Is image quality a function of contrast perception?

Article *in* Proceedings of SPIE - The International Society for Optical Engineering · March 2013

DOI: 10.1117/12.2008620

2 authors:

Andrew M Haun
University of Wisconsin–Madison

Eli Peli
Harvard Medical School

Is image quality a function of contrast perception?

Andrew M. Haun & Eli Peli
Schepens Eye Research Institute, Mass Eye and Ear,
Harvard Medical School, Boston MA

ABSTRACT

In this retrospective we trace in broad strokes the development of image quality measures based on the study of the early stages of the human visual system (HVS), where contrast encoding is fundamental. We find that while presenters at the Human Vision and Electronic Imaging meetings have frequently strived to find points of contact between the study of human contrast psychophysics and the development of computer vision and image quality algorithms. Progress has not always been made on these terms, although indirect impact of vision science on more recent image quality metrics can be observed.

1. CONTRAST PERCEPTION

Human spatial vision is complicated. We know certain facts about it: thresholds vary with spatiotemporal frequency and eccentricity, there is masking and adaptation etc., and all these are mediated by band-pass channels that analyze the image along different dimensions. It was thought and may still be held by many that by understanding the psychophysical properties of contrast perception, and by modeling these more and more precisely, measures of image quality could be designed. Results of new algorithms for encoding or compression or transmission or display of electronic images could be presented to a simulated human visual system that would then return a verdict on image quality consistent with human quality judgments.

The first HVEI meetings in the late 1980's witnessed some of the early attempts to blend the human contrast sensitivity function (CSF) with display modulation transfer functions (MTFs), to yield more-or-less linear measures of the "contrast displayed seen", in an approach dating back to Schade[1]. The underlying distortion of image quality at the time was a bandwidth limitation, i.e. the low-pass and noise characteristics of analog transmission and display systems. It was recognized that the low-pass characteristic of the visual system imposed even more of a premium on high frequency content[2]. There were debates over whether the nonlinear transduction of contrast by the visual system required incorporation of nonlinear measures in CSF-MTF summation methods [3]. The square-root integral method was proposed by Barten in 1989[4], and it remains the best exemplar of this type of direct display/HVS-characteristic synthesis for image quality evaluation.

At the same time, there were proposals that models of image quality should aim not to measure the quantity of image transmitted, but to discriminate visible differences between 'ideal'/original and degraded images, bringing image quality measurement closer to contemporary contrast psychophysics. These newer models incorporated ideas inspired by the multi-scale spatial transform performed by the visual system[5,6], and began to acknowledge the intrinsic, efficient connection between image statistics and visual/neural encoding [7,8]. Methodologies for evaluating subjective image quality were being developed against the backdrop of digital image compression development[9,10]; the first JPEG standard was published in 1992. Digital compression posed different problems for image quality assessment, making clear that the MTF-CSF approach would not be sufficient, since compression entails changes/artifacts in the structure of images[11], not just simple uniform contrast attenuation. The dominant problem was soon clear: algorithms sensitive to the changes imposed by digital compression were needed. Even as Barten honed his model of basic contrast sensitivity[12,13], the need was becoming clear for understanding of contrast sensitivity for specific compression artifacts. Watson and Ahumada, in particular, measured and modeled contrast sensitivity for compression artifacts[14,15]. Through the 90's, there were two parallel trends: study of DCT compression artifact visibility, and study of wavelet quantization error visibility[14-16].

Human Vision and Electronic Imaging XVIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8651, 86510C • © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 • doi: 10.1117/12.2008620

The image quality metrics introduced throughout the 90's were no longer strictly models of image quality; rather, visible difference from an ideal reference image had become the central direction of 'image quality'[18,19]. If compression squeezes out details or introduces artifacts, what matters is whether or not these changes would be visible to a human observer. Convolving an image with a CSF and then compressing the suprathreshold result to approximate human contrast sensitivity was no longer sufficient; spatial masking had to be incorporated, where contrast thresholds are raised by the presence of higher local contrast. Daly's Visual Differences Predictor[18], having only indirect evidence of how masking worked in complex images, incorporated a broad, and in retrospect quite good, educated guess at functional pattern masking. HVS-based models of artifact visibility had become ubiquitous by the mid-90's, and HVEI began to witness more basic contrast psychophysics aimed at bolstering these models. Important developments included the incorporation of contrast gain control processes into pattern masking models[14,20], inspired by recent findings in the visual neurosciences, and based solidly on measurements of human contrast sensitivity and discrimination.

Contrast is important in a trivial sense, in that images cannot be seen without it - contrast is the carrier or medium of visual information. However, above and beyond its importance to image detection, contrast seems to contribute qualitatively to perception. It is widely accepted that higher-contrast images look better, as shown in[21]. This could be due to higher contrast images bringing more content above threshold[2], or simply due to a psychological association between contrast and sharpness[22]. In a similar vein, it has been suggested repeatedly (cf. Kurihara et al[23] for a review) that simply adding high-frequency noise to an image improves its perceived quality (sharpness), in some cases, presumably because of a perceptual association between high frequency contrast and image sharpness. Over the years at HVEI, perceived contrast has been investigated for its links to image quality, although in our estimation, the returns have been of more basic than applied value. Arend[24] described experiments on perceived brightness and contrast in complex, abstract images, while Peli[25] explored just what the proper metric might be for measuring the contrast of a complex or even simple pattern. In an interesting 1994 paper, Pelah[26] described a method for 'inverting' the compression of brightness performed by the visual system, allowing for an observer to objectively view the 'subjective' image; Peli[27] also presented his approach to the simulation of contrast perception and a methodology for the evaluation of the validity of such simulations. This represented psychophysical testing of the concept underlying many of the visual difference metrics. Roufs et al[21] produced an interesting study of the relationship between display gamma and subjective image quality. However, the connection between these explorations of the subjective properties of contrast perception and the goal of image quality computation was always tentative and not explicit.

At the beginning of the last decade, the Modelfest project had just begun[28,29], as an attempt to establish a set of standard contrast sensitivity measures, for the purpose of regularizing the HVS component of image quality (and other vision) algorithms. One goal of the project was that a spatial standard observer (SSO) could be set up based on the Modelfest data. New algorithms meant to emulate visual processes should not violate the SSO, while new human subjects (basic science) studies should seek to violate the SSO in order to improve it. However, while the Modelfest project yielded interesting discussions, it does not seem to have succeeded in setting standards for contrast sensitivity. The main attempt at a Modelfest-based SSO was put forward by Watson and Ahumada in 2005[30]. However, most citations of this paper in the literature seem to be either regarding empirical statements about contrast sensitivity, or as an example of a modeling approach which for whatever reason is not applicable in the context of the citing paper; meanwhile, only a few published studies have used the SSO to benchmark performance[31] or to supply standard sensitivity values for a model observer[32], although Watson and colleagues have incorporated the SSO into tools for display metrology[33]. If the Modelfest project has had the impact intended, it has largely been underground. This is curious: if contrast sensitivity is of fundamental importance to vision algorithms, why has a centralized contrast sensitivity data resource been so neglected?

2. IMAGE QUALITY

Throughout this period, the standard of "image quality" was not particularly scrutinized, except for repeated admissions that image quality is a hard problem that must be, for the time being, practically approximated. Undoubtedly, many more papers were presented on models and algorithms for measuring or predicting image quality or fidelity, than on the question of just what image quality is in the first place. Relatively little work on what should constitute "quality" was presented: non-monotonic "no-reference" quality data such as that of Roufs et al[21] was never an objective of the various quality algorithms that emerged. It seemed to have been settled early on that we all know what image quality is, and that observers can therefore judge the distance of a processed image from our knowledge of the ideal image (note

Heynderickx and Bech's[34] warning against the potentially significant differences between what expert and naive viewers think of "image quality") 'Full reference' discrimination algorithms, computing a type of distance between ideal and processed images, therefore came to relatively dominate the image quality scene. This willingness to accept measure of visible difference may have been driven by the envisioned application of the time. If one wants to control the compression at the source before transmission (or storage) the original image is available and may be used in the way envisioned by these algorithms to achieve a level or degree of difference visibility. In recent years there is a growing interest in reduced- or no-reference measures and no reference measures of quality that necessarily will have to take a different direction in incorporating contrast perception.

By the late 90's, attempts were underway to define video quality metrics using similar approaches to what seemed to be succeeding with static image quality. Spatiotemporal contrast sensitivity measurements were employed to design video quality metrics which quantified visibility of error[35,36]. A popular idea in the late 90's was to make use of the visual system's declining-with-eccentricity contrast sensitivity and resolution to conserve bandwidth for digital video transmission [33-35], foreshadowing today's ROI video encoding. The work of the 90's on visual sensitivity to DCT and wavelet artifacts was recapitulated in the following decade with work on the visibility of MPEG artifacts [40,41,42].

The development of image quality metrics explicitly incorporating human contrast sensitivity has clearly slowed; or, rather, the basic HVS component has become informally standardized in the field and further incremental advances have not been incorporated. The psychophysics seen at HVEI in the recent decade is more likely to relate to attention or higher order visual perception. Consideration of multidimensional attentional salience has clearly superseded perceived contrast when it comes to modeling of suprathreshold stimulus strength [43]. The contrast-based discrimination models were, after all, being applied to a very specific question: what discriminable dimensions do humans interpret as decrements in quality, or as distance from the ideal image? It appears that precisely addressing this question may not, after all, require a more sophisticated or accurate model of the visual system. Although incorporation of attention due to saliency may represent a future way of improving coding (ROI coding), that will then require incorporation of attention measures into video quality algorithms.

Currently established image quality models such as the SSIM[44] and VIF[45] incorporate only very abstract and reduced HVS components, and are nonetheless very successful and consistent with human judgments of image quality – although not more so than HVS-based models (e.g. as shown by Laparra et al[46]). Instead, these models take the point of view that there is information or structure in an image which any system, human or automated, would naturally want to receive; they are thus more concerned with computation of the 'naturalness' of image statistics, rather than with their psychophysical visibility. This is entirely reasonable from a visual point of view: the visual system doesn't care about contrast per se, but uses it as a means of getting at the more complex structure of natural images. Must we reproduce the means if we want to reproduce the ends of image quality judgments?

Here, a theoretical problem emerges: why do different quality metrics with such disparate structures as the VSNR[47], VIF, and SSIM to name just a few, all work so well, giving similar results? What do they have in common, and what, if anything, do they have to do with human spatial vision? Seshadrinathan and Bovik[48] have pointed out that the VIF and SSIM structures both can be seen as carrying out divisive normalization that effectively emulate masking, and such divisive processes have long been fundamental to image quality metrics (going back to Daly's VDP). Another answer is that, however they are constructed, successful quality metrics put a premium on high frequency spatial structure. This can be accomplished through incorporation of a contrast sensitivity function based on human performance (as in the VSNR and countless other algorithms), but it also suffices simply to add white noise to the input (as in the VIF). This may sound like a strange way to emulate the complexities of the human visual system, until we consider that this approach – conceiving the CSF in terms of equivalent internal noise – has a long history in psychophysics[49].

3. CONCLUSION

So, is it important whether or not image quality can be related to contrast perception? Our opinion is that the question may have been effectively sidestepped when the decision was made to define 'image quality' as a measure of distance between two versions of the same image. For this restricted definition, the higher-order statistics of images, being of ultimate importance, seem to swamp the contributions of lower-order statistics. However, a 'complete' standard observer, which can make both qualitative and quantitative discriminations, between and within images, will have to incorporate machinery for making judgments at many levels, from the sharpness of an edge, to the correctness of color assignment, to the gist of a scene. Only with all these parts in place will the ultimate goal of a no-reference image quality metric may be within reach. When we have reached this stage, contrast perception will certainly play a role in image quality assessment, but it will be just one piece in a large and very complex puzzle.

REFERENCES

[1] Schade, O.H., “Optical and Photoelectric Analog of the Eye,” JOSA 46, 721–739 (1956).

[2] Roufs, J.A.J., “Brightness Contrast and Sharpness, Interactive Factors in Perceptual Image Quality,” in Proc. SPIE 1077, 66–72 (1989).

[3] Daly, S.J., “Application of a noise-adaptive contrast-sensitivity function to image data compression,” in Proc. SPIE 1077, 217–227 (1989).

[4] Barten, P.G.J., “Square root integral: a new metric to describe the effect of various display parameters on perceived image quality,” in Proc. SPIE 1077, 73–82 (1989).

[5] Zetzsche, C., and Hauske, G., “Multiple channel model for the prediction of subjective image quality,” in Proc. SPIE 1077, 209–216 (1989).

[6] Safranek, R.J., Johnston, J.D., and Rosenholtz, R.E., “Perceptually tuned sub-band image coder,” in Proc. SPIE 1249, 284–293 (1990).

[7] Field, D.J., “What the statistics of natural images tell us about visual coding,” in Proc. SPIE 1077, 269–276 (1989).

[8] Klein, S.A., and Carney, T., “‘Perfect’ displays and ‘perfect’ image compression in space and time,” in Proc. SPIE 1453, 190–205 (1991).

[9] De Ridder, H., “Subjective evaluation of scale-space image coding,” in Proc. SPIE 1453, 31–42 (1991).

[10] Stein, C.S., Watson, A.B., and Hitchner, L.E., “Psychophysical rating of image compression techniques,” in Proc. SPIE 1077, 198–208 (1989).

[11] Klein, S.A., Silverstein, D.A., and Carney, T., “Relevance of human vision to JPEG-DCT compression,” in Proc. SPIE 1666, 200–215 (1992).

[12] Barten, P.G.J., “Physical model for the contrast sensitivity of the human eye,” in Proc. SPIE 1666, 57–72 (1992).

[13] Barten, P.G.J., “Spatiotemporal model for the contrast sensitivity of the human eye and its temporal aspects,” in Proc. SPIE 1913, 2–14 (1993).

[14] Watson, A.B., Solomon, J.A., Ahumada, A.J., and Gale, A., “Discrete cosine transform (DCT) basis function visibility: effects of viewing distance and contrast masking,” in Proc. SPIE 2179, 99–108 (1994).

[15] Peterson, H.A., Albert J. Ahumada, J., and Watson, A.B., “Improved detection model for DCT coefficient quantization,” in Proc. SPIE 1913, 191–201 (1993).

[16] Pelah, A., and Campbell, F.W., “Visual detection of sampling and quantization artifacts on sinusoidal luminance gradients,” in Proc. SPIE 1913, 242–252 (1993).

[17] Chandler, D.M., and Hemami, S.S., “Additivity models for suprathreshold distortion in quantized wavelet-coded images,” in Proc. SPIE 4662, 105–118 (2002).

[18] Daly, S.J., “Visible differences predictor: an algorithm for the assessment of image fidelity,” in Proc. SPIE 1666, 2–15 (1992).

[19] De Ridder, H., “Minkowski-metrics as a combination rule for digital-image-coding impairments,” in Proc. SPIE 1666, 16–26 (1992).

[20] Teo, P.C., and Heeger, D.J., “Perceptual image distortion,” in Proc. SPIE 2179, 127–141 (1994).

[21] Roufs, J.A.J., Koselka, V.J., and Tongeren, A.A. van, “Global brightness contrast and the effect on perceptual image quality,” in Proc. SPIE 2179, 80–89 (1994).

[22] Haun, A.M., Woods, R.L., and Peli, E., “Perceived contrast of electronically magnified video,” in Proc. SPIE 7865, 21 (2011).

[23] Kurihara, T., Aoki, N., and Kobayashi, H., “Analysis of sharpness increase by image noise,” in Proc. SPIE, 724014 (2009).

[24] Arend, L., “Apparent contrast and surface color in complex scenes,” in Proc. SPIE 1453, 412–421 (1991).

[25] Peli, E., “Contrast of slightly complex patterns: computing the perceived contrast of Gabor patches,” in Proc. SPIE 2657, 166–174 (1996).

[26] Pelah, A., "Inverting the perceptual transform," in Proc. SPIE 2179, 2–9 (1994).
[27] Peli, E., "Contrast sensitivity function (CSF) and image discrimination," in Proc. SPIE 3644, 71–77 (1999).
[28] Carney, T., Klein, S.A., Tyler, C.W., Silverstein, A.D., Beutter, B., Levi, D., Watson, A.B., Reeves, A.J., Norcia, A.M., et al., "Development of an image/threshold database for designing and testing human vision models," in Proc. SPIE 3644, 542–551 (1999).
[29] Carney, T., Tyler, C.W., Watson, A.B., Makous, W., Beutter, B., Chen, C.-C., Norcia, A.M., and Klein, S.A., "Modelfest: year one results and plans for future years," in Proc. SPIE 3959, 140–151 (2000).
[30] Watson, A.B., and Ahumada, A.J., "A standard model for foveal detection of spatial contrast," Journal of Vision 5, 717–740 (2005).
[31] Mantiuk, R., Kim, K.J., Rempel, A.G., and Heidrich, W., "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," ACM Transactions on Graphics 30, 40:1–13 (2011).
[32] Baker, D.H., and Graf, E.W., "Natural images dominate in binocular rivalry," Proceedings of the National Academy of Sciences 106(13), 5436–5441 (2009).
[33] Watson, A.B., "The Spatial Standard Observer: A Human Vision Model for Display Inspection," in SID Symposium Digest of Technical Papers 37, 1312–1315 (2012).
[34] Heynderickx, I.E., and Bech, S., "Image quality assessment by expert and non-expert viewers," in Proc. SPIE 4662, 129–137 (2002).
[35] Westen, S.J.P., Lagendijk, R.L., and Biemond, J., "Spatiotemporal model of human vision for digital video compression," in Proc. SPIE 3016, 260–268 (1997).
[36] Watson, A.B., "Toward a perceptual video quality metric," in Proc. SPIE 3299, 139–147 (1998).
[37] Geisler, W.S., and Perry, J.S., "Real-time foveated multiresolution system for low-bandwidth video communication," in Proc. SPIE 3299, 294–305 (1998).
[38] Duchowski, A.T., and McCormick, B.H., "Gaze-contingent video resolution degradation," in Proc. SPIE 3299, 318–329 (1998).
[39] Daly, S.J., Matthews, K.E., and Ribas-Corbera, J., "Visual eccentricity models in face-based video compression," in Proc. SPIE 3644, 152–166 (1999).
[40] Watson, A.B., and Kreslake, L., "Measurement of visual impairment scales for digital video," in Proc. SPIE 4299 (2001).
[41] Moore, M.S., Foley, J.M., and Mitra, S.K., "Detectability and annoyance value of MPEG-2 artifacts inserted in uncompressed video sequences," in Proc. SPIE 3959, 99–110 (2000).
[42] Moore, M.S., Foley, J.M., and Mitra, S.K., "Comparison of the detectability and annoyance value of embedded MPEG-2 artifacts of different type, size, and duration," in Proc. SPIE 4299, 90–101 (2001).
[43] Wang, J., Chandler, D.M., and Le Callet, P., "Quantifying the relationship between visual salience and visual importance," in Proc. SPIE 7527, 75270K (2010).
[44] Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P., "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing 13, 600–612 (2004).
[45] Sheikh, H.R., and Bovik, A.C., "Image information and visual quality," IEEE Transactions on Image Processing 15, 430–444 (2006).
[46] Laparra, V., Muñoz-Marí, J., and Malo, J., "Divisive normalization image quality metric revisited," JOSA A 27(4), 852–864 (2010).
[47] Chandler, D.M., and Hemami, S.S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Transactions on Image Processing 16, 2284–2298 (2007).
[48] Seshadrinathan, K., and Bovik, A.C., "Unifying analysis of full reference image quality assessment," in ICIP 2008, 1200–1203 (2008).
[49] Ahumada, A.J., and Watson, A.B., "Equivalent-noise model for contrast detection and discrimination," JOSA A 2(7), 1133–1139 (1985).