EUROGRAPHICS 2014 / B. Lévy and J. Kautz
(Guest Editors)

Volume 33 (2014), Number 2

Supplemental Material: Perceptual Depth Compression for Stereo Applications

Dawid Pająk,[1] Robert Herzog,[2] Radosław Mantiuk,[3] Piotr Didyk,[4] Elmar Eisemann,[5] Karol Myszkowski,[2] and Kari Pulli[1]

[1]NVIDIA, [2]MPI Informatik, [3]West Pomeranian University of Technology, [4]MIT CSAIL, [5]TU Delft

Abstract

In this document we present results that are complementary to the ones described in Section 6 of the main paper. We also make several observations on depth compression that, while not being the core of the contributions we make, are still valuable and might serve as a basis for future research.

1. Efficient coding of depth values

In contrast to display devices, that use additive RGB color space to reconstruct the image on the screen, video coders represent and process image data in an opponent color space. The key property of such a representation is statistical decorrelation of color channels, which allows the encoder to compress them independently and focus the encoding on perceptually significant information. Additionally, pixel luminance L value is stored in a gamma-corrected format $Y \approx L^{1/2.4}$, which results in a non-linear luminance quantization that preserves equal visual difference ($\sim$ 1 JND) between two consecutive integer values. Such encoding ensures the optimal usage of the available integer range and improves the compression rate without affecting the visual quality.

Here we propose to follow a similar approach and transform depth frames into a perceptually uniform domain before the video coding process. Instead of depth, we encode disparity, which for stereo applications, where display parameters are usually known in advance, can be considered a measure of perceptually significant depth. The perceptual disparity data is quantized to fixed precision integers. This is because the majority of current video codes compress integer video data, and while specialized solutions for floating point representations exist, here we decide to rely on integer codecs as they are more efficient in both performance and compression.

Note that the transformation presented below is orthogonal to the method we cover in the main part of the paper. There, we rely on the sparsity of data in the residual frame, where most of the disparities are below our perception threshold, and relatively small amount of pixels need their values represented precisely. A non-linear quantization presented here could improve the compression ratio for these few percents of supra-threshold residual data, but would also complicate the coding and decoding procedure significantly. Because of the increased complexity and small compression gains, we have decided to exclude the horopter-dependent quantization of depth values from the main paper, and cover it here instead.

We begin by transforming physical depth denoted in meters to angular disparity, defined as a difference between horopter plane and pixel vergence angles. For typical real-life scenes the horopter plane position can vary, i.e., we can focus our eyes on objects at different depth within the same scene. In our investigations we consider depth compression for two distinct horopter positioning schemes: fixed horopter position (aggressive compression) and position varying within specified depth range (conservative compression). The former option can be useful in scenarios, where the horopter plane is known (compression guided by visual attention models) or fixed, for instance in video conferencing applications, where we pay attention to the user in the front.

To estimate the disparity-to-integer mapping, we adopt the methodology from [MKMS04] and start with the evaluation of R that maps the integer range $[0, 2^{nbits} - 1]$ to angular disparities. For a given integer q, the mapping should keep the quantization error $e(q) = \max\{|R(q+0.5) - R(q)|, |R(q) - R(q-0.5)|\}$ below the perceivable threshold level, which can be described by

$$e(q) \leq tvd(R(q)), \tag{1}$$

where $tvd(d)$ denotes a *threshold versus disparity* function and defines the minimum visibility threshold of a certain disparity stimuli presented against background disparity d. The above inequality can be rewritten as a parametric equation

$$e(q) = \lambda \cdot tvd(R(q)), \tag{2}$$

and by assuming $e(q) \approx 0.5\frac{dR(q)}{dq}$, we can write a differential equation

$$\frac{dR(q)}{dq} = 2 \cdot \lambda \cdot tvd(R(q)). \tag{3}$$

Together with boundary conditions $R(0) = 0arcmin$, $R(2^{nbits} - 1) = 60arcmin$, the above equation defines two-point boundary value problem which can be solved numerically with MATLAB's 'bvp4c' integrator. The quality of resulting mapping function $R(q)$ is described by a λ parameter, which tells us how much smaller the maximum quantization error is compared to the visibility threshold. As $R(q)$ is strictly monotonic, the inverse function that maps disparity values to quantization levels can be evaluated in a straightforward manner using lookup tables.

The perceptual response to angular disparity (positive or negative) is computed with a *transducer* function, which can be derived from the disparity-threshold discrimination data. Recently, [DRE*11] proposed a series of such transducer functions, however, as we quantize disparity in the spatial domain, we need to derive the mapping directly from discrimination data. This can be achieved in two steps. First, using original threshold discrimination function Δd, we estimate the threshold-versus-disparity function:

$$tvd(a) = \Delta d(a, \underset{f}{\operatorname{argmin}}\ \Delta d(a, f)), \tag{4}$$

that for each background disparity level a defines the minimum visibility threshold (among all frequencies f). Then, we integrate the tvd function to compute the final transducer.

Although it has been shown [LHW*10] that disparity comfort zones are not symmetric and we are more efficient at perceiving depth behind the horopter plane than in front of it, here we allocate equal integer range to both positive and negative disparities. We also assume that the input content has been preprocessed to minimize visual discomfort due to diplopia and blur. Effectively, we might be tempted to limit the range of the quantization function to cover disparity values only inside our perception limit, however, such an aggressive quantization would lead to visible geometry distortions during the warping. Therefore, to be on the safe side, we allocate about 15% of available integer range to disparities outside the comfort zone.

For a fixed horopter plane position, 5-bit quantization of response function (Fig. 1) results in $\lambda = 0.5345$, which indicates that the maximum quantization error for any disparity d is equal to $0.5345 \cdot tvd(d)$. This suggests the 6-bit (5-bit value + 1-bit sign) representation has more than enough precision to encode the entire range of visible disparities. This

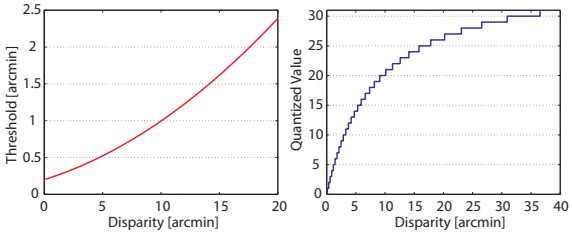

Figure 1: *(left) Threshold vs Disparity function derived from [DRE*11]). (right) Example quantization levels of disparity response function for 5-bit encoding.*

can be justified by the fact that compared to luminance contrast, the range of valid disparity values is relatively small and our visual system is not able to discriminate them very well. In the second scenario, where observer's eyes can verge to disparities in $[-40, 40]$ *arcmin* range , we quantize the disparity to full 8-bits, which produces $\lambda = 0.1385$.

2. Additional Results for the Disparity-Edge Masking Experiment

In addition to the findings in the main paper, here we present more results regarding the perceptual experiments. In Fig. 2 we show the edge discrimination thresholds (x-axis) for each experimental procedure (y-axis) for all eleven test subjects (color-coded points). For large masking amplitudes above 8 *arcmin* the variance of the thresholds increases drastically as some observers have trouble to properly fuse the left/right-eye images. Figure 3 shows 1D slices through the 2D disparity-edge discrimination function that was fit to the measured data in Fig. 2. The 2D function is depicted in the main paper.

References

[DRE*11] DIDYK P., RITSCHEL T., EISEMANN E., MYSZKOWSKI K., SEIDEL H.-P.: A perceptual model for disparity. *ACM Trans. on Graph. 30*, 4 (2011). 2

[LHW*10] LANG M., HORNUNG A., WANG O., POULAKOS S., SMOLIC A., GROSS M.: Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. on Graph. 29* (2010). 2

[MKMS04] MANTIUK R., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Perception-motivated high dynamic range video encoding. *ACM Trans. on Graph. 23*, 3 (2004), 733–741. 1

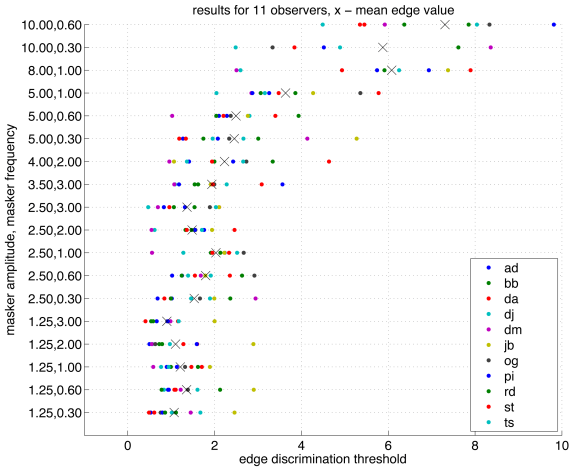

Figure 2: *The edge-disparity discrimination threshold measured during the experiment for individual observers. The OY axis shows the measurement points (masker amplitude in the p-norm scale and masker frequency in [cpd]). The black crosses depict the mean thresholds averaged for all observers.*

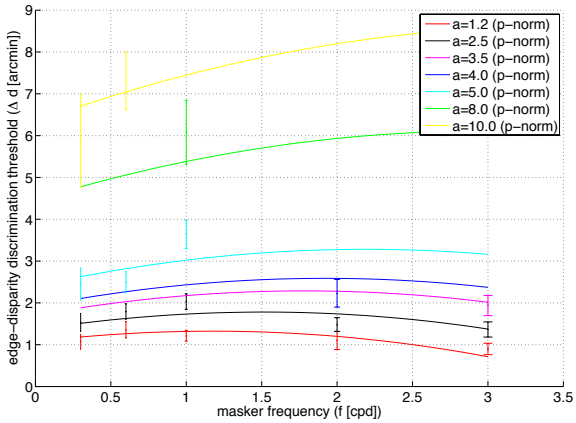

Figure 3: *The edge-disparity discrimination threshold as a function of the masker frequency for a range of masker amplitudes (a). The curves show slices of the 3D fitting function from the paper. The error bars depict the standard error of mean, their colors are consistent with the color of the curves.*