Image Halftoning by Error Diffusion: A Survey of Methods for Artifact Reduction

Vishal Monga[1], Niranjan Damera-Venkata[2]
and Brian L. Evans[1]

[1]Dept. of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX 78712-1084
vishal@ece.utexas.edu, bevans@ece.utexas.edu

[2]Hewlett-Packard Laboratories
1501 Page Mill Road, Palo Alto, CA 94304-1126
damera@exch.hpl.hp.com

September 10, 2003

ABSTRACT

Grayscale error diffusion introduces nonlinear distortion (directional artifacts and false textures), linear distortion (sharpening), and additive noise. In color error diffusion what color to render is a major concern in addition to finding optimal dot patterns. This article presents a survey of key methods for artifact reduction in grayscale and color error diffusion. The linear gain model by Kite *et al.* replaces the thresholding quantizer with a scalar gain plus additive noise. They show that the sharpening is proportional to the scalar gain. Kite *et al.* derive the sharpness control parameter value in threshold modulation (Eschbach and Knox, 1991) to compensate linear distortion. False textures at mid-gray (Fan and Eschbach, 1994) are due to limit cycles, which can be broken up by using a deterministic bit flipping quantizer (Damera-Venkata and Evans, 2001). Several other variations on grayscale error diffusion have been proposed to reduce false textures in shadow and highlight regions, including green noise halftoning (Levien, 1993) and tone-dependent error diffusion (Li and Allebach, 2002). Color error diffusion ideally requires the quantization error to be diffused to frequencies and colors, to which the HVS is least sensitive. We review the following approaches: color plane separable (Kolpatzik and Bouman 1992) design; perceptual quantization (Shaked *et al.* 1996, Haneishi *et al.* 1996) ; green noise extensions (Lau *et al.* 2000); and matrix-valued error filters (Damera-Venkata and Evans, 2001).

Keywords: L imit cycles, false textures, sharpening, adaptive filters, color spikes, human visual system

1. INTRODUCTION

Digital halftoning is the process of representing continuous-tone (a.k.a. grayscale and color) images with a finite number of levels for the purpose of display on devices with finite reproduction palettes. Examples include conversion of a 24-bit color image to a three-bit color image and conversion of an 8-bit grayscale image to a binary image. The resulting images are called halftones. Until the late 1990s, printing presses, ink jet printers, and laser printers were only able to apply or not apply ink to paper at a given spatial location. For grayscale printing, the ink dots were black. For color printing, a cyan, magenta, and yellow ink dot is possible at each spatial location. Many color printing devices can also produce a black ink dot. In these cases, the printer is a binary device capable of reproducing only two levels, where the presence of a dot on the paper may be indicated by the level -1, and the absence of a dot may be indicated by the level $+1$. In other applications, such as display on monochrome or color monitors, the levels available are usually more than two, but finite. In all cases, the goal of digital halftoning is to produce, via an ingenious distribution of dots, the illusion of continuous tone.

Halftoning is more complicated than simply truncating each multi-bit intensity to the lower resolution. Simple truncation would give poor image quality because the quantization error would be spread equally over

all spatial frequencies. Halftoning methods in current use may be categorized as classical screening, dithering with blue noise, direct binary search, and error diffusion. All four categories of halftoning methods may be found in modern desktop printers (the direct binary search method may be implicitly present as the method to design the screen being used). Of these halftoning approaches, error diffusion is the particular focus of this paper. Error diffusion is particular interesting to analyze because it involves a non-separable infinite impulse response filter in addition to the nonlinearity introduced by the severe quantization.

Classical screening, which is the oldest halftoning method in printing, applies a periodic array of thresholds to each color of the multi-bit image. Pixels can be converted to -1 (black) if they are below the threshold or $+1$ (white) otherwise. With the continuous-tone images taking pixel values from -1 to $+1$ inclusive, a mask of M uniform thresholds would be a permutation of the set $\{-\frac{M-1}{M+1}, -\frac{M-3}{M+1}, \ldots, 0, \ldots, \frac{M-3}{M+1}, \frac{M-1}{M+1}\}$ for M odd, or the set $\{-\frac{M-1}{M}, -\frac{M-3}{M}, \ldots, \frac{M-3}{M}, \frac{M-1}{M}\}$ for M even. A mask of M thresholds would support $M+1$ intensity levels. When applying a mask with uniform thresholds to a constant mid-gray image, half of the halftone pixels within the extent of the mask would be turned on, and half would be turned off. The ordering of the thresholds in the mask has a significant effect on the visual quality of the halftone. A clustered dot screen would cluster dots in a connected way, which helps mitigate ink spread when printed. A dispersed dot screen would spread out the dots, which is well suited for low-cost displays. Both classical clustered dot and dispersed dot screens suffer from periodic artifacts due to quantization by a periodic threshold array.

To a very rough approximation as a linear spatially-invariant system, the human visual system is lowpass to the luminance component of a color image or to a monochrome image with respect to spatial frequency. The human visual system is in general less sensitive to uncorrelated high-frequency noise than uncorrelated low-frequency noise. Dithering with blue noise (i.e. high-frequency noise)[1] attempts to place the quantization noise from the halftoning process into the higher frequencies. Noise shaping is a characteristic of error diffusion as described below, but large periodic masks of thresholds (e.g. 128 x 128 pixels) can be designed to produce halftones with blue noise.[2,3]

Direct binary search[4] produces blue noise halftones by iteratively searching for the best binary pattern to match a given grayscale image by minimizing a distortion criterion. The distortion criterion incorporates a linear spatially-invariant model of the human visual system as a weighting function.[5] The direct binary search method produces the halftones with the highest visual quality to date. Due to its implementation complexity, it is impractical for use as a halftoning method in desktop printers. However, direct binary search can be employed to design screens[6] and error diffusion parameters.[7]

Error diffusion produces halftones of much higher quality than classical screening, with the tradeoff of requiring more computation and memory.[8] Screening amounts to pixel-parallel thresholding, whereas error diffusion requires a neighborhood operation and thresholding. The neighborhood operation distributes the quantization error due to thresholding to the unhalftoned neighbors of the current pixel. The term "error diffusion" refers to the process of diffusing the quantization error along the path of the image scan. In the case of a raster scan, the quantization error diffuses across and down the image. "Qualitatively speaking, error diffusion accurately reproduces the graylevel in a local region by driving the average error to zero through the use of feedback".[9]

The two most common families of halftoning methods in desktop printers are screening and error diffusion. This paper reviews and analyzes a wide variety of error diffusion methods. Section 2 first describes classical error diffusion for grayscale images. Classical error diffusion sharpens the image and often generates false textures (called worms) in shadow, mid-gray, and highlight regions of the original image. This section shows how to linearize error diffusion using a scalar gain model for the quantizer, and applies the scalar gain model to compensate for frequency distortion in the error diffusion process. Frequency distortion compensation is used in objective quality measures. This section also presents several other flavors of error diffusion, including edge enhanced, green noise, tone dependent, and block. Section 3 describes error diffusion for color halftoning. The application of grayscale error-diffusion methods to the individual colorant planes fails to exploit the HVS response to color noise. Visually objectionable artifacts, e.g. false color spikes can result due to the same. The section describes alternative approaches using perceptual non-scalar (or vector) quantization and using matrix-valued error filters to reduce these artifacts. Section 4 summarizes the paper and concludes with open research

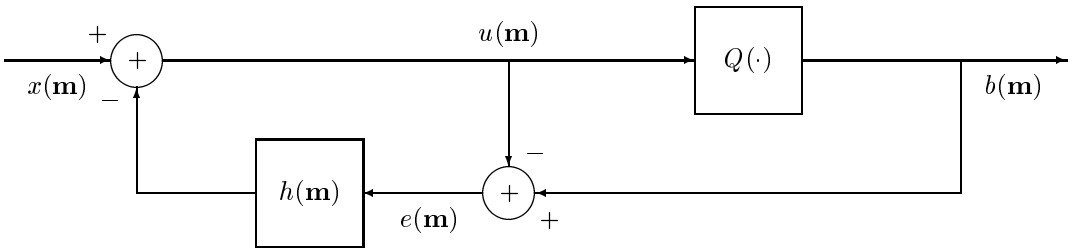


Figure 1. System block diagram for grayscale error diffusion halftoning where $\mathbf{m}$ represents a two-dimensional spatial index (m_1, m_2) and $h(\mathbf{m})$ is the impulse response of a fixed 2-D nonseparable FIR error filter having scalar-valued coefficients.

problems. We have implemented several error diffusion methods described in this paper in a freely distributable halftoning toolbox for Matlab:

http://www.ece.utexas.edu/~bevans/projects/halftoning/toolbox/index.html

2. GRAYSCALE ERROR DIFFUSION

Halftoning methods may be classified into three categories— amplitude modulation (AM), frequency modulation (FM) and AM-FM hybrid halftoning. In AM halftoning, the dot size is varied depending on the graylevel value of the underlying grayscale image while the dot frequency is held constant, e.g. clustered-dot ordered dither. FM halftones have a fixed dot size and shape, but the frequency of the dots varies with the graylevel of the underlying grayscale image. Conventional digital FM halftones have a fixed dot size of one pixel, e.g. those produced by dispersed-dot ordered dither and error diffusion. AM-FM halftones have variable dot shape/size, and variable dot frequency that depends on the graylevel value to be reproduced. Examples of AM-FM halftones include "green-noise" halftones by Levien,[10] halftones on space filling curves,[11] and halftones with texture control.[12]

Section 2.1 describes classical error diffusion, including Floyd-Steinberg error diffusion. Classical error diffusion suffers from false textures (worms) in shadow and highlight regions, and at mid-gray levels. Sections 2.2–2.4 reviews classical error diffusion and presents variations on error diffusion to correct for worms at mid-gray values. Sections 2.5 and 2.7 report variations on error diffusion to correct for worms in shadow and highlight regions. Two methods[7,13] report to remove worm artifacts for all tones.

2.1. Classical Error Diffusion Methods

The major 1976 advance in digital halftoning by Floyd and Steinberg[8] diffuses the quantization error over the neighboring continuous-tone pixels. As a grayscale image is raster scanned, the current pixel is thresholded against mid-gray to -1 (black) or 1 (white). The quantization error e is scaled and added to the nearest four grayscale (unhalftoned) pixels. The scaling factors are shown below, where $\times$ represents the current pixel[14]:

$$\begin{array}{ccc} & \times & \frac{7}{16} \\ \frac{3}{16} & \frac{5}{16} & \frac{1}{16} \end{array}$$

An alternate but equivalent implementation of error diffusion feeds back a filtered version of the quantization error to the input. This form, which is shown in Fig. 1, is also known as a noise-shaping feedback coder. In Fig. 1, $x(\mathbf{m})$ denotes the graylevel of the input image at pixel location $\mathbf{m}$, such that $x(\mathbf{m}) \in [-1, 1]$. The output halftone pixel is $b(\mathbf{m})$, where $b(\mathbf{m}) \in \{-1, 1\}$. Here, 1 is interpreted as the absence of a printer dot and -1 is interpreted as the presence of a printer dot. $Q(\cdot)$ denotes the standard thresholding quantizer function given by

$$Q(x) = \begin{cases} +1 & x \geq 0 \\ -1 & x < 0 \end{cases} \tag{1}$$

The error filter $h(\mathbf{m})$ filters the previous quantization errors $e(\mathbf{m}) \in [-1, 1]$:

$$h(\mathbf{m}) * e(\mathbf{m}) = \sum_{\mathbf{k} \in \mathcal{S}} h(\mathbf{k})\, e(\mathbf{m} - \mathbf{k}) \tag{2}$$

Here, $*$ means linear convolution, and the set $\mathcal{S}$ defines the extent of the error filter coefficient mask. The error filter output is fed back and added to the input. Note that $(0, 0) \notin \mathcal{S}$. The mask is causal with respect to the image scan.

To ensure that all of the quantization error is diffused, $h(\mathbf{m})$ must satisfy the constraint

$$\sum_{\mathbf{k} \in \mathcal{S}} h(\mathbf{m}) = 1 \tag{3}$$

This ensures that the error filter eliminates quantization noise at DC where the human visual system is most sensitive.[15] The quantizer input $u(\mathbf{m})$ and output $b(\mathbf{m})$ are given by

$$u(\mathbf{m}) = x(\mathbf{m}) - h(\mathbf{m}) * e(\mathbf{m}) \tag{4}$$

$$b(\mathbf{m}) = Q(u(\mathbf{m})) \tag{5}$$

Error diffusion halftones have significantly better quality over clustered and dispersed dot dither halftones, because they are free from periodic artifacts and shape the quantization noise into the high frequencies where the human eye is least sensitive. Since the halftone dots are of single pixel size, the illusion of continuous-tone is created by varying the dot frequency with graylevel. Thus, error diffusion is an example of FM halftoning.

The design of the error filter is the key to high quality error diffusion halftoning methods. The Floyd-Steinberg error filter was designed by trial-and-error to give four dyadic taps. Jarvis[16] and Stucki[17] proposed dyadic 12-tap error filters to reduce worms. Fig. 2 shows an example of a Floyd-Steinberg and Stucki error diffused halftones. For comparison purposes, a direct binary search halftone is also shown in Fig. 2. Recently, Ilbery[13] proposed a 25-tap Cauchy error filter to remove worm artifacts completely.

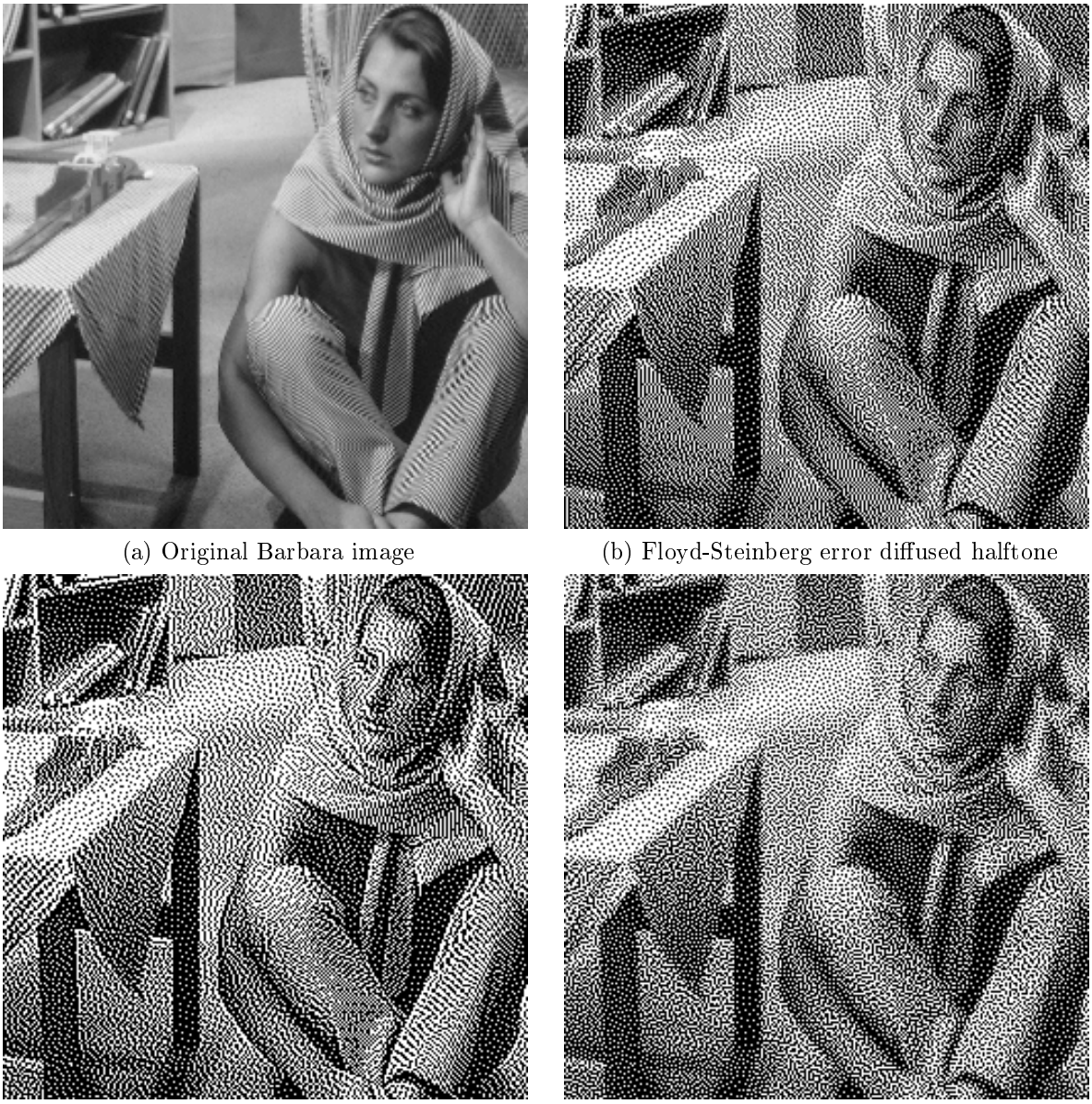
2.2. Analysis of Error Diffusion

Error diffused halftones suffer from several types of degradation. Knox[18] made the first major contribution to the analysis of error diffusion by mathematically showing that halftone quality can be improved by non-standard scanning techniques. For example, using a serpentine scan leads to a more symmetric error distribution. Moreover, he analyzed threshold modulation, which is the process of modulating the input of the quantizer, so as to break up objectionable artifacts in error diffusion.[19] Objectionable artifacts in error diffusion include "worm" artifacts in the very low and high graylevels and limit cycle artifacts in the mid-tones.

Hauck,[20] and later Anastassiou[21] and Bernard,[22] observed that error diffusion halftoning is two-dimensional sigma-delta modulation. Sigma-delta modulation is a popular method for A/D and D/A conversion in digital audio[23] that also employs feedback. This observation enabled others to analyze limit cycles and frequency responses of error diffusion methods by using results from sigma-delta modulation.

The mid-tone artifacts are the analogs of "limit cycles" in sigma-delta modulation.[24] Fan and Eschbach[25] analyzed constant input limit cycles in error diffusion and were able to predict which patterns which were most likely to occur. They also demonstrated how the error filter coefficients may be manipulated in order to reduce limit cycle behavior. Fan[26] showed that a sufficient condition for the stability of error diffusion is satisfied if the error filter weights are positive and sum to one. Fan also made the weaker conjecture that the error diffusion halftoning system is stable if the filter $1 - H(\mathbf{z})$ has only one zero at DC on the unit bicircle.

Knox[27] showed that error diffusion halftoning typically sharpens the original image. The amount of sharpening is proportional to the correlation of the input image with the error image. Closed-form frequency-domain analysis is enabled by a key result in one-dimensional sigma-delta modulation by Ardalan and Paulos.[28] They developed a linear gain model for the quantizer, which was accurate at low frequencies.



(a) Original Barbara image

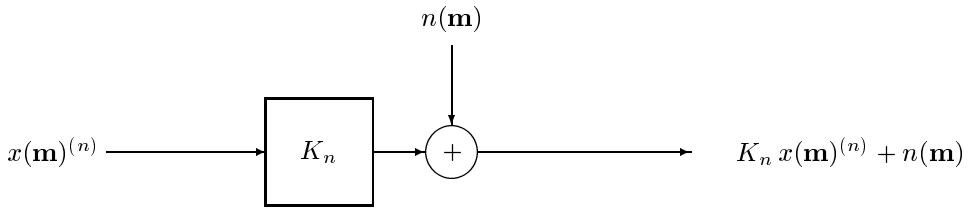
(b) Floyd-Steinberg error diffused halftone

(c) Jarvis error diffused halftone

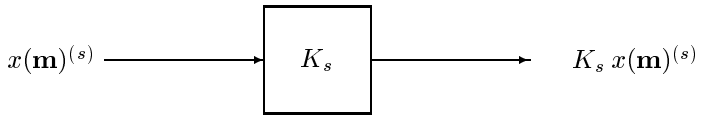
(d) Direct binary search halftone

Figure 2. Comparison of two classical error diffusion methods with the iterative direct binary search method. The direct binary search halftone is courtesy of Ti-chiun Chang and Jan Allebach of Purdue University.

Kite, Evans, Bovik, and Sculley[29, 30] applied the Ardalan and Paulos model for the quantizer. At the input to the quantizer, the signal component experiences a linear gain, and the quantization noise component undergoes a linear gain plus additive uncorrelated noise, as shown in Fig. 3. The gains are chosen to minimize the error incurred by using the model. For the noise path, $K_n = 1$, which is independent of the error filter.[29, 30] They also showed that the value of K_s is proportional to the amount of image sharpening. For example, $K_s \approx 2$ for Floyd-Steinberg, and $K_s \approx 4$ for Jarvis error diffusion. Jarvis halftones are visually sharper than Floyd-Steinberg



(a) Noise path.



(b) Signal path.

Figure 3. Linear gain model of the quantizer splits the quantizer input-output relationship into independent signal and noise paths.

halftones. They derived the signal transfer function (STF) and noise transfer function (NTF):

$$\mathrm{STF}(\mathbf{z}) = \frac{B_s(\mathbf{z})}{X(\mathbf{z})} = \frac{K_s}{1 + (K_s - 1)H(\mathbf{z})} \tag{6}$$

$$\mathrm{NTF}(\mathbf{z}) = \frac{B_n(\mathbf{z})}{N(\mathbf{z})} = 1 - H(\mathbf{z}) \tag{7}$$

The noise transfer function accurately predicts the highpass noise shaping when $H(\mathbf{z})$ is lowpass. When $H(\mathbf{z})$ is lowpass, the signal transfer function passes low frequencies and amplifies high frequencies. These transfer functions accurately predict the image sharpening and noise shaping effects in grayscale error diffusion halftoning.

2.3. Edge Enhancement Error Diffusion

Eschbach and Knox[31] modify conventional error diffusion to adjust halftone sharpness, as shown in Fig. 4. Edge enhancement error diffusion scales the current image pixel by a constant L and adds the result to the quantizer input. Hence, this may be regarded as an example of image-dependent threshold modulation.

As L increases, the sharpness of the resulting halftone increases. In a global sense, one value of L exists that minimizes sharpening, assuming that the image is wide sense stationary (WSS) and the input and output to the quantizer are jointly WSS.[30] Smaller values of L would cause blurring, and larger values would cause sharpening, with respect to the original grayscale image. Hence, L can be set to reduce linear distortion.

Kite *et al.*[9, 30] develop a formula for the globally optimal value of L that causes the signal components to be rendered in the halftone without sharpening when using a thresholding quantizer:

$$L_{opt} = \frac{1 - K_s}{K_s} \tag{8}$$

Here, the quantizer is modeled as a linear gain K_s for the signal path plus uncorrelated noise for the noise path,[30] as described in Section 2.2. Fig. 5(b) shows an unsharpened Floyd-Steinberg halftone using $L = -\frac{1}{2}$, i.e. $K_s = 2$. Fig. 5(c) shows an edge-enhanced Floyd-Steinberg halftone using $L = -\frac{3}{4}$, i.e. $K_s = 4$.

If the gain value K_s is chosen to be the linear minimum mean square error (LMMSE) estimator of the quantizer output vs. its input,[30, 32] then the error is guaranteed to be uncorrelated with the quantizer input. Since the model linearizes the quantizer, edge-enhancement error diffusion may be analyzed using linear system

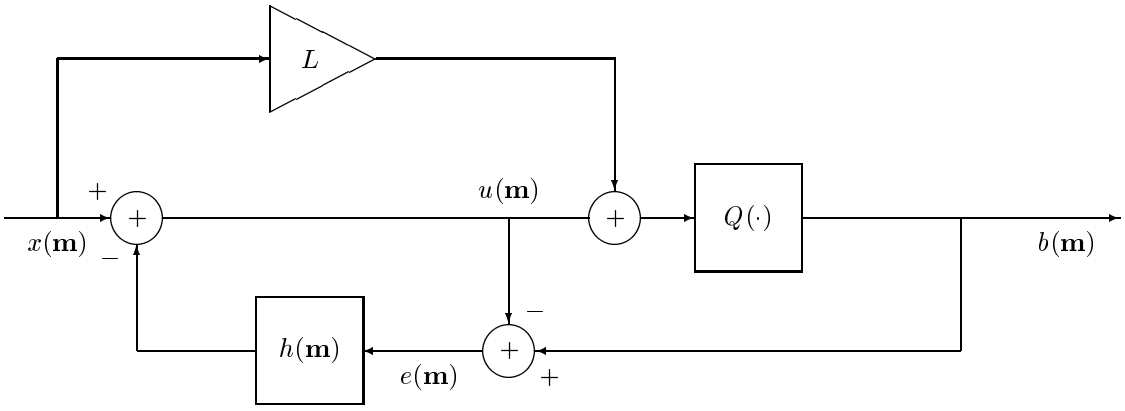


Figure 4. System block diagram for edge enhancement error diffusion where $h(\mathbf{m})$ represents a fixed 2-D non-separable finite impulse response error filter and L represents the scalar edge enhancement factor.

theory. The linear gain value affects signal shaping in error diffusion and the additive uncorrelated noise affects the noise shaping.[30,32] The linear gain value does not significantly affect the noise-shaping behavior of error diffusion.[29,30] This approach assumes that the input and output to the quantizer are jointly wide sense stationary stochastic processes. Since we must approximate statistical averages with sample averages, computing the LMMSE estimator on a per-image basis is computationally intensive.

Knox defined the *error image* in error diffusion to be the matrix of quantization errors scaled and displayed as an image.[27] Knox observed that the correlation of the *error image* with respect to the original is directly related to the frequency distortion produced by error diffusion. The greater the correlation of the original image with the *error image*, the sharper the halftone. Applying the linear gain model to edge enhancement error diffusion with $L = L_{opt}$ the signal component of the error image is shown mathematically to be equal to zero. Using L_{opt} in practice results in an error image that has low correlation with the input image. Thus, the sharpness cancelling methods are consistent with Knox's observations.

In the general case, the optimal value of L for sharpness compensation depends on (1) the error filter coefficients $h(\mathbf{m})$, (2) the quantizer function $Q(\cdot)$, and (3) the input grayscale image. Damera-Venkata and Evans[33] develop a least mean squares (LMS) framework for spatially adaptive algorithms using adaptive threshold modulation. The adaptive algorithms converge *in the mean* to the optimal value of L if the input and output of the quantizer are jointly WSS, and the input image is WSS. In a non-stationary environment, the algorithm tracks local variations in the input image. With this method, the linear gain that determines the sharpening of an error diffusion system need not be computed explicitly. The adaptive algorithm has a simple update of

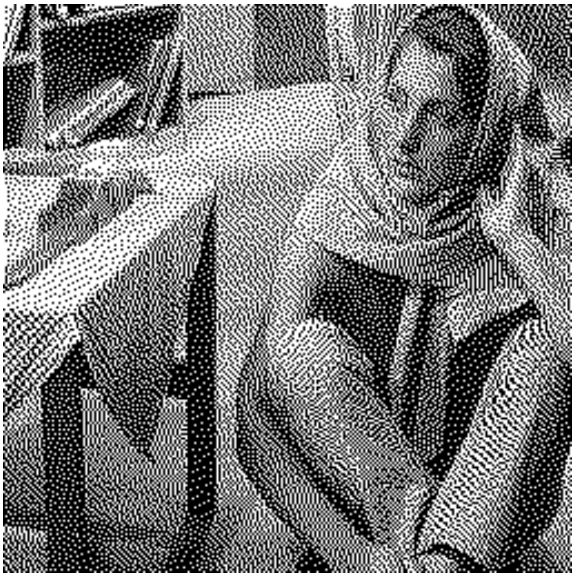
$$L^{(k+1)} = L^{(k)} - \lambda \left(o^{(k)} - x^{(k)} \right) x^{(k)} \tag{9}$$

where k denotes the previous location in the scan direction and $k+1$ is the current location. The scalar λ is the convergence parameter in the LMS algorithm.

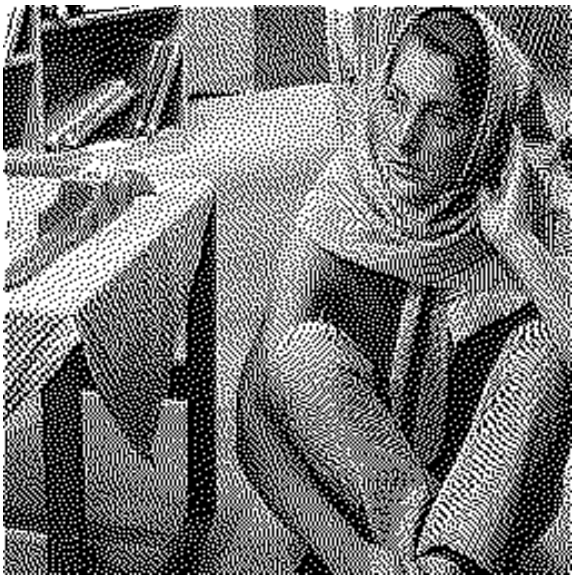
2.4. Block Error Diffusion

Conventional FM halftones (with single pixel halftone dots) suffer from physical printer imperfections such as dot gain and dot transfer issues.[34] The dot gain reduces the tonal range and causes loss of definition in the reproduced image.[12] Lau, Arce and Gallagher[12] further note that if the size/shape variation from printed dot to printed dot is small, then the effects of dot gain can be mitigated by dot gain compensation methods. If, however, there is a large variation in the size/shape of the printed dots, then clustering of the digital halftone dots adds robustness to the halftoning process, and in many cases becomes a necessity.[35]

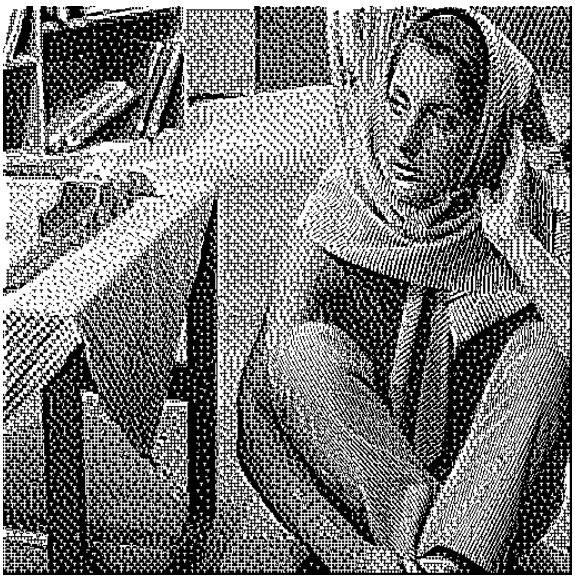
(a) Original Barbara image



(b) Edge enhanced halftoning with a flattened frequency response



(c) Edge enhanced halftoning with Stucki-like sharpening



(d) Block error diffused halftone (3×3 pixel clusters)

Figure 5. Variations on classical Floyd-Steinberg error diffusion halftoning.

AM-FM hybrid halftones do not suffer from periodic artifacts since they are not realizations of periodic stochastic processes. At the same time, the dots themselves are clustered, and hence these halftones are more robust to dot gain when printed. Lau, Arce and Gallagher[12] have extensively analyzed the spatial and spectral characteristics of these methods.

Velho and Gomez[11] produce AM-FM halftones by first dividing the image into cells along a space filling curve scan. Then, they compute the average intensity within a cell and generate a clustered dot pattern to approximate the average intensity. The difference between the dot pattern intensity and average intensity within

the cell is propagated to neighboring regions along the scan. The dot patterns are repositioned within the cell to introduce randomness and counteract periodic artifacts. The number of pixels in a cell controls the dot size.

Damera-Venkata, Monga and Evans[36] develop a framework for block error diffusion. Block error diffusion replaces a pixel in conventional error diffusion with a pixel block. In a pixel block, the quantization error at each pixel is diffused to select suitable pixels in neighboring blocks in properly selected proportions. Hence, an entire block of quantization error is diffused at a time. The model for block error distribution and operation of the block filter are shown in essentially the same as classical error diffusion block diagram in Fig. 1 except that the error filter has matrix-valued coefficients. Fig. 5(d) shows a block error diffused halftone. Fig. 6 explains the operation of the block error filter.

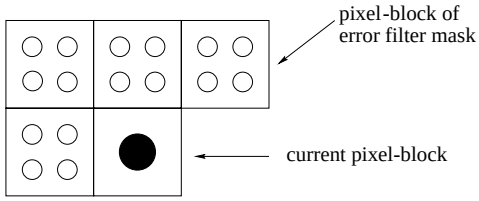


Figure 6. Block error filter operating on pixel-blocks of 2×2 pixels. The shaded circle indicates the current pixel-block. The unfilled circles indicate the error image pixels underlying the block filter mask. The pixels in the output pixel-block are computed using four linear combinations of all 16 error pixels within the error filter mask.

Fan[37] proposes a dot-to-dot error diffusion algorithm which combines traditional clustered dot dithering and error diffusion. He alleviates contouring with ordered dither by using a threshold matrix with linear/periodic thresholds. The basic idea is to perform error diffusion on dots instead of pixels so that coarse quantization artifacts can be reduced.[37] He uses periodic block thresholds and a scalar error filter.

Damera-Venkata, Monga and Evans[36] use aperiodic block thresholds and a matrix-valued error filter. They start with a scalar filter with the same support as the multifilter or block filter, and represent its coefficients by the row vector

$$\tilde{\gamma} = \left[g(0) \mid g(1) \mid \ldots \mid g(K^2 - 1) \right] \tag{10}$$

A multifilter $\tilde{\mathbf{\Gamma}}$ may be derived from it as follows:

$$\tilde{\mathbf{\Gamma}} = \tilde{\gamma} \bigotimes \tilde{\mathbf{D}} \tag{11}$$

Here, $\bigotimes$ denotes the Kronecker product operation and $\tilde{\mathbf{D}}$ is an $MN \times MN$ *diffusion matrix*. This is for pixel-blocks of size $M \times N$. Since the elements of $\tilde{\gamma}$ are the coefficients of a conventional error filter, they are non-negative and sum to one. Thus, to satisfy the constraints that all quantization error be diffused, the diffusion matrix must satisfy the constraints

$$\tilde{\mathbf{D}}\mathbf{1} = \mathbf{1} \tag{12}$$

$$\tilde{\mathbf{D}} \geq 0 \tag{13}$$

where $\mathbf{1}$ represents an $MN \times 1$ column vector with all of its elements equal to one. The physical meaning of deriving the block filter from a given conventional error filter via (11) is that the quantization error incurred at the current pixel block is diffused to the neighboring pixel-blocks in the same proportions that a conventional error filter diffuses error to its neighboring pixels. The constraints on the diffusion matrix simply indicate that *all* of the quantization error that is diffused to a pixel block must be diffused among pixels that compose the block. Thus, the pixel-blocks in the block-error diffusion framework are made to behave like pixels in conventional error diffusion and the block errors are diffused in much the same way as pixel errors in conventional error diffusion.

In determining whether a pixel-block is minority or majority, we first compare the quantization at mid-gray of the current quantizer input block and the current original grayscale block. If they are equal, then the block is a majority pixel-block; otherwise, it is a minority pixel-block. If the pixel block is determined to be a minority block, then the output pixel block is replaced with the desired dot-shape.

Using the block error diffusion framework, the authors generalize conventional error diffusion halftoning to produce FM halftones with user-controlled dot size and shape. The generated FM halftones can be designed to have very low dot size/shape variation, and the dot spacing is modulated depending on the underlying grayscale image.[36] FM halftones with clustered dots may be used to provide robust printed dots over all graylevel values.

Clustering pixels into blocks, i.e. as user-defined dot shapes, could be very useful for printers capable of rendering ink at the sub-pixel level, e.g. using sub-pixel laser pulse modulation in xerographic systems.[14] In this case, the proposed block error diffusion framework could render different dot shapes (depending on the underlying grayscale image) at the sub-pixel level, in which case the rendered block area would be the same as the rendered pixel area. Damera-Venkata and Yen[38] also use dot shape modulation within block error diffusion to embed information in hardcopy.

The block error diffusion framework may be efficiently implemented as a single-pass method that processes a neighborhood of blocks using a raster scan. Although the method could be extended to use other scans at the block level, a raster scan is preferred due to its simplicity. Based on a raster scan, the authors develop a fast parallel implementation of block error diffusion as a polyphase filterbank.[36] Block error diffusion[36] is shown to be approximately MN times faster than conventional error diffusion for pixel-block size of $M \times N$.

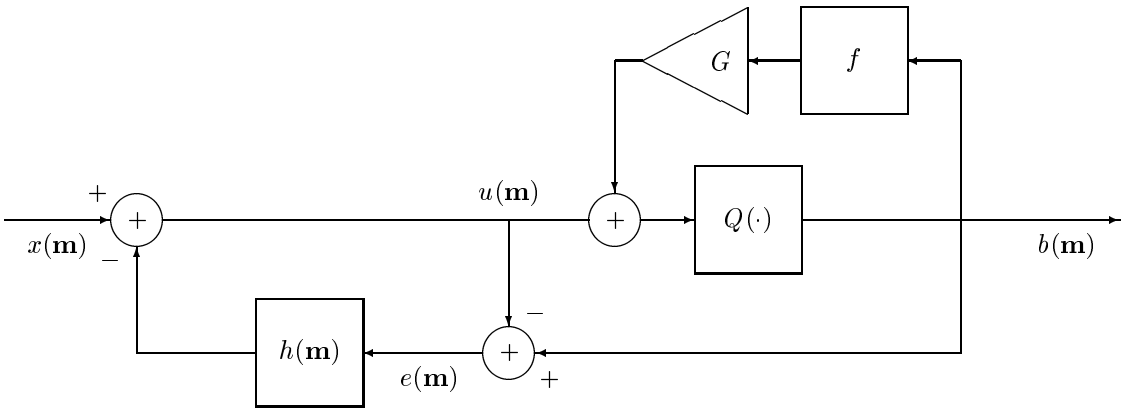


Figure 7. System block diagram for the green noise error diffusion of Levien. Here, $h(\mathbf{m})$ represents a fixed 2-D non-separable FIR error filter, and f represents a 2-D FIR hysteresis filter. The hysteresis constant G controls cluster size.

2.5. Green Noise Error Diffusion

Fig. 7 shows the setup for output-dependent feedback proposed by Levien.[10] The conventional error diffusion algorithm is a special case of Levien error diffusion when the hysteresis constant is $G = 0$. As with conventional error diffusion, the quantization error $e(\mathbf{m})$ is computed by subtracting the current modified input $u(\mathbf{m})$ (modified by incorporating past error terms via a linear weighting filter $h(\mathbf{m})$) from the current output $b(\mathbf{m})$.

Levien error diffusion differs from conventional error diffusion in that it filters the past output pixels by f, scales the result by a hysteresis constant G, and adds the result to the quantizer input. The halftoned output is

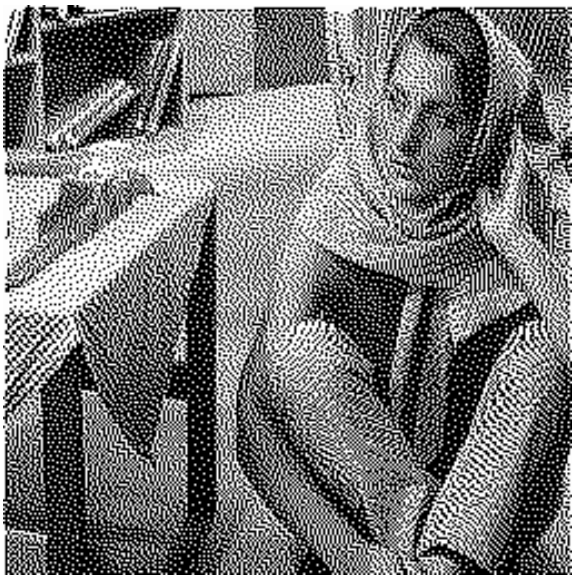
$$b(\mathbf{m}) = \begin{cases} 1 & u(\mathbf{m}) + G \sum_{\mathbf{k} \in \mathcal{O}} f(\mathbf{k}) b(\mathbf{m} - \mathbf{k}) > 0 \\ -1 & \text{otherwise} \end{cases} \tag{14}$$

Here $\mathcal{O}$ is a causal support set that does not include $\mathbf{0}$. Only past errors are included in the filtering.

The effect of adding a filtered version of the output to the quantizer input results in clustering of output pixels. Green noise halftoning produces an aperiodic clustered dot process which regulates dot reproduction

(a) Original Barbara image

(b) Green noise error diffusion with $G = 1.0$ using serpentine scan



(c) Error diffusion by Marcu

(d) Tone-dependent error diffusion

Figure 8. Variations on error diffusion to improve rendering in shadow and highlight regions. The halftone in (c) is courtesy of Gabriel Marcu of Apple, Inc. The halftone in (d) is courtesy of Ti-chiun Chang and Jan Allebach of Purdue University.

and tonal control on laser printers. More generally, green noise halftoning benefits printing processes with dot transfer and dot gain problems.[12] Lau, Arce and Gallagher[12] discovered that the pattern power spectrum of Levien error diffusion contained *mid-frequency* or *green noise* components between the blue noise patterns produced by regular error diffusion and ordered dither patterns. The hysteresis constant G controls the size of the dot clusters in the Levien algorithm. Lau, Arce and Gallagher[12] also suggested randomly perturbing the error filter coefficients to eliminate periodic artifacts in green noise error diffusion. Fig. 8(b) shows a green noise

halftone for $G = 1.0$. For green noise halftoning, serpentine scans generally give a significant visual improvement over raster scans.

Damera-Venkata and Evans[33] suggest an LMS adaptive algorithm to adapt the hysteresis filter coefficients in order to minimize mean square error between the output and the input. The LMS update equation is

$$\mathbf{f}^{(k+1)} = P_{\mathcal{C}} \left(\mathbf{f}^{(k)} - \lambda \left(o^{(k)} - x^{(k)} \right) \mathbf{o}_f^{(k)} \right) \tag{15}$$

where k denotes the previous location in the scan direction and $k+1$ is the current location. The scalar λ is the convergence parameter in the LMS algorithm, and $\mathbf{f}^{(k+1)}$ denotes the vector of hysteresis filter coefficients at the current pixel. $\mathbf{o}_f^{(k)}$ denotes the outputs used in the computation of the previous pixel, stacked into a vector. The projection operator $P_{\mathcal{C}}$ denotes a projection onto the convex set of filter coefficients that are non-negative and sum to unity. The adaptive algorithm converges *in the mean* to the optimal value of G if the input and output of the quantizer are jointly WSS, and the input image is WSS.[33]

For printers with pulse width modulation capability, He and Bouman[39] develop a green noise error diffusion algorithm (a.k.a. AM/FM halftoning) by simultaneously optimizing the dot density and dot size at a given graylevel. This is accomplished by first modeling the tone and distortion surfaces as functions of the dot size θ and the dot density ρ. They find the optimal θ and ρ by minimizing distortion $D(\theta, \rho)$ subject to the constraint that the tone $T(\theta, \rho)$ approximates the desired tone curve. The FM component of the algorithm lays out the dot centers while the AM part modulates the cluster sizes. The resulting dot pattern is the stochastic clustered dot pattern characteristic of green noise error diffusion.

2.6. Adaptive Error Diffusion

While many of the aforementioned approaches for improving error diffusion attempt to break up the patterning artifacts either use random perturbation or systematically change the way error is diffused, Kolpatzik and Bouman[40] use an explicit error criterion and then find the error-diffusion kernel (or error filter) that minimizes the error criterion under some assumptions on the characteristics of the quantization error in error diffusion. Wong[41] uses an adaptive method for dynamically adjusting the error filter to minimize a local frequency weighted error criterion that reflects the local characteristics of the image being halftoned (or binarized). The minimization procedure is performed simultaneously with the error diffusion procedure using the LMS algorithm. Further, Wong's framework makes no assumptions on the quantization error. The adaptive error diffusion scheme is illustrated in Fig. 2.6

2.7. Tone-Dependent Error Diffusion

Marcu[42] extended error diffusion to improve the halftone rendering at extreme levels of gray in the input image. In these shadow and highlight regions, the graylevel pixel is thresholded to a white or black dot. Then, the algorithm would search previously halftoned pixels to see if the same dot type has been rendered within a certain distance. If so, then the dot is suppressed. The search distance is larger for more extreme gray levels. Instead of using geometric distances, Marcu creates two lookup tables (requiring 338 bytes). The first lookup table indicates how many past (halftoned) pixels to search. This number indicates how many index offsets (with respect to the current pixel) to read from the second lookup table. On average, 10 past pixels are searched. For extreme gray levels of 1 and 254, 148 past (halftoned) pixels would be searched, which would reach up to nine rows above the current pixel row. Fig. 8(c) shows an example of Marcu's method.

Several tone dependent error diffusion methods have been developed. Methods included using error filters with different extents and coefficients according the the gray level of the current pixel being halftoned.[43–45] Ostromoukhov[46] designed optimal error filter weighting for a subset of gray levels based on blue noise spectra.

As mentioned in Section 1, direct binary search yields the best halftones to date. Analoui and Allebach[4] permute the current halftone using toggling and near-nearest neighbor swapping to improve a measure of visual quality with respect to the original image. Li and Allebach[7] employ direct binary search to choose the free parameters in error diffusion in a tone-dependent manner. For each gray level, a constant image at that gray level is input. Then, iterative search is employed to optimize the threshold and error filter coefficients. The approach is continued for all gray levels. When applied to an image at a given pixel, tone-dependent error

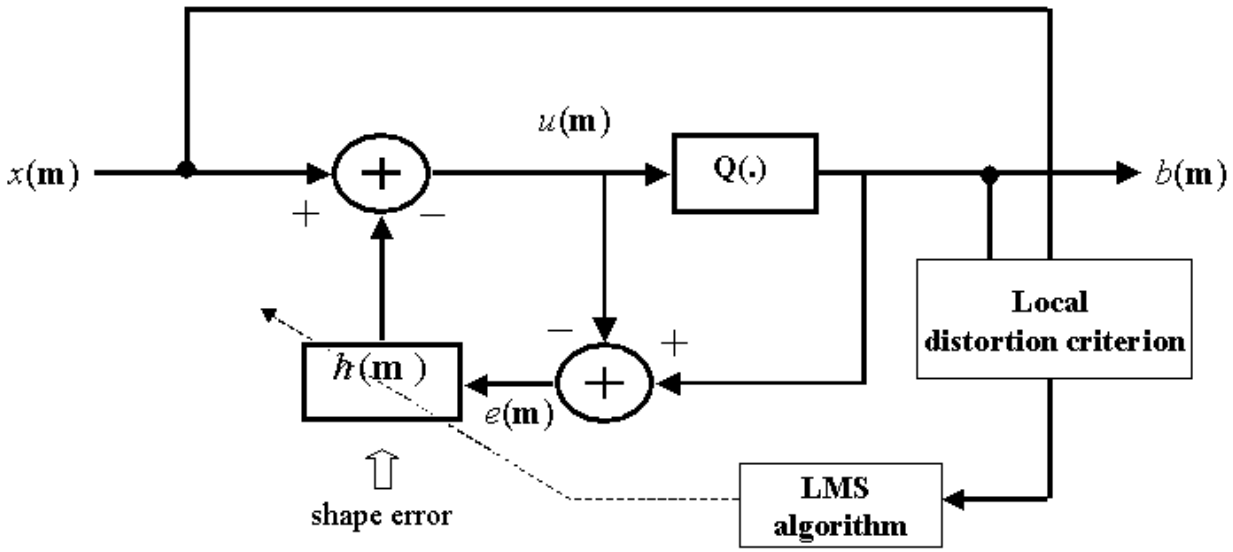


Figure 9. Adaptive error-diffusion system using the LMS algorithm for error minimization

diffusion looks up the threshold and error filter coefficients using the graylevel at the given pixel as the index. Fig. 8(d) shows an example of tone-dependent diffusion.

3. COLOR ERROR DIFFUSION

Color error diffusion is a high-quality method for color rendering of *continuous-tone* digital color images on devices with limited color palettes such as low-cost displays and printers. For display applications the input colorant space is a triplet of red, green, and blue (RGB) values, and the choice of output levels (i.e., the color palette) is a design parameter. For printing applications the input colorant space is a quadruple of cyan, magenta, yellow, and black (CMYK) values and the output levels are fixed. For example, for a bi-level CMYK printer there are 16 possible output colors. In this article we use an input RGB color space and discuss binary error diffusion. This allows us to concentrate the exposition to the essential properties of color error diffusion system design without having to focus on the issues of color palette design and device dependent nonlinear color transformations.

The application of grayscale error diffusion methods to the individual colorant planes fails to exploit the human visual system response to color noise. Ideally the quantization error must be diffused to frequencies and colors, to which the human visual system is least sensitive. Further it is desirable for the color quantization to take place in a perceptual space (such as Lab) so that the colorant vector selected as the output color is perceptually the closest to the color vector being quantized. We discuss each of the above two design principles of color error diffusion that differentiate it from grayscale error diffusion.

Section 3.1 describes color plane separable design and color plane separable implementation for error diffusion. The correlation of the color planes in error diffusion may be taken into account by using a (vector) quantizer based on a perceptual criterion, as discussed in Section 3.2.1 and 3.2.2, or matrix-valued error filters, as discussed in Section 3.3. These methods fall under the general umbrella of vector error diffusion.

3.1. Separable Methods

Kolpatzik and Bouman[40] use separable error filters in a luminance-chrominance to account for correlation among the color planes. Separate optimum scalar error filters are designed for the luminance and chrominance channels

independently based on a separable model of the human visual system. However, no constraints are imposed on the error filter to ensure that all of the red-green-blue (RGB) quantization error is diffused. Kolpatzik and Bouman model the error image as a white noise process and derive the optimum separable error filters to be used on the luminance and chrominance channels respectively. Such an approach implicitly assumes that there is no correlation between the luminance and chrominance channels. This implies that the transformation matrix from RGB to luminance-chrominance space is unitary. Damera-Venkata and Evans[32] solve for the optimum nonseparable error filter in the general case when all the error is required to be diffused, a non-separable color vision model is used and the linear transformation into the opponent color space is non-unitary. The luminance-chrominance separable error filters of Kolpatzik and Bouman are included in the general formulation of vector error diffusion.[32]

3.2. Quantization based on perceptual criteria

The use of the mean squared error (MSE) criterion in a colorant space is equivalent to uniform, separable, scalar quantization. The visual quantization error may be further reduced by performing the quantization according to perceptual criteria. Such methods typically aim to minimize colorimetric error, luminance variations, or a combination of the two.

3.2.1. Colorimetric quantization in error diffusion

Haneishi *et al.*[47] suggested the use of the XYZ and Lab spaces to perform quantization and error diffusion. In this case the rendering gamut is no longer a cube. The MSE criterion in the XYZ or Lab space is used to make a decision on the best color to output. The quantization error is a vector in the XYZ space is diffused using an error filter. (Lab space is not suitable for diffusing errors due to its nonlinear variation with intensity.) This method performs better than separable quantization but suffers from boundary artifacts[47,48] such as the "smear artifact" and the "slow response artifact" at color boundaries due to accumulated errors from neighboring pixels pushing quantizer input colors outside the gamut. This causes a longer lag in cancelling the errors. This effect may be reduced by clipping large color errors,[47,49] or by using a hybrid scalar-vector quantization method called semi-vector quantization.[48] This method is based on the observation that when errors in colorant space are small vector quantization does not produce the smear artifact. When large colorant space errors are detected, scalar quantization is used to avoid potential smearing. First, the colorants where the colorant space error exceeds a preset threshold are determined and quantized with scalar quantization. This restricts the possible output colors from which a color must be chosen using vector quantization in device independent color space.

3.2.2. Vector Quantization but Separable Filtering

One suggested improvement to separable color halftoning involves limiting the number of ink colors used to render a specific pixel. Shaked, Arad, Fitzhugh, and Sobel[50,51] suggest a method for using error diffusion for generating color halftone patterns that carefully examines each pixel's original color values simultaneously, distinct from past error, in order to determine potential output colors. By limiting the colors used, the authors argue that a smaller range of brightnesses in the colors are used to create each color area, which minimizes the visibility of the halftone pattern. This criteria, which is known as the minimum brightness variation criterion (MBVC), is based on the observation that the human eye is more sensitive to changes in "brightness" or luminance than to changes in chrominance, as summarized next.

A given color in the RGB cube may be rendered using the eight basic colors located at the vertices of the cube. Actually, any color may be rendered using no more than four colors, where different colors requiring different quadruples.[50] Moreover, the quadruple corresponding to a specific color is, in general, not unique. Suppose we want to print a patch of solid color, what colors should we use? Traditional work on halftoning addresses the issue of what pattern should the dots be placed in. The issue of participating halftone color was raised in,[52] but it served primarily as an example of how bad things can become. MBVC gives the issue a full answer. Based on the above arguments, the authors state the criteria they arrive at as follows: "To reduce halftone noise, select from within all halftone sets by which the desired color may be rendered, the one whose brightness variation is minimal."

The method proceeds by separating the RGB color space into Minimum Brightness Variation Quadrants. A RGB color space can be divided into six such quadrants.[51] The algorithm then works as follows:

Given a pixel value RGB($\mathbf{m}$) and error $e(\mathbf{m})$

1. Determine MBVQ based only on RGB($\mathbf{m}$)
2. Find the vertex v MBVQ tetrahedron closest to RGB($\mathbf{m}$) + $e(\mathbf{m})$
3. Compute the quantization error RGB($\mathbf{m}$) + $v(\mathbf{m}) - v$
4. Distribute error to the "future" pixels using standard error diffusion

A pixel whose original R, G, and B values are located within the WCMY tetrahedron will end up as one of those four colors, depending on which vertex its error places it closest to. The algorithm effectively reduces the number of pixel colors visible in a given solid region. It does not modify the color appearance when viewed from a significant distance away, however, since its average color should remain the same.

Another suggested improvement by Lau, Arce and Gallagher.[53] extended Levien's work in green noise error diffused halftones with clustered dots to color images. The extension employs an interference matrix to multiply the quantizer input prior to quantization. The interference matrix controls the overlap of dots of the individual primary colorants. Negative off-diagonal terms inhibit the overlap, and and positive off-diagonal terms encourage the overlap, of corresponding colorants.

3.3. Vector Error Diffusion

Separable methods for color error diffusion do not take into account the correlation among color planes. Vector error diffusion[54] represents each pixel in a color image as a vector of values. The thresholding step would threshold each vector component separately. The vector-valued quantization error (image) would be fed back, filtered, and added to the neighboring (unhalftoned) color pixels. A matrix-valued error filter could take correlation among color planes into account. For an RGB image, each error filter coefficient would be a 3×3 matrix.

As mentioned in Section 2.2, Kite *et al.*[29,30] quantify the sharpening and noise introduced by grayscale error diffusion by linearizing error diffusion. They replace the quantizer with the linear gain model developed by Ardalan and Paulos[28] for sigma-delta modulation. The model accurately predicts the noise shaping and image sharpening in error diffused halftones. Damera-Venkata and Evans[32] generalize the linear system model of grayscale error diffusion[30] to vector color error diffusion by replacing the linear gain model with a matrix gain model and by using properties of filters with matrix-valued coefficients.[32] The proposed "matrix gain" model includes the earlier linear gain model[29,30] as a special case. The matrix gain model describes vector color diffusion in the frequency domain, and predicts noise shaping and linear frequency distortion produced by halftoning.

The scalar gain model for the quantizer is shown in Fig. 3 in Section 2.2. As in the scalar case, the input to the quantizer is divided into signal and noise components. In the noise path, the gain is unity ($\tilde{\mathbf{K}}_\mathbf{n} = \mathbf{I}$), so the quantizer appears as additive uncorrelated noise. In the signal path, the gain is a matrix $\tilde{\mathbf{K}}_\mathbf{s}$. The matrix gain is related to the amount of sharpening, and the noise image models the quantization error. $\tilde{\mathbf{K}}_\mathbf{s}$ is chosen to minimize the error in approximating the quantizer with a linear transformation, in the linear minimum mean squared error sense:

$$\tilde{\mathbf{K}}_\mathbf{s} = \arg\min_{\tilde{\mathbf{A}}} E[\| \mathbf{b}(\mathbf{m}) - \tilde{\mathbf{A}}\ \mathbf{u}(\mathbf{m}) \|^2] \tag{16}$$

Here, $\mathbf{b}(\mathbf{m})$ is the quantizer output process (halftone), and $\mathbf{u}(\mathbf{m})$ is the quantizer input process. When $\mathbf{b}(\mathbf{m})$ and $\mathbf{u}(\mathbf{m})$ are wide sense stationary,[55] the solution for (16) is

$$\tilde{\mathbf{K}_\mathbf{s}} = \tilde{\mathbf{C}}_\mathbf{bu}\tilde{\mathbf{C}}_\mathbf{uu}^{-1} \tag{17}$$

where $\tilde{\mathbf{C}}_\mathbf{bu}$ and $\tilde{\mathbf{C}}_\mathbf{uu}$ are covariance matrices. The linearized vector error diffusion system has two inputs (original signal and quantization noise) and one output (the halftone), like its scalar counterpart. Using (17), the signal and noise transfer functions are[32]

$$\mathbf{B}_\mathbf{s}(\mathbf{z}) = \tilde{\mathbf{K}_\mathbf{s}} \left[\tilde{\mathbf{I}} + \tilde{\mathbf{H}}(\mathbf{z})(\tilde{\mathbf{K}_\mathbf{s}} - \tilde{\mathbf{I}})\right]^{-1} \mathbf{X}(\mathbf{z}) \tag{18}$$

$$\mathbf{B_n}(\mathbf{z}) = \left[\tilde{\mathbf{I}} - \tilde{\mathbf{H}}(\mathbf{z})\right] \mathbf{N}(\mathbf{z}) \tag{19}$$

In one dimension, (18) and (19) reduce to (6) and (7), respectively. The overall system response is given by

$$\mathbf{B}(\mathbf{z}) = \mathbf{B_s}(\mathbf{z}) + \mathbf{B_n}(\mathbf{z}) \tag{20}$$

For RGB vector error diffusion, matrix-valued error filter coefficients are adapted in[56] to reduce the mean squared error between the halftone and original. However, mean squared error does not have perceptual meaning in RGB space. Damera-Venkata and Evans[32] form an objective function J that measures the average visually weighted noise energy in the halftone. The output noise is computed by inverse transforming (19):

$$\mathbf{b_n}(\mathbf{m}) = \left[\tilde{\mathbf{I}} - \tilde{\mathbf{h}}(\mathbf{m})\right] \star \mathbf{n}(\mathbf{m}) \tag{21}$$

The noise energy is weighted by a linear spatially invariant matrix-valued HVS model, $\tilde{\mathbf{v}}(\mathbf{m})$, and form

$$J = E\left[\left\| \mathbf{v}(\tilde{\mathbf{m}}) \star \left[\tilde{\mathbf{I}} - \tilde{\mathbf{h}}(\mathbf{m})\right] \star \tilde{\mathbf{n}}(\mathbf{m}) \right\|^2\right] \tag{22}$$

Given a linear spatially invariant HVS model $\tilde{\mathbf{v}}(\mathbf{m})$, the problem is to design an optimal matrix-valued error filter

$$\tilde{\mathbf{h}}_{opt}(\mathbf{m}) = \arg \min_{\tilde{\mathbf{h}}(\mathbf{m}) \in \mathcal{C}} J \tag{23}$$

where the constraint $\mathcal{C}$ enforces the criterion that the error filter diffuses all quantization error[57]

$$\mathcal{C} = \left\{ \tilde{\mathbf{h}}(\mathbf{i}),\ \mathbf{i} \in \mathcal{S} \mid \sum_{\mathbf{i}} \tilde{\mathbf{h}}(\mathbf{i})\mathbf{1} = \mathbf{1} \right\} \tag{24}$$

$\mathcal{S}$ is the set of coordinates for the error filter support, i.e. $\mathcal{S} = \{(1,0),(1,1),(0,1),(-1,1)\}$ for Floyd-Steinberg. Damera-Venkata also analyzes vector color error diffusion with memory constraints using the matrix gain model.[58] Here the error filter is designed to minimize the visual effect of varying bit-allocations among the color error buffer channels. The human visual system model plays an important role in determining the performance of the designed optimum matrix-valued error filters.

We now explain the design of the linear human visual system model $\tilde{\mathbf{v}}(\mathbf{m})$. The linear color model employed by Damera-Venkata and Evans[32] is based on the pattern color separable model by Wandell *et al.*[59,60] They transfer device dependent RGB values into an opponent representation.[60,61] The three opponent visual pathways are white-black (luminance pathway) and red-green and blue-yellow (chrominance pathways). By x-y, we mean that in value, x is at one extreme and y is at the other.

Monga, Geisler and Evans[62] generalize this linear color model as a linear transformation $\tilde{\mathbf{T}}$ to a desired color space, which is not necessarily the opponent representation[59] but any one that satisfies pattern color separability, followed by appropriate spatial filtering in each channel. A complete HVS model is uniquely determined by the color space transformation and associated spatial filters. This generalization provides a platform for evaluation of different models in perceptual meaning and error filter quality obtained by minimizing (22).

The linear color model hence consists of (1) a linear transformation $\tilde{\mathbf{T}}$, and (2) separable spatial filtering on each channel. Each channel uses a different spatial filter. The filtering in the z-domain is a matrix multiplication by a diagonal matrix $\mathbf{D}(\mathbf{z})$. In the spatial domain, the linear HVS model $\tilde{\mathbf{v}}(\mathbf{m})$ is computed as

$$\tilde{\mathbf{v}}(\mathbf{m}) = \tilde{\mathbf{d}}(\mathbf{m})\tilde{\mathbf{T}} \tag{25}$$

Based on this framework, they evaluate four color spaces[62] in which to optimize matrix-valued error filters. The objective measure used for evaluation is the noise shaping gain NG of the optimal filter over the Floyd-Steinberg filter in decibels[32]:

$$NG = 10 \log_{10}\left(\frac{J_{fs}}{J_{opt}}\right) \tag{26}$$

Here, J refers to the value of the objective function given by (22). They also performed a subjective assessment procedure that evaluates the halftones based on a paired comparison task as described in.[62] The results of the subjective test corroborate the objective measures. The color spaces in order of increasing quality are (1) YIQ space, (2) YUV space, (3) opponent color space,[59,60] and (4) linearized CIELab color space.[63] These color spaces in conjunction with appropriate spatial filters as described in[62] form a unique HVS model. The color HVS model based on transformation to the linearized CIELab[63] color space and spatial filters for the luminance frequency response due to Nasanen and Sullivan[64] and the chrominance frequency response as given by Kolpatzik and Bouman[40] yields the best halftones. The subjective test is available online at

http://www.ece.utexas.edu/~vishal/cgi-bin/test.html.

3.4. Results

Fig. 10(a) shows the original *toucan* image. Fig. 10(b) shows a halftone generated by applying Floyd-Steinberg error diffusion separably. The green color impulses on the red toucan are easily visible on a color monitor. Fig. 10(e) and Fig. 10(g) show the green and blue planes of the Floyd-Steinberg halftone, respectively. The color impulses on the body of the red toucan are clearly visible in the green plane. Fig. 10(c) is a halftone generated using MBVC error diffusion as described in Section 3.2.2. Fig. 10(d) shows a halftone generated by applying an optimum matrix-valued error filter. The green color impulses are eliminated. Fig. 10(f) and Fig. 10(h) show the green and blue planes of the optimum halftone. The green channel (which contributes greatly to luminance) does not show spurious color impulses. However, since the error is shaped into the blue-yellow channel, the blue channel of the optimum halftone has several artifacts that are not easily visible in the optimum color halftone.

4. CONCLUSION

Two key results from 1-D sigma-delta modulation have been applied to error diffusion. The first result linearizes the thresholding quantizer, which enables linear system analysis and frequency distortion compensation. For the latter, the implementation complexity only increases by one multiplication and one addition per pixel. The second result replaces the thresholding quantizer with a deterministic bit flipping quantizer to break up midtone worms due to limit cycles. The implementation complexity would only increase by one comparison per pixel.

Two low-complexity approaches reduce false textures in shadow and highlight regions. Green noise halftoning employs feedback from the output to the quantizer input and uses a 2-D finite impulse response filter and a scalar gain in the feedback loop. The implementation complexity increase is one addition per filter tap per pixel, and one multiplication per pixel. Tone-dependent error diffusion applies a different threshold and set of error filter coefficients depending on the graylevel of the current pixel. The implementation complexity increase is five additional memory accesses per pixel, and 1,280 additional bytes of memory.

One can combine tone-dependent error diffusion, deterministic bit flipping quantization, and sharpening compensation to reduce false textures in shadow, midtone, and highlight regions as well as control the sharpening. Direct binary search could design the error filter coefficients for each gray level. A serpentine scan could be used to reduce directional artifacts.

Block error diffusion is a general framework for producing FM and AM-FM halftones with user controlled shape, size, and sharpness. The framework could also be used to embed FM halftones in other FM halftones, and has an efficient parallel implementation. Future work could explore the effect of other forms of the diffusion matrix on the halftone results. Another application is the design of AM-FM screens.

Future applications of the matrix gain model could include color halftone compression, inverse halftoning of color halftones, and optimal spatially varying sharpness control. For example, JPEG quantization tables or subband quantizer bit allocations in JPEG 2000 may be optimized based on the signal and noise transfer functions and the visual model developed in Section 3.3. One may allocate fewer bits to image frequencies where the color halftoning introduces high distortion. A method to design JPEG quantization tables based on empirically determined frequency distortions for grayscale image halftoning is described in.[65]

In color error filter design there are two major open problems. First, the constraints that are necessary for good image quality in terms of color smoothness (reduced irregularity in the halftone noise) and halftone dot

(a) Original image

(b) Floyd-Steinberg error filter

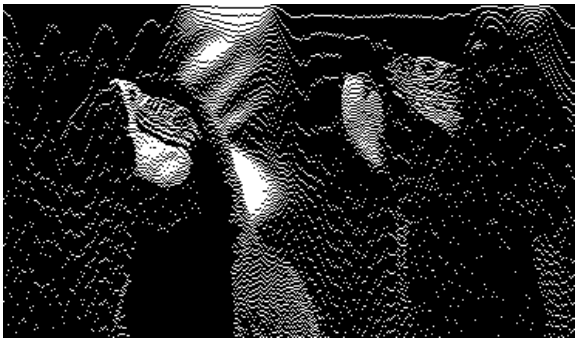
(c) MBVC error diffusion

(d) Optimum Error Filter

(e) Green plane of Floyd-Steinberg halftone

(f) Green plane of optimum halftone

(g) Blue plane of Floyd-Steinberg halftone



(h) Blue plane of optimum halftone

Figure 10. Role of human visual system model in color error diffusion

spacing have not yet been formulated. The second problem concerns stability of the vector diffusion framework. In[66] Fan analyzes stability of color error diffusion with separable color plane diffusion but vector quantization. He establishes constraints on both the color error filters and the quantization process that guarantee stability. These constraints are sufficient but not always necessary. Likewise, a necessary and sufficient condition for the numerical stability of color error diffusion with matrix valued filters has been elusive. The design process in Section 3.3 does not guarantee optimal dot distributions or numerical stability. The stability of the nonlinear feedback system governed by the error diffusion equations requires that the quantization error be bounded for all input colors in the input color gamut.

5. ACKNOWLEDGEMENTS

The authors would like to thank many people for great technical discussions on image halftoning over the years, including but not limited to Jan P. Allebach, C. Brian Atkins, Kathrin Berkner, Martin Boliek, Charles A. Bouman, Ricardo de Queiroz, Faouzi Kossentini, Peter W. M. Ilbery, Thomas D. Kite, Keith Knox, David J. Lieberman, Qian Lin, Gabriel Marcu, Bo Martins, Ramesh Neelamani, Victor Ostromoukhov, Eli Saber, Edward Schwartz, David A. D. Tompkins, Daniel Tretter, Robert Ulichney, Ping W. Wong, and Jonathan Yen.

REFERENCES

1. R. Ulichney, "Dithering with blue noise," *Proc. of the IEEE* **76**, pp. 56–79, Jan. 1988.
2. T. Mitsa and K. Parker, "Digital halftoning using a blue noise mask," *J. Opt. Soc. Am. A* **9**, pp. 1920–1929, Nov. 1992.
3. R. Ulichney, "The void-and-cluster method for dither array generation," *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV* **1913**, pp. 332–343, Feb. 1993.
4. M. Analoui and J. Allebach, "Model based halftoning using direct binary search," *Proc. SPIE Human Vision, Visual Processing, and Digital Display III* **1666**, pp. 109–121, Feb. 1992.
5. S. H. Kim and J. P. Allebach, "Impact of human visual system models on model based halftoning," *IEEE Trans. on Image Processing* **11**, pp. 258–269, Mar. 2002.
6. J. Allebach and Q. Lin, "FM screen design using the DBS algorithm," *Proc. IEEE Conf. on Image Processing* **I**, pp. 539–542, Sept. 1996.
7. P. Li and J. P. Allebach, "Tone dependent error diffusion," *SPIE Color Imaging: Device Independent Color, Color Hardcopy, and Applications VII* **4663**, pp. 310–321, Jan. 2002.
8. R. Floyd and L. Steinberg, "An adaptive algorithm for spatial grayscale," *Proc. Soc. Image Display* **17**(2), pp. 75–77, 1976.
9. T. D. Kite, *Design and Quality Assessment of Forward and Inverse Error-Diffusion Halftoning Algorithms.* PhD thesis, Dept. of ECE, The University of Texas at Austin, Austin, TX, Aug. 1998.
10. R. Levien, "Output dependent feedback in error diffusion halftoning," *IS&T Imaging Science and Technology* **1**, pp. 115–118, May 1993.
11. L. Velho and J. M. Gomez, "Digital halftoning with space filling curves," *Computer Graphics* **25**, pp. 81–90, July 1991.
12. D. L. Lau, G. R. Arce, and N. C. Gallagher, "Green-noise digital halftoning," *Proc. of the IEEE* **86**, pp. 2424–2442, Dec. 1998.
13. P. V. M. Ilbery, "Radially balanced error diffusion," *Proc. IEEE Conf. on Image Processing* **1**, pp. 645–648, Sept. 2002.
14. H. R. Kang, *Digital Color Halftoning*, SPIE Optical Engineering Press, 1999.
15. J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. on Info. Theory* **20**, pp. 525–536, July 1974.
16. J. Jarvis, C. Judice, and W. Ninke, "A survey of techniques for the display of continuous tone pictures on bilevel displays," *Computer Graphics and Image Processing* **5**, pp. 13–40, May 1976.
17. P. Stucki, "MECCA—a multiple-error correcting computation algorithm for bilevel hardcopy reproduction," Research Report RZ1060, IBM Research Laboratory, Zurich, Switzerland, 1981.

18. K. Knox, "Error diffusion: a theoretical view," *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV* **1913**, pp. 326–331, Feb. 1993.
19. K. Knox and R. Eschbach, "Threshold modulation in error diffusion," *J. Electronic Imaging* **2**, pp. 185–192, July 1993.
20. R. Hauck, "Binary coding techniques with emphasis on pulse density modulation," *SPIE Int. Optical Computing Conference* **700**, pp. 265–269, Sept. 1986.
21. D. Anastassiou, "Error diffusion coding for A/D conversion," *IEEE Trans. on Circuits and Systems* **36**, pp. 1175–1186, Sept. 1989.
22. T. Bernard, "From Σ-Δ modulation to digital halftoning of images," *Proc. IEEE Int. on Conf. Acoustics, Speech, and Signal Processing* , pp. 2805–2808, May 1991.
23. K. Pohlmann, *Principles of Digital Audio*, McGraw-Hill, 3rd ed., 1995.
24. S. Norsworthy, R. Schreier, and G. Temes, eds., *Delta-Sigma Data Converters*, IEEE Press, 1997.
25. Z. Fan and R. Eschbach, "Limit cycle behavior of error diffusion," *Proc. IEEE Conf. on Image Processing* **2**, pp. 1041–1045, Nov. 1994.
26. Z. Fan, "Stability analysis of error diffusion," *Proc. IEEE Int. on Conf. Acoustics, Speech, and Signal Processing* **5**, pp. 321–324, Apr. 1993.
27. K. Knox, "Error image in error diffusion," *Proc. SPIE Image Processing Algorithms and Techniques III* **1657**, pp. 268–279, Feb. 1992.
28. S. Ardalan and J. Paulos, "An analysis of nonlinear behavior in delta-sigma modulators," *IEEE Trans. on Circuits and Systems* **34**, pp. 593–603, June 1987.
29. T. D. Kite, B. L. Evans, A. C. Bovik, and T. L. Sculley, "Digital halftoning as 2-D delta-sigma modulation," *Proc. IEEE Conf. on Image Processing* **1**, pp. 799–802, Oct. 1997.
30. T. D. Kite, B. L. Evans, and A. C. Bovik, "Modeling and quality assessment of halftoning by error diffusion," *IEEE Trans. on Image Processing* **9**, pp. 909–922, May 2000.
31. R. Eschbach and K. Knox, "Error-diffusion algorithm with edge enhancement," *J. Opt. Soc. Am. A* **8**, pp. 1844–1850, Dec. 1991.
32. N. Damera-Venkata and B. L. Evans, "Design and analysis of vector color error diffusion halftoning systems," *IEEE Trans. on Image Processing* **10**, pp. 1552–1565, Oct. 2001.
33. N. Damera-Venkata and B. L. Evans, "Adaptive threshold modulation for error diffusion halftoning," *IEEE Trans. on Image Processing* **10**, pp. 104–116, Jan. 2001.
34. M. A. Coudray, "Causes and corrections of dot gain on press," *Screen Printing: The Journal of Technology and Management* **86**, pp. 18–26, Aug. 1996.
35. R. Ulichney, *Digital Halftoning*, MIT Press, Cambridge, MA, 1987.
36. N. Damera-Venkata, V. Monga, and B. L. Evans, "Clustered dot FM halftoning via block error diffusion," *IEEE Trans. on Image Processing* , submitted.
37. Z. Fan, "Dot-to-dot error diffusion," *J. Electr. Imaging* **2**, pp. 62–66, 1993.
38. N. Damera-Venkata and J. Yen, "Image barcodes," *Proc. SPIE Color Imaging: Processing, Hardcopy and Applications VIII* **5008**, Jan. 2003.
39. Z. He and C. Bouman, "AM/FM halftoning: A method for digital halftoning through simultaneous modulation of dot size and dot placement," *Proc. SPIE Color Imaging: Device-Independent Color, Color Hardcopy, and Applications VII* **4663**, Jan. 2002.
40. B. Kolpatzik and C. Bouman, "Optimized error diffusion for high quality image display," *Journal of Electronic Imaging* **1**, pp. 277–292, Jan. 1992.
41. P. Wong, "Adaptive error diffusion and its application in multiresolution rendering," *IEEE Trans. on Image Processing* **5**, pp. 1184–1196, July 1996.
42. G. Marcu, "An error diffusion algorithm with output position constraints for homogeneous highlight and shadow dot distribution," *SPIE Color Imaging: Device Independent Color, Color Hardcopy, and Graphic Arts III* , Jan. 1998.
43. R. Eschbach, "Reduction of artifacts in error diffusion by means of input-dependent weights," *J. Electronic Imaging* **2**, pp. 352–358, Oct. 1993.

44. J. Shu, "Adaptive filtering for error diffusion by means of input-dependent weights," *SID Digest of Technical Papers* , pp. 833–836, May 1995.
45. V. Ostromoukhov, "Enhanced error-diffusion method for color or black-and-white reproduction," 1998. US Patent No. 5,737,453.
46. V. Ostromoukhov, "A simple and efficient error-diffusion algorithm," *Proc. Int. Conf. on Computer Graphics and Interactive Tech.* , Aug. 2001.
47. H. Haneishi, T. Suzuki, N. Shimonyama, and Y. Miyake, "Color digital halftoning taking colorimetric color reproduction into account," *J. Electronic Imaging* **5**, pp. 97–106, Jan. 1996.
48. Z. Fan and S. Harrington, "Improved quantization methods in color error diffusion," *J. Electronic Imaging* **8**, pp. 430–437, Oct. 1999.
49. C. Kim, I. Kweon, and Y. Seo, "Color and printer models for color halftoning," *J. Electronic Imaging* , pp. 166–180, June 1997.
50. D. Shaked, N. Arad, A. Fitzhugh, and I. Sobel, "Ink relocation for color halftones," *HP Labs Technical Report, HPL-96-127R1* , 1996.
51. D. Shaked, N. Arad, A. Fitzhugh, and I. Sobel, "Color diffusion: Error-diffusion for color halftones," *HP Labs Technical Report, HPL-96-128R1* , 1996.
52. R. V. Klassen and R. Eschbach, "Vector error diffusion in a distorted color space," *Proc. IS&T Conf. 47th Annual Conf.* , 1994.
53. D. L. Lau, G. R. Arce, and N. C. Gallagher, "Digital color halftoning with generalized error diffusion and multichannel green-noise masks," *IEEE Trans. on Image Processing* **9**, pp. 923–935, May 2000.
54. H. Haneishi, H. Yaguchi, and Y. Miyake, "A new method of color reproduction in digital halftone image," in *Proc. IS&T Conf. 47th Annual Conf.*, (Cambridge, MA), May 1993.
55. H. Stark and J. W. Woods, *Probability, Random Processes and Estimation Theory for Engineers*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
56. L. Akarun, Y. Yardimci, and A. E. Cetin, "Adaptive methods for dithering color images," *IEEE Trans. on Image Processing* **6**, pp. 950–955, July 1997.
57. N. Damera-Venkata, *Analysis and Design of Vector Error Diffusion Systems for Image Halftoning.* PhD thesis, Dept. of ECE, The University of Texas at Austin, Austin, TX 78712 USA, http://www.ece.utexas.edu/~bevans/students/phd/niranjan/, Dec. 2000.
58. N. Damera-Venkata, "Incorporating memory constraints in the design of color error diffusion halftoning systems," *Proc. SPIE Color Imaging: Processing, Hardcopy and Applications VIII* **5008**, Jan. 2003.
59. A. B. Poirson and B. A. Wandell, "Appearance of colored patterns: Pattern-color separability," *J. Opt. Soc. Am. A* **10**, pp. 2458–2470, Dec. 1993.
60. X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color image reproduction," *SID Digest of Technical Papers* , pp. 731–734, 1996.
61. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley, 1998.
62. V. Monga, W. S. Geisler, and B. L. Evans, "Linear, color separable, human visual system models for vector error diffusion halftoning," *IEEE Signal Processing Letters* **accepted for publication**, 2003.
63. T. J. Flohr, B. W. Kolpatzik, R. Balasubramanian, D. A. Carrara, C. A. Bouman, and J. P. Allebach, "Model based color image quantization," *Proc. SPIE Human Vision, Visual Proc. and Digital Display IV* **1913**, pp. 270–281, 1993.
64. J. Sullivan, L. Ray, and R. Miller, "Design of minimum visual modulation halftone patterns," *IEEE Trans. on Systems, Man, and Cybernetics* **21**, pp. 33–38, Jan. 1991.
65. R. A. V. Kam, P. W. Wong, and R. M. Gray, "JPEG-compliant perceptual coding for a grayscale image printing pipeline," *IEEE Trans. on Image Processing* **8**, pp. 1–14, Jan. 1999.
66. Z. Fan, "Stability analysis for color error diffusion," *Proc. SPIE Sym. Electron. Imaging Sci. Technology* , Jan. 2000.