# Computation and Perception: Building Better Displays

Diego Gutierrez
Graphics & Imaging Lab
Universidad de Zaragoza

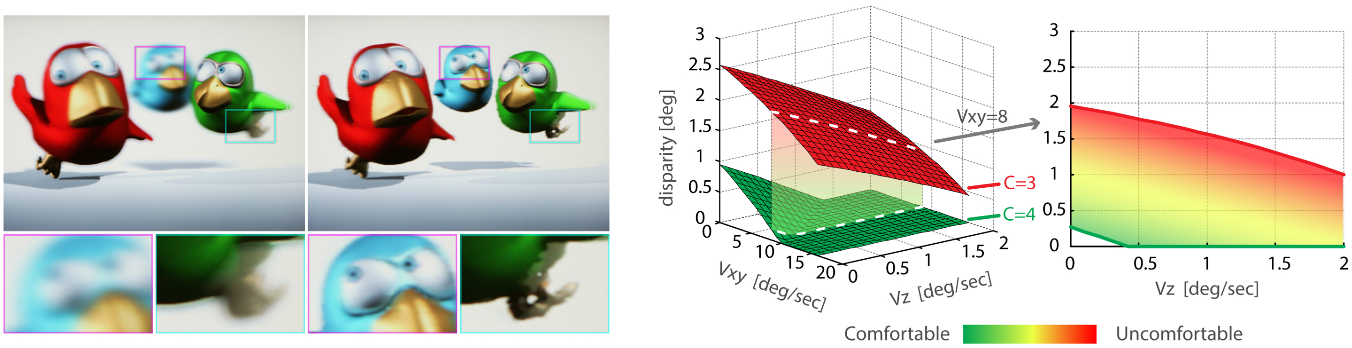


Figure 1: Applying perceptual considerations to improve the viewing experience (for the case of multiview and stereo content). Left: a computational model of human perception allows to overcome display-specific depth of field limitations for automultiscopic displays; details in blurred areas are recovered, while maintaining the apparent depth of the displayed objects [Masia et al. 2013a]. Right: a novel metric of visual comfort for stereo motion predicts comfort zones as a function of velocity, disparity and spatial frequency of the luminance; regions outside those zones should be avoided when creating stereo content [Du et al. 2013].

## Abstract

Computational displays have recently emerged as a fascinating new research area. By combining smart processing with novel optics and electronics, their ultimate goal is to provide a better viewing experience. This may be achieved by means of an extended dynamic range, a better color reproduction, or even glasses-free stereoscopic techniques. However, no matter what the improvements are, these will always be bounded by the limitations imposed by current technology. We argue that by adding *perceptual models of human vision* to the design of the displays, some of these hard limitations can be circumvented, providing an enhanced viewing experience beyond what should be physically and technically possible. In this paper we show examples of how such perceptually-based strategy is currently being applied in different prototype implementations.

**Keywords:** computational displays, applied perception

## 1 Introduction

*Computational displays* is an emergent, vibrant new field that studies the co-design of optics, electronics and computation to provide a more vivid, accurate and even immersive viewing experience to the user. Some of the proposed improvements include an extended dynamic range, better color reproduction, or improved parallax capabilities for 3D display [Masia et al. 2013b; Wetzstein et al. 2012a]. However, any of these advances is bounded by technological or physical limitations such as displayable color gamut (given by the display's primaries), maximum luminance or the angular-spatial resolution trade-off in automultiscopic displays. Take for instance automultiscopic displays; for these devices to be able to show a light field (a multiview representation of the same scene from slightly different perspectives) in real time, they would require a bandwidth of over one terabyte per second, which is way beyond what is feasible today. Instead, the newly introduced *compressive displays* explore the joint-design of optics, electronics, and computational processing, to exploit compressibility of the presented data (e.g., [Wetzstein et al. 2011; Wetzstein et al. 2012b]), based on the observation that all those images from slightly different points show a high degree of redundancy. Many other examples and strategies exist: Figure 2 shows a recent taxonomy based on which dimension of the plenoptic function each technique addresses.

So what role does perception play in all this? In the end, the goal of the displayed images is to be seen by a human observer. It thus makes sense to take this into account when showing visual information. This is in fact what has been done since, for instance, the adoption of RGB primaries for color TV, following the inner workings of the human eye and its three specialized color-sensitive cells.

Human perception is a fascinating topic in itself, which cannot be covered in any one single paper. It is an open research topic, involving many different fields. Nevertheless, without necessarily knowing all its intricacies, we can leverage what we do know and make it a building block of novel display technologies. For instance, color appearance models (CAMs) aim at predicting how colors will be *perceived* by an observer; from there, it is possible to adapt images and video to external specific viewing conditions such as environment illumination, to produce a more accurate perceived color reproduction [Reinhard et al. 2012].

In this paper we present an overview of some of our recent advances in the fields of computational imaging and displays, taking into account aspects of human perception. In particular, we focus on reverse tone mapping [Masia et al. 2009], 3D content remap-

| | Contrast and Luminance Range | Color Gamut | Spatial Resolution | Temporal Resolution | Angular Resolution I - Stereo | Angular Resolution II - Automultiscopic |
|---|---|---|---|---|---|---|
| Perceptual Considerations | Dynamic range of the eye (photopic-mesopic-scotopic vision), dynamic adaptation, CSF, Craik-O'Brien-Cornsweet illusion, visual masking | Dual-process theory, trichromatic & color-opponent stages, chromatic adaptation, standardized observers | Photoreceptor density, foveal and peri-foveal vision, SPEM | Temporal integration, Bloch's law, CFF, hold-type blur, SPEM | Panum's fusional area, zone of comfort, acc.-vergence conflict, DSF, Craik-O'Brien-Cornsweet illusion for depth, disparity models | Disparity models, motion parallax, accommodation, cue integration |
| Display Architectures | Two-layer optical modulation | Increasing the purity of primaries | Optical superposition | Backlight flashing | Spatial multiplexing (anaglyph, polarization) | Volumetric displays |
| | Local dimming | Multi-primary displays | Temporal superposition | | Temporal multiplexing (shutter glasses) | Light field displays |
| | Multi-projector systems | Projection systems | Optical Pixel Sharing | | Backward-compatible stereo | Compressive light field displays |
| | | | | | Autostereoscopic | Light field displays supporting accommodation |
| Content Generation | HDR image/video acquisition | Color Appearance Models | Super-resolution | Black data insertion | Camera parameters adjustment | Efficient image synthesis |
| | Apparent brightness enhancement | Gamut mapping | Sub-pixel rendering | Frame interpolation techniques | Disparity remapping | Light field retargeting |
| | Tone mapping | Radiometric calibration | Temporal integration | Motion compensated inverse filtering | Microstereopsis and Backward-compatible stereo | Disparity remapping |
| | Reverse tone mapping | | | Warping techniques | | Multiview from stereo |
| | | | | Leveraging information from the rendering pipeline | | |
| | | | | Mesh-based temporal upsampling | | |

Figure 2: A classification of modern display architectures and technologies (Table from [Masia et al. 2013b]).

ping for automultiscopic displays [Masia et al. 2013a], and a perceptual study of motion-induced discomfort when viewing stereo content [Du et al. 2013]. We refer the reader to the original papers for more details, and to the recent survey by Masia et al. [2013b] for a broader perspective on the topic.

## 2 Reverse Tone Mapping

Tone mapping scales down high dynamic range (HDR) content to fit the capabilities of a low dynamic range (LDR) display. Since the first methods were introduced to the graphics community [Tumblin and Rushmeier 1993; Ward 1994], many others have been proposed over the last couple of decades [Čadik et al. 2013]. *Reverse* tone mapping deals with the opposite problem: how to expand all the existing legacy LDR content for correct visualization on a modern HDR display (see Figure 3, left).

One of the first bit-depth (range) extension methods was proposed by Daly and Feng [Daly and Feng 2003], with several others following (such as [Banterle et al. 2006; Meylan et al. 2006; Rempel et al. 2007]). They also propose different strategies to enhance dynamic range, but under one common characteristic: the input LDR content was assumed to be well-exposed. We argue that the vast amount of LDR legacy content actually spans a large range of under- or over-exposures, due to different reasons, from bad setting choices to artistic intentions. We thus extend previous studies by taking into consideration varying exposure conditions [Masia et al. 2009].

We first evaluate some of the most popular existing reverse tone mapping operators. This allows us to identify a weakness in all of them: they do not deal well with overexposed content. In fact, most of the times the expanded content tends to look worse than the original LDR one. We make the following key observations: on the one hand, darker HDR depictions are usually preferred for bright input LDR images; on the other hand, in many cases contrast enhancements improve perceived image quality. Moreover, simple global reverse tone mappers, such as linear scaling and gamma boosting, never cause polarity reversals, ringing artifacts or spuriously boost regions well beyond their context. We then propose a global gamma-curve operator, with the particular desired characteristic that its specific gamma-value adapts to the degree of overexposure of the input image. Further extensions are proposed in [Masia and Gutierrez 2011], investigating other image metrics and statistics. This operator is shown to outperform all others when dealing with overexposed content, while being consistent throughout different exposure levels [Masia et al. 2009].

Additionally, our work reveals two important conclusions: First, a reasonably predictive evaluation of a reverse tone mapping algorithm can be made without directly testing on an HDR monitor. Second, the subjective opinions on the quality of HDR images that have been generated from LDR content seem to depend more on the presence or absence of disturbing spatial artifacts than on the exact intensities of different features. Figure 3, right, shows the results of our gamma-based operator, compared to some of the most popular existing ones: it can be seen how visible artifacts such as amplification of invisible contrast or contrast reversal are minimized with our method. Based on these findings, a novel selective reverse tone mapping operator was later introduced, where saliency information at object level allows to assign more dynamic range to predicted regions of interest in the image [Masia et al. 2010] (see Figure 4).

## 3 3D Content Remapping

Glasses-free automultiscopic displays are capable of producing the illusion of 3D content without the need of any additional eyewear. However, due to limitations in angular resolution, they can only show a limited depth of field, which translates into blurred-out areas whenever an object extrudes beyond a certain depth. The degree of blurriness is dependent on the type and specific characteristics of the display. Display-specific depth of field expressions have been derived for parallax barrier and lenslet-based systems [Zwicker et al. 2006], multilayer displays [Wetzstein et al. 2011], and directional backlit displays [Wetzstein et al. 2012b]. Figure 5 shows

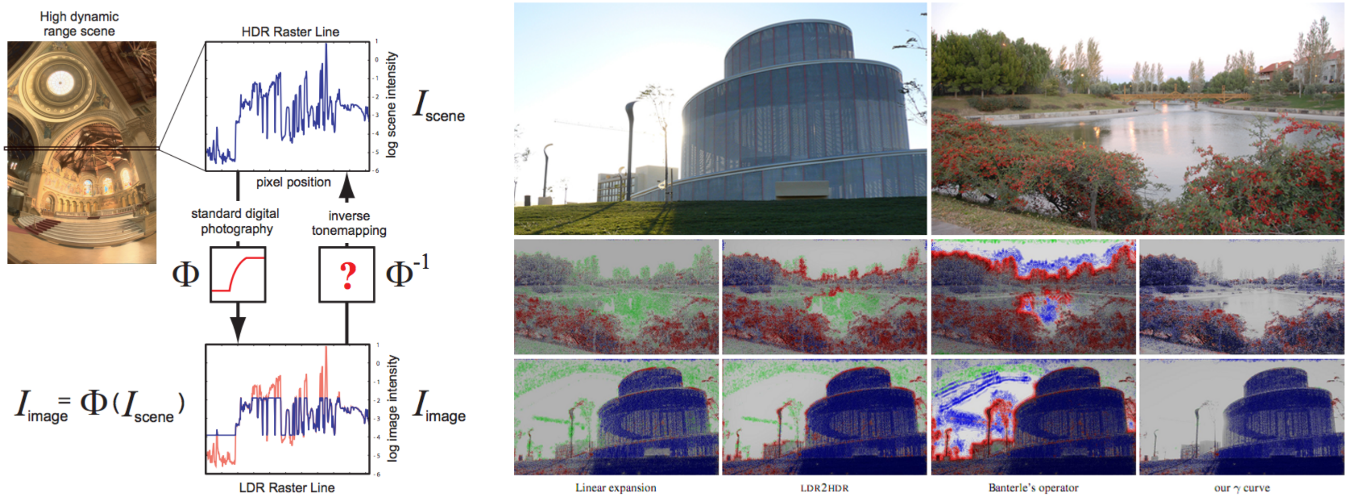


Figure 3: *Left:* The reverse tone mapping problem. Standard imaging loses dynamic range by transforming the raw scene intensities $I_{scene}$ through some unknown function Φ, which clips and distorts the original scene values to create the $I_{image}$ (clipped values shown in red). The goal of reverse tone mapping is to invert Φ to reconstruct the original luminance. *Right:* Comparing the results of several reverse tone mapping algorithms (input images from [Martin et al. 2008]). Green, blue and red identify loss of visible contrast, amplification of invisible contrast and contrast reversal respectively. Our gamma expansion does not lose any contrast, while minimizing gradient reversals. More importantly, it reveals more detail in the most significant areas of the images (trees, grass, bushes and buildings in the images shown). Both figures adapted from [Masia et al. 2009].

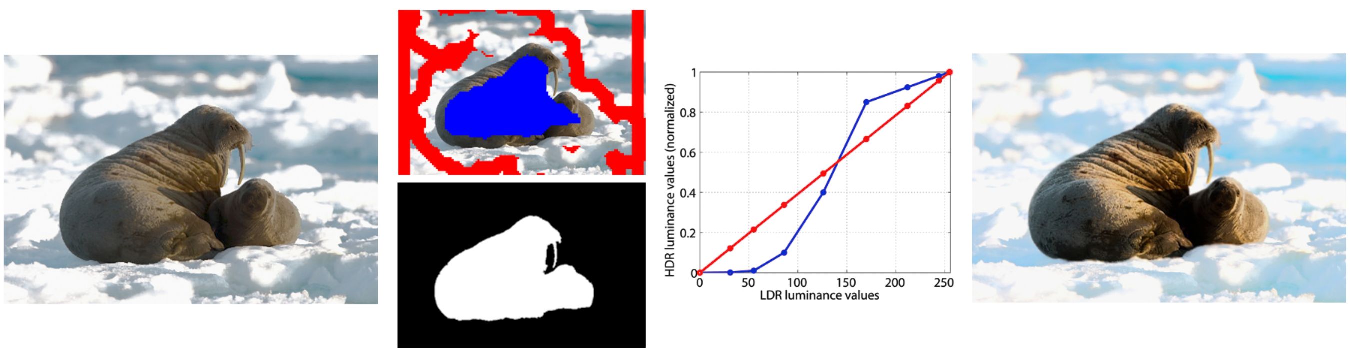


Figure 4: Selective reverse tone mapping [Masia et al. 2010]. *From left to right:* Input LDR image (copyright of National Geographic); auto-labeling of foreground and background and subsequent binary mask (see paper for details); expansion curves for foreground (blue) and background (red); final HDR image.

simulated views for a 3D scene, as seen on different displays.

In order to display an aliasing-free light field in an automultiscopic display, four-dimensional spatio-angular filters need to be applied. In practice, these filters model the depth-dependent blur of the individual displays and are described by a depth of field blur applied to the target light field. Intuitively, this approach fits the content into the DOF of the displays by blurring it as necessary; however, high frequency details are lost. On the other hand, simply reducing the apparent depth of the 3D content to the displayable range of the device (flat areas in the charts of Figure 5) maintains the original high frequency details, but the sensation of depth is reduced drastically.

We propose a solution to handle the intrinsic trade-off between the spatial frequency that can be shown in a display, and the perceived depth of a given scene [Masia et al. 2013a]. In particular, we combine exact formulations of display-specific depth of field limitations with models of human perception, to find an optimal solution. We take into account the frequency-dependent sensitivity to contrast of the human visual system, as well as the sensitivity to binocular disparity. We then obtain a function to be optimized, where a first objective term minimizes the perceived luminance and contrast difference between the original and the displayed scene, effectively minimizing blur, while a second term strives to preserve the perceived depth.

To take into account how frequency changes are perceived by a human observer, we rely on the fact that the visual system is more sensitive to near-threshold changes in contrast and less sensitive at high contrast levels. We adopt a conservative approach and employ sensitivities at near-threshold levels as defined by the contrast sensitivity function (CSF). To preserve perceived depth, we leverage the fact that the effect of binocular disparity in the perception of depth works in a manner similar to the effect of contrast in the perception of luminance. In particular, our ability to detect and discriminate depth from binocular disparity depends on the frequency and amplitude of the disparity signal, among other things.

We apply our method to the central view of the light field, and synthesize the rest of the views by warping. The end result (see Figure

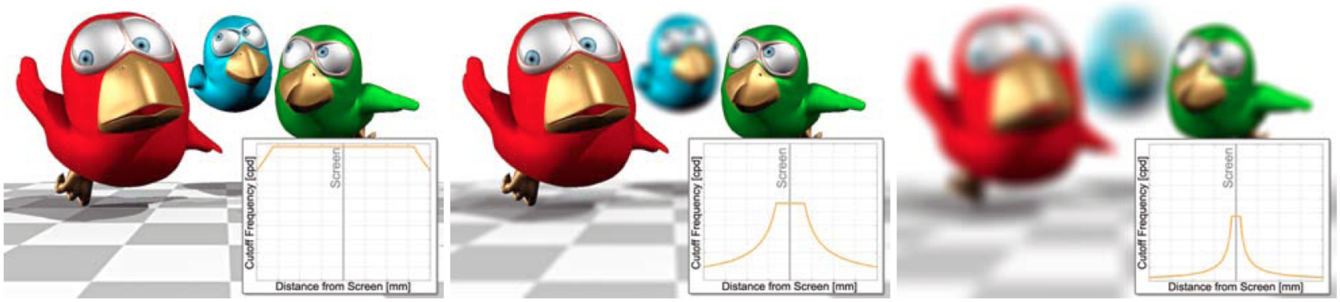


Figure 5: Simulated views of a 3D scene for three different displays. From left to right: Holografika HoloVizio C80, desktop and cell phone displays. Note how the last two displays fail to reproduce the scene properly, blurring out details due to the intrinsic depth-of-field limitations of the displays. The insets plot the depth vs. cut-off frequency charts for each display (image from [Masia et al. 2013a]).

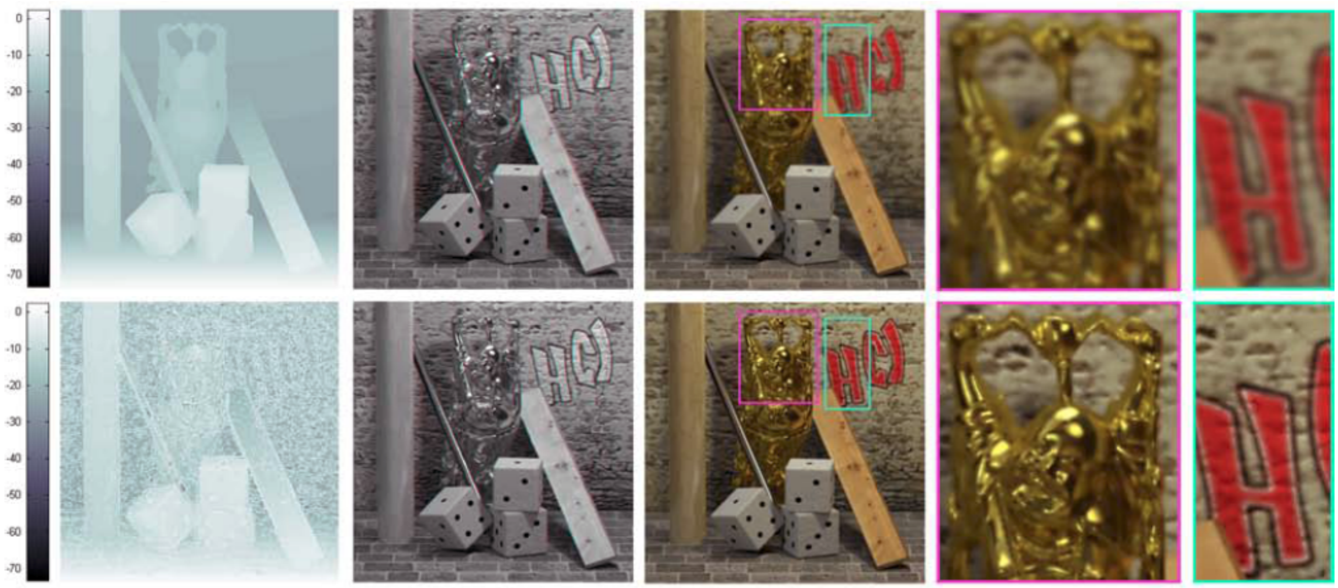


Figure 6: Result on a data set from the Heidelberg light field archive. The top row shows the original scene, while the bottom row shows our retargeted result. From left to right: depth map, anaglyph representation, central view image, and selected zoomed-in regions. Notice how our method recovers most of the high frequency details of the scenes, while preserving the sensation of depth (image from [Masia et al. 2013a]).

6) is a modified depth map that keeps both sharp details *and* the perception of depth (see Figure 1, left). The proposed framework can also be applied to retargeting disparities in stereoscopic image displays, supporting both dichotomous and non-dichotomous comfort zones [Shibata et al. 2011].

## 4 Stereo Motion

The recent popularity of stereo content has driven new applied research regarding the human visual system and stereo vision [Howard and Rogers 2002; Pollock et al. 2012]. In recent works, authors have analyzed the vergence-accommodation conflict [Shibata et al. 2011], a perceptual model for disparity [Didyk et al. 2011], or the influence of luminance contrast in perceived disparity [Didyk et al. 2012], to name a few works related to graphics.

It is well known that the vergence-accommodation conflict is a key factor when predicting the visual discomfort that stereo content may cause on the viewer. The safe area where stereo content can be placed (with objects appearing in front of and behind the screen) is called the comfort zone. However, it is also known that eye movement can also cause discomfort [Bahill and Stark 1975]: it follows that predicting comfort needs to take into account the *motion* of the displayed stereo content as well, and not just its depth relative to the screen. Here we summarize the results of our study in this regard: we run a psychophysical experiment where we systematically explore a large parameter space considering many aspects of stereo motion (the disparity $d$, planar onscreen velocity $v_{xy}$, velocity in depth $v_z$ and the spatial frequency of the luminance in the image $f_l$), and build a reliable measurement of visual comfort; from there, we derive a predictive metric to guide the placement of stereo content (see Figure 1, right). We refer the reader to the original publication for further details [Du et al. 2013].

Analyzing the data of our experiment, we determine a statistical measurement of visual comfort for stereoscopic motion. Figure 7 shows some representative slices of such measurement. Analyzing this carefully allows us to infer some important conclusions:

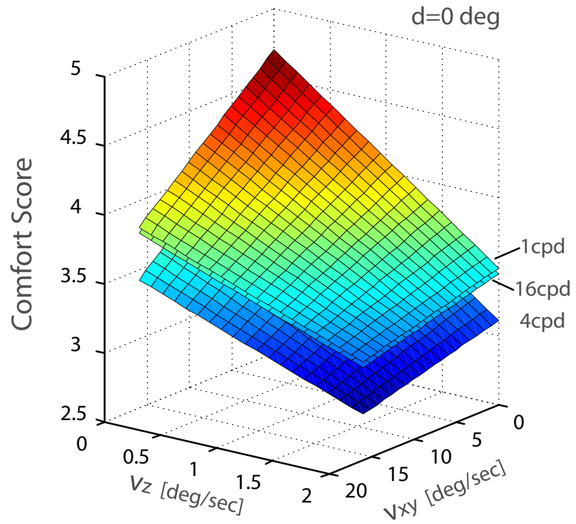


Figure 7: Example slices of our measurement function, for the case of $d = 0$. Each slice represents a different luminance frequency. Higher comfort scores yielded by the measurement predict better visual comfort (image from [Du et al. 2013]).

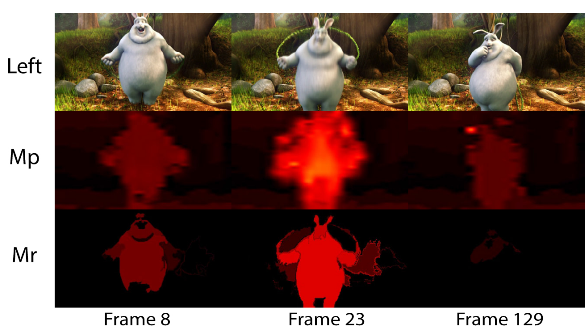


Figure 8: Comfort maps computed using our metric on three representative frames of the bunny movie (copyright Blender Foundation). From top to bottom: input frames, per-pixel results, and per-region results (brighter red indicates less comfort). Our metric predicts less comfort with faster movement (frame 23), in agreement with the perceptual experiments (adapted from [Du et al. 2013]).

- The sign of the disparity affects visual comfort. This was well-known for the case of static images, but had never been proven for the case of stereo motion
- The combination and interaction of all the parameters considered in the study affect visual comfort
- The spatial frequency of the luminance is a non-linear factor of visual comfort

Moreover, our measurement allows us to derive a metric to compute both a pixel-wise comfort map for each frame in a video sequence, and a global comfort score for the whole video. This can be seen in Figure 8, where brighter red areas show potential areas of visual discomfort. Notice how the metric correctly predicts more discomfort due to the faster movement present in frame 23.

## 5 Conclusions

The works discussed here represent just a cross-section of our research on perceptually-based computational displays; we again refer the reader to other sources for a more comprehensive taxonomy [Masia et al. 2013b; Wetzstein et al. 2012a]. We believe that the general field of computational displays will continue to grow, based on joint advances in optical design, electronics, or perceptual models. The field will thus benefit from a deeper understanding of the human visual system, and it will require a multidisciplinary effort including hardware specialists, physicists, neuroscientists, optics and image processing specialists. Last, advances in display technology will rely somehow on advances in capture technology and vice versa, since they both share the same basic problem: how to tame the high dimensionality of the plenoptic function.

Many avenues of future work remain open. In the context of this paper, reverse tone mapping is still an open problem, for which a definite solution valid for all content may not exist. Our 3D content remapping operator could be extended with more sophisticated models of human perception, including for instance motion parallax. Last, some applications of the comfort metric for stereo motion include: Stereoscopic production, using it as a guide to place stereo content when motion is present; Stereoscopic retargeting, defining zones of comfort to be incorporated as constraints in a retargeting operator; or visualization in virtual environments, for instance providing constraints to the user-defined scene navigation, to name a few.

## Acknowledgements

We would like to thank all the co-authors of the papers discussed here, as well as other collaborators that have helped in these works. This research has been partially funded by the European Commission (projects GOLEM and VERVE), and by the Gobierno de Aragón additionally (project TAMA).

## References

BAHILL, A. T., AND STARK, L. 1975. Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Exp Neurol 48*, 1, 95–106.

BANTERLE, F., LEDDA, P., DEBATTISTA, K., AND CHALMERS, A. 2006. Inverse tone mapping. In *Proc. of the 4th Intnl. Conf. on Comp. Graph. and Interactive Tech. in Australasia and Southeast Asia*, GRAPHITE '06, 349–356.

DALY, S. J., AND FENG, X. 2003. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, R. Eschbach and G. G. Marcu, Eds., vol. 5008 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 455–466.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2011, Vancouver) 30*, 4, 96:1–96:10.

DIDYK, P., RITSCHEL, T., EISEMANN, E., SEIDEL, H.-P., MYSZKOWSKI, K., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Transac-*

*tions on Graphics (Proceedings SIGGRAPH Asia 2012, Singapore) 31*, 5.

DU, S.-P., MASIA, B., HU, S.-M., AND GUTIERREZ, D. 2013. A metric of visual comfort for stereoscopic motion. *ACM Transactions on Graphics (SIGGRAPH Asia) 32*, 6.

HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth. Vol. 2: Depth Perception*. Oxford University Press.

MARTIN, M., FLEMING, R. W., SORKINE, O., AND GUTIERREZ, D. 2008. Understanding exposure for reverse tone mapping. In *Congreso Español de Informatica Grafica*, Eurographics.

MASIA, B., AND GUTIERREZ, D. 2011. Multilinear regression for gamma expansion of overexposed content. Tech. rep., Universidad de Zaragoza (Technical Report RR-03-11), July.

MASIA, B., AGUSTIN, S., FLEMING, R. W., SORKINE, O., AND GUTIERREZ, D. 2009. Evaluation of reverse tone mapping through varying exposure conditions. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia) 28*, 5.

MASIA, B., FLEMING, R. W., SORKINE, O., AND GUTIERREZ, D. 2010. Selective reverse tone mapping. In *Congreso Español de Informatica Grafica*, Eurographics.

MASIA, B., WETZSTEIN, G., ALIAGA, C., RASKAR, R., AND GUTIERREZ, D. 2013. Display adaptive 3d content remapping. *Computers & Graphics 37*, 8.

MASIA, B., WETZSTEIN, G., DIDYK, P., AND GUTIERREZ, D. 2013. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics 37*, 8.

MEYLAN, L., DALY, S., AND SSSTRUNK, S. 2006. The Reproduction of Specular Highlights on High Dynamic Range Displays. In *IS&T/SID 14th Color Imaging Conference (CIC)*.

POLLOCK, B. T., BURTON, M., KELLY, J. W., GILBERT, S., AND WINER, E. 2012. The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics 18*, 581–588.

REINHARD, E., POULI, T., KUNKEL, T., LONG, B., BALLESTAD, A., AND DAMBERG, G. 2012. Calibrated image appearance reproduction. *ACM Trans. Graph. 31*, 6 (Nov.), 201:1–11.

REMPEL, A. G., TRENTACOSTE, M., SEETZEN, H., YOUNG, H. D., HEIDRICH, W., WHITEHEAD, L., AND WARD, G. 2007. Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph. 26*, 3 (July).

SHIBATA, T., KIM, J., HOFFMAN, D., AND BANKS, M. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Jurnal of Vision 11*, 8.

TUMBLIN, J., AND RUSHMEIER, H. 1993. Tone reproduction for realistic images. *IEEE Comput. Graph. Appl. 13*, 6 (Nov.), 42–48.

ČADIK, M., HAJDOK, O., LEJSEK, A., FIALKA, O., WIMMER, M., ARTUSI, A., AND NEUMANN, L., 2013. Evaluation of Tone Mapping Operators. `http://dcgi.felk.cvut.cz/home/cadikm/tmo/`.

WARD, G. 1994. Graphics gems iv. Academic Press Professional, Inc., San Diego, CA, USA, ch. A contrast-based scalefactor for luminance display, 415–421.

WETZSTEIN, G., LANMAN, D., HEIDRICH, W., AND RASKAR, R. 2011. Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. *ACM Trans. Graph. (SIGGRAPH) 30*, 4, 1–12.

WETZSTEIN, G., LANMAN, D., GUTIERREZ, D., AND HIRSCH, M., 2012. Computational Displays. ACM SIGGRAPH Course Notes.

WETZSTEIN, G., LANMAN, D., HIRSCH, M., AND RASKAR, R. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH) 31*, 4, 80.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for Automultiscopic 3D Displays. In *Proc. EGSR*.