Vision-based approaches to digital halftoning

Jeffrey B. Mulligan
NASA Ames Research Center
Moffett Field, CA

Abstract

One way to describe the problem of digital halftoning is as a search for the quantized image that minimizes the visibility of artifacts. To apply this approach in practice, it is first to specify a computational model for computing visible error that can be used to rank candidate images automatically. The model may be incorporated directly into a search algorithm, or used after-the-fact to rank images produced by algorithms that are more heuristic. These approaches have been quite successful when applied to achromatic images, even when using a relatively simple visual model accounting for high-frequency contrast sensitivity but with no masking. This approach can be directly generalized to additional dimensions such as time and color. Visual models based on threshold measures may not be optimal for low bit rate conditions where quantization noise is visible. Instead, the degree to which the noise is effortlessly segmented through *perceptual scission* may influence the utility of the final image.

Introduction

There are many situations in which we wish to display a continuous tone image on a device with a more restricted range of output tones. The classic example is printing, where we have ink that is either present or not present. Ten or fifteen years ago, it was common for a computer graphics display to have 8 (or fewer) bits per pixel, which also prevented direct "truecolor" rendering of images. Even today, when full color 24 bit per pixel displays are the norm, there are still demanding applications such as medical imaging and visual testing where we may wish to render images which have too large or too small a dynamic range to be rendered properly on conventional hardware. In many cases, digital halftoning is the answer.

The underlying assumption of all halftoning processes is the display will be viewed at a distance such that the display device has a resolution higher than that of the visual system. Neighboring display elements will fall upon a single visual receptor, and their levels will be averaged. Correct implementation of visually optimized halftoning requires advance knowledge of the intended viewing conditions, and of course is only optimal when those conditions are met.

Visual Resolution

Perhaps the most familiar measure of visual resolution is Snellan acuity, measured with a letter chart and expressed by terms such as "20-20." The Snellan letters are presented at high contrast, and the clinician determines the smallest size that can be accurately read. A somewhat richer measure is provided by the Contrast Sensitivity Function (CSF). This function describes the sensitivity (inverse of threshold contrast) for a set of spatial frequencies. Human contrast sensitivity peaks at a spatial frequency of around 1 cycle per degree (cpd), at a value of around 100 (corresponding to a contrast of 1%).[1]

An initial limitation on contrast sensitivity is imposed by the optics of the eye. The optical performance of the system can be described by the point spread function (PSF) or optical Modulation Transfer Function (MTF). Normal (emmetropic) eyes roll off at around 60 cpd, which is well matched to the sampling rate of the array of photoreceptors[2] in the retina's center or fovea. Contrast sensitivity also shows a low-frequency fall-off that is not predicted by optical factors, but must arise from subsequent neural processing. This insensitivity to low frequencies allows us to be content with display monitors that are brighter in the center than at the edges, but is generally irrelevant for halftoning.

Another neural effect is manifested in the relative sensitivity to patterns of different orientations. While the optical MTF has approximate circular symmetry (for most eyes), contrast sensitivity is generally less for oblique orientations than for horizontal and vertical, known as the "oblique effect."[3-6]

Contrast sensitivity describes absolute threshold for a pattern in the absence of any other patterns. This is appropriate for describing the visibility of halftoning artifacts in large uniform areas, but is only approximately correct when there are other spatial patterns present. In general, the presence of other patterns reduces sensitivity somewhat, a phenomenon that is referred to as "masking." Masking has been found to be frequency selective, with a "critical band" of one or two octaves.[7] Masking data are often represented as threshold-versus-intensity (TVI) curves, in which the abscissa represents the strength (e.g., contrast) of the mask; the ordinate plots the strength (contrast) of the test stimulus which can just be detected in the presence of the given mask. Sometimes the test and mask are the same spatial pattern; in this case, the measurement is referred to as the "increment threshold." At high levels of the masking stimulus, the threshold rises with increasing masking level.[8,9]

Assessing Halftone Quality

It is not always easy to provide a quantitative definition of what constitutes the visual quality of a halftone image, but we know it when we see it. Exactly what is it that makes us like one halftone texture and not another? At this point, we should make an important distinction between what we will call "fine-scale" and "coarse-scale" halftones. Fine-scale halftoning refers to situations in which the display elements making up the halftone texture are very small (compared to visual resolution), as is often the case for printers. In this case, the halftone texture is likely to be invisible, or at least near threshold, and the threshold models referred to in the preceding section are appropriate. When the halftoning elements are large, however, the pattern is likely to be quite visible, and it is not obvious that threshold models are at all suitable for describing visual quality in this case. Ironically, most papers on this topic present example patterns in which the pattern is clearly visible (even if this requires magnification) so that the reader can see the detailed micropatterns produced by different algorithms. While it is true that this is necessary to see the fine-scale differences, it is perhaps sub-optimal for assessing visual quality.

There are some situations where coarse-scale halftoning is the only option, such as rendering images for display on low-resolution computer screens. In this situation, invisibility of the halftone pattern may be an impossible dream. In this case, minimizing the visibility of the texture may be less important than minimizing interference with the intended image content. While it is not obvious how to do this, one general principle might be to minimize the local spectral overlap between the source image and the error image. In the remainder of this paper we will restrict our attention to visibility-based approaches.

A common approach[10-13] to computing halftone quality is as follows: first, an error filter is defined, representing the low-pass characteristic of the visual system. It is convenient to define this filter in the space domain, where it can have a restricted region of support (3x3, 5x5, etc.). Next, an error image is computed by subtracting the desired values from the halftone values. This error image is then blurred by application of the filter. Finally, the sum of squares of the blurred values is computed to produce an overall measure of error. While the exponent of 2 is almost universally used for error summation, there is no firm theoretical basis for this. Using a higher exponent (such as 4) should give higher weighting to the largest errors, which might be desirable in a situation where the errors are mostly invisible, except for a few artifacts. Curiously, however, when we redo the analysis done in our 1992 paper[10] using the exponent of 4 instead of 2, we obtain the same condition for flipping a pixel, suggesting that these two metrics produce the same ranking of halftone patterns.

Attaining Quality

Halftoning methods can be divided into two classes: those that directly incorporate a visual model, and use it in an error minimization loop, and those that employ an effective heuristic, which although not directly based on a visual model nevertheless produces good results. Most of the early work, as exemplified by ordered dither[14] and error diffusion[15] falls into the latter category. These tend to be single-pass algorithms that are relatively easy to compute. More recent methods that directly incorporate the visual model[10-13,16] tend to be iterative procedures, and are much more computationally intensive. The general approach is to start with an initial image (which might be the output of another algorithm, random noise, or whatever you like) and sequentially visit individual pixels and try to improve the total error at that location. In the method we call "strict descent,"[10] only changes to the single pixel in question are considered. Alternatively, changes to the pixel and its neighbors can all be considered. In the method called "Direct Binary Search"[11] up to 8 changes are considered, corresponding to exchanging a pixel's value with each of its 8 nearest neighbors. Because this method only considers exchanges, the total number of on and off bits is not changed, so choice of the initial image is important. We have obtained good results using a 9-way search in which we consider both flipping the pixel in question, and exchanging with each of 8 neighbors.

Extensions to Color

The ideas from the preceding sections can be easily generalized to the case of color. Lights of different colors can be characterized by their "luminance," which is a spectrally weighted energy measure that captures how effective a given light is at evoking a sensation of flicker or motion. When the different colors that make up a pattern are matched in luminance, the resulting pattern is said to be "equiluminant" or "isoluminant." Vision for equiluminant stimuli is characterized by two main differences: first, there is diminished spatial and temporal resolution, and second the chromatic CSF does not show the low-frequency decline seen for achromatic patterns.[1,17]

The reduced spatial acuity of the chromatic system has suggested ways to improve halftone quality by moving error from the luminance component to the chromatic component, where it will presumably be less visible[12,13,18]. For an image represented by red, green, and blue components, we might begin by independently halftoning each color component down to 1 bit, then combining to form an image with 8 different basic pixel colors. (At least, this is reasonable for images to be viewed on an emissive display such as a computer monitor, where the color components combine additively. It may not be appropriate for the printing situation, depending on the characteristics of the inks.) In this case, we would expect the locations of "on" pixels in each of the color planes to be independent. For example, in a region with 50% red pixels on and 50% green pixels on, we would expect to find approximately

25% of the pixels colored with each category: red, green, yellow, and black. Clearly, we can attain the same space-average yellow with a pattern made up of black and yellow elements (red and green planes perfectly correlated) or of red and green elements (red and green planes perfectly uncorrelated). In the latter case, however, the local luminance variation will be less, so if that is indeed the most visible artifact then this pattern should be preferred. For arbitrary colors, however, it is generally impossible to anticorrelate the color bitplanes exactly, and the gains from these techniques are modest.

Extensions to the Time Domain

Visual resolution can be characterized in time as well as space, and while this is not relevant for printed halftones, some improvements can be realized for dynamic media such as video and computer displays. Contrast sensitivity in the time domain is qualitatively similar to what is observed for the space domain. The achromatic function is bandpass, peaking at a frequency of around 10 Hz,[19,20] while chromatic sensitivity is low-pass, with relatively poor sensitivity to high temporal frequencies.[1] "Color fusion" occurs for lights which are matched in luminance and exchanged at 20 or 30 Hz; if the luminances are mismatched, flicker is seen but no color variation. Eliminating the visible flicker in this situation is known as "heterochromatic flicker photometry" (HFP), which provides the operational definition of luminance.

Many of the previously discussed methods may be directly generalized to three dimensions. Filtered error minimization[10-13] can easily be reformulated in terms of error filtered in the time domain as well as two spatial dimensions, although the computational cost rises significantly. In addition, there are heuristic methods that perform reasonably well. Ordered dither[14] has been generalized to three dimensions.[21] Similarly, one can imagine generalizing error diffusion[15] to three dimensions, with the addition of one or more frames of storage to hold the errors to be diffused in time. In error diffusion, the weights normally sum to one, insuring that gray levels will be preserved. Carrying this principle over to a three-dimensional generalization would have the result that the within-frame weights have a sum less than one, however, resulting in distortion of the first frame. While this would be corrected eventually, a more flexible approach is that of purely temporal error diffusion.[21-23] In this approach, the first frame is processed with a chosen two-dimensional algorithm. The erorr image is computed by subtracting the desired values from the halftone values, and this error image is subtracted from the desired values for the next frame. Weighting the error by a constant less than one will insure that their effect is limited in time; alternatively, the errors from a small number of immediately preceding frames can be combined using a finite-impulse response filter. This scheme is quite general; in addition to halftoning, it can be applied to other forms of gray level quantization, such as MPEG or motion-JPEG compression.

Scission

Animated sequences produced with the temporal algorithms described in the previous section are often markedly superior to a static rendering. There are two reasons for this: first, assuming some temporal integration in the visual system, the desired signal is represented more faithfully, as intended. The second factor is less obvious: in an animated sequence, the halftone "noise" is dynamic (changes rapidly in time), while the underlying image to be presented is static. This allows a perceptual segregation or "scission" to occur, in which the target image is seen "through" the noise, much as one can see the road clearly through a dirty windshield. Perceptual scission occurs at the whims of individual subjects' idiosyncracies, but can be strongly influenced by factors such as stereo disparities.[24]

Previously, we have discussed the halftoning problem in terms of minimizing the visibility of the artifacts. This is appropriate for high-quality halftones in which we expect that the errors can be made invisible, but our observation of perceptual scission suggests a different approach for situations where halftone errors will be visible. To facilitate scission, instead of trying to minimize the visibility of the halftone noise, we might instead try to minimize its spectral overlap with the target image. Note that this is the opposite of approaches that rely on masking of artifacts by the target image itself.[25] In that case we allow more error in spectral bands where the target image has a high power level. Again, this is sensible when we expect that the artifacts will be invisible, but may impede perception of the target image in the presence of visible artifacts.

By analogy with the temporal case, we might imagine that to facilitate scission in a static image we should attempt to decorrelate the local noise spectrum with that of the target image. To incorporate this idea into methods based on filtered error minimization, we would want to use a space-variant filter whose properties depend on the local image content in place of a fixed filter based on detection data. Because the phenomenon of perceptual scission is less well understood than simple thresholds, there is not a good set of reference data from which to design the filter properties. Some trial-and-error experimentation will therefore be needed to explore this approach.

Summary

Computational visual models can be useful for halftoning algorithms, especially when plenty of preprocessing time is available, and high quality is the overriding priority. Previous implementations have, for the most part, been based on a single channel model. Multiple channel models may be better predictors of quality, particularly for the case of temporal halftoning, where a single channel model will be blind to coherent motion. Approaches using traditional models are best suited to high quality halftones where the artifacts are near threshold. For lower quality halftones, a criterion based on visual scission may be more appropriate.

References

1. D. H. Kelly, Spatiotemporal Variation of Chromatic and Achromatic Contrast Thresholds, J. Opt. Soc. Am., vol. 73, pp. 742-750, (1983).
2. David R. Williams, Topography of the Foveal Cone Mosaic in the Living Human Eye, Vision Res., vol. 28, pp. 433-454, (1988).
3. F. W. Campbell, and J. J. Kulikowski, The Effect of Orientation on the Visual Resolution of Gratings, J. Physiol., vol. 187, pp. 427-436, (1966).
4. Stuart Appelle, Perception and Discrimination as a Function of Stimulus Orientation: The "Oblique Effect" in man and animals, Psych. Bull., vol. 78, pp. 266-278, (1972).
5. M. A. Berkley, and F. Kitterle, Grating Visibility as a Function of Orientation and Retinal Eccentricity, Vision Res., vol. 15, pp. 239-244, (1975).
6. D. H. Kelly, No Oblique Effect in Chromatic Pathways, J. Opt. Soc. Am., vol. 65, pp. 1512-1514, (1975).
7. C. F. Stromeyer, and S. Klein, Spatial Frequency Channels in Human Vision as Asymmetric (Edge) Mechanisms, Vision Res., vol. 14, pp. 1409-1420, (1974).
8. Gordon E. Legge, and John M. Foley, Contrast Masking in Human Vision, J. Opt. Soc. Am., vol. 70, pp. 1458-1471, (1980).
9. Gordon E. Legge, A power law for contrast discrimination, Vision Res., vol. 21, pp. 457-467, (1981).
10. Jeffrey B. Mulligan and Albert J. Ahumada, Jr., Principled Halftoning Based on Human Vision Models, *Human Vision, Visual Processing and Digital Display III*, B. E. Rogowitz (ed.), *Proc*. SPIE, vol. 1666, pp. 109-121 (1992).
11. Mostafa Analoui and Jan P. Allebach, Model Based Halftoning Using Direct Binary Search, *Human Vision, Visual Processing and Digital Display III*, B. E. Rogowitz (ed.), *Proc*. SPIE, vol. 1666, pp. 96-108 (1992).
12. J. B. Mulligan, and A. J. Ahumada, Jr., Principled Methods for Color Dithering Based on Models of the Human Visual System, *SID Int. Symp., Dig. Tech. Papers*, pp. 194-197, (1992).
13. T. J. Flohr, B. W. Kolpatzik, R. Balasubramanian, D. A. Carrara, C. A. Bouman, and J. P. Allebach, Model Based Color Image Quantization, Human Vision, Visual Processing, and Digital Display IV, J. P. Allebach and B. E. Rogowitz (eds.), Proc. SPIE vol 1913, pp 270-281 (1993).
14. B. E. Bayer, An Optimum Method for the Two-Level Rendition of Continuous-Tone Pictures, IEEE Intl. Conf. Comm., vol. 1, pp. 26-11 – 26-15, (1973).
15. Robert W. Floyd, and Louis Steinberg, An Adaptive Algorithm for Spatial Greyscale, Proc. SID, vol. 17, pp. 75-77, (1976).
16. J. Sullivan, R. Miller, and G. Pios, Image Halftoning Using a Visual Model in Error Diffusion, *J. Opt. Soc. Am. A*, vol. 10, pp. 1714-1724 (1993).
17. K. T. Mullen, The Contrast Sensitivity of Human Color Vision to Red-Green and Blue-Yellow Chromatic Gratings, J. Physiol., vol. 359, pp. 381-400, (1985).
18. Jeffrey B. Mulligan, Digital Halftoning Methods for Selectively Partitioning Error into Achromatic and Chromatic Channels, Human Vision and Electronic Imaging: Models, Methods, and Applications, B. E. Rogowitz, and J. P. Allebach (eds.), Proc. SPIE vol 1249, pp. 261-270, (1990).
19. G. J. C. Van der Horst, and M. A. Bouman, Spatiotemporal Chromaticity Discrimination, J. Opt. Soc. Am., vol. 59, pp. 1482-1488, (1969).
20. Andrew B. Watson, Temporal Sensitivity, in Handbook of Perception and Human Performance, vol. 1: Snesory Processes and Perception, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), pp. 6-1 – 6-43, (1986).
21. J. B. Mulligan, Methods for Spatiotemporal Dithering, *SID Int. Symp., Dig. Tech. Papers*, pp. 155-158, (1993).
22. Jeffrey B. Mulligan, Application of Temporal Error Diffusion to Motion-JPEG, *Human Vision and Electronic Imaging II*, B. E. Rogowitz and T. N. Pappas (eds.), Proc SPIE vol. 3016, pp.288-295, (1997).
23. Jeffrey B. Mulligan, Application of M-JPEG Compresson Hardware to Dynamic Stimulus Production, *Spatial Vision*, vol. 11, pp. 19-32, (1997).
24. Barton L. Anderson, Stereoscopic Surface Perception, Neuron, vol. 24, pp. 919-928, (1999).
25. DCT Quantization Matrices Visually Optimized for Individual Images, *Human Vision, Visual Processing, and Digital Display IV*, J. P. Allebach and B. E. Rogowitz (eds.), Proc. SPIE, vol. 1913, pp. 202-216, (1993).

Biography

Jeffrey Mulligan received his B.A. degree in Physics from the Harvard University in 1980 and a Ph.D. in Psychology from the University of California at San Diego in 1986. During 1982 and 1983, he worked as a summer intern at AT&T Bell Laboratories at Murray Hill, New Jersey. Since 1986 he has worked at the NASA Ames Research Center, at Moffett Field, California, where he is a member of the Computational Vision Laboratory. His work has primarily focused the study of motion perception, using psychophysics, eye movement recording, and computational modeling. He is a member of the Association for Research in Vision and Ophthalmology, and the Optical Society of America.