pair was rated with equivalent results to Phong-shaded pair, and with higher quality than the Gouraud shaded pair. The explanation of this result is that the intensity contrast is increased when Phong shading is used instead of Gouraud shading, because the specular highlight is more visible in Phong shading.

These results indicate that mixed shading provides a viable alternative for stereoscopic rendering. Nevertheless, the situation will probably not be the same when “extreme” shading methods, such as flat shading, are used for one view. Since flat shading

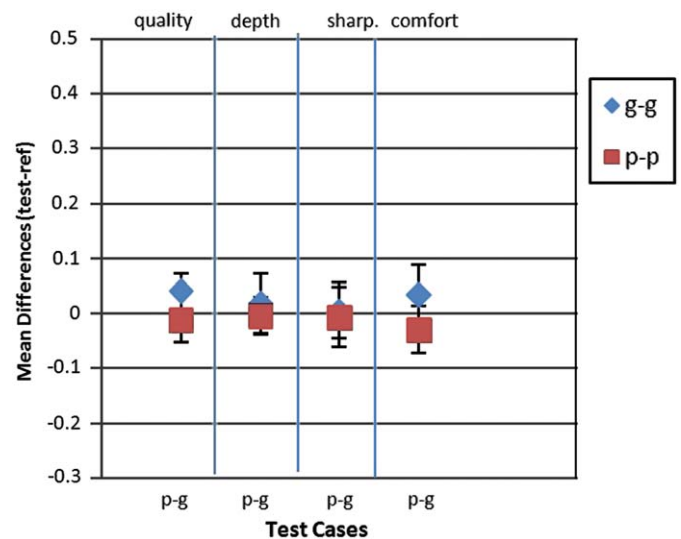


Fig. 14. Experimental results for mixed shading method. Levels: phong—phong shaded, gouraud—gouraud shaded (Error bars show the 95% confidence interval of the mean.).

increases the intensity contrast by resulting in a color discontinuity on the edges, flat shaded view will be dominant against the Phong shaded view on the edges.

#### 5.5. Mesh simplification

The results for the mesh simplification method are shown in Fig. 15. It is not possible to obtain a general inference by looking at the quality, depth, sharpness, and comfort results. This situation is not in conflict with our prediction, which is based on the idea that intensity contrast is higher on the edges (especially on the silhouette edges) and each mesh will probably dominate on its own edges. Therefore, the perceived 3D mesh is likely not one of the meshes, but an unpredictable combination of them.

As a result, it is not easy to predict the effect of a mixed pair on the combined percept, while using meshes of different level of detail for each view is not appropriate. However, simplifying the mesh for a single view may be applicable if the silhouette is preserved. One possible improvement may be application of a silhouette-preserving mesh simplification method which does not cause a significant increase in the intensity contrast of the simplified mesh.

#### 5.6. Texture resampling

Fig. 16 shows that the quality responses of the 1–2 and 1–3 pairs were close to the original 1–1 pair. All the mixed pairs were rated higher than the 3–3 pair, both for quality and sharpness. These results meet our expectations. On the other hand, the ratings for the 2–2 pair contradict our predictions. Our expectation was that the 2–2 pair would be rated lower than the mixed pairs, for which the quality is determined by the level 1 view according to our hypothesis. A likely explanation of this contradictory situation is as follows: In our experiment, we observed that the resolution of the level 2 texture maps were already sufficient for our objects since the area to cover is smaller than the size of the level 2 texture maps; so that the level 1 texture maps cannot provide higher quality than the level 2 texture maps. In this regard, if we consider the level 1 and level 2 texture maps as similar in final rendered image, the results seem to be consistent with our expectation. In conclusion, the texture resampling method provides a viable solution for stereoscopic rendering.

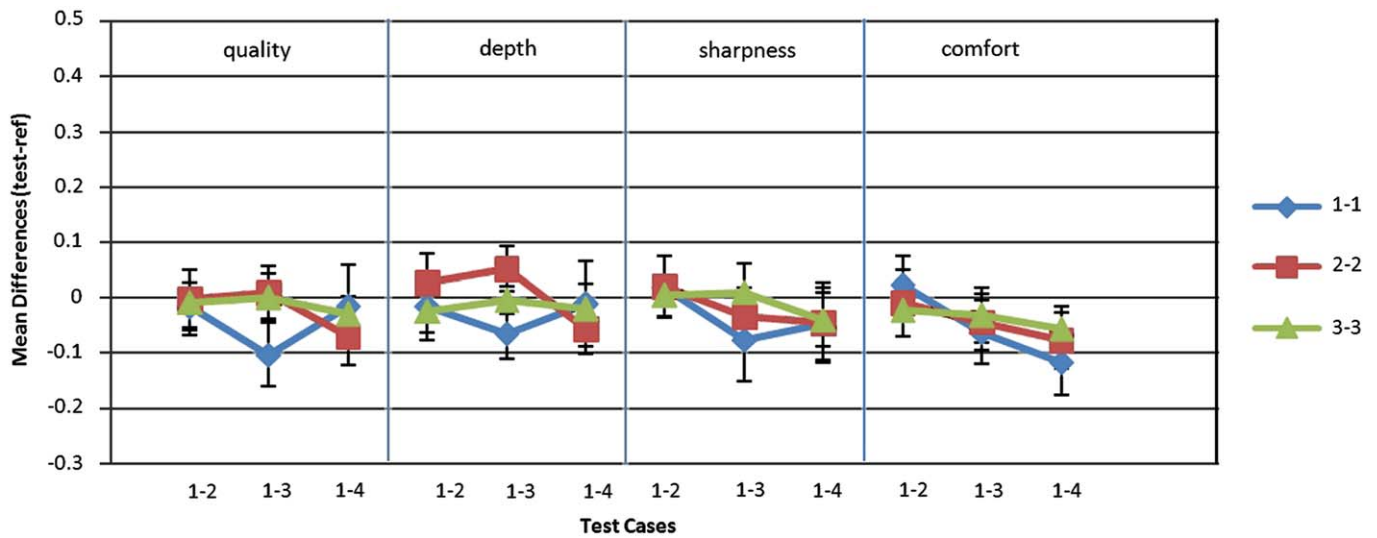


Fig. 15. Experimental results for mesh simplification method. Levels: 1—original mesh, 2—simplified to  $\frac{1}{2}$  of original face number, 3—simplified to  $\frac{1}{4}$  of original face number, 4—simplified to  $\frac{1}{8}$  of original face number (Error bars show the 95% confidence interval of the mean.).

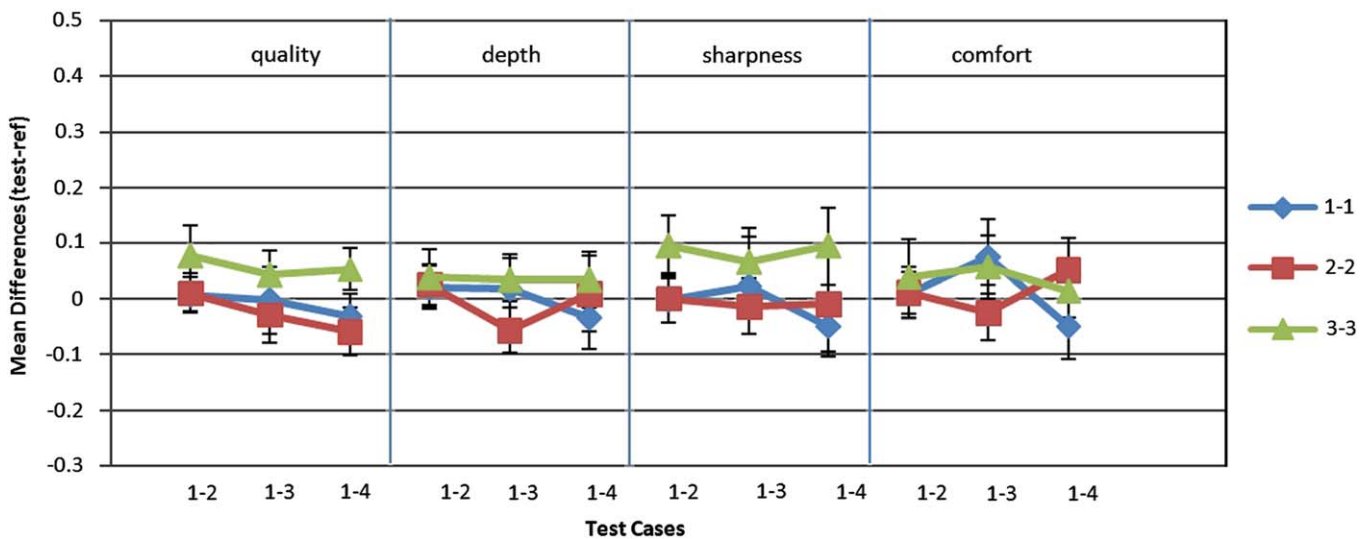


Fig. 16. Experimental results for texture resampling method. Levels: 1—texture size  $512 \times 512$ , 2—texture size  $256 \times 256$ , 3—texture size  $128 \times 128$ , 4—texture size  $64 \times 64$  (Error bars show the 95% confidence interval of the mean.).

### 5.7. Mixed shadowing

Fig. 17 illustrates that the mixed pair has closer ratings for quality to the case in which both views are not shadowed. Furthermore, the original reference pair has significantly higher quality ratings than the mixed pair. These two results imply that using shadows in only one view is not a feasible solution, as it affects the perceived quality.

Another result inferred from the sharpness ratings is that shadowed (on-on) and shadowless (off-off) reference pairs are sharper than the mixed pair, and the difference is more apparent while comparing the shadowed reference pair to the mixed pair. This situation may be explained as follows: in a mixed pair, right and left views may become dominant on different regions and this decreases the sharpness of the perceived stimulus.

Consequently, using shadow for a single view is not appropriate since it does not increase the quality and depth perception to a higher level than the pair without shadows. A mixed pair is rated to have lower comfort and sharpness than the reference pairs.

### 5.8. Discussion

Table 4 summarizes the feasibility of using the selected methods for mixed stereoscopic rendering, thus decreasing rendering complexity. Our experimental results show that it is possible to decrease the rendering cost of a 3D frame using methods: *framebuffer upsampling*, *specular highlight*, *mixed shading*, and *texture resampling*. However, this optimization approach is not feasible for effects such as *mixed-level antialiasing* and *mixed shadowing*.

Our hypothesis suggests that using different stimuli for each eye can be used for optimization purposes if the applied effect decreases the intensity contrast and as a result the high-quality view dominates in the mixed pair. In the meantime, one important point to consider is that the difference in levels between the two views should not be increased significantly. For example, higher levels of upsampling in one view decrease the perceived quality and depth, as the experiment results have shown.

## 6. Performance results

In this section, we further demonstrate the performance gain of the mixed-stereoscopic rendering approach, using the methods indicated as feasible in Table 4.

The performance gain of the mixed stereo approach over traditional stereoscopic rendering depends on the choice of the

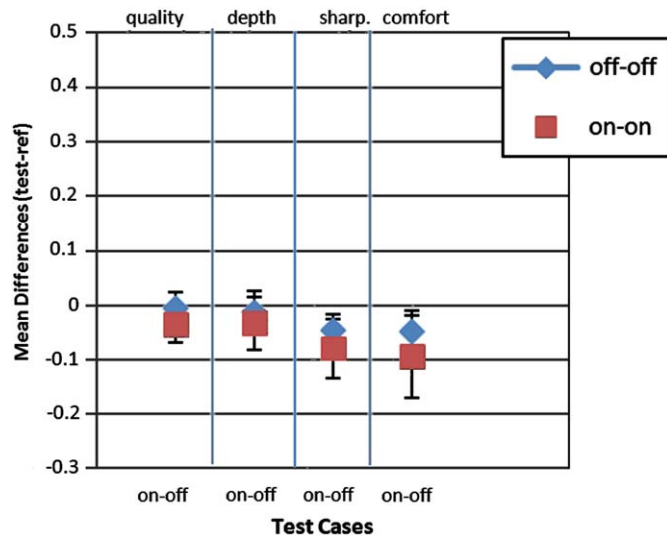


Fig. 17. Experimental results for mixed shadowing method. Levels: on—shadows are used, off—shadows are not used (Error bars show the 95% confidence interval of the mean.).

Table 4  
Summary of the experiment.

Method	Expectation	Applying to single view
Framebuffer upsampling	Upsampled view is suppressed	Feasible
Mixed-level antialiasing	Antialiased view is suppressed	Not feasible
Specular highlight	The view with specular highlight suppresses the other	Feasible
Mixed shading	Shaded with phong model suppresses the gouraud in general	Feasible (for phong–gouraud case)
Mesh simplification	Two meshes may not be perceived as a single mesh. Not appropriate to use	Not feasible
Texture resampling	Texture mapped with higher resolution image suppresses the other	Feasible
Mixed shadowing	Silhouette of shadows become apparent but may result in discomfort since brighter parts suppress inside the shadowed regions	Not feasible

Table 5  
Performance gains of the methods that are tested (The times shown in the table are the total times for rendering both left and right images of a frame. The gain columns indicate the stereoscopic rendering gain obtained by our mixed rendering approach over the traditional method.).

Method	Approach	Scene 1		Scene 2		Scene 3		Scene 4	
		Time (ms)	Gain	Time (ms)	Gain	Time (ms)	Gain	Time (ms)	Gain
Framebuffer upsampling	Traditional	6.8		9.48		63.9		445.1	
	Mixed	4.47	34%	7.74	18.3%	61.6	3.6%	441	0.9%
Framebuffer upsampling (Raytracing)	Traditional	7500		9820		17,400		40,600	
	Mixed	4950	34%	6520	33.6%	12,200	29.8%	27,300	32.8%
Mixed-shading	Traditional	6.8		9.48		64.5		445.1	
	Mixed	6.15	9.6%	8.65	8.7%	54	16.3%	361.2	18.8%
Specular highlight	Traditional	6.8		9.48		64.5		445.1	
	Mixed	6.57	3.4%	8.87	6.4%	56.1	13%	379	14.9%
Texture resampling	Traditional	8.53		16.08		60.68		95.24	
	Mixed	6.68	21.6%	12.67	21.1%	57.47	5.2%	93.2	2.1%
Texture resampling (Raytracing)	Traditional	16,900		20,840		46,680		75,320	
			5.5%		7.1%		8.3%		6.9%

method for rendering the scene. Using advanced rendering techniques, such as BRDF, area light sources, anisotropic texture filtering, etc. will increase the advantage of our approach over the traditional approach. However, in our performance measurements, we only use simple rendering for the traditional rendering case. Therefore, the performance gains demonstrated in this section can be considered as close to worst-case results. The best-case result of the performance gain for any method will be bound by 50%, as we render one of the views in high quality.

Table 5 contains the stereoscopic rendering times of a frame for traditional and mixed stereoscopic rendering approaches, along with the performance gains in percentages by the mixed stereoscopic rendering approach. We have measured the rendering times of each method for scenes of different densities. The complexity of the scenes increases from “scene 1” to “scene 4”; the number of polygons used in “scene 1” through “scene 4” are close to 4000, 40,000, 400,000 and 4 million, respectively.

**Framebuffer upsampling:** We have tested the performance gain obtained by the framebuffer upsampling method in both OpenGL rendered and ray-traced scenes, in which one view is rendered with full resolution and the other is rendered with quarter resolution. Afterwards, the smaller image is upsampled to match the original resolution. The ray-traced scene yields a significant amount of performance gain of 33%. In the OpenGL-rendering case, a similar performance gain of 34% is reached in scene 1, in which the rendering time is bound by the screen resolution. As the scenes get denser from scene 2 to scene 4, the scene complexity becomes more effective on determining the rendering time.

**Mixed shading:** We have implemented the mixed shading method using the GPU-based implementations of Phong and Gouraud shading models. In this method, an increase in the

performance gain is observed as the scene becomes denser to an extent. For denser scenes (scene 3 and scene 4), a performance gain of about 19% is reached; while for sparse scenes, approximately half of this gain is obtained.

**Specular highlight removal:** The specular removal method is tested on the Phong shading model – both views are rendered using the Phong shading model; while one of the views does not include the calculations for specular component. The performance gain with this method is 14% for dense scenes. As in the mixed shading case, the performance gain is more recognizable in the dense scenes.

**Texture resampling:** We have tested the texture resampling method in both OpenGL-rendered and ray-traced scenes; with textures with one-fourth width and height of the original texture for one view; and the original textures for the other view. The performance gains in the OpenGL case are higher for the sparse scenes in which each textured object covers larger area in the scene compared to the denser scenes. We infer that, as the number of different textures and the area of the textured objects in the scene increase, the performance gain obtained by our approach becomes more recognizable. In the ray-tracing case, the performance gain does not change much according to the scene complexity and is about at 7% over all of the scenes.

## 7. Conclusion & future work

In this paper, we have presented a perceptually-based optimization approach for stereoscopic rendering, which makes use of the binocular suppression mechanism of the human visual system. The proposed method exploits the fact that the 3D perception of the overall stereo pair in a region is determined by the dominant image on the corresponding region, instead of summation of the effect of two images. We have also introduced an estimate of the strength of a view, called *intensity contrast*, and used it to estimate whether the application of a method decreases the strength of that view. We have performed a subjective experiment on the selected methods and measured performance gains.

We conclude that decreasing the rendering cost for one view may be an effective technique to increase the rendering performance of 3D stereoscopic content, while retaining the depth, quality, and sharpness of the original 3D rendering. The following methods provide an effective solution: framebuffer upsampling, specular highlight, mixed shading, texture resampling. On the other hand, mixed-level antialiasing, mesh simplification, and mixed shadowing, produce unacceptably low levels of quality and sharpness.

Our research plans for the future include the investigation of further rendering and modeling solutions, such as silhouette preserving mesh simplification, techniques considering the effect of animation, methods for adding the effect of different methods, and the effects of longer term viewing of mixed sequences. Furthermore, our plans include the application of our solution in combination with other stereoscopic rendering solutions, which were described in Section 2.1. Lastly, this approach can be extended for multi-view rendering, considering the different multi-view display technologies and their challenges.

## Acknowledgments

The authors were supported by the European Commission FP7-213349 All 3D Imaging Phone project and TUBITAK. Also, we would like to thank all the participants of the subjective experiment described in this work and Onur Kucuktunc for his help on preparation of the test videos.

## References

- [1] Segal M, Akeley K. The OpenGL Graphics System: A Specification (Version 2.0); 2004.
- [2] Adelson SJ, Bentley JE, Chong IS, Hodges LF, Winograd J. Simultaneous generation of stereoscopic views. *Computer Graphics Forum* 1991;10(1): 3–10.
- [3] Hasselgren J, Akenine-Moller T. An efficient multi-view rasterization architecture. *Eurographics Symposium on Rendering* 2006.
- [4] Kalaiah A, Capin T. Unified rendering pipeline for autostereoscopic displays. *3DTV Conference* 2007.
- [5] Fu S, Bao H, Peng Q. Accelerated rendering algorithm for stereoscopic display. *Computers & Graphics* 1996;22:3–9.
- [6] Wan M, Zhang N, Qu H, Kaufman AE. Interactive stereoscopic rendering of volumetric environments. *IEEE Transactions on Visualization and Computer Graphics* 2004;15:28.
- [7] Fehn C. Depth-image based rendering (DIBR), compression and transmission for a new approach on 3DTV. *SPIE* 2004.
- [8] Howard IP, Rogers BJ. Binocular fusion and rivalry. *Binocular Vision and Stereopsis*. New York: Oxford Univ. Press; 1995.
- [9] Blake R, Logothetis NK. Visual competition. *Nature Reviews Neuroscience* 2002;13–21.
- [10] Adelson S, Hodges L. Stereoscopic ray tracing. *The Visual Computer* 1993;10(3):127–44.
- [11] Zhang L, TW J. Stereoscopic image generation based on depth images for 3D TV. *IEEE Transactions on Broadcasting* 2005;191–9.
- [12] Halle M. Multiple viewpoint rendering. *International Conference on Computer Graphics and Interactive Techniques*, 1998; New York.
- [13] He T, Kaufman A. Fast stereo volume rendering. *IEEE Visualization* 1996: 49–57.
- [14] Asher H. Suppression theory of binocular vision. *British Journal of Ophthalmology* 1953;37:37–49.
- [15] Blake R, Camisa J. The inhibitory nature of binocular rivalry suppression. *Journal of Experimental Psychology* 1979;5:315–23.
- [16] Hollins M, Leung EHL. The influence of color on binocular rivalry. *Visual Psychophysics and Physiology* 1978;181–90.
- [17] Perkins MG. Data compression of stereopairs. *IEEE Transactions on Communications* April 1992;40:684–96.
- [18] Stelmach L, Tam WJ, Meegan D, Vincent A. Stereo image quality: effects of mixed spatio-temporal resolution. *IEEE Transactions on Circuit and Systems for Video Technology* 2000;188–93.
- [19] Turkowski K, Gabriel S. Filters for common resampling tasks. *Graphics Gems I*. Academic Press; 1990.
- [20] Akenine-Moller T, Haines E, Hoffman N. In: *Aliasing And Antialiasing. Real-Time Rendering*, 3rd ed.. Wellesley: AK Peters; 2008.
- [21] Ngan A, Durand F, Matusik W. Experimental analysis of BRDF models. *Eurographics Symposium on Rendering* 2005.
- [22] Hearn D, Baker MP. Basic illumination models. *Computer Graphics With Opengl*, 3rd ed. New Jersey: Prentice Hall; 2004.
- [23] Phong BT. Illumination for computer generated pictures. *Communications of the ACM* 1975;18(6):311–7.
- [24] Gouraud H. Continuous shading of curved surfaces. *IEEE Transaction on Computer* 1971;20:623–9.
- [25] Garland M, Heckbert PS. Surface simplification using quadric error metrics. *Proceedings of the SIGGRAPH* 1997 1997.
- [26] Hearn D, Baker MP. Texture mapping. *Computer Graphics with OpenGL*, 3rd ed.. New Jersey: Prentice Hall; 2004.
- [27] Akenine-Moller T, Haines E, Hoffman N. *Shadows. Real-Time Rendering*, 3rd ed. Wellesley: Ak Peters; 2008.
- [28] Itti L, Koch C. Computational modeling of visual attention. *Nature Reviews Neuroscience* 2001;194–203.
- [29] Itti L, Koch CA. Saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* 2000;1489–506.
- [30] BT.500-11 IRR. Methodology for the subjective assessment of the quality of television pictures. Geneva 2002.
- [31] Seuntjens P, Meesters L, Ijsselstein W. Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation. *ACM Transactions on Applied Perception* 2006;3(2): 95–109.
- [32] Tam WJ, Stelmach LB, Coriveau P. Psychovisual aspects of viewing stereoscopic video sequences. *Stereoscopic Displays and Virtual Reality Systems*, 1998.
- [33] Cavedes J, Oberti FA. New sharpness metric based on local kurtosis, edge and energy information. *Signal Processing: Image Communication* February 2004;19(2):147–61.
- [34] Kooi FL, Toet A. Visual comfort of binocular and 3D displays. *Displays* 2004;25:99–108.
- [35] Yano S, Ide S, Mitsuhashi T, Thwaites H. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays* 2002;23:191–201.
- [36] Selst MV, Jolicoeur PA. Solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology* 1994;631–50.