Complexities of Complex Contrast

Andrew M. Haun & Eli Peli
Schepens Eye Research Institute, Massachusetts Eye and Ear Infirmary
Department of Ophthalmology, Harvard Medical School,
20 Staniford Street, Boston, MA, 02114

ABSTRACT

For the visual system, luminance contrast is a fundamental property of images, and is one of the main inputs of any simulation of visual processing. Many models intended to evaluate visual properties such as image discriminability compute perceived contrast by using contrast sensitivity functions derived from studies of human spatial vision. Such use is of questionable validity even for such applications (i.e. full-reference image quality metrics), but it is usually inappropriate for no-reference image quality measures. In this paper, we outline why the contrast sensitivity functions commonly used are not appropriate in such applications, and why weighting suprathreshold contrasts by any sensitivity function can be misleading. We propose that rather than weighting image contrasts (or contrast differences) by some assumed sensitivity function, it would be more useful for most purposes requiring estimates of perceived contrast or quality to develop an estimate of efficiency: how much of an image is making it past the relevant thresholds.

Keywords: Luminance contrast, contrast sensitivity function, perceived contrast, image quality, no-reference image quality, MTFA

1. CONTRAST

1.1 Primacy

For the visual system, luminance contrast is the fundamental carrier of information about images. Motion is perceived through temporal changes in luminance contrast; the most initial sensations of depth are formed from binocular combinations of monocularly sensed contrasts; chromatic variation is almost always correlated with luminance changes. All of these qualities (motion, depth, color) are important parts of normal visual experience, and thus of any full-quality representation of it; but a monochromatic, cyclopean still-image is a perfectly acceptable visual representation of a scene. Here we argue that in estimating the visual *quality* of an image, contrast thresholds are of principal importance; perceived (suprathreshold) contrast magnitudes although noticeable in side-by-side comparison are relatively less important; and that the specific sensitivity functions commonly used in standard practice to estimate perceived contrast and quality may be misapplied or inappropriate.

1.2 Measurement

Given the primary importance to vision of luminance contrast, it is of great practical and theoretical importance to have operational measures of it [1]. The simplest summary measures will usually fail in characterizing the apparent contrast of an image, and thus are not used except for the simplest of patterns. Michelson contrast, the absolute range of luminances in a pattern, is thus not widely used except for periodic grating patterns – it errs by ignoring too much of an image's spatial variation. A much more common measure of a complex image's contrast is the standard deviation of luminances in an image (RMS contrast), a measure of the average deviation in luminance from the image mean over a specified spatial area. This measure is less susceptible to extreme values in an image, and thus tracks better with perceptual appearance of image contrast. Still, RMS contrast is a relatively poor predictor of perceived contrast – it errs by equally weighting all of the image's spatial variation and has been shown repeatedly to fail in predicting perceived image quality.

1.3 Perception

The visual system responds to images through a system of overlapping neural networks, repeated across the visual field, which are sensitive to different spatial scales, and perceived contrast is related to the response of these networks. The

Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications,
edited by Reiner Eschbach, Gabriel G. Marcu, Alessandro Rizzi, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 8292, 82920E · © 2012 SPIE-IS&T · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.915365

network responses are not linear, however. The absolute sensitivity of the networks varies along multiple dimensions, most importantly spatial frequency (scale), with size and temporal envelope being other important factors [2]. In general, finer and finer details require higher and higher contrasts in order for an observer to detect them, until the acuity limit is reached and details are too fine to be sensed at any contrast. Extremely coarse details also can require higher contrasts for detection, although the exact nature of the behavior at that end of the range is less clear. This pattern of varying absolute sensitivity to contrast is called the contrast sensitivity function (CSF).

Once contrasts are detected, their perceived strength follows a compressive function of stimulus contrast [3,4]: as stimulus contrast is increased, the corresponding increase in perceived contrast lags. The compressive relationship between stimulus and perceived contrast has the result that at high physical contrasts, perceived contrasts of patterns at different spatial frequencies converge to a similar level, despite starting from very different thresholds, an effect known as 'contrast constancy' [5].

1.4 Quality

Because of this differential sensitivity to different spatial frequencies, most operational measures intended to replicate human perception of contrast involve a combination of the CSF with the input image contrast spatial frequency spectrum. For a given display method if the general form of a typical image's amplitude spectrum is known, and if the device's MTF is known, then an overall quantity of image contrast transmitted by the device to the observer can be computed. If the observer's CSF is also known, a combined quantity (e.g. the MTFA, the area between the image spectrum modulated by the display MTF from above, and the threshold function from below [6]), or something closely related (many VQ measures take a similar but more complex approach [7,8,9,10], including thresholds and other psychophysical nonlinearities based on human vision), is often taken to be indicative of the perceived quality of a displayed image, or of its discriminability versus an ideal reference. In the following sections, we detail why this may be a misguided notion at least in part, and particularly for non-reference quality measures.

2. THRESHOLD

2.1 Thresholds Matter

For vision the most important thing is that signals are sensed and available to the rest of the brain. Sensation of contrast in this context must be taken to include its phase (e.g. polarity) or position. Whether one signal is more or less represented than another is relatively less important. The strength of the representation is often due to interfering factors – fog, darkness, optical blur, and other factors all can contribute to weakly represented signals. The compression of suprathreshold contrast perception speaks to the diminishing returns for representation of stimulus magnitude. We therefore argue here that what matters most for visual perception is not the strength of a signal, but whether or not it is detected. This means that an integration of sensed signal magnitude cannot be a very meaningful measure of image quality. What matters to visual quality is not the amount of contrast which has exceeded the system's instantaneous established thresholds; rather, what matters is the *quantity of thresholds that have been exceeded*. So, if contrast is expected in some region of an image at some scale, but is not seen, perceived quality suffers. If this is correct, then the remaining question is: what thresholds are to be used in making such an analysis?

2.2 Thresholds are Elastic

Studies of contrast adaptation show that adaptation results in an adjustment of the detection threshold, so that it is dynamically set near the adapting contrast [11,12,13]. Experiments measuring contrast sensitivity against complex (broadband) image backgrounds, or immediately after adaptation to such backgrounds, show that the perceptual representation of a complex image is relatively flat and near-threshold, with significant suppression of perceived suprathreshold contrast observed especially towards lower spatial frequencies [14,15]. That is, although there may be a large difference between thresholds and typical image contrasts at the peak of the CSF, it is not true that there is a correspondingly large perceptual effect of this difference – i.e. high contrasts are never very far above the corresponding detection thresholds. The 'true' thresholds tend to be much closer to the image contrasts, so the actual perceptual effect must be correspondingly smaller.

2.3 What is the CSF good for?

The adaptation and broadband contrast sensitivity studies mentioned above demonstrate that the typical CSF – normally obtained during visual adaptation to a blank gray screen – used so often to define perceptibility of image contrast may be largely irrelevant to scene perception, except in the case of images whose contrast is so low as to be almost invisible. The CSF tells you if some detail *can* be seen under certain limited circumstances (usually a small stimulus surrounded by a uniform luminance field), but is not informative regarding *how* a similar detail will be perceived in a complex image, or even *whether* it is seen (detected) in a complex image. If contrast thresholds are adjusted through adaptation as described above, the instantaneous CSF is always hovering just below the current contrast distribution – and if contrast drops, the thresholds drop. The 'absolute' CSF simply defines the lower bounds of this process, but the distances of image contrasts from the these minimal threshold values are not especially important.

A use of the CSF that better reflects its role in perception would be as a limit to perceptual efficiency. For a given display device, if the MTF is smooth, a good measure of image quality might be simply the proportion of displayed contrasts above threshold, or perhaps the spatial frequency at which the band-limited efficiency define drops below a certain level.

Figure 1. Contrasts at and above 32 cycles per picture (see [1,17] for filter details) have been attenuated by 75%, but only at pixels where the band-limited contrast [1] was above a threshold value of 0.15. If the image is shown at the intended size (about 13cm across) and viewed at a normal reading distance (about 45cm), the contrast decrement begins at about 2 cycles per degree. Despite this severe distortion of contrast, the image retains most of its original quality. (Readers should note that these images are intended to be viewed on an electronic display with γ of about 2.0. Printing will linearize the images and cause both Figures 1 and 3 to appear washed-out).

2.4 What reduces quality?

According to what has been put forward so far, it could be suggested that it is better to sense more details weakly than to sense fewer details strongly. Once details are detected, the visual system adapts to current conditions, amplifies weak signals, and construct a normal representation of what is being sensed [16]; but once details fall below the threshold, there is nothing to amplify, and the result is a degraded, poor representation. Refer to the images in Figures 1 and 3 for an illustration of this concept. Figure 1's image has had its contrast attenuated by 75% at spatial frequencies above 32 cycles per picture (cpp; about 2 cycles per degree of viewing angle, cpd, from a 45 cm viewing distance on a standard printed page), but only if the original contrasts exceeded a threshold value of 0.15 (which was arbitrarily chosen to produce a clear demonstration), and in such a way as to keep all manipulated contrasts above the specified threshold, as described in Equation 1:

$$C_{x,f} = \begin{cases} \delta\left(c_{x,f} - t\right) + t, & if\left|c_{x,f}\right| \geq |t| \\ c_{x,f}, & if\left|c_{x,f}\right| < |t| \end{cases} \tag{1}$$

Here c represents contrast at a particular spatial position x and frequency f, δ is the contrast decrement equal to 0.25, and t is the threshold (which obtains the sign of the corresponding contrast). Despite the considerable distortion of contrast in this image (dotted colored lines in Figure 2) it is not immediately obvious that there is anything wrong with it.

Figure 3's image has been decremented at the same frequencies, but only at 'low' contrasts: i.e., whereas in Figure 1 contrasts above 0.15 were compressed towards that contrast level, in Figure 3 only contrasts *below* 0.15 have been compressed towards the local mean luminance, as described in Equation 2:

$$C_{x,f} = \begin{cases} c_{x,f}, & if\left|c_{x,f}\right| \geq |t| \\ \delta c_{x,f}, & if\left|c_{x,f}\right| < |t| \end{cases} \tag{2}$$

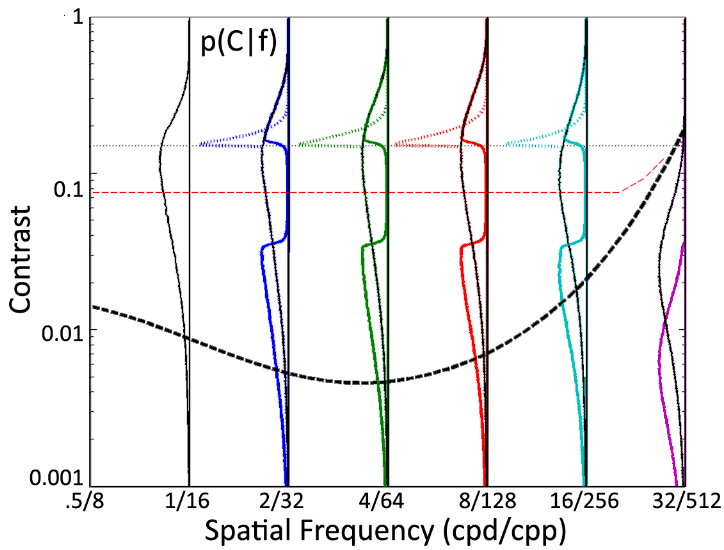

Figure 2. Contrast distributions in the bands of Figures 1 and 3. Original distributions are the smooth black lines. Dotted colored lines represent the contrast histograms for Figure 1. The 'threshold' was set at 0.15, so contrasts above this level were compressed towards the threshold level. Solid colored lines represent the histograms for Figure 3. Here contrasts above 0.15 remained unchanged, but lower contrasts were shifted downwards. The dashed black line is a 'standard observer' CSF [21]. The dashed red line represents a hypothetical adapted CSF, to illustrate why lowering the low contrasts in Figure 3 has such a drastic effect: the contrasts are lowered below some higher-than-standard CSF.

Figure 3 looks much more degraded than Figure 1, despite the fact that Figure 3 has lost much less contrast. In the affected bands, Figure 1's RMS contrast has been decreased by about 30%, while Figure 3's contrast has been decreased less than 10%. The important difference is that the lowered contrasts in Figure 1 are *still visible*, while those in Figure 3 have been pushed below the threshold of visibility (or quantized out of existence; this is unlikely since most contrasts in Figure 3, as shown in Figure 2, are still higher than the expected grayscale resolution of the images as displayed. We have confirmed that most – though not all – of the detail in the apparently smooth (blurred) areas of Figure 3 can be recovered even following 8-bit quantization of the image). It is important to note that if the CSF were where it is marked out in Figure 2 (dashed black curve), almost everything in Figure 3 would still be above threshold. However, as indicated in 2.2 and 2.3, the true thresholds should actually be much higher, perhaps even hugging the lower bound of the expected contrast distributions (dashed red line in Figure 2).

Figure 3. Similar to Figure 1, except that only contrasts *below* the set threshold have been decremented. The contrast loss in this image is much less than in Fig.1, but the degradation is far more severe.

It is interesting to note the situation at the highest frequency band. Near 32cpd, close to the human acuity limit, most of the contrasts in the original image fall below the 'absolute' threshold marked out by the standard CSF. Here, even an observer with better-than-normal visual acuity cannot possibly detect everything. The visual system therefore does not expect a 'complete' contrast distribution at the highest frequencies where it maintains sensitivity. In fact, removing all contrasts at this frequency does not contribute to any obvious degradation of the image, yet increasing contrast will cause the image to look sharpened or enhanced. This effect is interesting in itself, because a high resolution digital image usually will not look blurred, even though most of its highest frequency contrasts *must* be subthreshold. The sharpening effect of increasing very-high-frequency contrasts can therefore be unexpected and surprising.

2.5 Not even perceived contrast magnitude is a sum over the CSF

In another recent study we have shown that a blank-adapted CSF cannot predict human judgments of image contrast [18] independent of perceived image quality. Subjects decided which of two copies of an image, whose amplitude spectra had been randomly jittered, had higher contrast. Over several thousand such trials, using different images, subjects' choices were correlated with the jittered contrast values at different spatial frequencies, revealing a band-pass weighting function. Simulations demonstrated that a perceived contrast model which weights image contrast using a standard CSF (and nonlinear transducer functions; similar to the steps involved in many current quality metrics) yielded low-pass weighting functions unlike human performance. For the simulated observer to succeed in judging image contrast, gain control weights, stronger towards low spatial frequencies, must be included as predicted by the sensitivity experiments [14,15]. The result of incorporating such gain control is similar to what is described in the preceding sections: local (in space and scale) responses to image contrast tend to be small and just above the adjusted threshold.

3. CONCLUSIONS

The CSF is misused – so what?

What is missing at this time is psychophysical data on 'threshold capacity', i.e. how much structure is needed at a given scale to fulfill the visual system's expectations and how these expectations interact across scales. Since natural images with varying radially averaged spectrum slopes are all perceived to be in focus and clear [16], it is likely that the system's expectation at a given scale are affected by the contrasts seen at other scales (the slope of the amplitude spectrum) and their interactions.

We must, at last, address several more peripheral points. First, a caveat: all of these arguments are regarding how to interpret transmission of contrast at different scales, and the impact of such transmission on measures of image quality, but we have made our argument in the context of *amplitude*, not *phase* information. Attenuation of contrast as described above is decrease in amplitude, while optical blur and compression artifacts, two important factors affecting image quality, involve changes in the local spatial structure of an image beyond changes in amplitude. It is not clear whether our argument can be expanded to include, for example, a probabilistic approach to quantifying phase distortion in an image. Second, we have restricted our discussion entirely to static images, but this is not a problem: the contrast adaptation which affects thresholds in the manner we have described, putting much application of CSFs into question, operates rapidly [19]; even as a scene changes from second to second, the visual system is adapting, although there must be a temporal averaging of adaptive states over time; so, when viewing a dynamic scene such as a video, the situation (i.e. where the contrasts are, and where the thresholds are) is certainly not as clear-cut as illustrated in Figure 2, though the same principles should apply.

All of this can be taken together to make the claim that a summation of (suprathreshold) contrast magnitudes over the entire range of the CSF is not necessarily indicative of perceived quality or even contrast of an image. It must be important that contrasts within range of the CSF are sensed, but not how far above the marked thresholds they are; since most contrasts in an image will be above threshold, they will contribute redundantly to a summary measure using the baseline CSF. In fact, reducing contrast at mid-range spatial frequencies can yield computational benefits, for example freeing up dynamic range for application of high-frequency contrast enhancement [20] or simply by decreasing the required bit depth needed for smooth representation of spatial structure.

We close by proposing that a probabilistic measure – i.e. given the distribution of contrasts that have been sensed, what is the likelihood that *an expected proportion* has been sensed - would be an especially useful and meaningful measure of image quality. Such a measure would, in effect, specialize in sensing 'contrast gaps' or 'cliffs' (usually at higher retinal spatial frequencies) that indicate to the visual system that the image is degraded. This would be a more theoretically appealing, accurate, and even intuitive way of estimating the perceptual efficiency of a displayed image and its unreferenced quality.

4. ACKNOWLEDGEMENTS

Supported in part by NIH grant # EY05957 and a gift from Analog Devices Inc.

REFERENCES

1. Peli, E., “Contrast in complex images,” JOSA-A 7(10), 2032-2040 (1991).
2. Peli, E., “Contrast sensitivity to patch stimuli: effects of spatial bandwidth and temporal presentation,” Spatial Vision 7(1), 1-14 (1993).
3. Cannon, M.W., Fullenkamp, S.C., “A transducer model for contrast perception,” Vision Research 31(6) 983-998 (1991).
4. Legge, G.E., “A Power Law for Contrast Discrimination,” Vision Research 21(4) 457-476 (1981).
5. Georgeson, M.A., & Sullivan, G.D., “Contrast constancy: deblurring in human vision by spatial frequency channels,” Journal of Physiology 252, 627-656 (1975).
6. Barten, P.G.J., “Evaluation of subjective image quality with the square-root integral method,” JOSA-A 7(10), 2024 (1990).
7. Daly, S., “The Visible Difference Predictor: An Algorithm for the Assessment of Image Fidelity,” in A.B. Watson (Ed.), *Digital Images and Human Vision*, 179-206. The MIT Press, Cambridge MA, 1993.
8. Lubin, J., “A visual discrimination model for image system design and evaluation,” in E. Peli (Ed.), Visual Models for Target Detection and Recognition, 207-220. World Scientific Publishers, Singapore, 1995
9. Watson, A.B., “DCTune: A technique for visual optimization of DCT quantization matrices for individual images,” in Society for Information Display Digest of Technical Papers 24, 946-949 (1993).
10. Wang, Z., Bovik A.C., “Modern Image Quality Assessment,” Morgan and Claypool Publishers, 2006.
11. Greenlee, M.W., Heitger, F., “The functional role of contrast adaptation,” Vision Research 28(7), 791-797 (1988).
12. Foley, J.M., Chen, C.C., “Analysis of the effect of pattern adaptation on pattern pedestal effects: A two-process model,” Vision Research 37(19), 2781-2788 (1997).
13. Abbonizio, G., Langley, K., Clifford, C.W.G., “Contrast adaptation may enhance contrast discrimination,” Spatial Vision, 16(1), 45-58 (2002).
14. Bex, P.J., Solomon, S.G., Dakin, S.C., “Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure,” Journal of Vision 9(10):1 (2009).
15. Haun, A.M., Essock, E.A., “Contrast sensitivity for oriented patterns in 1/f noise: contrast response and the horizontal effect,” Journal of Vision 10(10):1 (2010).
16. Webster, M.A., Georgeson, M.A., Webster, S.M., “Neural adjustments to image blur,” Nature Neuroscience, 5(9), 839-840, (2002).
17. Peli E., “Contrast sensitivity function and image discrimination,” JOSA-A 18, 283-293 (2001).
18. Haun, A.M., Peli, E., “Measuring the perceived contrast of natural images,” Society for Information Display Symposium Digest of Technical Papers 42, 302-304 (2011).
19. Wilson, H.R., Humanski, R. “Spatial frequency adaptation and contrast gain control,” Vision Research, 33(8), 1133-1149 (1993).
20. Peli, E., “Limitations of image enhancement for the visually impaired,” Optometry and Vision Science 69(1), 15-24 (1992).
21. Watson, A.B., Ahumada, A.J., “A standard model for foveal detection of spatial contrast,” Journal of Vision 5(9), 717-740 (2005).