Stochastic Screen Halftoning for Electronic Imaging Devices

Qing Yu and Kevin J. Parker

Department of Electrical Engineering, University of Rochester, Rochester, New York 14627 E-mail: parker@ee.rochester.edu

Received June 12, 1997; revised September 12, 1997

For numerous digital imaging applications, there is a need to maintain the highest quality perceived image, while utilizing a printer or display that can only achieve a limited number of output states. Digital halftoning is the approach that has been widely used to meet this demand. In this focus paper, we provide a short summary of halftone techniques, then we concentrate on the newer and expanding roles of stochastic halftone screens—which are free of regular periodic structures and have numerous advantages in quality color rendering. We address some theoretical issues, design and optimality issues, printer compensation issues, and color quality issues that pertain to the development and use of stochastic screens for electronic imaging devices. © 1997 Academic Press

1. INTRODUCTION

In numerous digital imaging applications, there is a need to maintain the highest quality perceived image, while utilizing a printer or display that can only achieve a limited number of output states. Common examples of this include the use of ink-jet printers to make video hardcopy, and the display of images from web pages.

In the historic evolution of printed images, photographic halftone screens were used to render the illusion of gray scale with binary (black ink on white paper) printers. Gray regions were printed as a mosaic of black and white subregions, since the properties of the human visual system (HVS) would tend to create a perception of gray. In modern digital imaging applications, digital halftone screens or algorithms are employed to generate precisely defined patterns for halftoning. In cases where a limited number of intermediate states are achievable, the process is sometimes referred to as multitoning.

In this focus paper, we provide a short summary of halftone techniques and concentrate on the newer and expanding role of stochastic halftone screens—which are free of regular periodic structures and have numerous advantages in high quality color rendering. We address some theoretical issues, design and optimality issues, printer compensation issues, and color quality issues that pertain to the development and use of stochastic screens for digital imaging applications.

1.1. Ordered Dither

The history of halftoning technology can be dated back to the last century when physical screens and gauzes were used to generate halftone images. These techniques have been translated directly to digital halftoning. Some excellent comprehensive reviews have been published, including Ulichney [1], Roetling and Loce [2], Jones [3], and Kang [4]. A brief orientation is given here. Ordered dither is the natural digital solution, where a two-dimensional threshold array is designed and the halftoning process is accomplished by a simple pixelwise comparison of the gray scale image against the array (Fig. 1). This method is straightforward and requires little computation; thus ordered dither is the most popular and widely used technique. Depending on the progressive ordering of how halftone dots in a cell are turned on/off, ordered dither can be classified into clustered-dot and dispersed-dot. In clustered-dot ordered dither, adjacent pixels are turned on as gray level changes to form a cluster in the halftone cell. Clustered-dot dither is primarily used for printing devices that have difficulty printing isolated single pixels. Obviously, this congregation of pixels will result in noticeable low-frequency structures in the output image. On the other hand, in disperseddot ordered dither, halftone dots in a cell are turned on individually without grouping them into clusters. Therefore, sharp edges can be better rendered compared to clustered-dot dither. However, dispersed-dot techniques are more susceptible to dot gain, a problem that is considered in later sections. Figure 2 gives an example of clustereddot and dispersed-dot patterns.

1.2. Stochastic Processes

Many problems of ordered dither, including susceptibility to Moiré patterns and highly visible texture, can be traced back to its rigid regular structures. To break these regular structures, researchers sought less obtrusive halftone patterns. Blue noise halftoning, also called stochastic



FIG. 1. Ordered dither halftoning technique.

screening or frequency modulated (FM) screening, has been the most active research field in digital halftoning in recent years. These terms have been loosely applied to both algorithm approaches and the screen approach. Error diffusion [5] is the algorithm approach that has been most extensively studied, whereas the Blue Noise Mask (BNM) [6, 7] is the term first applied to a screen or threshold array that produces unstructured, visually appealing halftone patterns. In order to follow a precise definition from now on, the term "stochastic screening" applies to a threshold array. Also, "mask" and "screen" will be used interchangeably when both will refer to a threshold array.

1.2.1. Error Diffusion

Error diffusion is an adaptive algorithm that produces patterns with different spatial frequency content de-



FIG. 3. Flowchart for standard error diffusion.

pending on the input image value. It forces average tone content to remain the same and attempts to localize the distribution of tone levels. Figure 3 shows the flowchart for error diffusion. This approach was first presented by Floyd and Steinberg back in the 1970's [5]. Subsequently, many modifications and derivations have been proposed in the design of error filter [8], threshold value [9], feedback loop [10], as well as processing sequence [11]. Although all these algorithms require intensive computation and some artifacts exist, the quality of the halftone image, particularly the sharp edges and many image details, is generally considered excellent [12]. The success of error diffusion lies in the fact that it is a "good blue-noise generator," as pointed out by Ulichney [1]. In the academic literature, the nature of noise is often described by a color name; i.e., white noise is so named because of its flat power spectrum. Blue noise, on the other hand, has most of its energy located at high spatial frequencies with very little lowfrequency component. A typical blue-noise radial average power spectrum (RAPS) is shown in Fig. 4. Patterns with blue-noise characteristics generally enjoy the benefits of aperiodic uncorrelated dot patterns without low-frequency graininess.

1.2.2. Stochastic Screen

Stochastic screen halftoning is the subject of active research. It combines the simplicity of ordered dither with the



FIG. 2. Halftoned image from clustered-dot halftoning (left) and dispersed-dot halftoning (right).



FIG. 4. A blue-noise radial average power spectrum.

blue-noise quality of error diffusion (see Fig. 5). Stochastic screen halftoning is a point comparison process, so it is easy to implement. Thus, devices currently using ordered dither technique may be switched to stochastic screen halftoning simply by replacing the original dither array with a stochastic screen. The halftone image from a stochastic screen will have the typical visually pleasing blue-noise characteristics, which is guaranteed when screens are generated from blue-noise dot patterns of individual gray levels. The Blue Noise Mask, proposed by Mitsa and Parker, was the first stochastic screen to realize the above scheme [6, 7, 13].

The following sections will concentrate on the design of stochastic screens and their applications in black-and-white halftoning, multitoning, and color halftoning. Our review focuses on the scientific literature published in peer-reviewed forums. The organization is the following: In Section 2, the construction of the prototypical stochastic screen, the Blue Noise Mask, is outlined. Section 3 details the common filter approaches in screen construction, and various filter design techniques are examined. In Section 4, the optimality of blue-noise binary pattern in terms of screen design is pursued. In Section 5, various modifications of stochastic screens in order to meet special applications are introduced, such as screens with dot-gain compen-

sation and screens for multilevel-output devices. In Section 6, color halftoning is investigated, and different stochastic color halftone schemes are presented, followed by an evaluation based on a human visual model. Finally in Section 7, a summary is given, and current problems with stochastic screen halftoning are identified and future research is proposed.

2. THE CONSTRUCTION OF A STOCHASTIC SCREEN—BLUE NOISE MASK

In this section, the algorithm [6, 7, 13, 14] to generate a Blue Noise Mask is presented. First, an initial blue-noise binary pattern b[i, j, g] (two-dimensional binary pattern at gray level g) for some intermediate level g (0 < g < 255, assuming an 8-bit mask) is required. Using the filtering and swapping technique presented in Section 3, such a pattern with a blue-noise characteristics is obtained and used as the initial pattern. From this initial pattern, an initial mask m[i, j] is generated, which when used to halftone the constant gray image of level g, produces the initial binary pattern b[i, j, g].

Once level g is completed, level g + 1 is processed (Fig. 6). For this level, the blue-noise pattern is created by converting the appropriate number (the total number of pixels in the binary pattern divide by the total number of levels) of 0's to 1's in the previous pattern g. At the same time, the mask m[i, j] is updated. This process is repeated until the mask has been updated for all the levels above g to level 255. Analogous procedures are used to construct the mask for all the levels below g to level 0. The resulting two-dimensional array m[i, j] will be the final Blue Noise Mask (Fig. 7).

There is a significant constraint on the converting and swapping operation in this mask construction. In making a mask, the binary patterns at different levels are dependent. For example, in the upward construction process, all the 1's in the binary pattern for level g are contained in the binary pattern g + 1, so when converting and swapping



FIG. 5. Halftoned image from error diffusion (left) and Blue Noise Mask (right).

one black is turned to white



FIG. 6. Blue Noise Mask construction: from level g to g + 1.

1's and 0's, these common 1's shared by the two neighboring levels cannot be changed.

The construction technique outlined above is quite general and has enabled the generation of masks with different properties such as 8-bit depth (level 0–255) and 12-bit depth (level 0–4095), small size (64 by 64) and large size (256 by 256), isotropic and anisotropic.

There are two critical parts in designing a Blue Noise Mask: the digital filters and the optimality issue. These topics will be covered in Sections 3 and 4, respectively.

3. COMMON FILTER APPROACHES

As described in the previous section, the algorithm for generating the Blue Noise Mask recognizes that in a halftone array, the binary pattern at any gray level g + 1 can be thought of as being "built up" from the binary pattern at level g. And furthermore, the Blue Noise Mask algorithm utilizes the concept of filtering a binary pattern from level g to select the location of pixels that would be the "best candidates" for addition of majority pixels required for level g + 1. Once binary patterns for all gray levels have been sequentially produced in this way from some "seed" level, the binary patterns can be summed to produce a threshold array or Blue Noise Mask. The role of filters in this procedure warrants close attention, and a summary of some important results is given below.

Mitsa and Parker [6, 7, 13] selected the location of pixels that should be changed from minority to majority values by finding the extremes of an error function e[i, j]. This error function was generated by directly filtering the binary pattern of level g and then subtracting the binary pattern from the filtered pattern. Specifically, in the image domain

$$e(i,j) = [h_{hp_s}(i,j) * b_g(i,j)] - b_g(i,j),$$
(1)

where $h_{hp_g}(i, j)$ is a high pass filter designed for level g, $b_g(i, j)$ is the binary pattern, and "**" denotes convolution with circular "wrap-around" properties. In the transform (or spatial frequency) domain:

$$E(k, l) = B(k, l)[H_{hp_a}(k, l) - 1].$$
 (2)

Since $H_{hp_g}(k, l)$ is chosen as a blue-noise (highpass) filter, then the overall filter $[H_{hp_g}(k, l) - 1]$ is a lowpass one. Thus, an essential feature of this filtering is that binary pattern "clumps" in the image domain (corresponding to low-frequency energy in the transform domain) can be located by directly filtering the binary pattern. Also, the highpass region (or lowpass region) can be related to the principal frequency f_g [1], which is a function of the gray level g. Mitsa and Parker also suggested that the filter H_{hp_g} could be made adaptive to directly shape the RAPS of the binary pattern for level g. That is,

$$H_{hp_g}(\rho) = \sqrt{D(\rho)}/B_g(\rho), \qquad (3)$$

where $D(\rho)$ is the desired blue-noise RAPS for level g + 1 and $B_g(\rho)$ is the known RAPS for level g. As with any filtering approach, care should be taken to avoid unwanted discontinuities in the transform domain that produce "ring-ing" in the image domain. This approach is depicted in Figs. 8 and 9. Note that this specific filter is computationally more involved than simple lowpass filters, but the approach demonstrates the central requirement of providing a lowpass filter with a cutoff frequency linked to gray level g by the principal frequency f_g [15]:



FIG. 7. Adding up dot patterns of each gray level to make a Blue Noise Mask.



FIG. 8. A desired blue-noise RAPS $D(\rho)$ (solid line) and an actual blue-noise RAPS $B_g(\rho)$ (dotted line) at gray level 210.

$$f_g = \begin{cases} \sqrt{g}, & \text{for } g \le 1/2\\ \sqrt{1-g} & \text{for } g > 1/2, \end{cases}$$
(4)

where g is the gray level normalized to 1. This relationship is plotted as the solid line in Fig. 10. The dashed line in Fig. 10 corresponds to the factor of $1/\sqrt{2}$ (in principal frequency, or $\sqrt{2}$ in average separation) that was discussed by Mitsa and Parker [16] as an empirical choice of transition cutoff frequency for some filters. In a later paper, Parker, Mitsa, and Ulichney [17] demonstrated how, at a single gray level, changes in the filter cutoff frequency could produce different types of final halftone patterns. Specifically, for f_c representing the cutoff frequency of a filter, let $f_c = Kf_g$, where K is an adjustable scaling factor. While K varies from 0.5 to 1.0, binary patterns of different "textures" were produced, with $K = 1/\sqrt{2}$ employed for general use.

Ulichney [18] further explored the filter issue, choosing a two-dimensional Gaussian filter implemented in the image domain for direct operation on the binary pattern. He demonstrated that, for small size arrays (less than 32 by 32, for example), even a single Gaussian filter that is not adjusted for gray levels (as in the previous work) could produce a useful Blue Noise Mask for some applications. Of course, the seed pattern and Gaussian filter width require careful choice in order to produce a desirable result.

In general, it is beneficial to vary the cutoff frequency



FIG. 10. Principal frequency (solid line) and cutoff frequency of the highpass filter (dashed line) for each gray level.

(or Gaussian width parameter) with gray level [14, 16]. Yao and Parker [14] also demonstrated that a variety of lowpass filter shapes could produce desirable halftone patterns, so long as the filter parameters were adjusted for appropriate cutoff with respect to the principal frequency. Mitsa and Brothwarte [19] further developed the concept of a filter bank (for different gray levels), where lowpass filters were each adjusted for a specific cutoff below the principal frequency (K = $\frac{1}{2}$), and this was related to a wavelet-type filter bank. Dalton [20] described the use of bandpass filters to produce textured binary patterns.

Thus, the general lessons from these works are that direct filtering of binary patterns can be useful in selecting pixels that can be changed to produce a desired result in the image domain, and correspondingly approximate a desired power spectrum in the transform domain. To illustrate the varieties of filters that have been used, Figs. 11 and 12 depict the transform domain and corresponding image domain filters from Mitsa and Parker [6, 16], Ulichney [18], and Yao and Parker [14], in each case the filter represented is the one that would be used for gray level 210 out of 256. Only isotropic filters are shown, but anisotropic filters have also been used [14].

Note that the specification of a filter as a lowpass filter in the image domain provides perhaps the easiest way to explain the algorithm to new designers. However, the benefit of specifying the filter in the transform domain and, as an initial highpass operation, is that the final RAPS of



FIG.9. Overall lowpass filter $[H_{hp_g}(k, l) - 1]$ and original highpass filter $H_{hp_g}(k, l)$.



FIG. 11. Different filters in frequency domain.



FIG. 12. Different filters in spatial domain.

the binary pattern will generally approximate the shape of the highpass filter that is specified in Eq. (2) or (3) [14, 16, 17]. Thus, a halftone designer considering the final RAPS of the binary pattern can envision changes resulting from different filter shapes by considering the filter as an initial highpass filter in transform domain as taught in the early references [6, 7, 13, 16].

Modifications to the filters may be further explored. One interesting filter application is the manipulation of the spectral peak at the principal frequency. In filtering the binary pattern with a simple lowpass filter, such as a Gaussian, the low frequencies are minimized by the algorithm which changes certain identified pixels [14]. By careful choice of a bandpass filter, one can enhance selective frequencies in the transform domain, through the proper identification of certain pixels in the image domain.

For example, consider a starting white-noise binary pattern at level 210 shown in Fig. 13. Now apply the following two different filters individually to that pattern:



FIG. 13. Upper: white noise starting pattern; Left: final pattern from filtering and swapping with a Gaussian filter; Right: final pattern from filtering and swapping with a Gaussian filter which has a bandreject component.



FIG. 14. Two filters in frequency domain. A small "dip" is seen near frequency sample 110 in one curve.

1. Gaussian,

$$F(u,v) = e^{-(u^2 + v^2)/2\sigma^2};$$
(5)

2. Gaussian with a band-reject component,

$$F(u,v) = e^{-(u^2 + v^2)/2\sigma^2} - ae^{-((\sqrt{u^2 + v^2} - f_g)^2)/2\sigma'^2}.$$
 (6)

Figure 14 shows the two filters in transform domain, where the parameter "a" of Eq. (6) is set as 0.02 for the second filter. Each of these two filters was repeatedly applied to the starting pattern with pixel swapping until the perceived mean square error (MSE) between the binary pattern and its corresponding uniform gray pattern stopped decreasing. The resulting patterns are given in the bottom part of Fig. 13 and the corresponding RAPS in Fig. 15.

Thus, selective enhancement (or suppression) of regions of the power spectrum can be designed by proper selection of filters. Note also that the difference in lowpass filters (Fig. 14) is very subtle but the resulting binary patterns have a recognizably different RAPS. The examples given here demonstrate the utility in designing binary patterns with particular characteristics.

The properties of the human visual system can also be incorporated into the filter. Important research combining the power spectrum and human visual system (HVS) concepts with binary patterns was reported by Sullivan, Ray, and Miller [10]. Utilizing a model of the low-contrast phot-



FIG. 15. RAPS for the two final blue-noise patterns.



FIG. 16. Spatial Frequency Response of the human visual model by Sullivan *et al.* [10].

opic modulation transfer function as illustrated in Fig. 16, they were able to generate a locally unstructured "tileable" binary pattern, 32 by 32 square, for each gray level, and they used a cost function with HVS weighting to guide a Monte-Carlo approach with simulated annealing in the creation of individual binary patterns. As a general rule, the human eye has maximum sensitivity in the horizontal and vertical directions and minimum sensitivity in the diagonal direction. Therefore, an anisotropic filter may be designed so that the resulting pattern will have more energy in the diagonal direction than in the horizontal and vertical directions. For example, an original two-dimensional Gaussian filter F(u, v) may be modified in the following way:

$$F'(u, v) = [1 + c * \cos(4\theta)] * F(u, v), \tag{7}$$

where θ is the central angle around the DC point and c is a constant between 0 and 1 which controls the amount of energy distribution.

4. OPTIMALITY OF BLUE-NOISE BINARY PATTERN AND STOCHASTIC SCREEN

The filtering techniques presented in Section 3 produce visually pleasing blue-noise binary patterns. However, are these binary patterns also optimal for their corresponding gray levels and for the construction of a mask? These questions will be addressed in this section.

Yao [21] has given a detailed mathematical analysis of Blue Noise Mask construction based on a human visual model (similar to the model used by Sullivan *et al.*), which provides insights to the filtering process and also prescribes the locations of the dots that, when swapped, will result in a binary pattern with minimum perceived MSE with respect to the original uniform gray pattern. The analysis of the filtering technique put a lower bound on the achievable perceived MSE, assuming that a filter based on the human visual system is also used to measure the perceived MSE between the gray and binary patterns. As Yao pointed out, the difference between the local filtered output of the largest white clump and the largest black clump must be greater than a certain value T in order for the perceived MSE to be further reduced. T is given by

$$T = 1/4\lambda\sigma^2,\tag{8}$$

where σ is the standard deviation of a Gaussian filter based on a human visual model.

In another way of speaking, a nonzero lower limit on the perceived MSE will be reached when the filtering technique is employed.

To exceed this limit, a postfiltering algorithm [22] is introduced below. By locally enforcing a vector process after filtering, perceived MSE is further reduced and more visually pleasing binary patterns are obtained. This new algorithm will be presented first. Then, a series of binary patterns for a certain gray level are generated varying from a white noise pattern to a highly structured pattern. By studying these patterns in both image and transform domain, the optimality issue is investigated.

4.1. Electrostatic Force Algorithm

This force algorithm is based on the model of electrostatic force between electric charges. Point charges of same polarity repel each other while charges of different polarity attract. The force is proportional to $1/r^2$, where r is the distance between the two charges. Therefore, in case of a binary pattern whose pixel values are either 1 or 0, if all the majority pixels are treated as point charges with "+" polarity and all the minority pixels are treated as point charges with "-" polarity, then there will be interactive force between all the pixels. Furthermore, if those minority pixels are allowed to move freely under the net electrostatic force from pixel charges in a certain neighborhood W, after some number of iterations, a nearly homogeneous distribution of those minority pixels should be expected. Since every pixel has some force acting on it, a threshold value should be set such that only when the net force on a pixel surpasses the threshold value, a pixel move in the direction which would minimize the net forces on that minority "charge." Remember that the starting pattern for this force algorithm is the one we obtain from the filtering process, which is already free of clumps. Therefore, the histogram of the net force (either in horizontal or vertical direction) on every minority pixel will be highly peaked near 0. In this case, it is reasonable to assume the mean value of net force on every minority pixel to be 0, and



FIG. 17. Perceived MSE vs iteration in filtering and swapping process.

the threshold value (TH) can be related to the standard deviation (SD) of the net force as

$$TH = V \cdot SD. \tag{9}$$

V is a variable that is adaptive to the gray level as well as the iteration number. As binary patterns are two-dimensional, the force calculation and pixel movement must be done in the horizontal and vertical directions, respectively.

A different force-relaxation model for adaptive halftoning of images was proposed by Eschbach and Hauk [23].

4.2. A Progressive Series of Binary Patterns

To illustrate the previous procedure, a white-noise pattern at level 245 is used as the initial pattern, then the transform-domain filtering is applied Fig. 17 shows the perceived MSE drop versus iteration number and Fig. 18 shows the difference between the largest white clump and the largest black clump (DWB) for each iteration. Since the initial pattern is a white-noise one, the DWB is quite large and the perceived MSE decreases in each iteration. After a certain number of iterations, the DWB approaches the lower bound T set in Eq. (8) (in this case approximately 0.0137), then the filtering process can no longer improve the binary pattern (Fig. 19). Figure 20 shows the binary pattern (P2) obtained from the filtering process with perceived MSE of 0.263.

From this pattern (P2), the force algorithm is carried out. The neighborhood W, which is used to calculate the net force on each pixel, is set as 13 by 13 and the starting value of V is set as 1.5. Figure 20 shows the binary pattern (P3) after just five iterations with perceived MSE of 0.165.

It is quite obvious that by locally enforcing the vector



FIG. 18. Difference between the largest white clump and the largest black clump of filter output vs. iteration in the filtering and swapping process.



FIG. 19. Electrostatic force model.

force process, the perceived MSE are further reduced and more uniform patterns are generated.

4.3. Optimality Issue

Without strict proof here, it is noted that the force algorithm does converge after further iterations. Figure 20 shows the final pattern (P4) obtained when the force algorithm converges after 75 iterations. The perceived MSE of P4 is 0.087 and this pattern has a highly ordered structure.

Since all the binary patterns in a mask are constructed



FIG. 20. Patterns P1 to P4 (clockwise from upper left corner).



FIG. 21. RAPS of patterns P1, P2, P3, and P4.

from a "seed" pattern, the choice of "seed" pattern should be based on its suitability for mask generation. In another way of speaking, an optimal binary pattern should not degrade the quality of its neighbor levels (g + 1 and g - 1 and so forth for one level g).

Obviously, a white-noise pattern cannot be optimal. However, the highly structured pattern is not optimal either. If this highly structured pattern is used as an initial pattern and neighboring levels are constructed, those neighboring binary patterns are generally visually annoying due to noticeable disruption of the semi-regular patterns established by the specific initial pattern. This leads to the question: What pattern between white noise and highly structured patterns constitutes an optimal bluenoise "seed" pattern for Blue Noise Mask construction?

Figure 21 shows the RAPS of all patterns presented in Figure 20. P1 has the typical white-noise characteristics, P2 and P3 have the typical blue-noise characteristics, and P4 has a very high peak at the principal frequency with an emerging "second harmonic" peak. The trend is very obvious that spectrum starts from white-noise shape, gradually switches to blue-noise shape, and ends up in a shape with energy concentrated around the principal frequency. Therefore, in carrying out the force algorithm, a criterion should be set that will enable the computer to terminate the process once an excessive concentration of energy at principal frequency emerges.

Another experiment has been carried out as follows. For an intermediate gray level (level 245), another series of binary patterns are generated with the force algorithm. In this case, pattern 0 corresponds to the output from the filtering technique, pattern 5 is the output after five iterations of the force algorithm on pattern 1 and pattern 50 is the output after 50 iterations of the force algorithm on pattern 1. Then for each of these three patterns, its neighboring gray levels are constructed. Using the HVS model by Sullivan *et al.* (Section 3), a plot of the perceived MSE between the binary pattern and corresponding uniform gray pattern for each gray level is generated, respec-



FIG. 22. MSE vs gray level (partial) for Blue Noise Masks constructed from three "seed" patterns of different blue-noise RAPS characteristics.

tively. These plots are shown in Fig. 22. The plots for pattern 0 and pattern 50 show very large discontinuities around the initial level (245), which means that they are not optimal for mask construction. Therefore, the smoothness of the perceived MSE transition could serve as a parameter to design an optimal binary pattern.

5. SPECIAL APPLICATIONS

So far, all the discussions have considered the design of an ideal stochastic screen for an ideal device. However, since real printers and displays are not ideal, a special screen can be designed to meet individual application requirement. Although these requirements could be met with different pre/post processing techniques, by incorporating the device characteristics into screen design, both rendering time and memory can be reduced.

5.1. Dot-Gain Compensation

In digital printing, one major concern is dot gain, which can be attributed to ink spread or dot overlap and usually is a combination of both (Fig. 23). With stochastic screens, isolated and dispersed halftone dots are typically generated, and therefore dot-gain compensation will be necessary. In practice, dot-gain compensation usually is per-



FIG. 23. Dot gain model.



FIG. 24. Printer characteristic curve and lookup table curve for compensation.

formed using lookup tables before halftoning. With stochastic screen halftoning, dot-gain compensation can be actually included in the screen design process [24]. In general, these approaches can be classified into two categories. One is by printing fewer black dots than required in ideal case, and the other is by printing black dots in a preferred way while keeping the number of black dots for each level untouched.

5.1.1. Printing Fewer Black Dots

The nonlinearity of a specific printer can be directly accounted for in the construction of a mask. Gray patches of certain levels are first printed to get the printer inputoutput characteristics curve. Then, a corresponding curve to compensate for this nonlinearity is generated. This curve will show how many dots are actually needed to correctly render a gray level. Thus, instead of converting a pre-set number pairs of 1's and 0's to move up/down one level, a variable number pair of dots are converted according to the compensation curve. Figure 24 shows the printer characteristic curve, an ideal (linear) mask curve and a lookup table curve for a printer.

Sometimes, if the printer characteristic curve is not available during mask design or if several masks have to be designed, a mask with 12-bit depth can be built first instead of the typical 8-bit ones. The mask construction is exactly the same as in Section 2, except that 4096 gray levels are considered. Once the 12-bit mask is completed, it could be easily mapped back to an 8-bit mask when the printer curve is available. Also, with different mapping strategies, many different 8-bit masks might be generated from this 12-bit one in a short period.

5.1.2. Printing Dots in a Preferred Way

The two methods presented above try to reduce dot gain by printing fewer black dots than in an ideal case, i.e., the number of black dots printed corresponds to a lighter level than the desired one, but the desired level is achieved due to ink spread or dot overlap. Another way to look at dot gain is that it is related to the area-to-perimeter ratio of printed dots. The area of paper covered by a dot is measured in pixels, while the perimeter is the total length of travel around the outside of a printed dot. It is easy to see that the smaller this ratio, the bigger the dot gain. For an isolated dot, this ratio is 0.25 assuming each dot has a unit diameter. In clustered-dot dither where the halftone dots in a cell are connected, this ratio is bigger than 0.25, which is the reason that clustered-dot dither generally shows less dot gain. Thus, another approach to reduce dot gain is to increase the area-to-perimeter ratio.

The nonsymmetric mask. Generally speaking, dot gain can be severe for dark gray levels since black dots are the majority. In the construction of a Blue Noise Mask, certain white dots have to be replaced with black ones to go from level g to level g-1. Normally, with a lowpass filter picking up those white dot candidates (as specified in Section 2), connected white dots are more likely to be selected than isolated white ones. Therefore, as the total number of white dots is decreasing, the number of isolated white dots is actually increasing, which leads to a decreased area-toperimeter ratio and a potential dot-gain problem. To avoid this, isolated white dots can be eliminated first as we move to lower levels, while keeping the connected white dots intact until all isolated white dots are removed. With this modified algorithm, dot gain could be reduced to a certain degree. Because the resulting binary patterns at darker levels contain connected white dots in small "clusters," the use of this nonsymmetric mask is analogous to automatically switching the printer to coarser resolution in dark regions where dot gain is a problem, as illustrated in Fig. 25.

The checkmask. In making this mask, a mid-gray checkerboard pattern is generated first where each pixel is replicated to a 2 by 2 dot. This replicated checkerboard is used as the binary pattern for level 128, hence the name "checkmask" for the final mask built from this initial pattern. There are two reasons to use a replicated checkerboard as a starting pattern. First, the area-to-perimeter ratio is increased, compared to a regular blue-noise pattern. Second, it will be suitable to certain printers having difficulty at the highest spatial frequency, i.e., those printers whose modulation transfer functions (MTF) have low cutoff frequency. For an 8-bit mask, a binary pattern at level 128 will have its principal frequency at the longest spatial frequency and most of its energy is actually concentrated around this frequency. A replicated 2 by 2 checkerboard has its energy shifted to lower frequencies, thus avoiding the highest spatial frequency and easing the requirement on a printer. A partial gray ramp of a checkmask is shown in Fig. 26.

5.2. Multitoning Stochastic screen halftoning could easily be generalized for use with devices having multilevel output. Typically,

(top) and a nonsymmetric mask (bottom).

such techniques are referred to as multitoning. Figure 27 shows a generalization of the stochastic screen technique for application to multilevel devices. It can be seen that this is equivalent to the binary implementation, except that a quantization operation replaces the threshold operation.

Assume an input image I(x, y) has p different levels and the output device has q possible levels, also assume the mask M(x, y) is 256 by 256 in size and has 256 levels. Then, the multitoning process can be given by







FIG. 27. Multitoning flowchart.

$$O(x, y) = \text{INT}\left[\frac{I(x, y) + \frac{M(x_m, y_m)}{256}\frac{p-1}{q-1}}{\frac{p-1}{q-1}}\right], \quad (10)$$

where O(x, y) is the output value, $x_m = x \text{ MOD } 256$ and $y_m = y \text{ MOD } 256$, INT() indicates an integer truncation, and MOD stands for the modulation operation.

It can be seen that the screen value is scaled and added to the input image before a simple threshold operation.

Although any stochastic screen designed for black-andwhite halftoning could be used with the above implementation for multitoning, an improvement could be made when screens are optimized specifically for multitoning. Spaulding and Ray [25] have investigated methods that minimize a visual cost function within a quantization interval; they have also reported using nonuniform quantization functions in the multitoning implementation.

6. COLOR HALFTONING

Color imaging normally requires mixing of three additive primary colors (RGB) for CRT display or subtractive primary colors (CMY) for print. Printing technology can also utilize a fourth primary (K) to provide a better black hue, enlarge the color gamut and improve image quality. Additional colors can be added to further enlarge the color gamut. Color halftoning is the process of generating halftone images for the different color planes for a printing or display device. Color image halftoning is significantly more complicated than halftoning for a gray scale image. All the qualities required of blackand-white halftone images apply to color halftone images that are composed of multiple color planes, but, in addition, the interactions between color planes must be precisely controlled.

6.1. Conventional Color Halftoning Approach

In conventional halftoning, the same clustered-dot screen can be used to halftone the C, M, Y, K planes separately to obtain four halftone images, which are then used to control the placing of color on paper. One immediate problem of this scheme is the appearance of Moiré patterns, which are caused by the low-frequency components of the interference of different color planes. When combining periodic signals, such as two color halftone screens of vector frequency f_1 and f_2 , interference produces a "beat" at the vector difference frequency $f_b = f_1 - f_2$ [2]. If the individual color screens were made at the same angle and frequency, any slight spatial frequency modulation due to misregistration in the printing process forms a low-frequency visually objectionable beat. To reduce Moiré patterns, the screens are typically oriented at different angles, usually 30° apart. At 30°-separation, the frequency of the Moiré is about half the screen frequency, thereby producing a high-frequency "rosette"-shaped beat pattern. Color errors caused by misregistration only appear at high frequencies and therefore are not easily detected. To achieve the highest frequency and least visible rosette patterns, it is typical practice to orient cyan at 105°, magenta at 75°, yellow at 0°, and black at 45° (Fig. 28). Because yellow and black have the least and most impact on visual sensitivity, respectively, they are oriented at angles where human eyes are most and least sensitive. Although yellow is at 15° relative to the other color planes, its impact on intercolor Moiré is not objectionable because of low contrast. One recent direction of research on color printing is the introduction of more colors to expand the color gamut [4]. Rotation angles of these colors must be assigned to reduce Moiré patterns. However, there is a limit to the number of angle selections. Therefore, it can be difficult to apply the conventional color halftoning technique to high-fidelity color printing.

Stochastic halftoning, such as error diffusion and stochastic screens, eliminates the Moiré concern completely,



FIG. 28. Screen angle rotation in conventional color halftoning.

since the halftone dots created by these processes are relatively unstructured. This removes the constraints of the rotation angle. Therefore, high-fidelity color printing is more realizable with stochastic halftoning.

6.2. Color Halftoning Using Error Diffusion

One characteristic of error diffusion is that it produces correlated dot patterns. Therefore, if error diffusion is applied to the individual color planes (so-called "scalar error diffusion"), the halftone images for different color planes will be highly correlated. To improve the visual quality of color halftoning using error diffusion, Miller and Sullivan [26] process the color image in a vector color space with each image pixel treated as a color vector. This technique is called "vector error diffusion." The color image is first converted to a nonseparable color space, and each pixel is assigned with the closest halftone color. The resulting vector halftone error is distributed to neighboring pixels in the same manner as scalar error diffusion.

Klassen *et al.* [27] also proposed a vector error diffusion technique that minimized the visibility of color halftone noise. It is a well-known property of the human visual system that the contrast sensitivity decreases rapidly with increasing spatial frequency. Thus the minimum threshold above which patterns are visible rises rapidly with increasing spatial frequency. One approach to increase spatial frequency, so as to minimize the visibility of color halftone noise, is to select low-contrast color combinations wherever possible and to generate the finest possible "mosaic" pattern without large clumps. Therefore, light gray is printed using nonoverlapping cyan, magenta, yellow, and unprinted white pixels, as opposed to printing occasional black clusters on a large white background. This concept of utilizing the finest possible patterns also serves as a fundamental rule for designing schemes in using stochastic screens for color halftoning.

6.3. Color Halftoning Using Stochastic Screens

The following schemes have been proposed to apply the Blue Noise Masks to halftone color images [28].

6.3.1. The Dot-on-Dot Scheme

The simplest application of the mask utilizes the same mask for each color plane; this is known as the dot-on-dot technique. Although this approach is easiest to implement, it is rarely used in practice because it results in the highest level of luminance modulation and the output is most sensitive to misregistration.

Suppose a gray patch is to be rendered with equal levels of cyan, magenta, yellow. With this dot-on-dot approach, the halftone images for each color plane are identical. Therefore, whenever a cyan dot is printed, a magneta dot and a yellow dot also are printed. As a result, the final output patch will be formed with dispersed black dots on the white background. This produces a larger level of luminance modulation than other cases where color dots are not coincident, resulting in more visible halftone patterns. Another possible outcome under the same scenario is that, if one color plane is misregistered relative to others, the annoying color "banding" effect will appear with possible color shifts.

6.3.2. The Shifted Mask Scheme

To decrease correlation of the color planes, spatially shifted masks are employed for halftoning each color plane. This will also increase the spatial frequency of the printed dots. For example, one mask can be used for the cyan plane; then it can be shifted in the horizontal and vertical directions in a wrap-around manner and applied to the magenta plane. Similarly, the mask can be shifted by different amounts and applied to yellow and then to black. This technique will tolerate misregistration and, therefore, is more robust for a real printing process. However, if the set of shifts are not chosen carefully, lowfrequency structures may appear when a color pattern is overlapped with its shifted version. Generally, the shift values must be tested first by printing some overlapping gray patches halftoned with the shifted masks and a decision must be made if these shift values are optimal.

6.3.3. The Inverted Mask Scheme

In this strategy, one mask is applied to one color plane and its inverted version is applied to another. Inverting a mask means taking the 255th complement of a mask:

$$m_i[i,j] = 255 - m[i,j]. \tag{11}$$



FIG. 29. A color patch halftoned with different schemes (from left to right, top to bottom: dot-on-dot, shift, invert, four-mask, and error diffusion).

In light regions, this scheme results in the nonoverlapping arrangement of color dots with high spatial frequency. However, this scheme is only applicable for two color planes (typically cyan and magenta), and some other scheme has to be used to determine other color planes.

6.3.4. The Four-Mask Scheme

This scheme is actually an extension of the inverted technique. It is based on the same idea: increasing the spatial frequency of the printing dots and minimizing the low-frequency energy introduced by the overlapping of color planes. Unlike the inverted technique, this scheme will not limit the number of masks that can be used, so it is more appropriate for generating high-fidelity color halftone images (Fig. 29).

In the case of the four CMYK color planes, four anticorrelated masks are generated from four mutually exclusive seed patterns. The steps are given below:

1. Generate one Blue Noise Mask that produces visually pleasing unstructured binary patterns.

2. Assuming that the mask values are from 0 to 255, for all the mask locations that have a value in the interval 0 to 63, a binary pattern is defined by setting the pixels corresponding to these locations to black dots, while the remaining pixels are set to white.

3. Three more binary patterns are made in the same way by picking the location of pixels with values in the range 64–127, 128–191, and 192–255.

4. This construction ensures that these binary patterns exhibit blue-noise characteristics. The filtering and swapping technique can be further used to eliminate any residual periodic structures.

5. These four binary patterns are used as initial patterns to generate four masks.

When these four masks are applied to different color planes, they generate color halftone dots that are maximally dispersed, therefore achieving the highest spatial frequency, especially at areas of highlight levels.

6.3.5. Evaluation Using a Human Visual Model in CIELAB Space

In this part, a human visual model is used to evaluate the effects of the four schemes presented above on the luminance and chrominance of a color test patch [29]. Figure 30 shows the block diagram of this evaluation; the HVS model by Sullivan *et al.* (Section 3) is used in this case. Both the halftoned and the original color patch are passed through the human visual model and then converted to the CIELAB space. The color difference in CIELAB space is given by



FIG. 30. Flowchart for color halftone schemes evaluation.

$$\Delta E = \sqrt[2]{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2},$$
 (12)

where ΔL^* , Δa^* , Δb^* are corresponding differences between two colors. This color difference can be further broken up into components of luminance error ΔL^* and chrominance error ΔC^* , which is given by

$$\Delta C^* = \sqrt[2]{(\Delta a^*)^2 + (\Delta b^*)^2}.$$
 (13)

Our analysis [29] shows that different perceived errors are produced by different mask techniques. In general, the dot-on-dot scheme results in minimum chrominance error but maximum luminance error and the four-mask scheme results in minimum luminance error but maximum chrominance error, while the result from the shift scheme falls in between.

6.3.6. Adaptive Color Halftoning

Beyond the previous methods, one solution to reduce perceived colorimetric error is to apply two mutually exclusive masks on two color planes first and then to apply an adaptive scheme on other planes. Another advantage of



FIG. 31. Flowchart of adaptive color halftone scheme.



FIG. 32. Details of adaptive decision step.

this adaptive scheme is that color reproduction could be taken into account [30]. Figures 31 and 32 show the flow-chart of this new scheme.

For this method, it is necessary to know the $L^*A^*B^*$ values of the eight primary colors of the destination printer. Thus, solid patches of the primary colors are first printed, then measured, and a lookup table (LUT) of the $L^*A^*B^*$ values for each primary color is generated.

Next, one mask is applied to the cyan plane and its inverted version on the magenta plane. In this way, the highest spatial frequency is achieved. At each image pixel, there are only two possible values for the yellow plane, either 255 or 0. Therefore, only one from two possible primary colors (c_1 and c_2) for that image pixel should be selected. To do that, the corresponding $L^*A^*B^*$ values for c_1 and c_2 as well as the $L^*A^*B^*$ values for original pixel (c_0) have to be identified. Then the distances in CIELAB space between c_1 and c_0 and c_2 and c_0 are calculated, and the primary color which gives the smaller distance is selected. Finally, the luminance and chrominance error between the chosen halftone color and original color are calculated and passed to neighboring pixels in the error diffusion sense.

To compare the performance of this adaptive scheme with other schemes mentioned earlier, a real image is halftoned with all the schemes presented so far and the perceived luminance and chrominance errors of the resulting halftone images are evaluated [30]. Assuming the viewing distance at 10 inches and printer resolution at 300 dpi, the perceived MSE between the original image and each color halftone image for the luminance channel and chrominance channel are calculated. As the results show [30], the lowest colorimetric error (especially luminance error) is achieved using the adaptive scheme. The tradeoff is computational complexity. However, by having one color channel be adaptive, increased flexibility is obtained to manipulate the output so as to reduce colorimetric error while permitting customization to specific printing hardware. It can be seen that the approach is easily extended to black and other high-fidelity color inks.

7. CONCLUSION

The introduction of stochastic screens in recent years has opened up new possibilities for rendering images on printing or display devices with limited output states. The Blue Noise Mask, the prototypical stochastic screen, combines the speed of threshold arrays with the desirable unstructured patterns of error diffusion. Critical topics for analysis and research include the questions of filtering, optimality, color strategies, and application-dependent design. These topics have been discussed in the previous sections. Further research will need to consider refinements to the concept of optimal design for color rendering, and more advanced models of the human visual perception.

REFERENCES

- 1. R. Ulichney, Digital Halftoning, MIT Press, Cambridge, MA, 1987.
- P. G. Roetling and R. P. Loce, Digital halftoning, in *Digital Image Processing Methods* (E. R. Dougherty, Ed.), Chap. 10, pp. 363–413, Dekker, New York, 1994.
- P. R. Jones, Evolution of halftoning technology in the United States patent literature, J. Electron. Imaging 3, No. 3, 1994, 257–275.
- H. Kang, Color Technology For Electronic Imaging Devices, SPIE, Bellingham, WA, 1997.
- 5. R. W. Floyd and L. Steinberg, An adaptive algorithm for spatial greyscale, *Proc. Soc. Inform. Display* **17**, No. 2, 1976, 75–77.
- T. Mitsa and K. J. Parker, Digital halftoning using a blue noise mask, in *ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, Toronto, Canada, pp. 2809–2812, IEEE, Los Alamitos, CA, 1991.
- T. Mitsa and K. J. Parker, "Digital halftoning using a blue-noise mask," *Journal of the Optical Society of America A*, vol. 9, pp. 1920– 1929, Nov. 1992.
- P. W. Wong, "Error diffusion with dynamically adjusted kernel," in *ICASSP-94: 1994 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, (Adelaide, South Australia), pp. 113–116, IEEE, Apr. 1994.
- K. T. Knox, "Threshold modulation in error diffusion on non-standard rasters," in *Proceedings, SPIE—The International Society for Optical Engineering: Human Vision, Visual Processing, and Digital Display V* (B. E. Rogowitz and J. P. Allebach, eds.), vol. 2179 (San Jose, California), pp. 159–169, SPIE, Feb. 1994.
- J. Sullivan, L. Ray, and R. Miller, "Design of minimal visual modulation halftone patterns," *IEEE Transaction on Systems, Man, and Cybernetics*, vol. 21, pp. 33–38, Jan./Feb. 1991.
- G. Marcu and S. Abe, Halftoning by back error compensation, in Proceedings, NIP12: International Conference on Digital Printing Technologies, pp. 132–135.
- K. T. Knox, Error diffusion: A theoretical view, in Allebach and Rogowitz [32, pp. 326–331].
- T. Mitsa and K. J. Parker, Digital halftoning using a blue noise mask, *Image Processing Algorithms and Techniques III, SPIE 1452*, pp. 47–56, 1991.
- M. Yao and K. J. Parker, Modified approach to the construction of a blue noise mask, J. Elec. Imag. 3, No. 1, 1994, 92–97.
- 15. R. A. Ulichney, "Dithering with blue noise," *Proc. IEEE* **76**, 1988, 56–79.
- 16. T. Mitsa and K. J. Parker, Power-spectrum shaping of halftone patterns and its effect on visual appearance, in *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, San Francisco, California*, pp. 193–196, IEEE, Mar. 1992.
- K. Parker, T. Mitsa, and R. Ulichney, A new algorithm for manipulating the power spectrum of halftone patterns, SPSE's 7th Int. Congress on Non-Impact Printing, 1991, pp. 471–475.
- R. Ulichney, The void-and-cluster method for dither array generation, in Allebach and Rogowitz [32, pp. 332–343].
- 19. T. Mitsa and P. Brathwaite, Wavelets as a tool for the construction of a halftone screen, in Rogowitz and Allebach, [31, pp. 228–238].
- 20. J. Dalton, Perception of binary texture and the generation of stochastic halftone screens, in Rogowitz and Allebach [31, pp. 207–220].
- 21. M. Yao, L. Gao, and K. J. Parker, "An analysis of the blue noise mask based on a human visual model," in *Proceedings*, *NIP12: International Conference on Digital Printing Technologies* [34].
- 22. Q. Yu, K. J. Parker, and M. Yao, "Optimality of blue noise mask

binary patterns," in *Proceedings*, *NIP12: International Conference on Digital Printing Technologies* [33, pp. 66–69].

- R. Eschbach and R. Hauck, "A 2-D pulse density modulation by iteration for halftoning," *Optics Communications* 62, 1987, 300–304.
- M. Yao and K. J. Parker, Dot gain compensation in the blue noise mask, in Rogowitz and Allebach [31, pp. 221–227].
- K. Spaulding and L. A. Ray, Method and apparatus for generating a halftone pattern for a multilevel output device, U.S. Patent No. 08/131,801. [Assigned to Eastman Kodak Company]
- R. Miller and J. Sullivan, Color halftoning using error diffusion and a human visual system model, SPSE's 43rd Annual Conference, pp. 149–152, 1990.
- R. V. Klassen, R. Eschbach, and K. Bharat, "Vector error diffusion in a dispersed colour space," in *Proceedings, IS&T's 47th Annual Conference/ICP*, pp. 489–491, IS&T, 1994.
- M. Yao and K. J. Parker, "Color halftoning using blue noise masks," in *Proceedings*, *NIP12: International Conference on Digital Printing Technologies* [34].
- Q. Yu, K. J. Parker, and M. Yao, "Color halftoning with blue noise masks," in *Proceedings, Fourth Color Imaging Conference: Color Science, Systems, and Applications, Scottsdale, AZ*, pp. 77–80, IS&T/SID, Nov. 1996.
- 30. Q. Yu and K. J. Parker, "Adaptive color halftoning for minimum perceived error using the blue noise mask," in *Proceedings, SPIE— The International Society for Optical Engineering: Color Imaging: Device-Independent Color, Color Hard Copy, and Graphic Arts II* (G. B. Beretta and R. Eschbach, eds.), vol. 3018, (San Jose, California), pp. 272–277, SPIE, Feb. 1997.
- B. E. Rogowitz and J. P. Allebach, eds., Proceedings, SPIE—The International Society for Optical Engineering: Human Vision, Visual Processing, and Digital Display VI, vol. 2411, (San Jose, California), SPIE, Feb. 1995.
- J. P. Allebach and B. E. Rogowitz (Eds.), Proceedings, SPIE—The International Society for Optical Engineering: Human Vision, Visual Processing, and Digital Display IV, 1913, SPIE, San Jose, CA, 1993.
- 33. Proceedings, NIP12: International Conference on Digital Printing Technologies, IS&T, San Antonio, TX, 1996.
- Proceedings, IS&T's Eleventh International Congress on Advances in Non-Impact Printing Technologies, IS&T, 1995.



QING YU received the B.S. degree in Physics (1994), *magna cum laude*, from University of Houston and the M.S. degree in Electrical Engineering (1995) from University of Rochester. He is currently a Ph.D. candidate at the Department of Electrical Engineering, University of Rochester. His research interests include digital printing, color image processing, and human visual system.

In the summer of 1996, he was employed by Xerox Corporation as a software engineer in the Office Document Product Division; in the summer of 1997, he was employed by Eastman Kodak Company as a professional intern in the Image Science Division.

Mr. Yu is a member of IEEE and IS&T.



KEVIN J. PARKER received the B.S. degree in engineering science, summa cum laude, from SUNY at Buffalo in 1976. Graduate work in electrical engineering was done at MIT, with M.S. and Ph.D. degrees received in 1978 and 1981. From 1981 to 1985 he was an assistant professor of electrical engineering and radiology. Dr. Parker has received awards from the National Institute of General Medical Sciences (1979), the Lilly Teaching Endowment (1982), the IBM Supercomputing Competition (1989), the World Federation of Ultrasound in Medicine and Biology (1991). He is a member of the IEEE Sonics and Ultrasonics Symposium Technical Committee and serves as reviewer and consultant for a number of journals and institutions. He is also a member of the IEEE, the Acoustical Society of America, and the American Institute of Ultrasound in Medicine. He has been named a fellow in both the IEEE and the AIUM for his work in medical imaging. In addition, he was recently named to the Board of Governors of the AIUM. Dr. Parker's research interests are in medical imaging, linear and nonlinear acoustics, and digital halftoning.