A Perceptual Evaluation of 3D Unsharp Masking

Matthias Ihrke[a,b], Tobias Ritschel[c], Kaleigh Smith[c]
Thorsten Grosch[c], Karol Myszkowski[c], Hans-Peter Seidel[c]

[a]Bernstein Center for Computational Neuroscience, Bunsenstrasse 10, Göttingen, Germany;
[b]Max-Planck-Institute for Dynamics and Self-Organization, Bunsenstrasse 10, Göttingen, Germany;
[c]Max-Planck-Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany;

ABSTRACT

Much research has gone into developing methods for enhancing the contrast of displayed 3D scenes. In the current study, we investigated the perceptual impact of an algorithm recently proposed by Ritschel et al.[1] that provides a general technique for enhancing the perceived contrast in synthesized scenes. Their algorithm extends traditional image-based Unsharp Masking to a 3D scene, achieving a scene-coherent enhancement. We conducted a standardized perceptual experiment to test the proposition that a 3D unsharp enhanced scene was superior to the original scene in terms of perceived contrast and preference. Furthermore, the impact of different settings of the algorithm's main parameters enhancement-strength (λ) and gradient size (σ) were studied in order to provide an estimate of a reasonable parameter space for the method. All participants preferred a clearly visible enhancement over the original, non-enhanced scenes and the setting for objectionable enhancement was far above the preferred settings. The effect of the gradient size σ was negligible. The general pattern found for the parameters provides a useful guideline for designers when making use of 3D Unsharp Masking: as a rule of thumb they can easily determine the strength for which they start to perceive an enhancement and use twice this value for a good effect. Since the value for objectionable results was twice as large again, artifacts should not impose restrictions on the applicability of this rule.

Keywords: 3D Unsharp Masking, scene-coherent enhancement, Cornsweet Illusion

1. INTRODUCTION

Contrast enhancement is a standard and relatively well-researched 2D image processing operation, which can greatly improve the overall image appearance.[2] In the context of 3D computer graphics, and in particular 3D image synthesis, such contrast enhancement has been traditionally performed in the post-processing stage over rendered images, typically as a part of the so-called tone mapping operation.[3] Recently, some successful efforts towards developing algorithms working directly in the 3D object space (instead of traditional 2D image space), which lead to contrast enhancement have been proposed.[1,4-6] Enhancing the contrast of 3D virtual scenes in a way that aids to their understanding and interpretation as well as increases their visual appeal is a worthwhile and challenging task. Especially because of the many potential applications (e.g. medical imaging, computer games, computer-aided design), a better understanding to what extent perceptual improvements in terms of better contrast perception can be achieved using scene-enhancement is of great interest. However, controlled psychophysical studies specifically investigating the perceptual effect of those inherently 3D techniques have not been performed so far. In previous studies for 2D image enhancement, Lin et al.[7,8] could show that local contrast enhancement does result in a preferred overall impression of the modified images. They evaluated the perceptual quality of images in terms of a "most eye-pleasing sharpness" which they defined as the parameter setting

---

Figure 1. Traditional vs. 3D Unsharp Masking. From left to right: Original Image; Enhanced version by traditional unsharp masking ($\lambda = 1.0, \sigma = 8.0$); Enhanced version by 3D Unsharp Masking. Note that traditional Unsharp Masking overdoes the enhancement in some regions breaking scene-coherence (right). The 3D version preserves scene-coherence.

resulting in the strongest positive deviation of the enhanced image from the original. Images where enhanced by the Laplacian-of-Gaussian (LoG) method and it was shown that their perceptual quality was highest when edge sharpness was increased 2.6 times the just-noticeable difference. Other studies in visual perception show that in addition to sharpness, perceived image contrast depends significantly on attributes such as lightness, hue, and saturation.[9] In any case, image preference increases when perceived contrast increases up to an upper threshold, but then decreases when the enhancement is deemed too strong.[10]

In the current paper, we intended to systematically evaluate the perceptual impact of a recent algorithm proposed by Ritschel et al.[1] that provides a general technique for enhancing the contrast in synthesized scenes. The algorithm is based on traditional image-based Unsharp Masking and extends this technique to 3D scenes (see section 1.1). The technique is particularly interesting from a perceptual perspective, because its effect can be related to a well-known illusion from psychophysical literature, the Cornsweet effect[11] (see section 1.2) which describes an enhancement in perceived contrast in regions separated by a gradient of a specific shape.

In this study we selected 3D Unsharp Masking, which is more general than other 3D techniques that specifically manipulated normal vectors,[4] local light source positions,[5] and depth maps.[6] For more extensive analysis of these methods as well as visual comparison of resulting image enhancement please refer to the paper by Ritschel et al.[1]

1.1 3D Unsharp Masking

3D Unsharp Masking is an extension of the well-known image-based Unsharp Masking technique[12] that is directly operating on 3D scenes. Technically, generic Unsharp Masking describes the process of increasing the high-frequency components of a signal by adding back the high-frequency components of the signal itself. Formally, Unsharp Masking $\mathcal{U}$ of a signal $\mathcal{S}$ is defined as:

$$\mathcal{U}(\mathcal{S}) = \mathcal{S} + \lambda \mathcal{C}(\mathcal{S})$$

where $\mathcal{C}(\mathcal{S}) = \mathcal{S} - \mathcal{S}_\sigma$ is the contrast signal containing a high-pass filtered version of the original signal. Here, λ is the gain, controlling the strength of the adjustment. σ is the size of the kernel applied for smoothing the blurred image used to generate the contrast signal. Therefore, σ can be interpreted as the range of frequencies that is enhanced by Unsharp Masking. In traditional Unsharp Masking, the signal $\mathcal{S}$ is simply an image, while Ritschel et al.'s method implements generic Unsharp Masking on the reflected light over the mesh surface. This allows the inclusion of information about the structure of the displayed scene (e.g. occlusion, shadows etc.) into the calculation of the enhancement, potentially enabling a more scene-coherent enhancement.

For illustrative purposes, consider the comparison of the two methods given in Figure 1. While traditional Unsharp Masking (b) overemphasizes details that are not relevant to understanding the scene in (a), 3D Unsharp Masking does not suffer from these pitfalls, but rather adapts the enhancement to the underlying 3D structures, thus avoiding over-enhancements. An important advantage of the 3D technique is its suitability for handling animations, while 2D Unsharp Masking may lead to numerous artifacts and inconsistencies when applied to video. The paper by Ritschel et al.[1] and the accompanying Web page illustrate in the full extent all issues discussed here: `http://www.mpi-inf.mpg.de/resources/3DUnsharpMasking`.

Figure 2. The Cornsweet Illusion. The actual gradient (right), when added to the border between two nearly equiluminant regions induces the illusion of contrast (right). When the gradient is covered by a patch, the contrast is reduced again (middle).

1.2 Perceptual Background

The apparent perception of contrast is modulated by many factors. The actual difference in luminance between adjacent regions used to objectively define patterns is perceptually modulated by context-cues (e.g. simultaneous contrast[13]) and the characteristics of the border between the two regions.[14] Among the multifaceted effects that borders of different profiles can produce, the Cornsweet profile takes a special role because it can achieve the impression of a brightness difference between two regions that are actually of the same luminance (see Fig. 2). It is achieved by introducing opposing gradients at the border where two areas meet, thereby sharpening the edges. This effect is exclusively due to the border but affects even relatively large areas. In a perceptual study by Purves and Lotto,[15] it was argued that this effect was a consequence of the human visual system (HVS) being optimized to make sense of 2D information from the retina in a 3D context: the HVS interprets the 2D retinal information in terms of how this information could be acquired from a three-dimensional scene. In agreement with this hypothesis, the authors could show, that in settings where the gradient can be interpreted as stemming from differences in illumination or reflectance, the perceptual effect is much stronger.

In previous work in Computer Graphics literature, it was recognized, that image-based unsharp masking produces profiles at edges that are similar to the Cornsweet profile.[16] By calculating the Unsharp Masking on the lighting-function, the Cornsweet gradient is effectively adapted to 3D scene features in the image plane. For example, the gradient becomes less broad with distance to viewer depending on the orientation of the objects in the scene. This can be thought of as an implementation of an optimized enhancement: the HVS evaluates the brightness of areas depending on the shape of the edge gradient that is assumed to stem from the arrangement and illumination of structures in the 3D scene. By letting the gradient vary according to the lighting function it is optimized to produce the impression of being caused by illumination. In view of Purves et al.'s[15,17] theory, the prediction that such gradients are more effective than non-adaptive gradients can be derived.

1.3 Goal

The purpose of the current study was to empirically evaluate how the scene-enhancement method discussed in Sec. 1.1 modifies the perceived quality of images of 3D scenes. Of main interest was the impact that the parameters of the algorithm has on the enhancement. In the next sections, we describe our psychophysical experiment followed by a detailed summary, analysis and discussion of the results. Finally, we discuss and conclude.

2. METHOD

In this paper, we present a perceptual experiment that was conducted to test the proposition that a 3D unsharp enhanced scene was superior to the original scene in terms of perceived contrast and preference. Furthermore, the impact of different settings of the algorithm's main parameters enhancement-strength (λ) and gradient width (σ) were studied in order to provide an estimate of a reasonable parameter space for the method. Supporting this goal, we pushed the algorithm to its limits by overenhancing images until their appearance was objectionable to most observers to find an upper bound for the enhancement. Finally, to assess whether the enhancement was indeed perceived at crucial locations (edges, figure-ground separation etc.) we asked subjects to indicate which features of the scenes benefited most from the enhancement.

2.1 Study Design

The standard psychophysical procedure "method of adjustment" was used[18] in the study. This was deemed to be the most appropriate choice for the current purposes since it allows both for an efficient implementation of the experiment and provides a single means of estimating all thresholds and preference values of interest. In this framework, the kernel width σ and the gain value λ (see Sec. 1.1) were considered by allowing the subjects to adjust λ for a given scene under different tasks and varying σ. Of special interest for determining a reasonable parameter range are those values for λ which result in a barely visible enhancement (λ_{jn}) and objectionable artifacts (λ_{obj}), respectively. This is because these values give a lower and an upper bound between which the parameters are free to vary assuming a monotonic relationship between parameter and perceived impact on the image. To find out about the absolute preference of the users, as well as to test the reliability of the procedure, the subjectively "best contrast" setting of the gain value (λ_{best}) for a given scene and a given σ was also obtained.

In our experimental setup, two stimuli were presented side-by-side, where both images depicted the same scene seen from the same viewpoint. One of the two images showed an enhanced version of the scene while the other showed the original, non-enhanced scene for comparison. The actual spatial layout of the displays was randomized across subjects in order not to confuse the subjects. The space of the kernel-width σ was discretized in three steps $\sigma \in \{\sigma_{\text{low}}, \sigma_{\text{medium}}, \sigma_{\text{high}}\}$ which were chosen in a pilot study to produce a perceptually low, medium and strong effect, respectively. For the 3 different settings of σ, the subjects were required to adjust λ according to the tasks outlined above and explicitly given in Appendix A.

Thus, the study implemented a 4×3 within-subject design by varying 4 scenes and the kernel size σ (low, medium, high) as independent variables. The gain values λ_{jn}, λ_{best} and λ_{obj} adjusted by the user under the 3 different instructions functioned as dependent variables. The experiment was divided into 32 experimental trials. The three steps of the kernel width σ were presented 2 times in randomized order for each scene in order to provide the possibility to assess the reliability of the method. The sequence of the scenes was also randomized across subjects. In addition, the first two trials for each scene were repeated as they functioned as practicing phase for each scene. The data from these first two trials did not enter in the final analysis.

2.2 Participants

15 participants (9 male, 6 female) with normal or corrected-to-normal vision took part in the experiments. Subjects were compensated for their efforts with a small fee (15 USD). Participants were recruited from the university campus, were mostly students of computer science and their mean age was 24 years (range 19 to 30 years). Care was taken to eliminate effects stemming from extensive training by ensuring that subjects were inexperienced in the field of computer graphics and photography (subjects with more than casual experience in any of these fields were excluded from the experiment). All subjects were naïve regarding the goal of the experiment.

2.3 Materials and Apparatus

All stimuli were presented at a resolution of 2048×1536 on a 20.8 inch (diagonal) Barco Coronis Color 3MP (MDCC 3120-DL) display that was connected to a personal computer running the Psychtoolbox software.[19] The monitor was viewed by the subjects orthogonally at a distance of 80 cm, the whole trial-display occupying 29.68 visual degrees. The stimuli were generated by rendering 4 different scenes and then enhancing them using the 3D Unsharp Masking algorithm.[1] The images were displayed on the screen at a resolution of 960×720 each. The four tested scenes are depicted in Figure 4 (first row). The increment of λ in two consecutive steps of the adjustments were chosen to be below the visibility threshold, that is subjects were not able to distinguish two images rendered with these values (as empirically verified in pilot studies). Furthermore, between two consecutive steps, a mask was displayed for 300 ms. A lowpass version of the image, blurred with a gaussian filter of size 50×50 pixels, was used as an effective mask. The mask was used in order to prevent subjects from judging the adjustment temporally (between successive, rapid steps of the enhancement) instead of comparing to the original image. This method was particularly necessary for the just-noticeable threshold, as the perceptual threshold would probably be lower when visual sensory (iconic) memory could have been used by the subjects.[20]
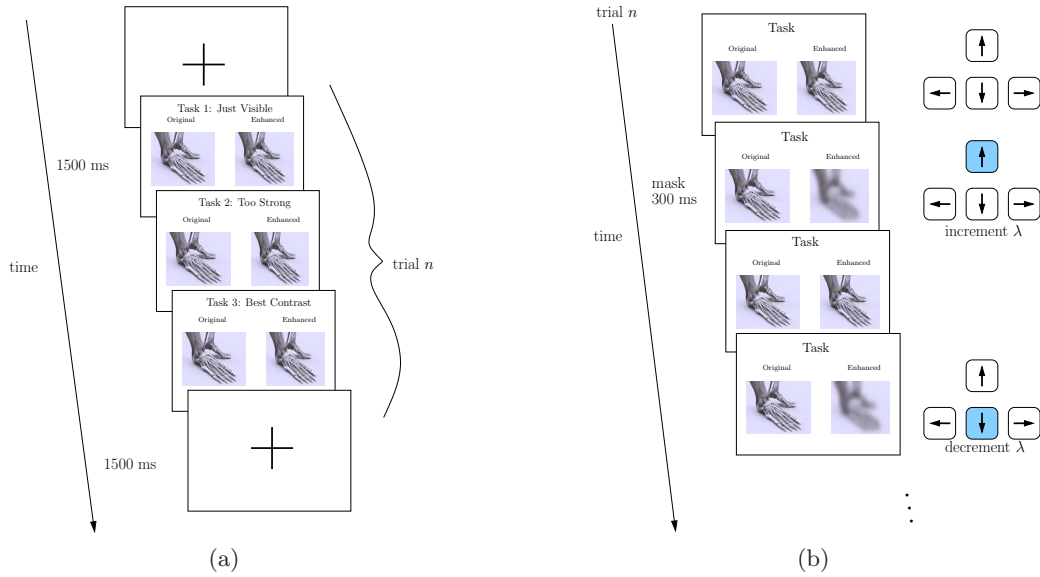
Figure 3. Experimental procedure for a single trial. (a) Each trial began with a fixation-cross after which subjects were required to adjust the contrast as indicated by λ for a given scene. The task for which λ was to be adjusted was indicated on top of the display. Subjects were not under time-pressure during this task. (b) When adjusting λ, the interval between consecutive images was masked. After each adjustment, participants indicated their choice by pressing a button and the next task appeared.

2.4 Procedure

On entering the laboratory, subjects were asked for their experience in the areas of computer graphics and/or photography. Subjects that judged their experience within one of these fields as "above average" were excluded from the study. The participants were seated in front of a monitor running the experimental software in an otherwise darkened room. They received standardized on-screen instructions regarding the procedure of the experiment and potentially remaining questions were answered by the experimenter. Before the actual experiment started, the participants were shown all scenes that were to appear in the course of the session in randomized order.

Each trial started with a fixation cross in the middle of the screen (1500 ms) that was followed by the trial-display, consisting of images of the original and the enhanced scene (see Fig. 3). The location of the enhanced image (left or right) was balanced across subjects. At the beginning of each trial, these images were identical as the gain value λ was initially set to 0. Subjects were asked to adjust the enhancement by pressing the up- and down-keys. A discrete beep indicated that the subjects reached either $\lambda = 0$ or the upper limit $\lambda = 3$. The participants confirmed their choice of the image by pressing the Return-key. For each display, the subject had to complete 3 tasks (exact wording of the instructions, see Appendix A):

1. adjust the image until there was a just noticeable difference in contrast between the original and the enhanced image,

2. adjust the image until it looked too strong and artifacts appeared,

3. find the subjectively best-contrast setting (this task was made more comprehensible by giving the analogy of adjusting the contrast of a TV-set).

The current task was indicated above the two images and changed whenever the subject completed a task. After the participants had completed the final 8th trial for each scene, they were presented with a hardcopy showing the original, non-enhanced image and asked to mark the parts that they used most frequently to evaluate the strength of the enhancement. The complete experimental session took approximately one hour. In spite of this

rather long time required to complete the experiment, subjects did not complain about a loss of concentration, an effect that might be due to the reportedly good sense of control induced by the adjustment procedure.

3. RESULTS

For a descriptive summary of the data, please refer to tables 1 and 2 where mean values of the adjustment strength λ computed over subjects are reproduced for all scenes and settings of σ. In addition, thumbnails of the images chosen by the subjects for all tasks are depicted in Figure 5 along with the corresponding values for the parameters.

Table 1. Summary of the results from the study. User-chosen λ values are given for each setting of σ, scene and task. Absolute values for λ are scene-dependent.

	User-adjusted λ								
	λ_{jn}			λ_{best}			λ_{obj}		
	σ_{low}	σ_{medium}	σ_{high}	σ_{low}	σ_{medium}	σ_{high}	σ_{low}	σ_{medium}	σ_{high}
Dice	0.05	0.06	0.05	0.11	0.11	0.10	0.22	0.22	0.23
Keys	0.25	0.22	0.21	0.34	0.31	0.26	0.75	0.63	0.56
Chamfer Plane	0.34	0.36	0.32	0.46	0.55	0.55	0.96	1.04	1.05
Feet	0.23	0.22	0.20	0.43	0.27	0.33	0.85	0.69	0.71

Table 2. Summary of the results from the study. User-chosen, rescaled $\hat{\lambda}$ values are given in JND units for each setting of σ, scene and task.

	User-adjusted $\hat{\lambda}$								
	$\hat{\lambda}_{\text{jn}}$			$\hat{\lambda}_{\text{best}}$			$\hat{\lambda}_{\text{obj}}$		
	σ_{low}	σ_{medium}	σ_{high}	σ_{low}	σ_{medium}	σ_{high}	σ_{low}	σ_{medium}	σ_{high}
Dice	1.00	1.00	1.00	1.81	1.82	2.28	4.83	4.20	5.55
Keys	1.00	1.00	1.00	1.48	1.46	1.09	3.39	2.89	2.73
Chamfer Plane	1.00	1.00	1.00	1.78	1.83	1.91	3.52	3.22	3.63
Feet	1.00	1.00	1.00	2.06	1.39	1.76	3.99	3.38	3.73

Investigating the reliability of our method-of-adjustment task was done by comparing the choice of the parameters for the two repetitions of the same task. We found large intra-rater correlations (all $r > .75$, all $p < .001$), indicating that subjects were consistent in their judgment and that our method was quite reliable. In order to decrease the impact of random fluctuations, the mean of the two repetitions is used in the further analysis as the dependent variable. Also, to correct for outliers, all values whose difference to the mean value (for each scene, σ and task) exceeded that of two times the standard deviation were replaced by the mean of the remaining values (less than 5% of the data). Because the gain value λ scales differently for all scenes, independent 3×3 ($\sigma \times$ task) ANOVAs were computed for each scene, treating both factors as repeated measures because they were obtained from the same subject.

The ANOVA for scene *Chamfer Plane* revealed only a main effect of task, $F(2, 28) = 58.99, p < .001$, all other main and interaction effects were non-significant. Using adjusted pairwise contrasts (following Holm's[21] proposal), the found effect can readily be shown to stem from the best contrast setting being larger than the lower threshold ($\lambda_{\text{jn}} < \lambda_{\text{best}}$: $t(14) = 2.14, p < .05$) and the higher threshold being larger than the preferred setting ($\lambda_{\text{best}} < \lambda_{\text{obj}}$: $t(14) = 4.03, p < .01$). As an indicator for the size of the effect, Cohen's d[22] was computed for all of the pairwise contrasts, yielding medium to strong effects according to Cohen's conventions[22] ($d(\lambda_{\text{jn}}, \lambda_{\text{best}}) = .78$, $d(\lambda_{\text{best}}, \lambda_{\text{obj}}) = 1.47$). The corresponding ANOVA for the *Dice* scene also reveals a single main effect of task, $F(2, 28) = 32.06, p < .001$. Again, pairwise comparisons indicate the pattern of results
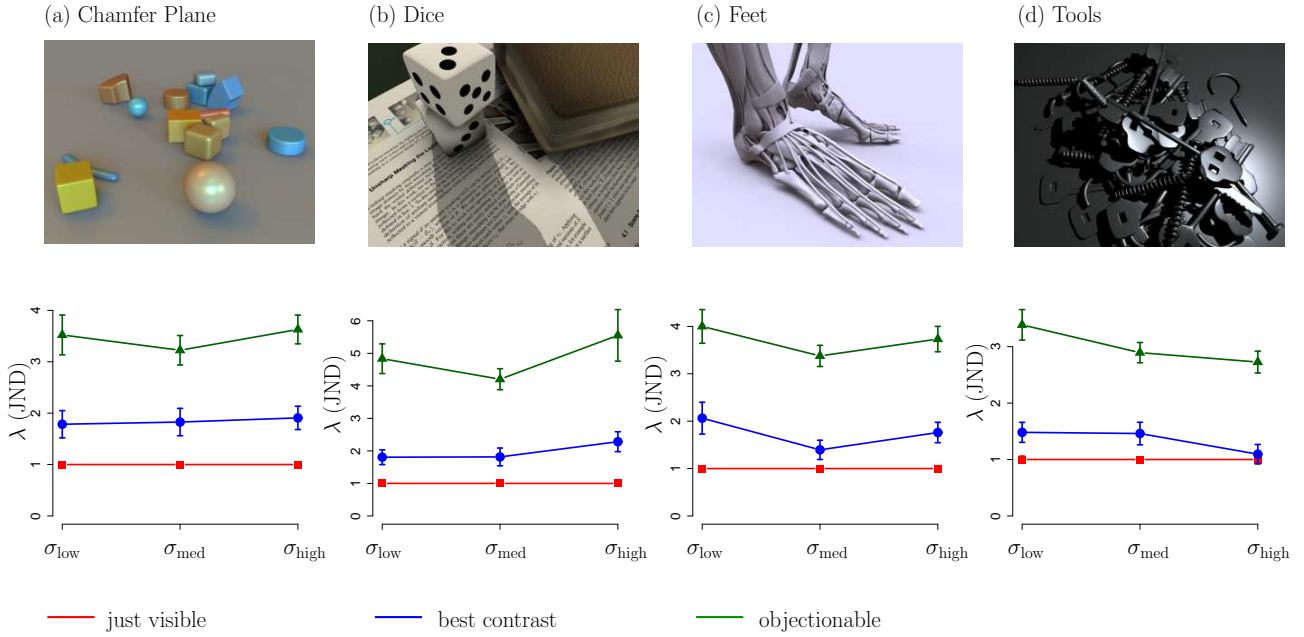
(a) Chamfer Plane   (b) Dice   (c) Feet   (d) Tools

Figure 4. User adjustments of the gain value λ varied as a function of gradient size σ for the 4 diverse scenes used in the study in JND units. Error-bars indicate the standard error of the mean (SEM).

outlined above ($\lambda_{\text{jn}} < \lambda_{\text{best}}$: $t(14) = 1.83, p = .08$; $\lambda_{\text{best}} < \lambda_{\text{obj}}$: $t(14) = 2.75, p < .03$). Again, the effect sizes are moderate to strong ($d(\lambda_{\text{jn}}, \lambda_{\text{best}}) = .67$, $d(\lambda_{\text{best}}, \lambda_{\text{obj}}) = 1.01$).

The ANOVA for the *Feet* scene reveals again the same main effect of task, $F(2, 28) = 49.01, p < .001$ but also a main effect of the kernel width σ, $F(2, 28) = 16.46, p < .001$ as well as an interaction $\sigma \times$ task, $F(4, 56) = 7.53, p < .001$. The main effect of task can be shown to stem from the same pattern discussed above, $\lambda_{\text{jn}} < \lambda_{\text{best}}$: $t(14) = 1.16, p < .03$ ($d = 0.88$); $\lambda_{\text{best}} < \lambda_{\text{obj}}$: $t(14) = 4.51, p < .01$, ($d = 1.65$). From table 1, the expectation that the main effect for σ is caused by a preference for higher λ-values for low σ and only for the preferred and upper threshold, can be derived. However, the planned contrasts do not reach significance for the pairwise comparisons ($\sigma_{\text{low}}, \sigma_{\text{medium}}$: $t(14) = 1.77, p = .09$, $\sigma_{\text{low}}, \sigma_{\text{high}}$: $t(14) = 1.38, p = .18$). The interaction effect is due to the visibility threshold's independence on σ (see Fig. 4), because the λ_{jn}-values did not show any effect of σ ($F(2, 28) = 0.05, p > .9$).

For scene *Keys*, again, the ANOVA yields a main effect of task ($F(2, 28) = 32.15, p < .001$) and σ ($F(2, 28) = 7.85, p < .002$) as well as the interaction $\sigma \times$ task, $F(4, 56) = 4.54, p < .05$. The effect of task however does not manifest itself in a significant difference between lower threshold and preferred value ($t(14) = 1.28, p = .21$), so users preferred a rather subtle enhancement for that scene that was barely above visibility threshold. There is however still a significant range of parameter values before the occurrence of artifacts ($\lambda_{\text{jn}} < \lambda_{\text{obj}}$: $t(14) = 4.86, p < .001, d(\lambda_{\text{jn}}, \lambda_{\text{obj}}) = 1.70$). The overall σ-effect as well as the interaction effect from the ANOVA cannot be tracked down to single differences, due to missing statistical power.

In summary, the reported analyses revealed a strong main effect of task, which was shown to stem from an increasing trend $\lambda_{\text{jn}} < \lambda_{\text{best}} < \lambda_{\text{obj}}$ in all scenes. The moderate to strong effect-sizes indicate that the parameter-range between lower and upper threshold was rather large. In addition, an effect of the kernel size σ for only two of the scenes, *Feet* and *Keys* was found while for scenes *Dice* and *Chamfer Plane* it had a zero-effect. A possible explanation for this result could be that the scenes differed according to the mean distances between objects, thereby producing a higher amount of "cluttering" for the *Feet* and *Keys* scenes (in the other two scenes, larger shapes that are relatively distant from each other dominate). For higher values of σ, parts of objects close to each other in the scene could have contributed to the local enhancement, producing objectionable artifacts already for relatively low values of λ.

(a) Dice

(b) Feet

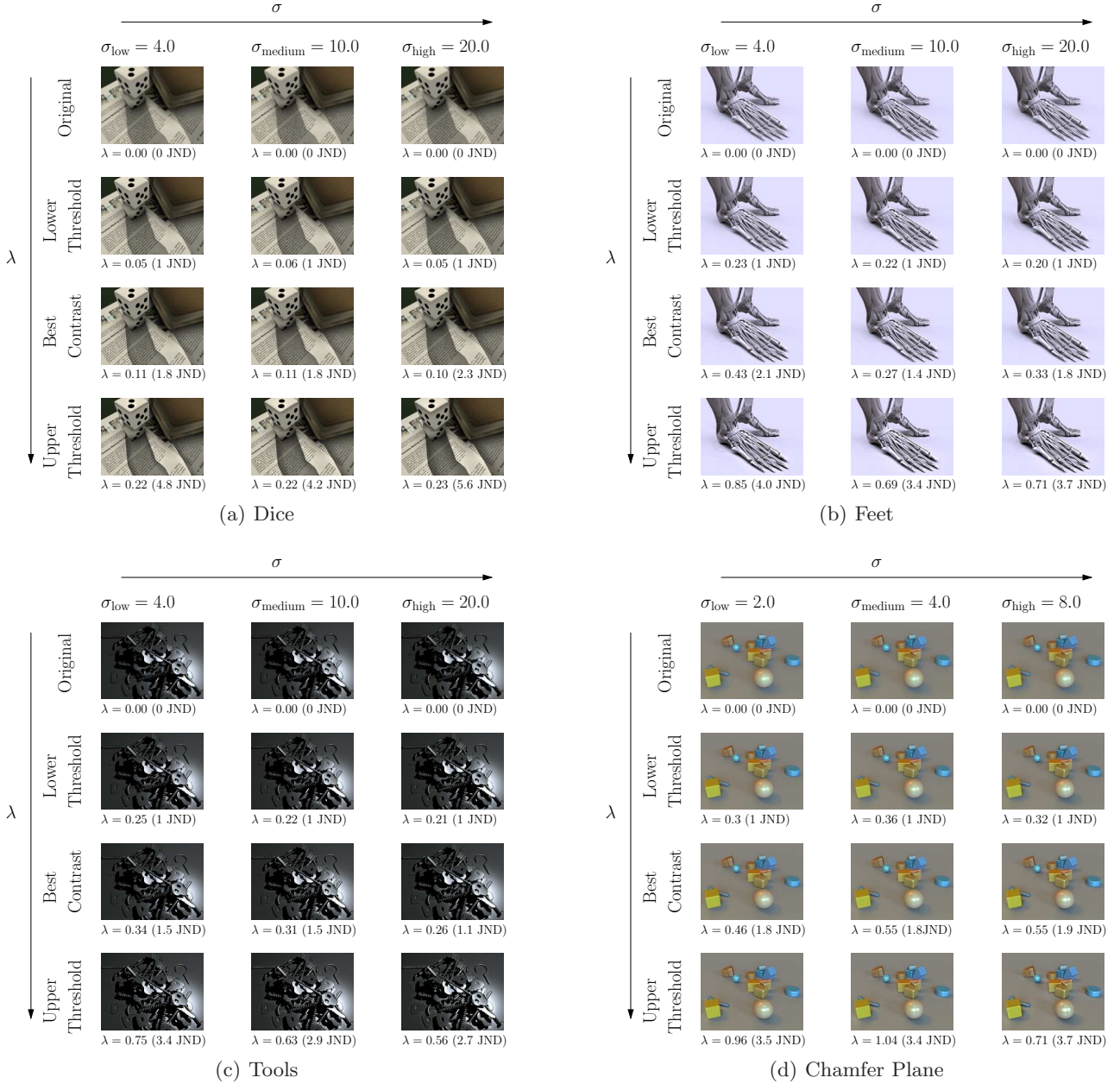(c) Tools

(d) Chamfer Plane

Figure 5. User-adjusted parameters for the four scenes. The images show the results of the parameter-adjustment carried out by the participants for lower and upper threshold and best contrast, respectively.

3.1 Just-Noticeable Differences

A problem inherent to the reported analyses is that scenes cannot directly be compared since λ scales differently for different scenes. Therefore, to enable a cross-scene comparison, we rescaled our raw data to just-noticeable-difference (JND) units, based on the value of the lower threshold by

$$\hat{\lambda}_i = \frac{\lambda_i}{\lambda_{\mathrm{jn}}}, \qquad i \in \{\mathrm{jn}, \mathrm{best}, \mathrm{obj}\}. \tag{1}$$

This conversion implicitly assumes that perceived contrast and λ-value are linearly related. While this is probably not true in general, we assume this approximation to be valid because the range of interesting values was relatively
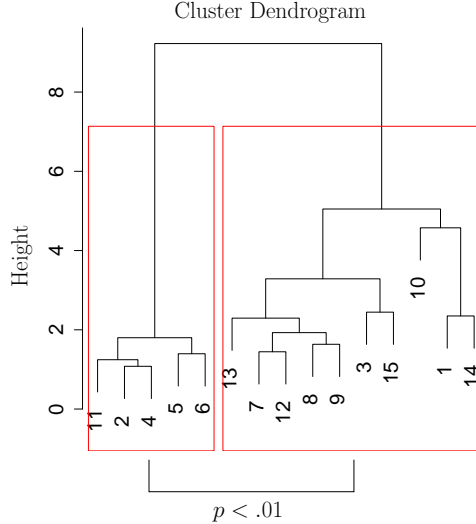
Figure 6. Agglomerative Cluster Analysis for the subjects using Ward's method and Euclidean distances. Numbers indicate different subjects, clusters (K-Medoids) are marked in red.

small. After this adjustment, $\lambda_{\text{jn}} = 1.0$ for all scenes because it was used for rescaling.

To determine whether the best-contrast setting shows a dependency on the scenes, an overall $4 \times 3 \times 3$ (scenes $\times \ \sigma \ \times$ task) ANOVA was conducted with the rescaled $\hat{\lambda}$-values. A main effect of scene reached significance ($F(3, 42) = 5.09, p < .005$). Pairwise comparisons reveal however, that only scene *Dice* and *Keys* differ in a statistically significant way ($t(14) = 2.89, p < .05$) while all other comparisons did not approach significance. Because these two scenes were very different from each other in terms of lighting, assembly and spatial frequency on the mesh. While the *Keys* scene is rather cluttered, large and simple shapes dominate the *Dice* scene. Gradient effects in high-frequency scenes can easily impair one another and should generally be used with more caution. We therefore assume that the enhancement parameter's space is perceptually rather uniform and only diverges when scenes show very different characteristics.

The other statistical comparisons did not produce any new conclusions, the pattern of results already discussed remains stable also for the transformed values. The transformation of the λ-values to JND-units emphasizes the finding, that the preferred setting of λ is relatively stable across scenes and settings of σ. The emergence of artifacts however, is apparently more scene-dependent. As a rule of thumb it can be formulated, that coming from a just visible enhancement, the strength can be doubled to yield close-to-optimal results and multiplied by four to yield an approximation for the upper threshold that still results in tolerable results.

3.2 Inter-individual Differences

In the previous analyses, a large interindividual variance was apparent. We hypothesized, that this variance was largely due to interindividual differences in how much contrast is preferred. To investigate these differences, we conducted a hierarchical cluster analysis applying Ward's method with Euclidean distances[23] to vectors for each subject containing the means of the adjustments over all tasks, scenes and settings of σ. The resulting cluster-dendrogram (see Fig. 6) reveals two groups of subjects, distinguished by their preference of the adjustment (clusters computed using K-Medoids with $K = 2$). We assume, that the two subgroups are separating persons with strong and weak preference for contrast, respectively. To test this assumption, we ran a statistical test comparing the two groups of subjects according to their mean adjustment of λ_{jn}, λ_{best} and λ_{obj}. The test showed that the subjects assigned to cluster 1 preferred lower values of λ for all tasks compared to subjects assigned to cluster 2 (λ_{jn}: $t(14) = -3.23, p < .01$, λ_{best}: $t(14) = -7.12, p < .01$, λ_{obj}: $t(14) = -4.71, p < .01$).

This result shows that a universal, "perfect" setting of the parameters does not exist since individual preference plays a huge role in how much enhancement is accepted. Most subjects preferred rather strong enhancements
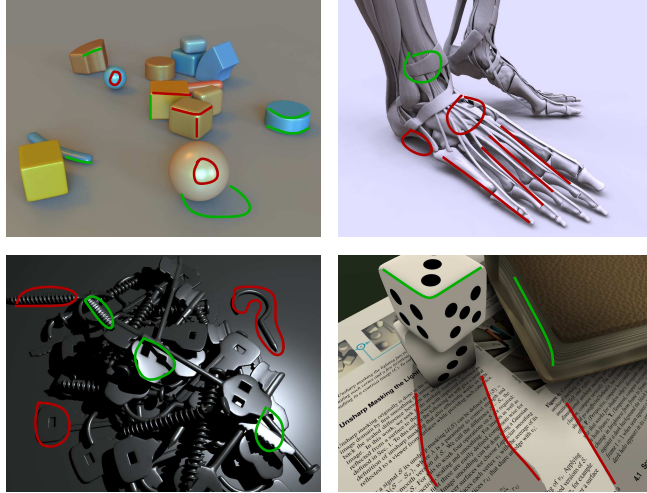
Figure 7. Regions of Interest. Subjects were asked to mark the features and regions that they considered to be most directly related to their given responses on a hardcopy of the scenes. Responses are marked, color differentiates different groups of users.

since cluster 2 comprised two thirds of our subjects, nevertheless a more conservative choice of the parameter is advisable in real-world applications in order not to put off people preferring less contrast.

3.3 Regions of Interest

The regions indicated by the subjects as being the most important for their decision of the enhancement strength are depicted in Figure 7. Features used for the enhancement by more than 5 of the subjects are marked. Colour indicates groups of subjects estimated from the hardcopy. All subjects focussed on regions that were expected to be sensitive to the algorithm (e.g. the shadows in the *Dice* scene or the edges of the objects in the *Chamfer Plane*). However, some subjects based their ratings on very few or even one feature at least for the estimation of the upper threshold where artifacts started to appear. It is therefore concluded, that (1) subjects were able to grasp the meaning of the enhancement and to relate it appropriately to their response and (2) that they were well able to locate the regions where the enhancement was first producing artifacts ensuring the validity of their responses.

4. SUMMARY AND DISCUSSION

Statistical tests confirmed, that all participants preferred a clearly visible enhancement over the original, non-enhanced scenes and that the setting for objectionable enhancement was far above the preferred settings. The effect of the gradient size σ was negligible. The absolute values of the adjusted parameters varied strongly depending on the scenes' content. However, when scaling the values based on their scene-dependent visibility thresholds to JND (Just Noticeable Difference) units, a general pattern emerged. Users typically rated enhancements twice the visibility-threshold (2 JND) as preferred while an enhancement of 4 JND was mostly perceived as objectionable (see Fig. 5). A close inspection of the scene-regions that were rated as benefiting most from the procedure showed that enhancement was perceived where it was expected, namely at edges and highlights (Chamfer Plane), shadow-boundaries (Dice) and highly-structured regions (Feet).

In general, our results are well in line with those obtained by Lin et al.[8] who found that an enhancement of about 2.6 JND was best using their LoG method. The minor differences to the results from this study may be due to the different methodologies used to obtain an estimate for the just-noticeable difference. While Lin et al. derived the JND using assumptions about noticeability of single-pixels using background luminance adaptation and local activity masking, we directly measured the JND for each of our scenes. This fact, along with the assumption of linearity implicit in the conversion in Eq. (1) might be responsible for this deviation but, of course, also the used techniques differed vastly.

It was observed, that judgments differed vastly between individual subjects. However, the inter-rater variance was shown to be reducible by computing within-cluster values for the ratings, based on a hierarchical cluster analysis. The resulting clusters clearly differentiated between participants that favored weak ($\approx 33\%$) and strong enhancements ($\approx 66\%$), respectively indicating that designers applying the method should be cautious not to overdo the enhancements.

In conclusion, the results from our study confirm the hypothesis that 3D Unsharp Masking increased the perceived contrast of a scene in a way that was perceived as attractive over a wide range of possible parameter-settings. The general pattern found for the parameters also provides a useful guideline for designers when making use of 3D Unsharp Masking: as a rule of thumb they can easily determine the strength for which they start to perceive an enhancement and use twice this value for a good effect. Since the value for objectionable results was twice as large again, artifacts should not impose restrictions on the applicability of this rule.

In further studies, we plan to investigate how different methods of enhancing scenes relate to one another. It is also a challenging task to explicitly address the issue of whether adaptation of the gradient as proposed by 3D Unsharp Masking is indeed the optimal method to trick the HVS into perceiving higher contrast. Studies applying artificial, low-structured stimuli would be necessary to further illuminate and characterize this effect.

APPENDIX A. INSTRUCTIONS

The actual wording of the instructions was as follows[*]:

"In this experiment, we are interested in finding out about how well a contrast enhancement procedure works. You will therefore be asked to adjust the strength of the enhancement in a way that the resulting image looks best to you. You will be presented with two images side-by-side, where the left image will be the original scene, while the right image will be enhanced. You can control the strength of the enhancement by pressing <up> or <down> on the keyboard. When you are done adjusting the strength, please press <Enter> to confirm your choice."

"We are interested in your opinion about several aspects of the enhancement:

1. At first, we would like you to adjust the enhancement such, that you can just see a difference in contrast.

2. In a second task, please adjust the strength so that the contrast is too strong (artifacts appear).

3. In a last step, please find the enhancement that you think produces the best contrast, while still not being too strong."

"In summary, we would like you to find the setting for the following requirements:

1. Enhancement just Visible

2. Looks too strong (objectionable)

3. Overall best Look (best contrast).

The required task will be indicated on the top of the screen. Please take your time in finding the values you prefer, there is no time pressure. Please note that there is no "correct" solution and that your answer only reflects your personal taste."

"The experiment will now start. If you have any questions, please contact the experimenter now."

ACKNOWLEDGMENTS

---

[*]Instructions are given for an example subject as the wording differs according to the randomization schedule. Line breaks indicate new display on the screen.

REFERENCES

[1] Ritschel, T., Smith, K., Ihrke, M., Grosch, T., Myszkowski, K., and Seidel, H.-P., "3D Unsharp Masking for Scene Coherent Enhancement," *ACM Trans. Graph. (ACM SIGGRAPH Proceedings)* **27**(3) (2008).

[2] Bovik, A. C., ed., [*Handbook of Image and Video Processing*], Academic Press, Inc., Orlando, FL, USA (2005).

[3] Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P., [*High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*], Morgan Kaufmann (2005).

[4] Cignoni, P., Scopigno, R., and Tarini, M., "A simple Normal Enhancement technique for Interactive Non-photorealistic Renderings," *Computer & Graphics* **29**(1) (2005).

[5] Rusinkiewicz, S., Burns, M., and DeCarlo, D., "Exaggerated shading for depicting shape and detail," *ACM Trans. Graph. SIGGRAPH* **25**(3), 1199–1205 (2006).

[6] Luft, T., Colditz, C., and Deussen, O., "Image enhancement by unsharp masking the depth buffer," *ACM Trans. Graph.* **25**(3), 1206–1213 (2006).

[7] Lin, W., Dong, L., and Xue, P., "Visual Distortion Gauge Based on Discrimination of Noticeable Contrast Changes," *IEEE Transactions on Circuits and Systems for Video Technology* **15**(7), 900 (2005).

[8] Lin, W., Gai, Y., and Kassim, A., "Perceptual impact of edge sharpness in images," *Vision, Image and Signal Processing, IEE Proceedings -* **153**, 215–223 (April 2006).

[9] Calabria, A. and Fairchild, M., "Perceived image contrast and observer preference I: The effects of lightness, chroma, and sharpness manipulations on contrast perception," *The Journal of Imaging Science and Technology* **47**, 479–493 (2003).

[10] Calabria, A. and Fairchild, M., "Perceived image contrast and observer preference II: Empirical modeling of perceived image contrast and observer preference data," *The Journal of Imaging Science and Technology* **47**, 494–508 (2003).

[11] Cornsweet, T., [*Visual Perception. Chapter II: The Experiment of Hecht, Schlaer, and Pirenne*], Academic Press (1970).

[12] Pratt, W., "Digital Image Processing," *New York: Wiley* (1991).

[13] Purves, D., Williams, S., Nundy, S., and Lotto, R., "Perceiving the Intensity of Light," *Psychological Review* **111**(1), 142–158 (2004).

[14] Kingdom, F. and Moulden, B., "Border Effects on Brightness: A Review of Findings, Models and Issues," *Spatial Vision* **3**, 255–262 (1988).

[15] Purves, D., Shimpi, A., and Lotto, R. B., "An empirical explanation of the Cornsweet effect," *J. of Neurosci* **19**, 8542–8551 (1999).

[16] Krawczyk, G., Myszkowski, K., and Seidel, H.-P., "Contrast Restoration by Adaptive Countershading," in [*Computer Graphics Forum (Proceedings of Eurographics 2007)*], **26** (2007).

[17] Purves, D. and Lotto, R. B., [*Why We See What We Do: An Empirical Theory of Vision*], Sinauer, Sunderland, MA, USA (2003).

[18] Gescheider, G. A., [*Psychophysics: The Fundamentals*], Lawrence ErlbaumAssociates (1997).

[19] Brainard, D., "The Psychophysics Toolbox," *Spatial Vision* **10**, 437–442 (1997).

[20] Sperling, G., "The information available in brief visual presentations," *Psychological Monographs: General and Applied* **74**(11), 1–30 (1960).

[21] Holm, S., "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics* **6**, 65–70 (1979).

[22] Cohen, J., [*Statistical power analysis for the behavioral sciences*], Hillsdale, NJ: Lawrence Earlbaum Associates (1988).

[23] Hastie, T., Tibshirani, R., and Friedman, J., [*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*], Springer (2001).