Human Visual Models and Binary Image Rendering

John Dalton
Apple Computer, Cupertino, California

Abstract

There is a long history of various techniques used to render continuous tone images on binary marking engines. Models of visual perception of images can aid in a unified analysis of these binarization techniques and their subsequent visual appearance and perceived image quality.

Introduction

Visual models used in the design and optimization of halftoning algorithms have typically been based on visually weighted distortion metrics. Psychophysical measurements of contrast sensitivity at the threshold of detection are used to derive the visual weightings in these models.[16] The halftone design procedure tries to optimize the visually weighted mean square error between the original continuous tone image and the binary halftoned image. Reducing this error leads to improved halftone renditions.

Another way to think about the problem of binary image rendering is that the binary rendition is an attempt by a user to convey some kind of visual information. The halftoning research community has focused on the admirable goal of achieving photographic realism in binary prints. However, artistic rendering is typically based on visual abstraction of content. This can be seen in technical illustration, which is generally not photo-realistic. Abstraction of the visual image is used to convey more information. Sketches or line drawings are able to convey significant meaning without the use of tone information. Caricature is another example where distortion of a visual image actually increases recognition.

The traditional approach to visually weighted halftone design is heavily based on the notion of threshold detection of visual difference. Visually weighted detection of noise is very different then modeling abstraction of visual information, which is the primary concern of a graphic artist or an observer perceiving the content of a visual image. Binary image rendering models that seek to model and generate information content or artistic style need different visual models then the traditional visually weighted noise models discussed above.

What is needed are visual models that can be used to successfully predict visual appearance of images. This includes notions of image similarity as well as information content. These models can then be used to derive binary image renderers that allow for artistic expression of style or content information maximization as well as optimization of visual appearance. They also have applicability to content-based image retrieval.

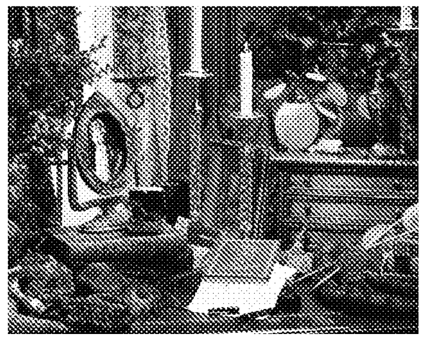


Figure 1a. Halftone rendering

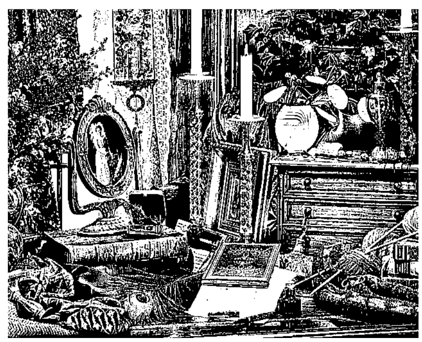


Figure 1b. Pseudo-lithographic rendering

Visually Weighted Halftone Optimization

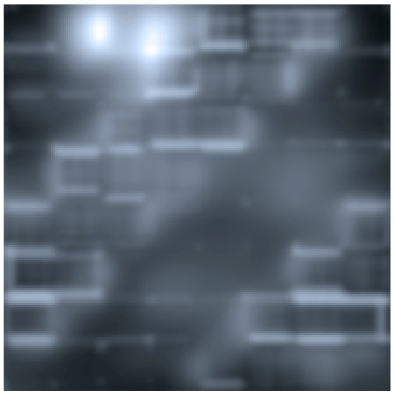
Channel models have a long history in the psychophysical literature, and visual distortion metrics[16] typically incorporate some type of oriented multi-scale spatial-frequency channels. Lubin10 has developed a very complete model of early vision that is designed to predict visually weighted error in images. The model has been used with some success at NASA Ames in the design of flat panel displays and to predict the effects of quantization in images. Dalton[4] investigated the use of models of this kind to aid in the design of stochastic screens.

Figure 2 shows the visually weighted error maps predicted by the visual model for five different halftoning techniques and a simulated viewing distance of 1 meter. The visually weighted maximum and rms errors for a simple test image from the comparative study are shown in Table 1.

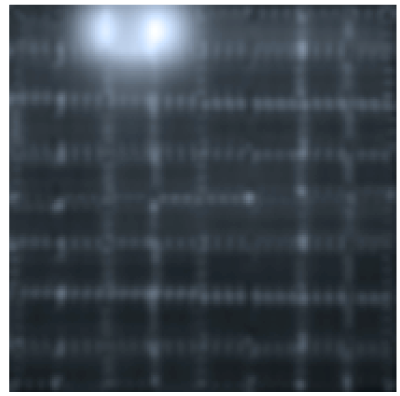
Table 1. Visually weighted errors from comparative study

halftone technique	max error	rms error
direct binary search	1.346	0.528
error diffusion	1.925	0.242
clustered dot	1.434	0.497
grating dot	1.310	0.308
Stochastic screen	1.844	0.679

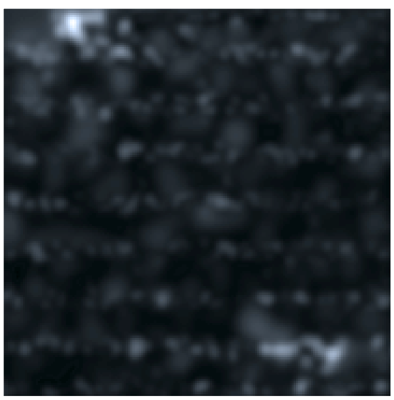
One problem with the clustered dot and grating dot dithered images for this particular test pattern is very noticeable edge distortion in the binary images. Examination of the single digit visually weighted errors doesn't really convey any information about the noticeable edge distortion associated with these two techniques. However, the visually weighted error maps in Figure 2 do



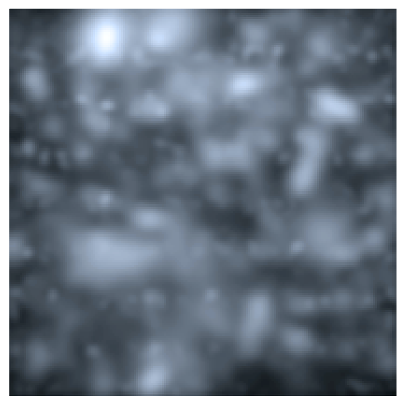
2a



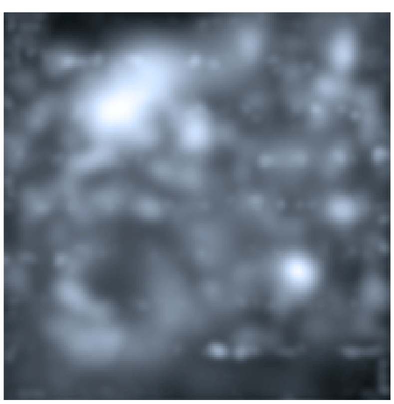
2d



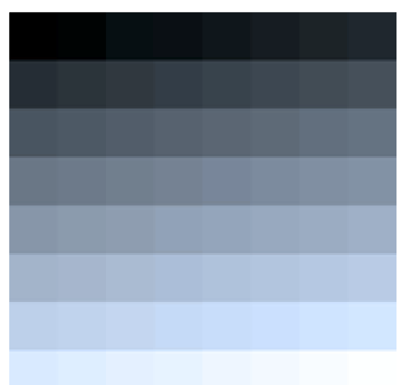
2b



2e



2c



2f

Figure 2a-2f. Visually weighted error maps for five different halftoning techniques. The grayscale source image is displayed on the bottom right. 2a. (top left) clustered dot ordered dither; 2b. (top right) grating dot ordered dither; 2c. (mid left) error diffusion dither; 2d. (mid right) stochastic screen dither; 2d. (bot left) direct binary search ; 2e. (bot right) continuous tone original successfully predict the perceivable edge distortion at the appropriate positions in the test images.

The performance of error diffusion for this test pattern and simulated view distance is also interesting, as it has the worst max error and best rms error. Examination of the error map image for this technique indicates that the visual model is successfully predicting the worm-like artifacts typical of error diffusion in the shadows and highlights.

The results above show that single number visually weighted errors are not necessarily good predictors of the appearance of errors in halftoned images, or at least should be used with caution. On the other hand, the visually weighted error maps do successfully predict the noticeable halftone distortions mentioned above. However, there is nothing in the model that understands the structure associated with the error maps for the clustered and grating dot techniques, something that is readily observable by a human viewer.

Many halftoned images are viewed well above the threshold of detection, where the binary structure of the halftone image is readily apparent and is a significant factor in the perception of the binary image. Design of visually weighted halftones could benefit from visual models that understand more about visual appearance. Algorithms such as direct binary search,[1] which utilize a visual model to optimize a halftone image, act like an observer pushing down high error spots on a visually weighted error map. They don't have any direct understanding of visual perception of image structure other then the detectability of different frequency grating patterns.

A traditional lithographic artist on the other hand, builds up a binary image rendering in a very different fashion, that is highly linked to visual perception of image structure.[14] Can one build visual models that incorporate a higher form of visual representation then just detectability of visual patterns? And if so, how could they be incorporated into the design of binary rendering algorithms?

Artistic Perception and Visual Processing

Many artists going back to Leonardo and more recently Speed have differentiated mass and line as parallel streams of visual representation.[15] A common theme in texts on drawing is how to appropriately render tone and outlines to achieve a specific type of visual style.[7,14] The quality of the pen strokes used to build up these two separate forms influences the perception of the rendered image. Other characteristics of visual images such as perspective, lighting and shadows, and negative space are also treated as separate properties of the visual image.[5] How these properties are represented directly influences the viewers perception of the rendered image.

Analogies of the modes of artistic representation discussed above can be seen in scientific observations of the visual system. The Gestalt laws of Pragnanz are concerned with the perception of visual form and organization. Lowe's[9] nonaccidental properties (smooth continuation, cotermination, parallelism, and symmetry) are derived from the Gestalt laws and relate to specific features in images that define viewpoint invariant object properties.

Biederman's[2] Geon theory of object recognition is based on structural primitives derived from non-accidental features as opposed to grayscale edge information. Adequate representation of the nonaccidental properties of images has been shown to directly influence object recognition rates.

Studies of figure-ground phenomena have shown the presence of a low temporal/high spatial frequency figure system and a high temporal/low spatial frequency ground system. Hayes[8] also discusses two separate processes for perception of images. They include a fine scale process used for detection of contour structure that is insensitive to sign and a coarse scale process used for shape from shading that is sensitive to sign.

Ramachandran's[17] studies of shape from shading phenomena show that there is a single light source assumption for shape from shading as well as an assumption that the light source is from above. Adelson's transparency illusions can also be analyzed as shadows in the context of separate processing of object and lighting properties of a scene. Freeman[6] designed a shape from shading model that is based on steerable filters.

Many models of object recognition typically start from extraction of line drawings from edges. Hayes[8] outlines the development of theories of line drawing perception, which developed historically as follows:

- Kepes/Goodman—line drawings follow a set of conventions that are learned.
- Gibson—lines have point-to-point correspondence to the original scene.
- Attneave—lines indicate points where information is maximal.
- Ratcliff—lines represent scene contours where there is a local change in brightness.
- Marr—lines are in correspondence with natural symbols computed by the brain.

Interestingly, studies with luminance ramps show that shape from shading can be generated with illusory contour, but not with edges.[17] Illusory contour is also oftentimes a stronger determinant for object segmentation in images then edge information. This may be because occlusion is a stronger indication of an object boundary then brightness edges, which are influenced by light and shadow.

Burr and Morrone[3] suggest that the phase relationship of visual channels are important to the perception of structure in images. They develop a model based on the local energy profile of a set of even/odd symmetry multi-scale filters. Their study of mach bands suggests that congruence of phase across visual channels determines visual features.

Robbins and Owens[11] suggest that 2D phase congruency determines the detectability of 2D image features which contain significant variation in more then one orientation. They derive an algorithm to detect 2D feature points from the local 2D energy profile. Simoncelli[13] has derived a technique for generating a rotation invariant local image structure signature. This technique is based on steerable wedge filters and can be used to locate junctions in grayscale imagery. 2D image feature or junction points share a close correspondence with cotermination properties.

The computer graphics community has recently begun to investigate rendering computer graphics images in a variety of artistic styles that simulate traditional artistic techniques such as the use of pen and ink in technical drawing.[12] The quest for photographic realism in the halftoning and digital printing communities has obscured investigation into alternative rendering techniques that simulate artistic techniques or alternative media directly on the marking engine.

Simulated artistic rendering from natural images as opposed to computer graphic models should ideally be based on visual models that capture the perceptual representational phenomena that artists encode and examine in their work. These models need to address higher order perception of structure and similarity in images rather then just detectability of spatial patterns.

Figure 1 shows a comparison of a traditional halftone screen rendering of a test image and an alternative pseudo-lithographic rendering. Rather then simulating photographic realism, the pseudo-lithographic rendering is trying to maximize information content in a stylized artistic rendering The binary rendering algorithm is based on a simple visual model that has parallel processing for mass and outline, and is implemented with multi-resolution steerable filters.

Models based on cotermination properties should allow for further image abstraction in binary renderings. Nonaccidental features could also be used to detect recognizable structure in visually weighted error maps. Direct binary search utilizes visual distortion metrics within the context of an exchange based optimization algorithm. Incorporation of a visual weighting for tone with the models discussed above could provide an alternative to algorithms such as dbs that would also allow for stylistic variation in addition to realistic tonal rendering.

References

1. J. P. Allebach, T. J. Flohr, D. P. Hilgenberg, "Model-based halftoning via direct binary search", *IS&T's 47th Annual Conference/ICPS 1994*, pp. 476–482; ***see Recent Progress in Digital Halftoning, Vol. I, pg. 106.***
2. I. Biederman, "Visual Object Recognition," in Visual Cognition, edited by S. M. Kosslyn and D. N. Osherton, MIT Press, 1995, pp. 121–165.
3. D. C. Burr and M. C. Morrone, "Feature detection in biological and artificial visual systems," in *Vision: coding and efficiency*, edited by Colin Blakemore, Cambridge University Press, Cambridge, 1990, pp. 185–194.
4. J. Dalton, "Perception of binary texture and the generation of stochastic halftone screens", *1995 SPIE/IST Symposium on Electronic Imaging Science and Technology, Proceedings Human Vision, Visual Processing, and Digital Display VI*., San Jose, CA, 95.
5. B. Edwards, "Drawing on the Right side of the brain," G. P. Putnam's Sons, N.Y., 1989.
6. W. T. Freeman, and E. H. Adelson, "The design and use of steerable filters", *IEEE Trans. Pat. Anal. and Machine Intell*., Vol. **13**, No. 9, 891-906, Sept. 1991.
7. A. L. Guptill, "Rendering in Pen and Ink," Watson-Guptill, N.Y. 1976.
8. A. Hayes and J. Ross, "Lines of sight," in *The Artful Eye*, edited by R. Gregory et. al., Oxford University Press,N.Y, 1995, pp. 339-352.
9. D. G. Lowe, "Perceptual organization and visual recognition", Kluwer, Norwell MA, 1985.
10. J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital images and human vision*, edited by Andrew Watson, Bradford Books, 1993, pp.163-178.
11. B. Robbins and R. Owens, "2D feature detection via local energy" to appear in *Image and Vision Computing*. (download at http://www.cs.uwa.edu.au:80/~robyn)
12. M. Salisbury, S.E. Anderson, R. Barzel, and D.H. Salesin, "Interactive Pen-and-Ink Illustration", *Siggraph 94 Computer Graphics Proc., Annual Conference Series*, 1994.
13. E. P. Simoncelli and H. Farid, "Steerable wedge filters for local orientation analysis", *IEEE Trans on Image Processing*, Sept. 1996.
14. G. Simmons, "The Technical Pen," Watson-Guptill, N.Y., 1992.
15. H. Speed, "The practice of drawing," Seeley, London, 1923.
16. P. C. Teo and D. J. Heeger, "Perceptual Image Distortion", *SPIE/IST Symposium on Electronic Imaging Science and Technology, Proceedings Human Vision, Visual Processing, and Digital Display VI*., San Jose, CA, 94, pp.
17. V. S. Ramachandran, "2-D or not 2-D, that is the question," in *The Artful Eye*, edited by R. Gregory et. al., Oxford University Press, N.Y, 1995, pp. 249-267.

❋ Previously published in *OSA/IS&T Optics & Imaging in the Information Age*, pp. 191–194, 1996.