CONTENT AWARE QUANTIZATION: REQUANTIZATION OF HIGH DYNAMIC RANGE BASEBAND SIGNALS BASED ON VISUAL MASKING BY NOISE AND TEXTURE

Jan Froehlich[1,2,3], Guan-Ming Su[1], Scott Daly[1], Andreas Schilling[2], Bernd Eberhardt[3]

[1]Dolby Laboratories Inc., Sunnyvale, CA, USA
[2]University of Tübingen, Tübingen, Germany
[3]Stuttgart Media University, Stuttgart, Germany

ABSTRACT

High dynamic range imaging is currently being introduced to television, cinema and computer games. While it has been found that a fixed encoding for high dynamic range imagery needs at least 11 to 12 bits of tonal resolution, current mainstream image transmission interfaces, codecs and file formats are limited to 10 bits. To be able to use current generation imaging pipelines, this paper presents a baseband quantization scheme that exploits content characteristics to reduce the needed tonal resolution per image. The method is of low computational complexity and provides robust performance on a wide range of content types in different viewing environments and applications.

Index Terms— Image quantization, image encoding, masking, high dynamic range, video encoding.

1. INTRODUCTION

While the most recent advances in image quality previously focused on increasing spatial resolution [1], the focus nowadays shifts to extending dynamic range [2], tonal resolution (i.e., bit-depth) [3] and color gamut [4]. There has been extensive research on the most efficient quantization scheme for encoding high dynamic range (HDR) baseband signals [5]. The resulting 'SMPTE ST.2084' [6] encoding curve needs 12 bits of tonal resolution to quantize any content in its 0-10,000cd/m^2 range without visual artifacts. In the following discussions, this encoding will be referred to as 'PQ' for Perceptual Quantizer. The bit-depths and tonal nonlinearity of PQ were determined by threshold visibility criteria and the most demanding imagery (low gradients having zero noise and no texture). Current monitors can reproduce this tonal resolution of 12 bits by using a 10-bit LCD panel with a local backlight array [7].

While 12 bits is the goal, there are key technologies requiring a lower bit-depth. As an example, interfaces like Display-Port, HDMI and HD-SDI only support 10 bits of tonal resolution for typical applications. File formats like MXF and DPX are also limited to 10 bits in their most used flavors, as are compression codecs like H.264 and popular H.265 (HEVC) Main 10 profile. Therefore it would be desirable to have a quantization scheme available that can quantize any content at 10 bits or lower without introducing visible artifacts. Figure 1 shows the locations for the proposed bit depth reduction and expansion in the image processing chain.

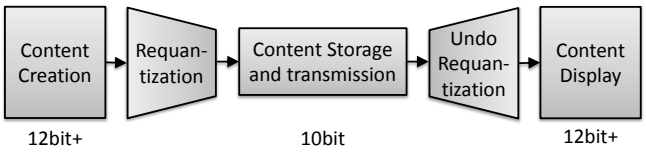

Figure 1. Block diagram for the proposed HDR image storage and transmission flow.

When aiming to reduce the number of code values needed for HDR image quantization below 12 bits, either knowledge about the content, the observer or the viewing environment must be exploited [5]. Prior approaches to this problem either assume limited parameters for content and viewing environment [8] [9] or prioritize tonal resolution on those parts of the tone scale with the largest amount of pixels [10]. Other approaches operate in the frequency-domain [11] [12] and are therefore only applicable to compression, whereas our goal is uncompressed baseband transmission and file storage or pre-processing for compression.

In the following paragraph we will first examine the dependency of needed tonal resolution on image properties by conducting a fundamental study on quantization. From the findings of this study a re-quantization method named CAQ (for Content Aware Quantization) will be derived and compared to determining quantization thresholds by using a current state of the art image difference metric as well as an image based verification study.

2. METHODS

It is known from the work on image difference metrics [13] [14] that the visibility of small differences (like they occur in quantization) depends on the local noise and texture. Since PQ was intentionally designed to exclude noise, CAQ will be designed to exploit the phenomenon of 'masking' [15] [16] of small differences by noise and texture to reduce the needed tonal resolution per image. For camera-captured images, Poisson distributed photon shot noise [17] [18] is a

strong contributor to masking of small differences. To explore the relationship between noise and required quantization we first performed a fundamental study using shallow gradients as test target. The slopes of these gradients were calculated so that the spacing of the quantization steps covered a range of frequencies around the peak contrast sensitivity of the human eye at the respective luminance and the viewing distance of one picture height. A variable amount of spatially uncorrelated white Gaussian noise was added to the gradient, starting with zero noise. As can be seen in Figure 2 the same tonal resolution can show visible contouring (2.b) or be visually lossless (2.j) depending on the noise amplitude on the gradient prior to quantization.

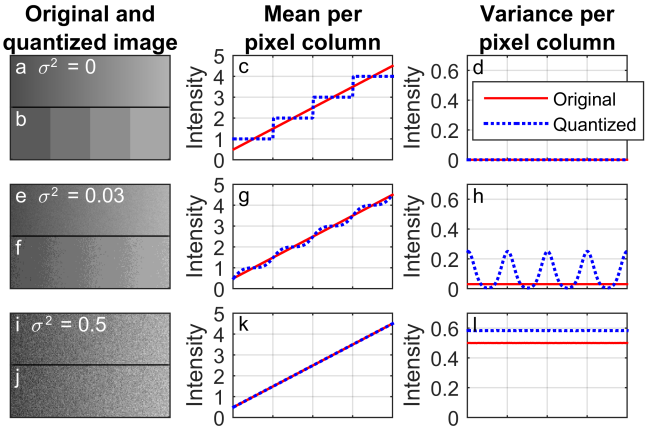

Figure 2. Quantization study test pattern for evaluating the visibility of false contours. The variance of the noise σ^2 is denoted relative to the magnitude of one quantization step.

The three participants of this study were asked to identify the coarsest quantized image for which no change could be spotted when self-toggling [19] between the original image (2.a, 2.e, 2.i quantized at 12bit in PQ) and the same image at lower quantization (2.b, 2.f, 2.j). The investigated variable parameters are listed in Table 1.

Parameters for fundamental quantization study	
Mean Luminance	0.01, 0.1, 1, 10, 100, 300cd/m^2
Temporal frequency	0fps (still image), 24fps
Spatial bandwidth	20, 10, 5 cycles per degree
Amplitude	0, 1, 2, 4, 8, 16, 32, 64 standard deviation σ in 12 bit code-values
Quantization (tonal resolution)	q = 5 to 12 for 2^q code values to encode the full PQ range

Table 1. Variable parameters for quantization study.

All combinations of the parameters in Table 1 were evaluated. Figure 3 shows the minimum bit depth needed for visually lossless quantization of the gradients. While luminance (3.a) as well as temporal- and spatial frequency of the noise (3.b, 3.c) have a low impact on required tonal resolution, it can be seen in (3.d) that the required bit depth is inversely correlated to the amplitude of the noise. As an example, PQ-encoded image areas containing white Gaussian noise with a standard deviation of four 12 bit code values can be quantized using 9 bits without showing any visual difference for all the other parameter combinations.

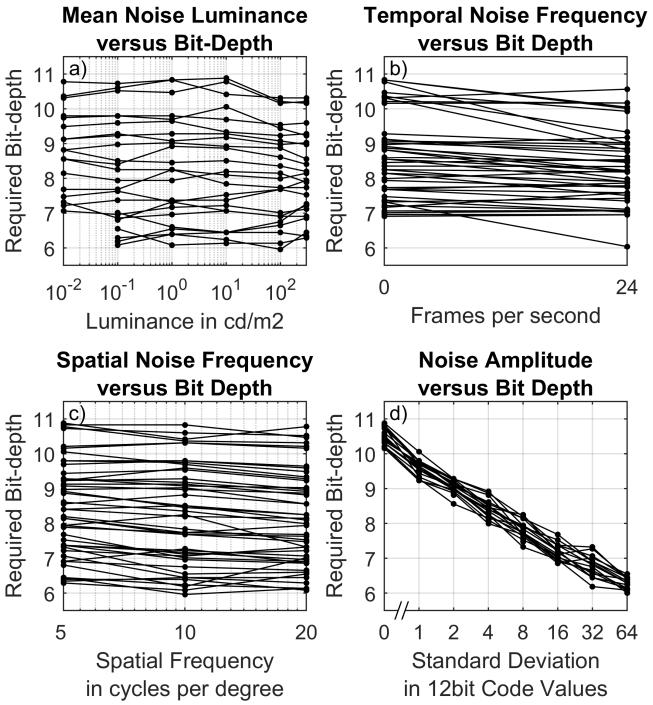

Figure 3. Quantization study results. To illustrate the correlation of the needed bit-depth for visually lossless encoding to the parameter on the x-axis, lines connect study results that only differ in this parameter.

2.1. Content aware quantization (CAQ)

The CAQ quantization scheme introduced in this paragraph is designed to exploit the masking of quantization artifacts by noise and texture as observed in Figure (3.d) to reduce the number of code values needed to quantize an individual image. As for camera-captured images the photon shot noise is related to luminance, CAQ predicts the required quantization for 8 equally spaced intensity segments from 0 to 8/9 in the PQ domain (0 to 3524cd/m^2). Since CAQ predicts the required quantization per intensity segment, but spatially global, it can be applied to the image and removed by means of a simple one-dimensional lookup table (LUT). To undo this variable tone-curve at the receiver side codecs like H.265/HEVC can embed the reverse LUT as SEI message [20] while most file formats and signal interfaces support user defined ancillary data [21].

The block diagram for the CAQ analysis is shown in Figure 4. The input for the algorithm is the PQ-encoded intensity channel. To estimate local noise and texture an isotropic Gaussian high-pass filter (4.b) with a standard deviation of 2.5 pixels for HD images is applied to the PQ

encoded intensity image (4.a). After rectification (4.c), it is blurred (4.d) again to get a robust estimate of the local masking of quantization artifacts per pixel. These steps are consistent with models [22] [23] [24] of the HVS phase uncertainty properties [25].

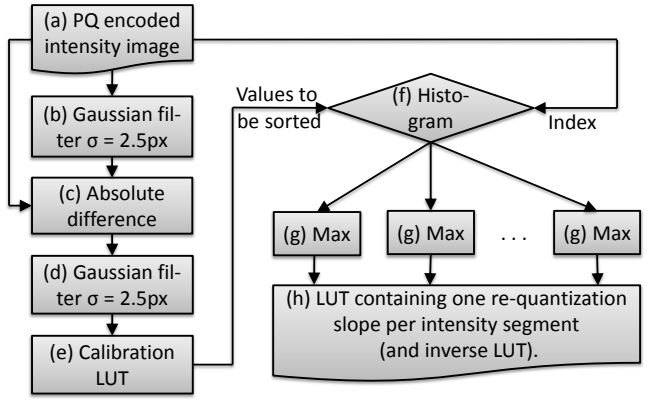

Figure 4. CAQ noise and texture estimation block diagram.

To calculate the minimum allowed quantization level per pixel, a calibration LUT (4.e) is applied. This LUT is derived from the fundamental quantization study as illustrated in Figure 5.

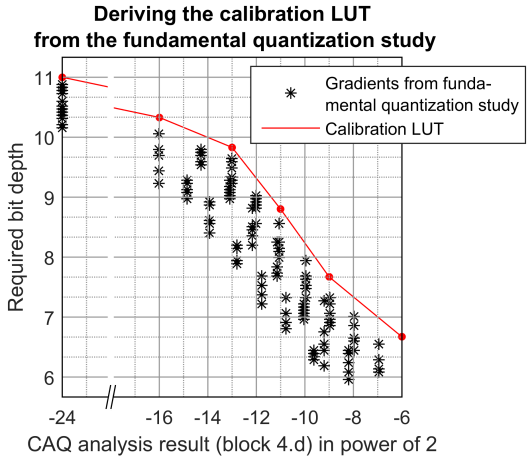

Figure 5. Calibration LUT calculation by determining the minimum required bit-depth for each CAQ analysis results.

Each of the data points in Figure 5 corresponds to one gradient from the fundamental quantization study. The x-axis location shows the CAQ analysis result from block (4.d) while the location on the y-axis corresponds to the needed bit depth for this gradient as found in the fundamental quantization study. Applying the LUT in (4.e) assigns each pixel a bit depth that is sufficient to quantize this pixel without a visual difference. To find the minimum allowed quantization per intensity range per image, the needed bit depth predictions are sorted by the intensity of the corresponding original image pixel into image-dependent histogram bins (4.f). Finally, the minimum allowed quantization for each segment is determined by calculating the maximum (4.g) of the needed bit depths for each bin.

2.2. HDR-VDP-2.2 based quantization

The CAQ design is inspired by the visual difference predictor. A new version of the visual difference predictor for HDR imagery is available as 'High Dynamic Range Visual Difference Predictor 2.2' (HDR-VDP-2.2) [26]. To evaluate how CAQ performs compared to the HDR-VDP 2.2, we also determined the quantization for each of the 8 intensity segments by calculating the coarsest quantization that stays below HDR-VDP 2.2's 50% visibility threshold for all pixels. We opted to use HDR-VDP 2.2 over SSIM [14] with PU curve [27] because HDR-VDP can be used in calibrated luminance mode.

2.3. Image based verification study

To verify CAQ and compare it with re-quantization based on HDR-VDP-2.2, an image based verification study was performed. For this study achromatic still frames were shown on a dual modulation HDR LCD display [7]. Still frames were selected as a worst case because uncorrelated temporal noise and motion also mask quantization artifacts [28] (also see Figure 3.b). During evaluation, the user could adjust the quantization per intensity segment in real-time using physical sliders. To help spot the quantization artifacts, a temporal linearly increasing offset of 1/10 of the current quantization step was added per frame before quantization and subtracted after quantization. This phase-shift of the quantization threshold resulted in continuously moving contouring artifacts. It helped to find the exact detection threshold because of the "pop-out" effect [29] of motion. It also simulates a typical worst case for quantization when an image is followed by a slightly darker version of itself as occurs in 'fade to black' dissolves.

The study was performed in a dark room to keep veiling glare at a minimum level and the viewing distance was one picture height. Eight postproduction experts from the TV- and film industry, who perform image evaluation tasks every day, adjusted the quantization for the 8 intensity segments on 12 images. These images were selected to originate from different technologies (analog film, digital cameras and computer animation) and included Hollywood movies, commercials and TV programs. All participants had 20/20 eyesight. 3 users performed the study without eyesight correction, 4 with glasses, and one user wore contact lenses.

3. RESULTS

The correlation between the image based verification study and CAQ as well as the HDR-VDP-2.2 based quantization is compared in Figure 6. Predictions below the dashed line

may result in visible quantization errors, because the predicted tonal resolution is less than required by the observers of the verification study. Except for the darkest segment CAQ performs better compared to the HDR-VDP-2.2 based re-quantization, as the CAQ predictions tend to be closer to the visibility threshold and mostly above, as needed for worst-case design. For the darkest segments CAQ often over-predicts the required bit-depth compared to the verification study because CAQ does not include an observer glare model [30] [31]. This is by design because observer glare depends on the position, orientation and age of the observer [32]. In addition to the uncertainty about the observer, cropping parts of the image, as performed in pan and scan operations or in the processing for small mobile screens, can fully remove sources of glare from the image. Consequently, relying on observer glare for re-quantization would have worked for our controlled study set up, but will fail for the intended usage scenarios. Therefore, CAQ only exploits image inherent local texture for re-quantization, to stay independent of the observer and display conditions.

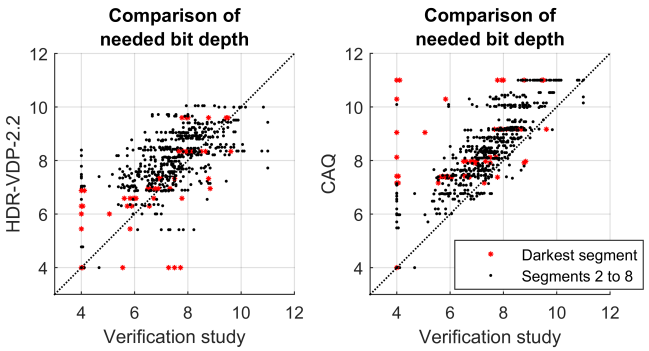

Figure 6. Tonal resolution prediction by HDR-VDP-2.2 and CAQ compared to the image-based verification study.

The Spearman rank order correlation between all results from the verification study was 0.78 for CAQ and 0.74 for the HDR-VDP-2.2 based quantization. When omitting the darkest segment, the correlation for CAQ improved to 0.87 while HDR-VDP-2.2 yielded 0.70.

The typical results for camera-captured images that are subject to photon shot noise can be observed in Figure 7. In the 'Fantasy Flight' image (7.a) the darkest segment requires a tonal resolution of ~6 bits while the brighter segments need ~8 bits of tonal resolution. The 'Showgirl' image (7.b) has a bimodal histogram. In this case, CAQ predicts a much higher needed tonal resolution for the darkest segment compared to the user study and HDR-VDP-2.2. This is due to the intended lack of glare prediction in CAQ. By a using a higher criterion than 50% detection, the prediction of HDR-VDP-2.2 would be lower and likely provide a better statistical fit, but the results would conflict with worst-case design goals of having no images requiring a higher bit depth than the model predictions.

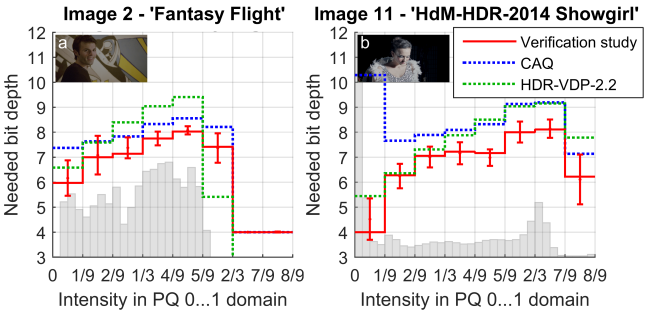

Figure 7. Predicted minimum quantization for two images.

The results from the image based verification study confirm the expectation from the gradient and noise based study. When using *variable* quantization, most camera captured HDR images need 50-200 code values, while computer generated noiseless content typically needs 200 to 500 code values for visually lossless representation. Using CAQ, all images from the verification study could be quantized to less than 1024 code values. For an extended image set of 200 low-noise video frames, only two computer-generated images containing very smooth gradients over the full luminance range needed more than 10 bits of tonal range according to CAQ. To avoid visible artifacts in these cases, the local masking prediction map (4.e) can be used to apply local dithering solely to those areas where the needed quantization bit depth cannot be reached. On an Intel 'Xeon E5 1620 v3' processor CAQ runs about three orders of magnitude faster compared to HDR-VDP-2.2.

4. CONCLUSION

High dynamic range imagery with a tonal resolution of 12 bits and more needs to be quantized at bit depths of 10 bits or less to fit into current image storage and transmission pipelines. We present a robust and fast method to determine the needed tonal resolution of high dynamic range images by exploiting local noise and texture. Our method allows to re-quantize images by means of a simple one dimensional lookup table to 10 bits or less. In addition to the lookup table our method delivers a map of the needed bit depth per pixel. This map can be used to locally apply dithering for those rare cases where more than 10 bits of tonal resolution are needed for visually lossless representation. Compared to using a state of the art image difference predictor, our method performs better for estimating needed quantization and is of significantly lower computational complexity.

5. ACKNOWLEGEMENTS

We thank the study participants and Philipp Kraetzer in conducting the verification study and acknowledge Robin Atkins, Scott Miller, Timo Kunkel, and Peng Yin for their thoughtful comments in proofreading and Pat Griffis for the name 'CAQ'.

6. REFERENCES

[1] International Telecomunication Union, "Parameter values for the HDTV standards for production and international programme exchange," *ITU-R Rec. BT.709-6,* 2015.

[2] Society of Motion Picture and Television Engineers , "Study Group Report: High-Dynamic-Range (HDR) Imaging Ecosystem," *SMPTE,* pp. 1-52, 2015.

[3] A. Luthra, E. François and W. Husak , "Requirements and Use Cases for HDR and WCG Content Distribution," *ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N15084,* 2016.

[4] International Telecommunication Union, "Parameter values for ultra-high definition television systems for production and international programme exchange," *ITU-R Rec. BT.2020-2,* pp. 1-6, 2015.

[5] M. Nezamabadi, S. Miller, S. Daly and R. Atkins, "Color signal encoding for high dynamic range and wide color gamut based on human perception," *IS&T/SPIE Electronic Imaging,* 2014.

[6] Society of Motion Picture & Television Engineers, "High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays," *ST.2084:2014,* pp. 1-14, 2014.

[7] A. Ninan, "Impacting the Display Industry Through Advances in Next Generation Video," *Society for Information Display - Display Week,* 2014.

[8] T. Borer, "WHP 283 - Non-linear Opto-Electrical Transfer Functions for High Dynamic Range Television," British Broadcasting Corporation, 2014.

[9] T. Borer and A. Cotton, "WHP 309 - A “Display Independent” High Dynamic Range Television System," British Broadcasting Corporation, 2015.

[10] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory,* vol. 28, no. 2, pp. 129-137, 1982.

[11] R. Rosenholtz and A. B. Watson, "Perceptual adaptive JPEG coding," *International Conference on Image Processing,* vol. 1, pp. 901-904, 1996.

[12] M. J. Nadenau, J. Reichel and M. Kunt, "Wavelet-based color image compression: exploiting the contrast sensitivity function," *IEEE Transactions on Image Processing,* vol. 12, no. 1, pp. 58-70, 2003.

[13] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology,* pp. 2-15, 1992.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 4, pp. 600-612, 2004.

[15] D. G. Pelli, "Effects of visual noise," *University of Cambridge,* 1981.

[16] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *JOSA Optical Society of America,* vol. 70, no. 12, pp. 1458-1471, 1980.

[17] T. Seybold, C. Keimel, M. Knopp and W. Stechele, "Towards an Evaluation of Denoising Algorithms with Respect to Realistic Camera Noise," *IEEE International Symposium on Multimedia (ISM),* pp. 203-210, 2013.

[18] M. Schoeberl, A. Brueckner, F. Siegfried and A. Kaup, "Photometric limits for digital camera systems," *Journal of Electronic Imaging,* vol. 21, no. 2, p. 020501 1, 2012.

[19] D. M. Hoffman and D. Stolitzka, "A new standard method of subjective assessment of barely visible image artifacts and a new public database," *Journal of the Society for Information Display,* vol. 22, no. 12, pp. 631-643, 2014.

[20] International Telecommunication Union, "Recommendation H.265 - High efficiency video coding," *Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video,* 2015.

[21] Society of Motion Picture & Television Engineers, "Ancillary Data Packet and Space Formatting," *ST.291-1:2011,* pp. 1-17, 2011.

[22] S. Daly, "A visual model for optimizing the design of image processing algorithms," *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference,* vol. 2, pp. 16-20, 1994.

[23] F. Xiao-fan and S. Daly, "Automatic JPEG Compression Using a Color Visual Model," *IS&T PICS: Image Processing, Image Quality, Image Capture Systems Conference,* 2003.

[24] A. Lukin, "Improved visible differences predictor using a complex cortex transform," *International Conference on Computer Graphics and Vision,* 2009.

[25] T. Caelli, M. Huebner and I. Rentschler, "The detection of phase shifts in two-dimensional images," *Perception & Psychophysics,* vol. 37, no. 6, pp. 536-542, 1985.

[26] M. Narwaria, R. K. Mantiuk, M. P. Da Silva and P. Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging,* vol. 24, no. 1, pp. 010501-010501, 2015.

[27] A. Tunc, R. Mantiuk and H.-P. Seidel, "Extending quality metrics to full luminance range images," *Electronic Imaging,* pp. 68060B-68060B, 2008.

[28] A. J. Ahumada Jr, B. L. Beard and R. Eriksson, "Spatiotemporal discrimination model predicts temporal masking functions," *Photonics West - Electronic Imaging,* pp. 120-127, 1998.

[29] P. McLeod, J. Driver, Z. Dienes and J. Crisp, "Filtering by movement in visual search," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 17, no. 1, p. 55, 1991.

[30] A. Rizzi and J. J. McCann, "Glare-limited appearances in HDR images," *Journal of the Society for Information Display,* vol. 17, no. 1, pp. 3-12, 2009.

[31] Commission Internationale de l'Éclairage, "CIE 146:2002 - CIE equations for disability glare," *Color Research & Application,* no. 27, p. 457–458, 2002.

[32] J. J. Vos, "On the cause of disability glare and its dependence on glare angle, age and ocular pigmentation," *Clinical and Experimental Optometry,* vol. 86, no. 6, pp. 363-370, 2003.