Apparent Stereo: The Cornsweet Illusion Can Enhance Perceived Depth

Piotr Didyk[1] Tobias Ritschel[1,2,3] Elmar Eisemann[2]
Karol Myszkowski[1] Hans-Peter Seidel[1]

[1]MPI Informatik, Saarbrücken, Germany;
[2]Télécom ParisTech / CNRS, Paris, France;
[3]Intel Visual Computing Institut, Saarbrücken, Germany

ABSTRACT

It is both a technical and an artistic challenge to depict three-dimensional content using stereo equipment and a flat two-dimensional screen. On the one hand, the content needs to fit within the limits of a given display technology and at the same time achieve a comfortable viewing experience. Given the technological advances of 3D equipment, especially the latter increases in importance. Modifications to stereo content become necessary that aim at flattening or even removing binocular disparity to adjust the 3D content to match the comfort zone in which the clash between accommodation and vergence stays acceptable. However, applying such modifications can lead to a reduction of crucial depth details. One promising direction is backward-compatible stereo, for which the disparity is low enough that overlaid stereo pairs seem almost identical. It builds upon the Craik-O'Brien-Cornsweet effect, a visual illusion, which uses so-called Cornsweet profiles to produce a local contrast that leads to a perceived brightness increase. Similarly, Cornsweet profiles in disparity can lead to an illusion of depth. Applying them skilfully at depth discontinuities allows for a reduction of the overall disparity range to ensure a comfortable yet convincing stereo experience. The present work extends the previous idea by showing that Cornsweet profiles can also be used to enhance the 3D impression. This operation can help in regions where the disparity range was compressed, but also to emphasize parts of a scene. A user study measures the performance of backward-compatible stereo and our disparity enhancement.

Keywords: stereo, perception, 3D, backward-compatible stereo, Cornsweet, depth

1. INTRODUCTION

Perspective, occlusion, texture gradients or shading are often easily realizable on common displays and can serve as depth cues to convey the geometric configuration of a scene. The human visual system (HVS) is trained at deriving spatial information from such monocular depth cues,[1,2] which has been exploited by artists to faithfully depict space.[3] Nonetheless, the HVS also employs binocular cues, which are more difficult to reproduce with standard equipment and require specialized hardware.

While previously only anaglyph stereo was accessible on the consumer-level market, today, we find a variety of techniques to produce stereo effects ranging from polarization or shutter glasses to autostereoscopic displays.[4,5] This trend is underlined by the increasing amount of stereo content in form of TV broadcasts, feature films, and computer games. Although the quality of stereo equipment is constantly improving, the reproducible depth range is smaller than what is observable in the real-world.[6] These limitations should be considered when producing stereo content. While it might seem natural to simply render two views, one for each eye, such an approach is not always sufficient. For example, the distance between the virtual cameras might not correspond to the actual eye distance of the observer. Similarly, one might have made assumptions concerning the distance to the screen, or even the type of screen itself, which can substantially differ. Especially for movies, where stereo equipment, observers and their position are unknown, it is important to take these considerations into account.

While the general creation of stereo image pairs is scene- and artist-dependent, one important rule is to avoid large disparities. In general, these can result in a viewing discomfort or even fail to produce stereo when the two images can no longer be fused. In order to avoid such problems, the range of disparities often needs to be significantly reduced.[7,8] An extreme example of such an operation is microstereopsis,[9] where the camera distance is reduced to a minimum, meaning that a stereo image pair has just enough disparity to create a 3D impression. Recently, it was shown that disparity can be reduced even further without sacrificing too much of the stereo impression by computing so-called backward-compatible stereo.[10] Stereo-image pairs computed with this technique appear almost ordinary to the naked eye but convey a stereo impression when special equipment is used. The method relies on the observation that the Craik-O'Brien-Cornsweet effect[11,12] is applicable to binocular disparity. By introducing disparity only at depth discontinuities, artifacts that appear when the stereo-pair images are overlaid, which is usually what one perceives when stereo glasses are unavailable, are significantly reduced.

In this paper, we evaluate the backward-compatible stereo approach. We illustrate its effectiveness and usefulness by showing that Cornsweet illusion, as previously applied to brightness, can increase stereo perception without introducing a large overall disparity. We present a way of respecting potential limits of a given display technology improving at the same time depth impression. Our technique controls depth perception in a multi-scale manner taking into account the sensitivity of the human visual system to disparity signal. Furthermore, we make use of the machinery of backward-compatible stereo to produce expressive stereo rendering where certain elements and details are enhanced, hereby steering the attention of the observer and providing more information about spatial details without violating the comfort constraints when needed.

The paper is organized as follows: We review human depth perception (Section 2) and give an overview of previous work (Section 3). Then, we describe our general approach building on the Cornsweet illusion in the context of stereovision (Section 4). We further present possible disparity manipulations (Section 4.2) and evaluate the impact on perceived stereo images (Section 5). We discuss strengths as well as limitations (Section 6) before concluding (Section 7).

2. DEPTH PERCEPTION

The HVS relies on a large variety of depth cues to compensate for the fact that each retinal image is two dimensional. Those depth cues can be categorized[1] as pictorial information (occlusions, perspective foreshortening, relative and familiar object size, texture and shading gradients, shadows, aerial perspective), dynamic information (motion parallax), ocular information (accommodation and convergence), and binocular information (disparities). These depth cues are integrated and interpreted by the HVS to derive observer-object and inter-object distances. The importance of the cues may strongly depend on the object's distance to the eye[13] and dominant cues may prevail or a compromise (in terms of likelihood of the cues) is perceived [1, Chapter 5.5.10].

Stereopsis refers to the HVS determining the depth of presented objects by measuring displacement (binocular disparity) between their images created on the retinas of the left and right eye [1, Chapter 5.3]. Using a vergence mechanism, the eyes fixate at one point in the scene, for which the binocular disparity becomes zero. Assuming such a given vergence angle there is a set of points in the scene for which disparity is equal zero. This set is called *horopter*. All points lying in front of it lead to non-zero *crossed* (negative) disparity, which increases as their distance to the observer is reduced. Similarly, all points behind the horopter feature *uncrossed* (positive) disparity, which increases with the distance to the observer. Stereopsis is a strong depth cue,[13] as can be conveniently studied in isolation from other depth cues by means of *random-dot stereograms*.[14]

Vergence-accommodation Conflicts arise because of the clash between the accommodation, which maintains the display's screen within the range of ± 0.3 diopters around the of field (DOF),[6] and the displayed stereo indications. With increasing screen disparity, vergence drives the fixation point away from the screen plane. Hence, a conflict between the fixation point and focusing point is established. This incompatibility can be tolerated to a certain degree, beyond which it leads to a visual discomfort.[6]

On the one hand, these observations suggest that disparity reduction is desirable. On the other hand, simply reducing disparity results in a loss of depth information. Moreover, while naturalness and quality of depth are highly correlated, people generally prefer (judge of higher quality) slightly exaggerated depth,[15] even at a possible expense of naturalness. Such enhancement can also increase the sense of presence in a virtual environment, but, in this case, images should still appear natural.[16] In this work, we will show that backward-compatible stereo is a good candidate to reach these goals.

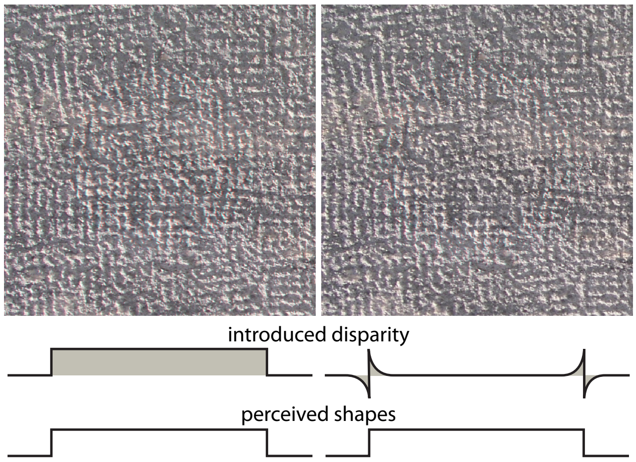

Figure 1. Top: A circle with depth due to disparity and apparent depth due to Cornsweet disparity profiles in anaglyph. Texture was required to provide disparity cues. Bottom: The corresponding disparity profiles as well as perceived shapes. The solid area depicts the total disparity, which is significantly smaller when using the Cornsweet profiles.

Contrast and Contours The Craik-O'Brien-Cornsweet illusion is a well-known luminance-contrast phenomenon where two regions with the same luminance are separated by a sharp discontinuity with luminance gradually decaying towards equiluminant regions.[17] The two different lightness levels at the discontinuity are propagated by the filling-in mechanisms of the HVS which results in the impression that one region is brighter. Thus, the illusion creates an apparent brightness difference between both regions, which leads to similar appearance as the introduction of physical differences by means of a step function separating the regions, but without the loss of dynamic range.[18,19] Different shapes/profiles can be used to produce such a local contrast.[17]

Cornsweet Illusion for Depth exists and Anstis et al.[11] found that a depth Cornsweet profile adds to the perceived depth difference between real textured surfaces, as confirmed by Rogers and Graham[12] for random-dot stereograms. This effect is illustrated in Fig. 1 for textured stereograms, where the Cornsweet profile is applied directly to the screen disparity.

Rogers and Graham observed that the induced depth difference over the whole surfaces amounted up to 40% with respect to the depth difference at the discontinuity. They further measured that the effect is stronger along the horizontal (i.e. eye separation) direction, but recent results indicate no significant difference with respect to the orientation.[20] The great advantage of the Cornsweet disparity is its locality that enables depth cascading (refer to the depth Mondrians in Fig. 8) without accumulating screen disparity as it would usually be required. The effect is remarkably strong, and we will exploit it to enhance depth impression and to reduce physical screen disparity.

3. PREVIOUS WORK

We cast the problem of distortion-free stereo vision as a *range-mapping* problem, that is, finding an *operator* which maps a given high dynamic range into a low dynamic range of a certain output medium in such a way, that it minimizes the perceived difference between the signal in the two ranges. Another functionality that we consider within the same computational framework is *depth quality enhancement* for any existing operator.

Operators can be either global or local. A *global* operator maps one value to another value in the range, independent of where this value occurs in the signal. For *local* operators, the result also depends on the location in the signal, i. e. its context. Due to the local nature of Cornsweet illusion our range mapping operator as well as depth quality enhancement are inherently local.

While this work addresses mapping of stereo cues to the limited range of a flat screen, several other media have been considered by computer graphics before.

Luminance The most prominent example of range mapping is called *tone mapping* in High Dynamic Range Imaging where the signal is luminance in a digital image and the target range is the limited screen luminance.[21] Many existing local operators[22,23] are tuned for the best use of dynamic range and enable multi-resolution manipulation of detail visibility. Our computational framework also relies on multi-resolution, per band depth processing. Notably, Krawczyk et al.[19] suggest a local operator based on the Cornsweet illusion. In a perceptual framework, they analyze the distortion (i. e. loss) in contrast caused by an arbitrary operator in various bands and re-introduce contrast when possible via Cornsweet profiles. In our depth quality enhancement we apply a similar principle to binocular stereo cues.

Geometry Compressing arbitrary three-dimensional geometry into the limited range of an almost flat object like a coin is called *bas-relief* and was addressed by Weyrich et al.[24] They apply a non-linear global operator and a local gradient-domain decomposition into frequency bands. In principle, their manual artistic controls enable the addition of a Cornsweet profile into the compressed depth. In this work, we aim for an automatic method to add the depth Cornsweet illusion, but will enable its use also for expressive means.

Disparity For stereo images, Lang et al.[8] map a range of disparities into a limited range that matches certain viewing conditions such as screen size or viewer distance. They formulate an optimization process that guides the warping of stereo image pairs while respecting constraints imposed on the resulting disparity and their temporal changes, as well as saliency-driven image distortions. Nonetheless, these operations can reduce, distort or remove disparities completely. In our approach, we want to avoid losing important details. Didyk et al. developed a general framework to study the perceptual effect of disparity.[10] Their model predicts the frequency-dependent visibility of disparity and offers a way to transform given disparities into a perceptually uniform space. They present several applications of their method, including compression, backward-compatible stereo, and depth-perception transfer. We will make use of their model to control the effect of the disparity enhancement.

Image Enhancement Several techniques exist to enhance color images based on depth information to make a spatial layout more apparent.[25–28] Our approach orthogonally works on depth cues and could be combined with such approaches.

4. OUR APPROACH

In this section, we will detail the processing pipeline of our approach.

Overview An overview of our approach is shown in Fig. 2. As input of our algorithm we use a linearized depth buffer that has a corresponding color image. Based on this depth information, we derive, as an output, a disparity map that defines the stereo effect.

To compute the disparity map, we first convert the linearized depth into pixel disparity based on a scene to world mapping. The pixel disparity is converted to a perceptually uniform space,[10] which also provides a decomposition into different frequency bands. Our approach will act on these bands to yield the output pixel disparity map which defines the enhanced stereo image pair. Given the new disparity map, we can then warp the color image according to this definition. Our approach is orthogonal to the technique used for this warping (we adopt the efficient real-time approach[29] that accounts for high-frequency disparity structures and uses a warping grid to prevent holes).

Figure 2. From left to right: Starting from an original depth map a pixel disparity map is computed and then a disparity pyramid is built. After multi-resolution disparity processing, the dynamic range of disparity is adjusted and the resulting enhanced disparity map is produced. The map is then used to create enhanced stereo image.

4.1 Disparity Linearization

First, we need to fix a scene unit that scales the scene such that one scene unit corresponds to a world unit. Then, given the distance to the screen and the eye distance of the observer, we can convert this depth into pixel disparity. Next, we make use of a perceptual disparity space following the approach by Didyk et al.[10] The advantage of this space is that all modifications are predictable and uniform because the perceptual space provides a measure of disparity in just-noticeable units. It, hence, allows a convenient control over possible distortions that we may introduce. In particular, any changes below 1 JND should be imperceptible. One interesting observation is that the perceptual model relies on a Laplacian decomposition. As illustrated in Section 4.2, such a perceptual decomposition itself is in fact well suited for the purposes of adding Cornsweet profiles.

4.2 Modifying Disparity

This section presents the various manipulations we apply to the initial disparity map. Depending on the purpose (retargeting, enhancement, backward-compatible stereo...), the applied operations differ. In Section 5, we present a perceptual study that evaluates our depth enhancement as well as backward-compatible stereo.[10]

4.2.1 Retargeting

One of our main applications is to retarget stereo content. Hereby, we mean modifying the pixel disparity to fit into the range that is appropriate for the given device and user preferences (distance to the screen and eye distance). Typically, such retargeting implies that the original reference pixel disparity D^r is scaled to a smaller range D^s. Consequently, in D^s some of the information may get lost or become invisible during this process. Inspired by previous work[19] in the field of tone-mapping, we want to compensate this loss by adding Cornsweet profiles P_i to enhance the apparent depth contrast.

As the perceptual decomposition is performed using a Laplacian pyramid, the bands correspond to Cornsweet profile coefficients (each level is a difference of two gaussian levels, which remounts to unsharp masking). Hence, modifying higher bands in the pyramid remounts to modifications in form of Cornsweet profiles. E.g., adding the sum of these higher bands would directly yield unsharp masking. In practice, it is a good choice to only involve the top five bands of the perceptual decomposition to add the lost disparities.We estimate the *loss* of disparity in D^s with respect to D^r by comparing the disparity change in each band of a Laplacian pyramid:

$$R_i = C_i^r - C_i^s$$

where R_i are the corrections in a given band i, C_i^r and C_i^s are the bands of the reference and distorted disparity respectively.

In theory, one might be tempted to simply add all R_i directly on top of D^s. Effectively, this would add Cornsweet profiles to the signal, but care has to be taken that the resulting pixel disparity does not create disturbing deformation artifacts and remains within the given disparity bounds. In order to prevent disturbing distortions, we limit the Corsnweet profiles directly in the perceptual space, as detailed in the following.

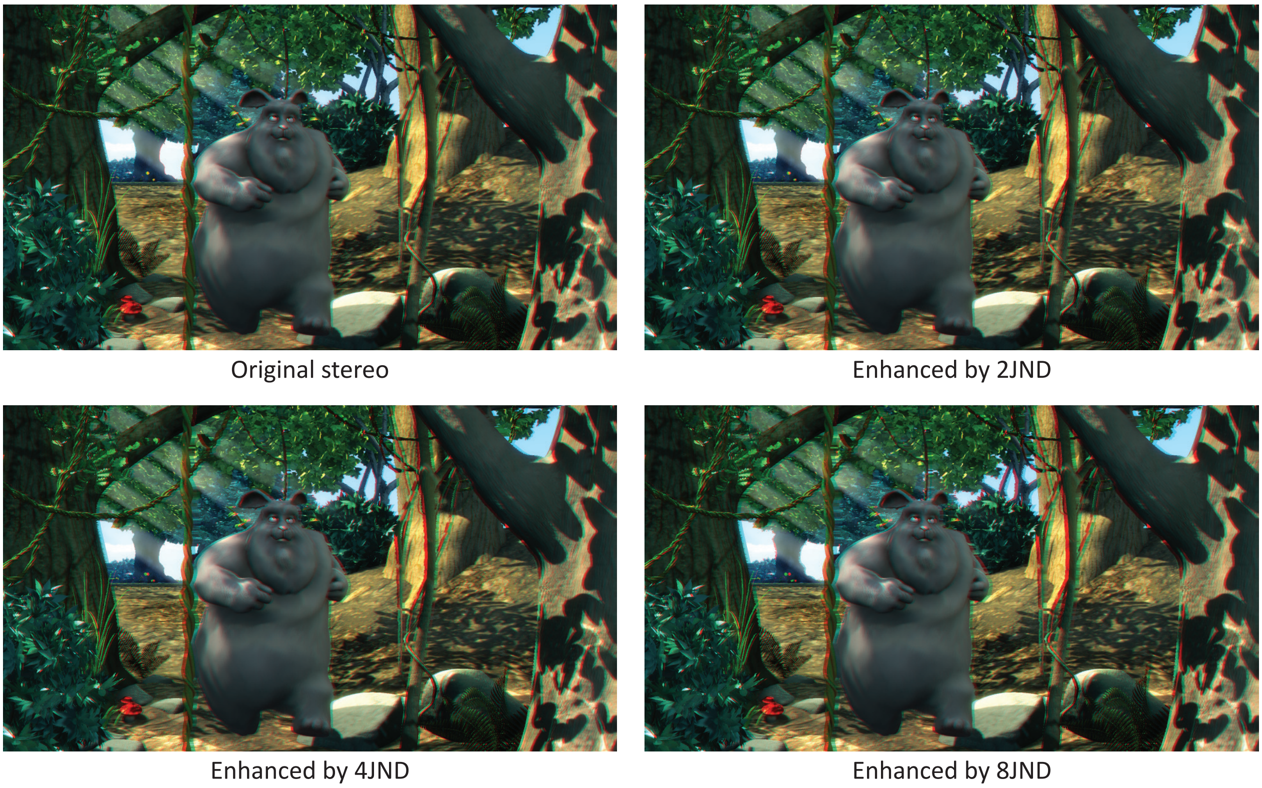

Figure 3. We can change the effect of depth perception by increasing JNDs. In this way, we can uniformly exaggerate the depth impression ("Big Buck Bunny" © by Blender Foundation).

Limiting Cornsweet profiles To assure that added Cornsweet profiles do not yield a too large disparity range, we manipulate the corrections R_i. A first observation is that all values are in JND units, hence, we can limit the maximum influence of the Cornsweet profiles, by clamping individual coefficients in R_i so they do not exceed a limit given in JND units. Clamping is a good choice, as the Laplacian decomposition of a step function exhibits the same maxima over all bands situated next to the edge, is equal zero on the edge itself, and decays quickly away from the maxima. Because each band has a lower resolution with respect to the previous, clamping of the coefficients lowers the maxima to fit into the allowed range, but does not significantly alter the shape. The combination of all bands together leads to an approximate smaller step function, and, consequently, choosing the highest bands leads to a Cornsweet profile of limited amplitude. In Fig. 3, we show how different limits result in different enhancement strength.

Unfortunately, this will not yet ensure that the enhancement layer R (composed of all R_i) combined with D^s will not result in too large value. Clamping is a straightforward way of limiting the profiles R, but it results in flat areas whenever the disparity bounds are exceeded. The second possibility is to scale profiles using a monotonic mapping function. Here, a good mapping seems to be a logarithmic function that favors small variations, which we do not need to clamp as they usually do not result in an exceeded disparity range. Nonetheless, an important observation is that some parts of D^s might allow for more aggressive Cornsweet profiles than others without exceeding the comfort zone. Therefore, instead of using a global method, we propose to locally scale the Cornsweet profiles to best exploit local disparity variations and to make sure that most of the lost contrast is restored. Wherever the limits are respected, these scaling factors are simply one, otherwise, we ensure that the multiplication resolves the issue of discomfort. Scaling is an acceptable operation because the Cornsweet profiles vary around zero.

Deriving a scale factor for each pixel independently is easy, but if each pixel were scaled independently of the others, the Cornsweet profiles might actually disappear. In order to maintain the profile shape, scaling factors should not vary with higher frequencies than the scaled corresponding band. Hence, we compute scale

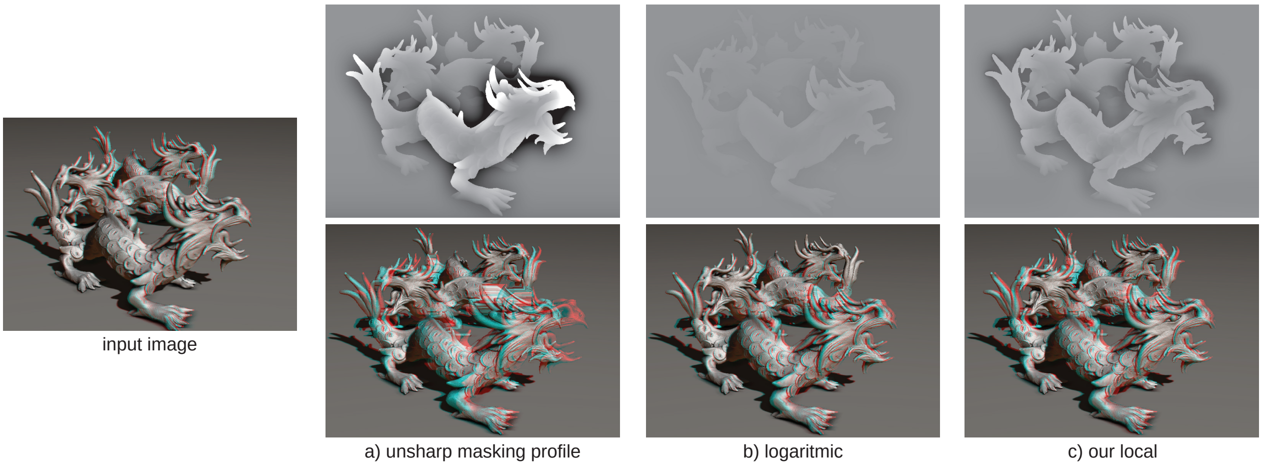

Figure 4. Different approaches for limiting Cornsweet profiles. All of examples were limited to the range $[-0.5, 0.5]$. A simple unsharp-masking profile can exceed the range of possible disparities while the image is enhanced (a). Logarithmic suppression (b) limits big profiles but at the same time those that could stay bigger (in the far plane) get almost invisible. Our local method (c) limits profiles locally preserving small ones.

factors per band.

One observation is that we relied on a pyramidal decomposition, consequently, R_i has a two times higher resolution than R_{i+1}. This is important because when deriving a scaling S_i per band, it will automatically exhibit a reduced frequency variation. Hence, we derive per-pixel-per-band scaling factors S_i that ensures that each band R_i when added to D^s would not exceed the limit. Next, these scaling factors are "pushed down" to the highest resolution from the lowest level by always keeping the minimum scale factor of the current and previous levels. This operation results in a high-resolution scaling image S. We finally divide each S by the number of bands to transfer (here, five). This ensures that $D^s + \sum_i R_i S$ respects the given limits and maintains the Cornsweet profiles. Figure 4 illustrates our local scaling in comparison to other approaches and shows that it best preserves the Cornsweet profiles, while reproducing most of the original contrast.

4.2.2 Artistic enhancement

Our previously described retargeting ensures that contrast is preserved as much as possible. Although this enhancement is relatively uniform, it might not always reflect an artistic intentions. E.g., some depth differences between objects or particular surface details might be considered important, while other regions are judged unimportant. Fig. 5 (right) shows an example where the distance between the two dragons in the background has been enhanced, as well as the details in the foreground where the dragon scales appear more detailed. Fig. 5 (right) shows an example where the distance between the two dragons in the background has been enhanced, as well as the details in the foreground where the dragon scales appear more detailed. It is also possible to increase the overall depth impression in the scene by increasing disparity scaled in JNDs units (see Fig.3).

To give control over the enhancement, we developed a simple interface that allows an artist to specify which scene elements should be enhanced and which ones are less crucial to preserve. Precisely, we allow the user to specify weighting factors for the various bands which gives an intuitive control over the frequency content. Using a brush tool, the artist can directly draw on the scene and locally decrease or increase the effect. By employing a context-aware brush, we can achieve ensure edge-stopping behavior to more easily apply the modifications.

4.2.3 Backward-compatible Stereo

Using our technique, we can produce backward-compatible stereo that "hides" 3D information from observers without 3D equipment. The observation is that a zero disparity leads to a perfectly superposed image for

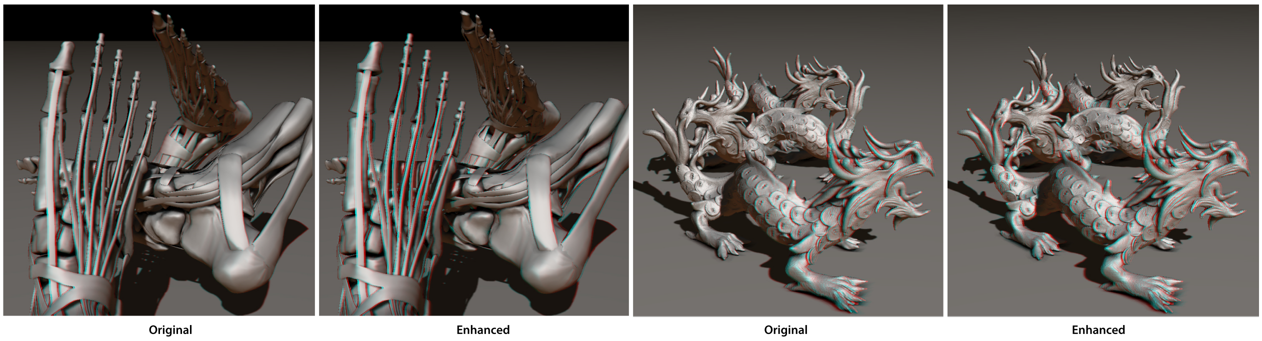

Figure 5. Depth enhancement using the Cornsweet illusion. Original and enhanced anaglyph images are shown for two different scenes with significant depth range. Note a better separation between the foreground and background objects and a more detailed surface structure depiction.

both eyes. Unfortunately, this also implies that no 3D information is experienced anymore. Therefore, our goal is to reduce disparity where possible to make both images converge towards the same location, hereby it appears closer to a monocular image.

In particular, this technique can transform anaglyph images and makes them appear close to a monocular view (teaser image).

The implementation follows the same process as for the retargeting, but we do not add the scaled disparity. In this case, the Cornsweet profiles will create apparent depth discontinuities, while the overall disparity remains low. This is naturally achieved because Cornsweet profiles are centered around zero.

The solution is very effective, and has other advantages. The reduction leads to less ghosting for imperfect shutter or polarized glasses (which is often the case for cheaper equipment). Furthermore, more details are preserved in the case of anaglyph images because less content superposes. This is particularly visible for the grass and sky in the foreground of Fig.7. Furthermore, it is important to realize that much of the scene structure remains understandable because the HVS is capable of propagating some of the perceived differences over the neighboring surfaces. When comparing to an image of equivalent disparity (scaled to have the same mean), almost all depth cues are lost. In contrast, to produce a similar relative depth perception, the disparity can become very large in some regions even causing problems with eye convergence. Finally, our backward-compatible approach could be used to reduce visual discomfort for cuts in video sequences that exhibit changing disparity.[8]

4.2.4 Photo Manipulation

Finally, converting 2D photos into 3D[30] is never perfect. To minimize and facilitate the user interaction, we can concentrate on local discontinuities and avoid a global depth depiction. According to our findings even a localized depth representations can deliver a good scene understanding (refer to Fig. 7). This is not surprising, as it is an observation that has been used for centuries in the form of bas-relief depictions. In fact, again the Cornsweet profile seems to be a very effective shape in this context.

5. RESULTS

We implemented our method on a standard Geforce GTX 480 GPU using OpenGL. As all operations are realizable on a GPU and applied to textures, the solution performs almost independently of the geometric complexity of the scene.

The depth conversion is similar to a tone mapping operator, and we store the multi-scale perceptual space as a MIP map. Hence, in each band the resolution is halved along each axis. This choice proved sufficient and made all operations very cheap. Consequently, all applications reach real-time performance. Coupled with the real-time view synthesis following,[29] the overall rendering time remains at 30 to 100 Hz in all scenes depicted in the paper.

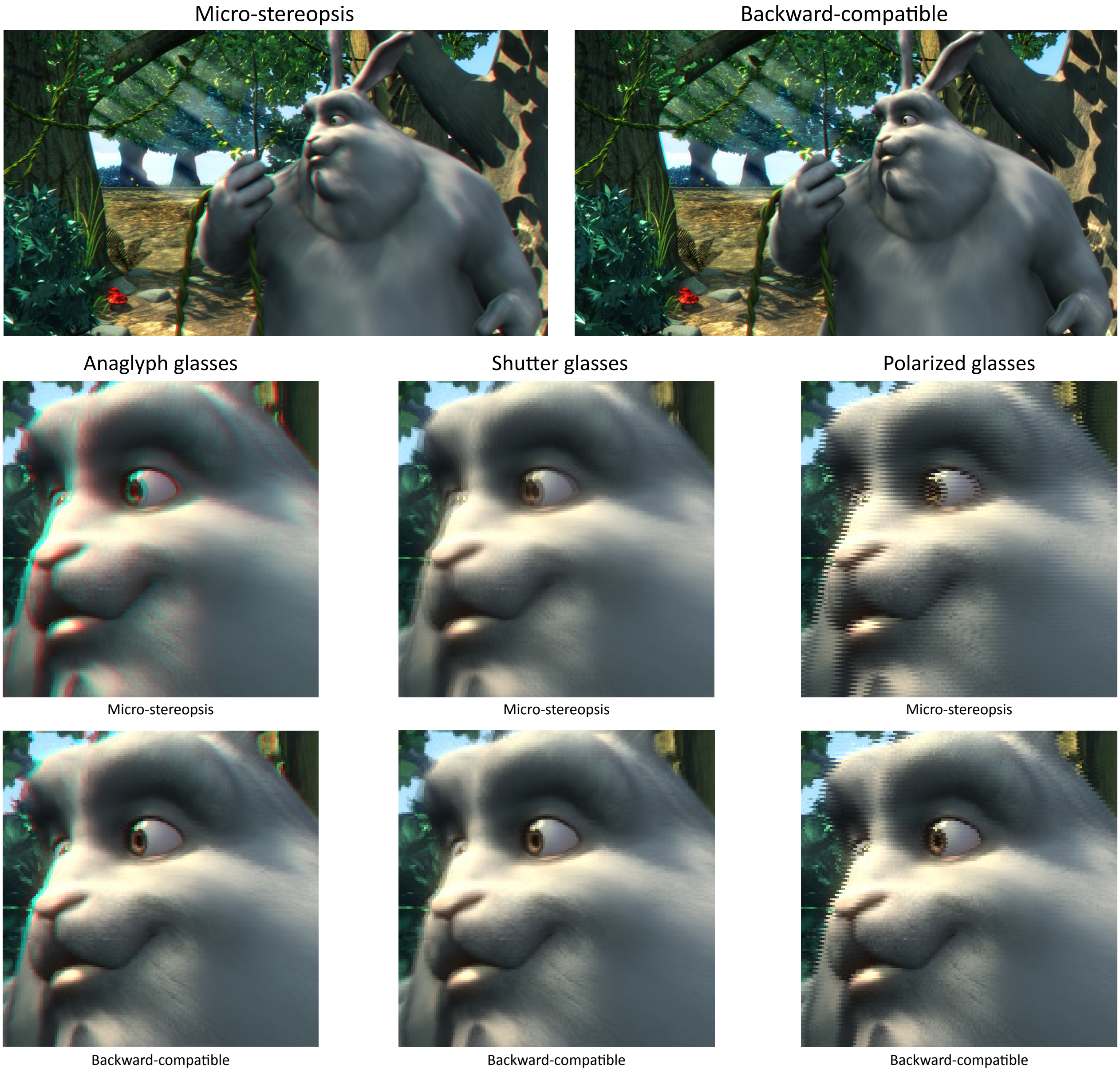

Figure 6. The figure presents a comparison between the backward-compatible stereo and the micro-stereopsis technique. The second method was adjusted in a way that both versions exhibit the same depth impression. The insets present zoomed-in versions of images displayed using differed stereo equipment. It can be seen that although the depth impression in both cases is very similar, the backward-compatible version reveals less disturbing artifacts while watched without stereo equipment. This is well visible especially for areas without depth discontinuities such as body of the bunny (notice in particular the shadows), or inside the tree (“Big Buck Bunny” © by Blender Foundation)

To evaluate the backward-compatible approach, we performed two user studies. Since the 3D effect depends on the size of the image and distance to the screen,[7] the images in this paper are optimized to cover about 20 - 25 visual degrees. We used NVIDIA’s 3D Vision active shutter glasses with a Samsung SyncMaster 2233RZ display (1680 $\times$ 1050 pixels). Subjects that took part in our studies had correct or corrected-to-normal vision.

In our first study, ten naive subjects participated. First, we investigated the overall quality of our method. For this, we handed a backward-compatible stereo image with “hidden” anaglyph content. We asked each subject for flaws or particularities in the image. None of those that received our output reported the artifacts produced by the stereo information within the first minute. Furthermore, only two subjects reported this observation within two minutes. After two minutes, the subjects received anaglyph glasses and were asked to report their observation concerning the stereo impression of the backward-compatible stereo image and the standard 2D image shown side by side. All 10 subjects agreed that the backward-compatible stereo image exhibits a 3D effect whereas the standard image does not. Obviously, such results depend on the underlying image content, but the findings give a clear indication that 3D content can be hidden to a large extent.

The second study was conducted to measure the depth effect of our solution and to show that it reduces disturbing artifacts when not using special equipment. To this extent, we let six participants compare the depth percept of two stereo images, one with our backward-compatible stereo and one with standard stereo. We then asked them to adjust the disparity in the standard stereo image (by approaching the two cameras), such that the depth impression was equivalent to our backward-compatible version. Such an adjustment of camera distances is similar to performing micro-stereopsis.[9] In Fig. 6, we show comparison of the backward-compatible version and the average result.

6. DISCUSSION

Our approach shows that the Cornsweet effect is a practical tool to manipulate stereo content convincingly. It is effective enough to even achieve cascading (multiple profiles on top of each other, Fig. 8). One limitation is that similar to the Cornsweet illusion in luminance, the manipulation might change the appearance of the shape or even material to some extent. On the other hand, we do not manipulate the colors in the rendered image itself, which means that we preserve many of the original cues (lighting, material) that are particularly helpful in conveying an overall satisfactory appearance. This is particularly visible in complex stimuli (Fig. 6) where the spatial layout is convincingly captured without introducing very high disparities. These properties make backward-compatible stereo an interesting trade-off.

Our approach is general in the sense that it is orthogonal the way that input images were captured, be it 3D rendering, a depth camera or a multi-view surface reconstruction, further it is orthogonal to the display modality used to present the stereo color image pair.

7. CONCLUSIONS

In this work, we proposed to use a visual illusion to manipulate the spatial impression. The approach is computationally simple and was validated in a perceptual study.

There are many interesting avenues for future research. Not all stereo cues are equally important for all distances and other stereo cues, besides disparity, could be enhanced, when disparity becomes ineffective. For example, warm-cold shading might distort colors, but helps in conveying spatial organization. In fact, generating exactly those stereo cues that are actually used for a certain depth, while minimizing their distorting effect, would allow to save time and maximize the perceptual effectiveness.

8. ACKNOWLEGMENTS

We would like to thank the Stanford 3D Scanning and aim@shape repository for the models. We express our gratitude to the Blender Foundation for the scenes from the movie Big Buck Bunny. Further, we would also like to thank the Intel Visual Computing Institute (IVCI) at Saarland University and the French National Research Agency (iSpace&Time) for partial funding of this work.

REFERENCES

[1] Palmer, S. E., [*Vision Science: Photons to Phenomenology*], The MIT Press (1999).
[2] Howard, I. P. and Rogers, B. J., [*Seeing in Depth*], vol. 2: Depth Perception, I. Porteous, Toronto (2002).
[3] Livingstone, M., [*Vision and Art: The Biology of Seeing*], Harry N. Abrams (2002).
[4] Onural, L., Sikora, T., Ostermann, J., Smolic, A., Civanlar, M. R., and Watson, J., "Assessment of 3DTV technologies," in [*NAB Broadcast Engineering*], 456–467 (2006).

Figure 7. Converting a photo (Left) into a 3D image (Middle, anaglyph), just by using the blue channel as depth. Our enhancement (Right, anaglyph) can be used to put stereo cues only where depth contrast exists, minimizing the global error due to the naïve 3D reconstruction, but with locally plausible cues. The insets on the right depict the corresponding disparity maps.

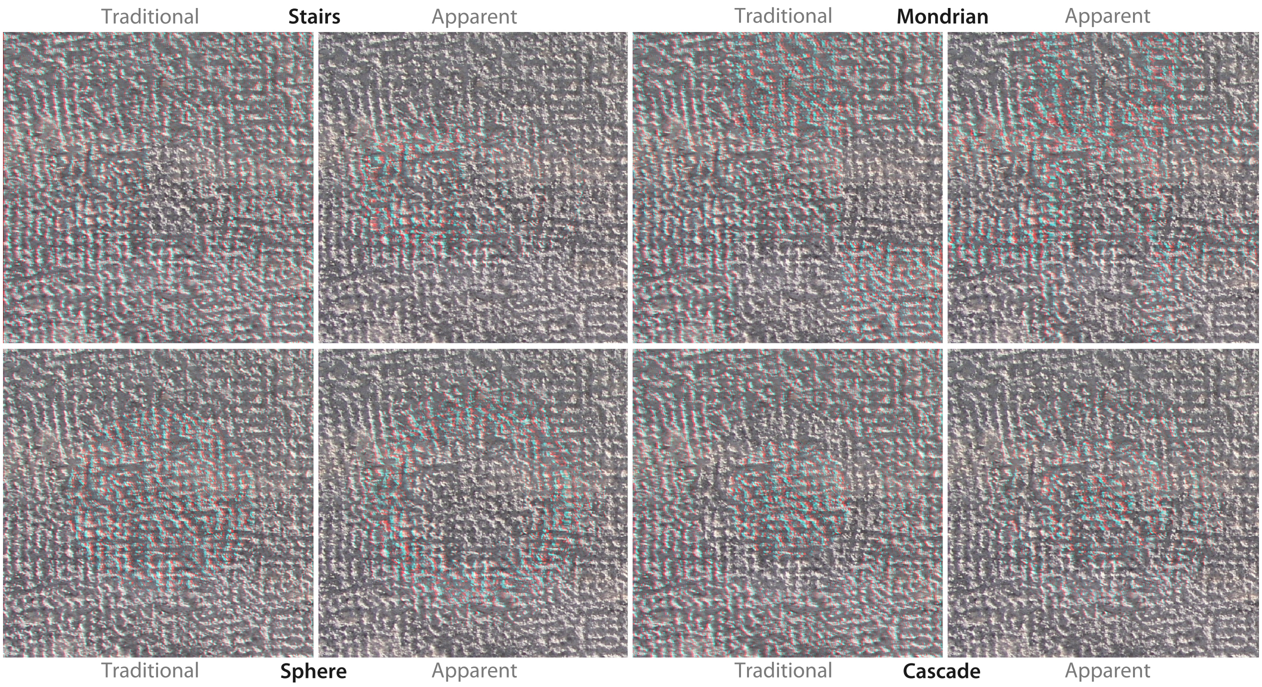

Figure 8. Four stereo images in anaglyph are considered where no stereo cue other than stereopsis is present, similar to random dot stereograms.[14] Note that cascading the local Cornsweet profiles still conveys a consistent impression of discrete depth changes, while in the traditional approach disparity accumulation is required for proper stereoscopic effect.

[5] Matusik, W. and Pfister, H., "3DTV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Transactions on Graphics* **23**(3), 814–824 (2004).

[6] Hoffman, D., Girshick, A., Akeley, K., and Banks, M., "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision* **8**(3), 1–30 (2008).

[7] Jones, G., Lee, D., Holliman, N., and Ezra, D., "Controlling perceived depth in stereoscopic images.," SPIE (2001).

[8] Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., and Gross, M., "Nonlinear disparity mapping for stereoscopic 3D," *ACM Transactions on Graphics* **29**(4), 75:1–10 (2010).

[9] Siegel, M. and Nagata, S., "Just enough reality: comfortable 3-d viewing via microstereopsis," *Circuits and Systems for Video Technology, IEEE Transactions on* **10**(3), 387–396 (2000).
[10] Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., and Seidel, H., "A perceptual model for disparity," *ACM Transactions on Graphics (TOG)* **30**(4), 96 (2011).
[11] Anstis, S. M. and Howard, I. P., "A Craik-O'Brien-Cornsweet illusion for visual depth," *Vision Research* (18), 213–217 (1978).
[12] Rogers, B. and Graham, M., "Anisotropies in the perception of three-dimensional surfaces," *Science* **221**, 1409–1411 (Sept. 1983).
[13] Cutting, J. and Vishton, P., "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in [*Perception of Space and Motion (Handbook Of Perception And Cognition)*], Epstein, W. and Rogers, S., eds., 69–117, Academic Press (1995).
[14] Julesz, B., "Binocular Depth Perception without Familiarity Cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis," *Science* **145**, 356–362 (July 1964).
[15] IJsselsteijn, W. A., de Ridder, H., and Hamberg, R., "Perceptual factors in stereoscopic displays: The effect of stereoscopic filming parameters on perceived quality and reported eyestrain," in [*Human Vision and Electronic Imaging III*], Rogowitz, B. E. and Pappas, T. N., eds., 282–291, SPIE 3299 (1998).
[16] Ijsselsteijn, W., de Ridder, H., Hamberg, R., Bouwhuis, D., and Freeman, J., "Perceived depth and the feeling of presence in 3DTV," *Displays* **18**(4), 207–214 (1998).
[17] Kingdom, F. and Moulden, B., "Border effects on brightness: A rreview of findings, models and issues," *Spatial Vision* **3**(4), 225–62 (1988).
[18] Pratt, W. K., [*Digital Image Processing*], John Wiley & Sons (1991).
[19] Krawczyk, G., Myszkowski, K., and Seidel, H.-P., "Contrast restoration by adaptive countershading," *Computer Graphics Forum* **26**, 581–590 (Sept. 2007).
[20] Sato, M., "A psychophysical study on the anisotropy and individual differences in human depth perception," *International Congress Series* **1269**, 97–100 (2004). Brain-Inspired IT I.
[21] Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., Heidrich, W., and Myszkowski, K., [*High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*], Morgan Kaufmann, Second Edition (2010).
[22] Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R., "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics* **27**(3), 67:1–67:10 (2008).
[23] Li, Y., Sharan, L., and Adelson, E. H., "Compressing and companding high dynamic range images with subband architectures," *ACM Transactions on Graphics* **24**(3), 836–844 (2005).
[24] Weyrich, T., Deng, J., Barnes, C., Rusinkiewicz, S., and Finkelstein, A., "Digital bas-relief from 3D scenes," *ACM Transactions on Graphics* **26**, 32 (July 2007).
[25] Luft, T., Colditz, C., and Deussen, O., "Image enhancement by unsharp masking the depth buffer," *ACM Transactions on Graphics* **25**(3), 1206–13 (2006).
[26] Bruckner, S. and Gröller, E., "Enhancing depth-perception with flexible volumetric halos," *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1344–51 (2007).
[27] Ritschel, T., Smith, K., Ihrke, M., Grosch, T., Myszkowski, K., and Seidel, H., "3D unsharp masking for scene coherent enhancement," *ACM Trans. Graph.* **27**(3), 90:1–8 (2008).
[28] Bezerra, H., Eisemann, E., Décoret, X., and Thollot, J., "3d dynamic grouping for guided stylization," in [*NPAR '08: Proceedings of the 6th International Symposium on Non-photorealistic Animation and Rendering*], 89–95, ACM (2008).
[29] Didyk, P., Ritschel, T., Eiseman, E., Karol, M., and Seidel, H.-P., "Adaptive image-based stereo view synthesis," in [*Proc. VMV*], (2010).
[30] Saxena, A., Chung, S. H., and Ng, A. Y., "Learning depth from single monocular images," in [*In NIPS 18*], MIT Press (2005).