

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 been translated directly to digital halftoning. Some excellent advantages in quality color rendering. We address some theoretical issues, design and optimality issues, printer 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

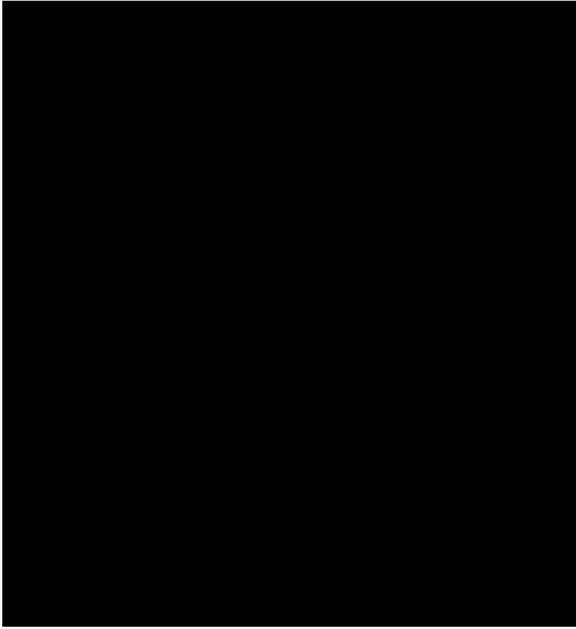


FIG. 1. Ordered dither halftoning technique.

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-

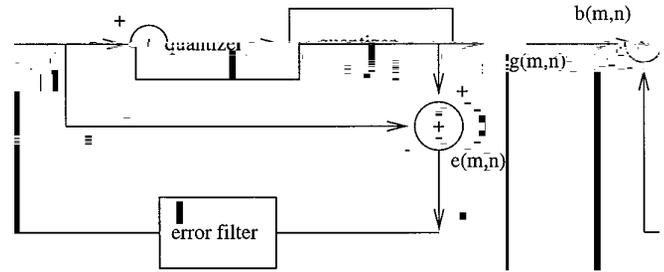


FIG. 3. Flowchart for standard error diffusion.

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 low-frequency 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).

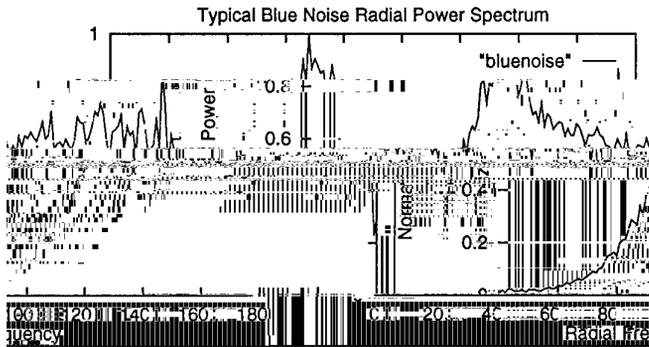


FIG. 4. A blue-noise radial average power spectrum.

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, multitone, 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

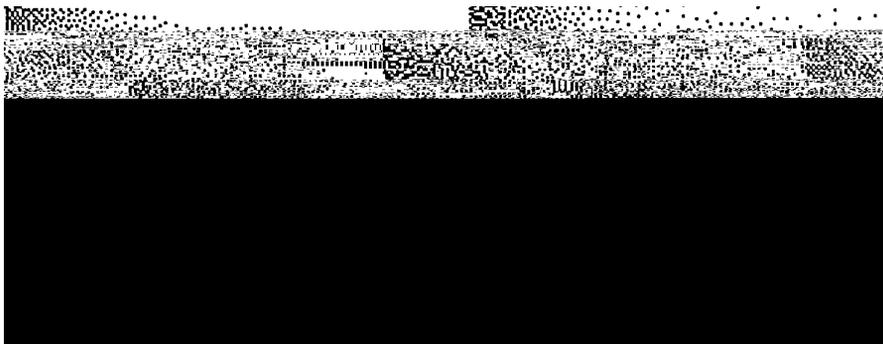


FIG. 5. Halftoned image from error diffusion (left) and Blue Noise Mask (right).

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 convolu-

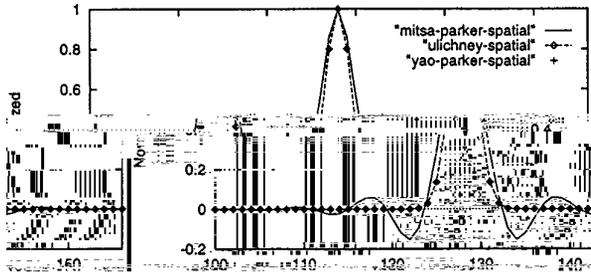


FIG. 12. Different filters in spatial domain.

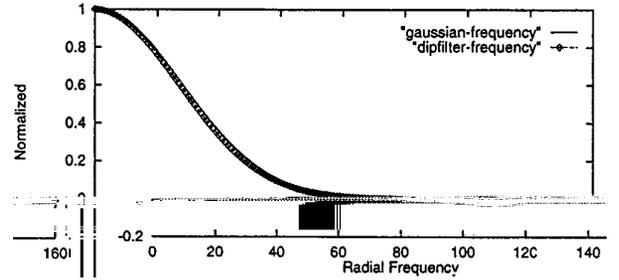


FIG. 14. Two filters in frequency domain. A small "dip" is seen near frequency sample 110 in one curve.

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

1. Gaussian,

$$F(u, v) \approx e^{-(u^2 + v^2)/2s^2}; \quad (5)$$

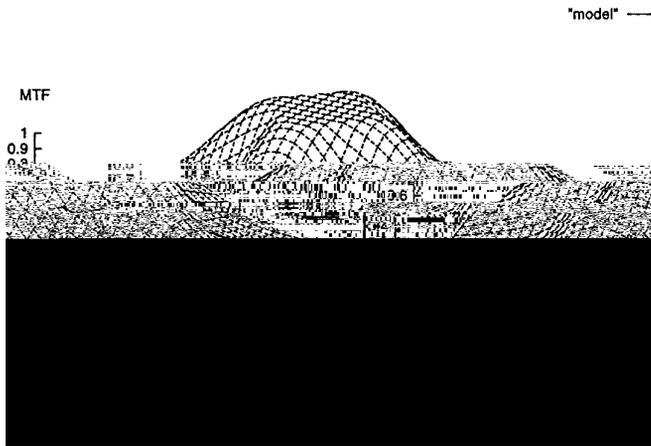


FIG. 16. Spatial Frequency Response of the human visual model by Sullivan et al. [10].

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

respect to the original uniform gray pattern. The analysis of the dithering technique put a lower bound on the achievable perceived MSE, assuming that a dither 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 dithered 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

$$T \geq 1/4s^2, \quad (8)$$

where s is the standard deviation of a Gaussian dither based on a human visual model.

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

To exceed this limit, a post-dithering algorithm [22] is introduced below. By locally enforcing a vector process after dithering, perceived MSE is further reduced and more visually pleasing binary patterns are obtained. This new algorithm will be presented next. Then, a series of binary

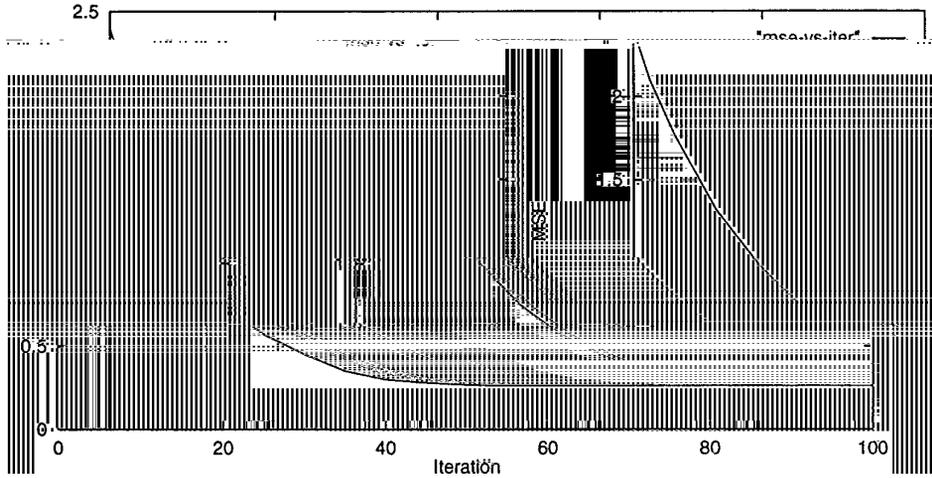


FIG. 17. Perceived MSE vs iteration in filtering and swapping process.

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 6 iterations with perceived MSE of 0.165.

It is quite obvious that by locally enforcing the vector

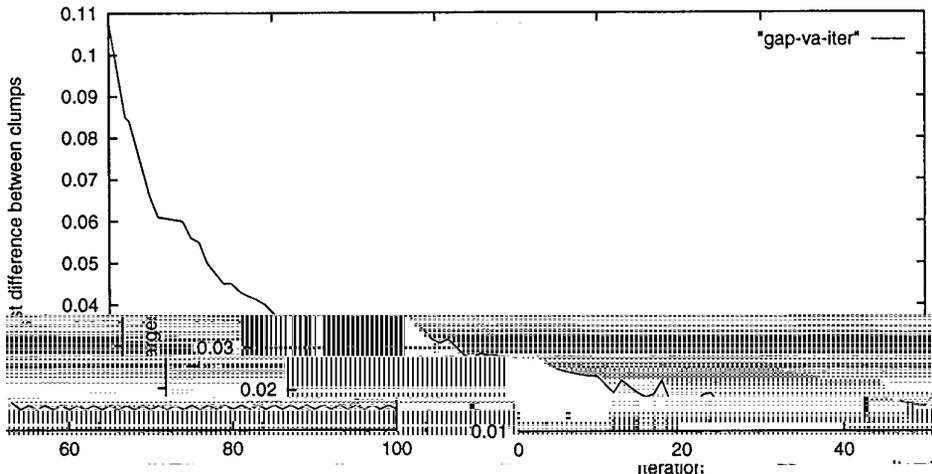


FIG. 18. Difference between the largest white clump and the largest black clump of filter output vs. iteration in the filtering and swapping process.

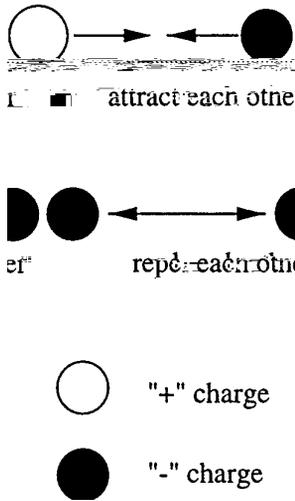


FIG. 19. Electrostatic force model.

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

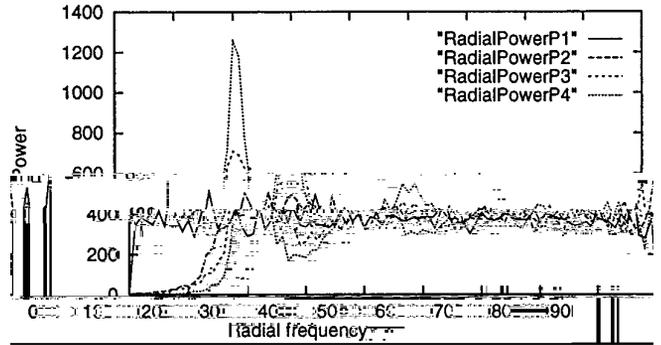


FIG. 21. RAPS of patterns P1, P2, P3, and P4.

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 ($g \pm 1$ and $g \pm 2$ 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

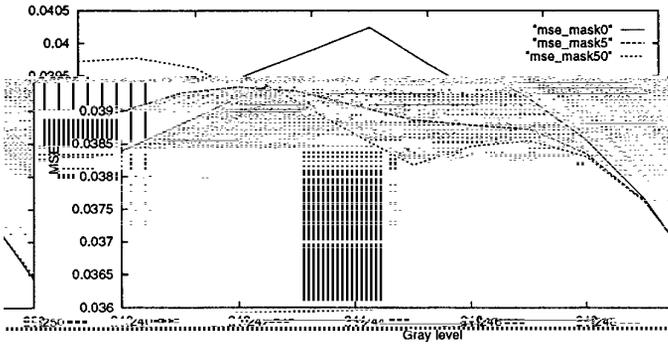


FIG. 22. MSE vs gray level (partial) for Blue Noise Masks constructed from three "seed" patterns of different blue-noise RAPS characteristics.

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-

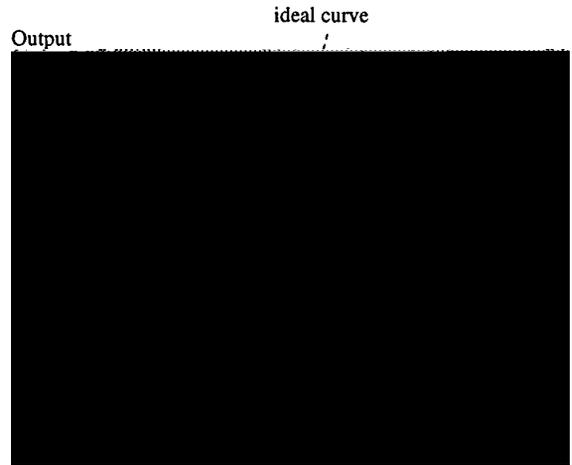


FIG. 24. Printer characteristic curve and lookup table curve for compensation.

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.

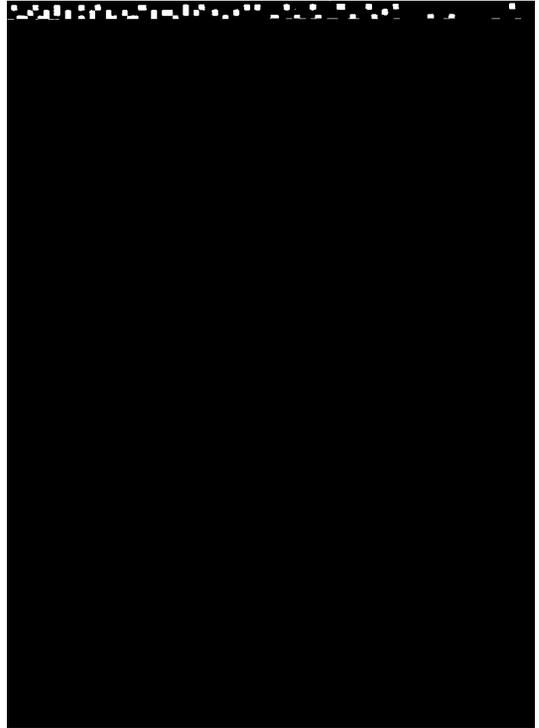
5.1.1. Printing Fewer Black Dots

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 input-output 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 810(12-ne190E85e)-420oical6e,ax-400(tyTd [(t

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



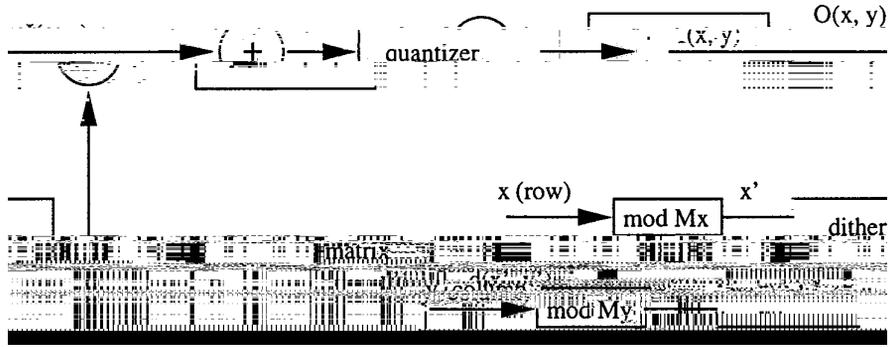


FIG. 27. Multitone flowchart.

$$O(x, y) = \text{INT} \left\{ \frac{x + y + \frac{M(x_m, y_m) p \geq 1}{256} + \frac{q \geq 1}{2} + 1}{p \geq 1} \right\} \quad (10)$$

where $O(x, y)$ is the output value, $x_m = x \text{ MOD } 256$ and $y_m = y \text{ MOD } 256$, $\text{INT}(\cdot)$ 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-and-white halftone could be used with the above implementation for multitone, an improvement could be made when

6.1. Conventional Color Halftone Approach

In conventional halftone, 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

This concept of utilizing the most possible patterns also serves as a fundamental rule for designing schemes in using



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

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 maxi-

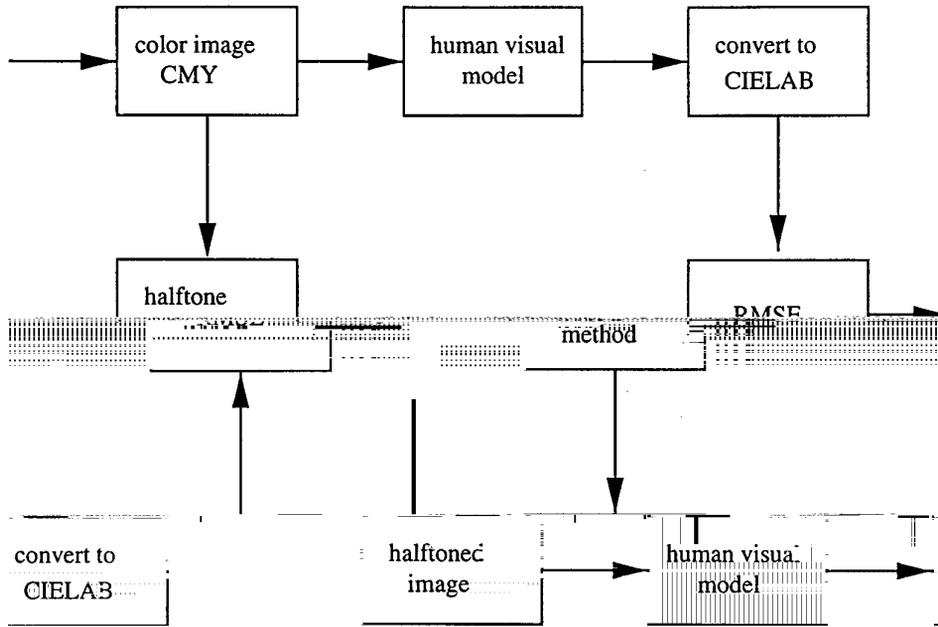


FIG. 30. Flowchart for color halftone schemes evaluation.

$$DE = \sqrt{(DL^*)^2 + (Da^*)^2 + (Db^*)^2}, \quad (12)$$

where DL^* , Da^* , Db^* are corresponding differences between two colors. This color difference can be further broken up into components of luminance error DL^* and chrominance error DC^* , which is given by

$$DC^* = \sqrt{(Da^*)^2 + (Db^*)^2}. \quad (13)$$

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

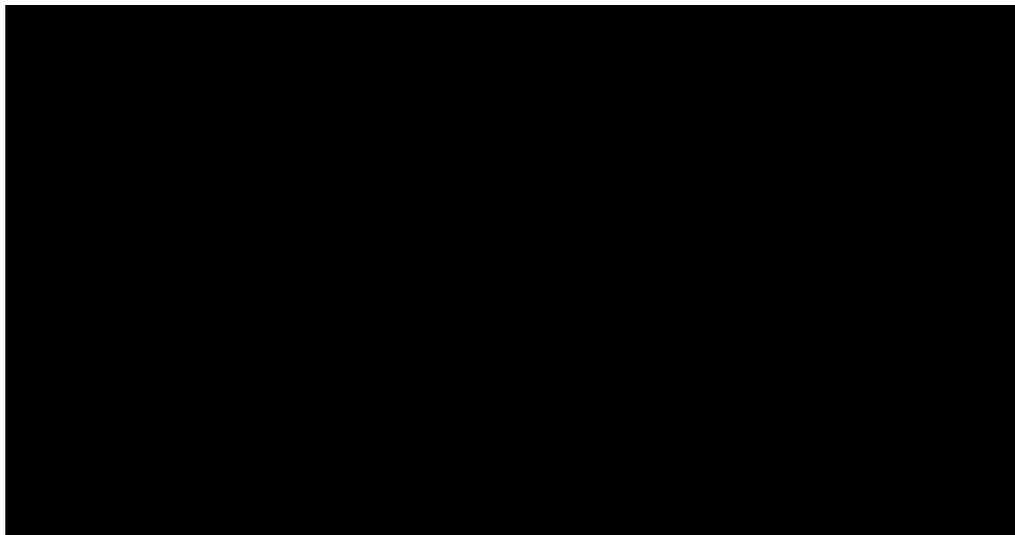


FIG. 31. Flowchart of adaptive color halftone scheme.

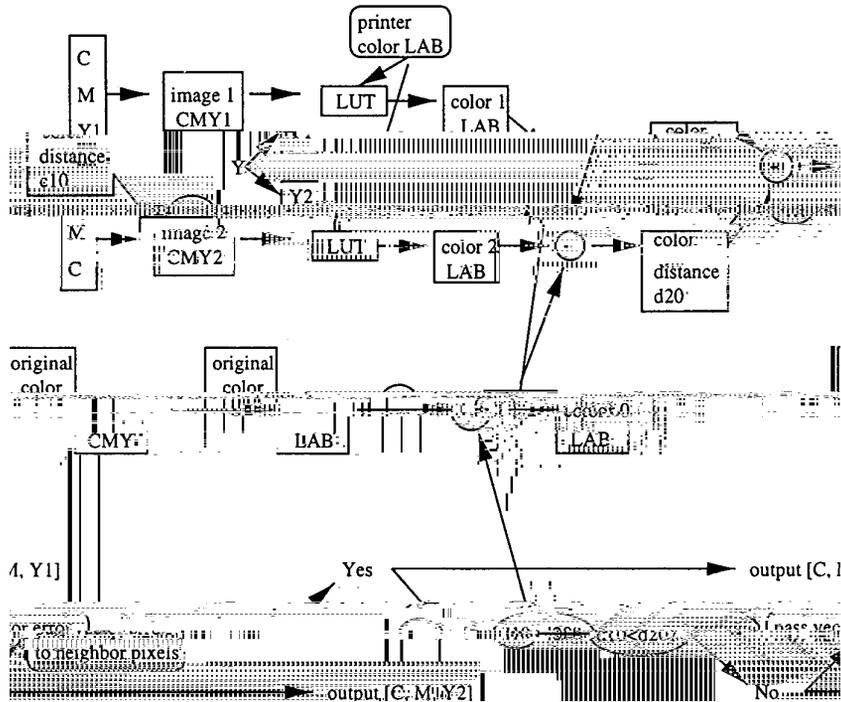


FIG. 32. Details of adaptive decision step.

this adaptive scheme is that color reproduction could be taken into account [30]. Assuming the viewing distance at 10 inches and printer resolution at 300 dpi, the perceived MSE between the

REFERENCES

1. R. Ulichney, Digital Halftoning , MIT Press, Cambridge, MA, 1987.
2. 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.
3. P. R. Jones, Evolution of halftoning technology in the United States (1925-1965), JEB (R) 350 (Page 3), No. (Digit 1992 2867-129) chap. 432i5-0187Tation, nates1.pat16-1 for L. R. 13425 Eschba (R) 3750 (Envis CE
4. R. Eschbach and R. Hauck, "A 2-D pulse density modulation by iteration for halftoning," Optics Communications **62**, 1987, 300±304.
5. M. Yao and K. J. Parker, Dot gain compensation in the blue noise mask, in Rogowitz and Allebach [31, pp. 221±227].
6. K. Spaulding and L. A. Ray, Method and apparatus for generating binary patterns," in Proceedings, NIP12: International Conference on Digital Printing Technologies [33, pp. 66±69].

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