Properties of Improved Dot Diffusion for Image Halftoning

Murat Mese▲ and P. P. Vaidyanathan

Department of Electrical Engineering, California Institute of Technology, Pasadena, California

The dot diffusion method for digital halftoning has the advantage of parallelism unlike the error diffusion method. The method was improved by optimization of the so-called class matrix so that the resulting halftones are comparable to the error diffused halftones. First, 8×8 class matrices were used for dot diffusion method. However, there is a problem with this size of class matrix: enhancement of images is necessary before halftoning. It was found later that if the size of the class matrix was increased to 16×16, then there is no need for the enhancement step. In this article, we will review the dot diffusion. This will be followed by a discussion on special cases of dot diffusion. Then, we will show how the optimization is done to get the class matrices.

Journal of Imaging Science and Technology 45: 291–296 (2001)

Introduction

Digital halftoning is the rendition of continuous-tone pictures on displays that are capable of producing only two levels. There are many methods for digital halftoning,[1] such as ordered dither, error diffusion, dot diffusion, and direct binary search (DBS). Ordered dithering is a thresholding of the continuous-tone image with a spatially periodic screen. In error diffusion, the error is 'diffused' to the unprocessed neighbor points.[2] The dot diffusion method for halftoning introduced by Knuth[3] is an attractive method that attempts to retain the good features of error diffusion while offering substantial parallelism. This method will be reviewed in the section entitled "Dot Diffusion". The method was improved by optimization of the so-called class matrix by Mese and Vaidyanathan[4] and inverse halftoning algorithms for dot-diffused images were proposed in Ref. 5.*

A mathematical description of dot diffusion algorithm was also given in Ref. 6. In Knuth's original method, the images are enhanced with the help of a high pass filter before dot diffusion.[3] The enhancement step was found to be necessary in order to reduce artificial periodicity patterns in the dot diffused image. But the enhancement step also results in very noticeable sharpening of edges. It was shown later[7] that enhancement is unnecessary if the size of the class matrix is increased sufficiently. We shall elaborate on this in the section entitled "Dot Diffusion without enhancement".

In this article, the description of dot diffusion is reviewed in the "Dot Diffusion" section. In the section entitled "Special cases of Dot Diffusion", we will show that error diffusion and binary quantization are special cases of dot diffusion. The optimization without the enhancement step is discussed in the "Optimization of the class matrix" section. Then the problems with the enhancement step will be pointed out and the 16×16 class matrix is optimized for dot diffusion without enhancement in the section entitled "Dot Diffusion without enhancement". A preliminary version of this article was presented in Ref. 7.

Dot Diffusion

The dot diffusion method for halftoning has only one design parameter, called the class matrix C. It determined the order in which the pixels are halftoned. To be specific the pixel positions (n_1, n_2) of an image are divided into IJ classes according to $(n_1 \bmod I, n_2 \bmod J)$ where i and j are constant integers. Let $x(n_1,n_2)$ be the contone image with pixel values in the normalized range [0,1]. Starting from class $k = 1$, we process the pixels for increasing values of k. For a fixed k, we take all pixel locations (n_1,n_2) belonging to class k and define the halftone pixels to be

$$h(n_1,n_2) = \begin{cases} 1 & \text{if } x(n_1,n_2) \geq 0.5 \\ 0 & \text{if } x(n_1,n_2) < 0.5 \end{cases} \quad (1)$$

We also define the error $e(n_1,n_2) = x(n_1,n_2) - h(n_1,n_2)$. We then look at the eight neighbors of (n_1,n_2) and replace the contone pixel with an adjusted version for those neighbors which have a higher class number, i.e., those neighbors that have not been halftoned yet). To be specific, neighbors with higher class numbers are replaced with $x(n_1,n_2) + 2e(n_1,n_2)/w$ for orthogonal neighbors and $x(n_1,n_2) + e(n_1,n_2)/w$ for diagonal neighbors where w is such that the sum or errors added to all the eight neighbors is exactly $e(n_1,n_2)$. The extra

Original manuscript received August 1, 2000

▲ IS&T Member

* Class matrix is defined in the "Dot Diffusion" section.

TABLE I. A Part of the Class Matrix (Defined in Eq. 2) Which Is Not Close to the Boundaries of the Class Matrix

…	$y(n_1-2)+n_2-1$	$y(n_1-2)+n_2$	$y(n_1-2)+n_2+1$	…
…	$y(n_1-1)+n_2-1$	$\mathbf{y(n_1-1)+n_2}$	$y(n_1-1)+n_2+1$	…
…	yn_1+n_2-1	yn_1+n_2	yn_1+n_2+1	…

factor of two for orthogonal neighbors, i.e., vertically and horizontally adjacent neighbors, because vertically or horizontally oriented error patterns are more perceptible than diagonal patterns. The contone pixels $x(n_1,n_2)$ which have the next class number $k+1$ are then similarly processed. The pixel values $x(n_1,n_2)$ are of course not the original contone values but the adjusted values according to earlier diffusion steps. (2) When the algorithm terminates, the signal $h(n_1,n_2)$ is the desired halftone. Notice that the pixels in the same class can be processed simultaneously.

Usually an image is enhanced[3] before dot diffusion is applied. For this the continuous image pixels ($C(i,j)$ are replaced by

$$C'(i,j)=\frac{C(i,j)-\alpha\overline{C}(i,j)}{1-\alpha} \quad \text{where} \quad \overline{C}(i,j)=\frac{\sum_{u=i-1}^{i+1}\sum_{v=j-1}^{j+1}C(u,v)}{9}.$$

Here the parameter α determines the degree of enhancement. If $\alpha = 0$, there is not enhancement, and the enhancement increases as α increases. If $\alpha = 0.9$ then the enhancement filter can be further simplified.[3]

Special Cases of Dot Diffusion

Commonly used sizes for the class matrix are 8×8 and 16×16. A trivial special case of the dot diffusion algorithm arises when the class matrix is of size 1×1. In this case the algorithm reduces to direct pixel by pixel binary quantization.

Next, what happens when the class matrix is made arbitrarily large? Assume that the image is of size $x \times y$ and let the class matrix of size $x \times y$ be defined as follows:

$$C(n_1,n_2) = (n_1-1)y + n_2 \text{ for } n_1 = 1, \ldots, x,\ n_2 = 1, \ldots, y. \quad (2)$$

Because of the structure of this class matrix, the pixels are processed in raster scan order the same way the pixels are processed in error diffusion. In Table I we have shown a part of the class matrix which is not close to the boundaries of the class matrix. In step $(n_1-1)y + n_2$, the error due to quantization at (n_1,n_2) is diffused only to the pixels at locations (n_1,n_2+1), (n_1+1, n_2-1), (n_1+1, n_2) and (n_1+1, n_2+1). Furthermore, the error will be diffused to the four neighbor pixels with the diffusion coefficients shown in Table II. Thus, the dot diffusion method becomes identical to error diffusion if the class matrix is as large as the image itself, and defined as in Eq. 2. The only difference between this special case and the Floyd-Steinberg error diffusion is in the values of the filter coefficients as summarized in Tables II and III. The filter in Table II is referred to as the DD (dot diffusion) filter to distinguish it from the FS (Floyd-Steinberg) filter in Table III. Using these two sets of filters in error diffusion, we found that the image qualities are nearly identical. There may be slight differences in the implementation complexities (e.g., the denominators in FS filters are powers of two which makes divisions very easy). The main point of this discussion, though, is to make the conceptual connection between dot diffusion and error diffusion.

TABLE II. Diffusion Coefficients for the Case Where Dot Diffusion Reduces to Error Diffusion (Infinite Size Class Matrix). These are called the DD filter coefficients.

	$\times$	$\frac{2}{6}$
$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

TABLE III. Floyd–Steinberg Error Diffusion Coefficients. These are called the FS filter coefficients.

	$\times$	$\frac{7}{16}$
$\frac{3}{16}$	$\frac{5}{16}$	$\frac{1}{16}$

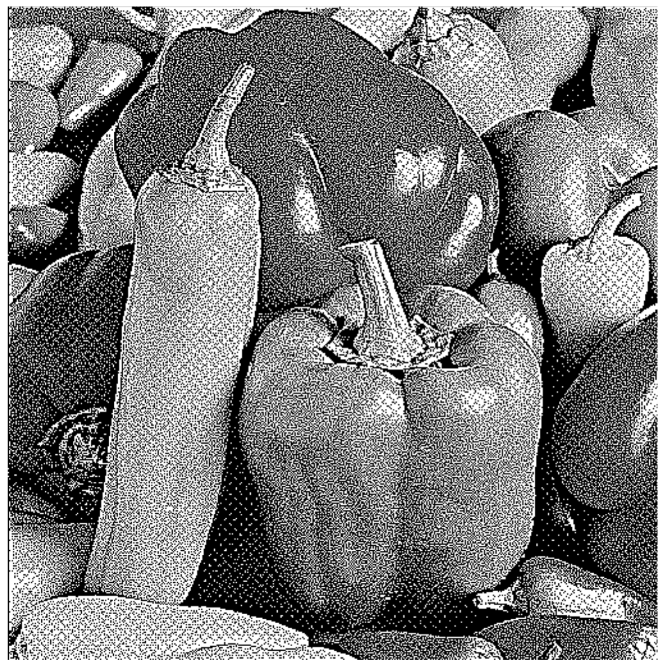

Figure 1. Dot diffusion with Knuth's class matrix (8×8 class matrix).

Optimization of the Class Matrix

Knuth introduced the notion of **barons** and **near-barons** in the selection of his class matrix. A baron has only low-class neighbors, and a near-baron has one high-class neighbor. The quantization error at a baron is not distributed to neighbors, and the error at a near-baron is distributed to only one neighbor. Knuth's idea was that the number of barons and near-barons should be minimized. He exhibited a class matrix with two barons and two near-barons. The resulting halftones still exhibit periodic patterns similar to ordered dither methods (see Fig. 1). Knuth has also produced a class matrix with one baron and near-baron, but unfortunately these were vertically lined up to produce objectionable visual artifacts. In our experience, the baron/near-baron criterion does not appear to be the right choice for optimization.

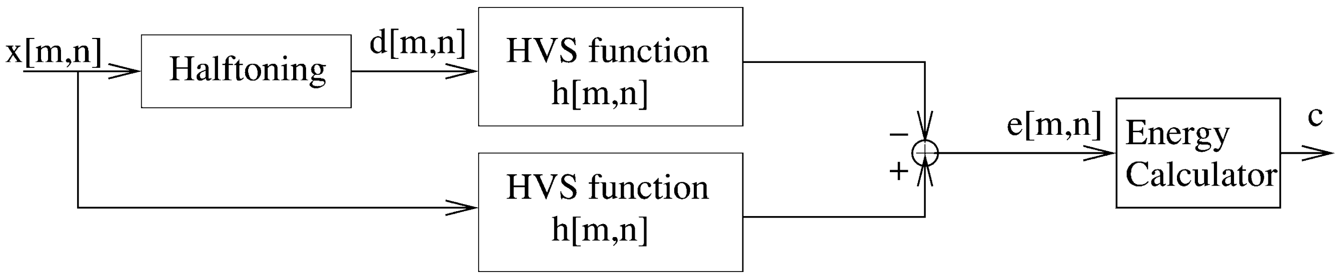

Figure 2. Perceived halftoning error of an image for a given HVS function.

We used a different criterion for quantifying the quality of halftone images.[4] In this criterion, the Human Visual System (HVS) is taken into account. The images are passed through a model of the HVS function. Since our model is linear, we apply the HVS function to the difference between the original and halftone images. The energy of the resulting image is defined to be the perceived halftoning error (PHE). The calculation of PHE of a given image $x[m,n]$ for a given HVS function $h[m,n]$ is depicted in Fig. 2. In the figure, the output of the energy calculator is

$$c = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \left| e[m,n] \right|^2 .$$

The image used in the optimization should be chosen wisely. For example in this article we have chosen a gray scale ramp because the cost of a gray scale ramp is the average value of the costs of gray scales which exist in the gray scale ramp.

We will use a specific HVS model in the optimization. In the frequency domain the HVS model is defined as follows:

$$H_c(u,v) = aL^b e^{-\frac{1}{s(\phi)} \frac{\sqrt{u^2+v^2}}{c\log(L)+d}} .$$

Here, u and v are the frequency variables in cycles/degree subtended at the retina and L is the average luminance in $candela/m^2$. The quantity

$$\sqrt{u^2+v^2}$$

is therefore the radial frequency. The quantity ϕ is the angular frequency defined as

$$\phi = atan\left(\frac{u}{v}\right).$$

The various constants and the function $s(\phi)$ are defined as follows:

$$a = 131.6,\ b = 0.3188,\ c = 0.525,\ d = 3.91.$$

$$s(\phi) = \frac{1-w}{2}\cos(4\phi) + \frac{1+w}{2} \text{ where } w = 0.7.$$

The dependence on radial frequency

$$\sqrt{u^2+v^2}$$

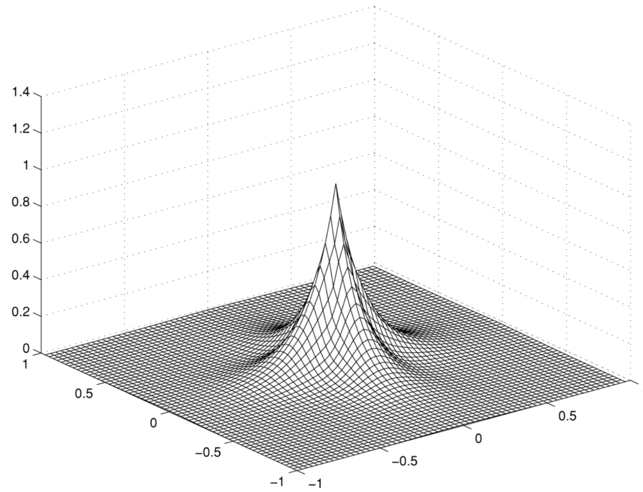

Figure 3. Normalized HVS function $H(w^1,w_2)$ for T = 0.0165 (RD = 300 dpi × 11.5827 in). The axes are

was developed in Ref. 8. Then the angular dependence of the model, i.e., $s(\phi)$ was introduced in Ref. 9 following Daly's model.[10] We used L = 10 $candela/m^2$ in our experiments. With $h_c(x,y)$ denoting the inverse Fourier transform of $H_c(u,v)$, the discretized version $h[m,n] = h_c(Tm,Tn)$ is used in the calculations. The relation between $H(w_1,w_2)$ (the discrete Fourier transform of $h[m,n]$) and $H_c(u,v)$ is as follows:

$$H(w_1,w_2) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} H_c\left(\frac{w_1 - 2\pi k}{2\pi T}, \frac{w_2 - 2\pi l}{2\pi T}\right) \quad (3)$$

Sampling the inverse transform at interval T = 0.0165 corresponds to a certain printer resolution, R dpi, viewed at a specific distance, D inches. Because a length x viewed at a distance D subtends an angle of $\Theta = \tan^{-1}(x/D) \approx x/D$ radians for $x << D$, the spacing of the dots will be:

$$T = \frac{1}{RD} \text{ radians} = \frac{180}{\pi} \frac{1}{RD} \text{ degrees} \quad (4)$$

This clarifies the relation between T, D and R. In particular, T = 0.0165 corresponds to RD = 300 dpi × 11.5827 in. In Fig. 3, the normalized HVS function is shown for this value T.

In the optimization we are looking for a class matrix that minimizes the cost function. Notice that the optimization is equivalent to finding a combination of num-

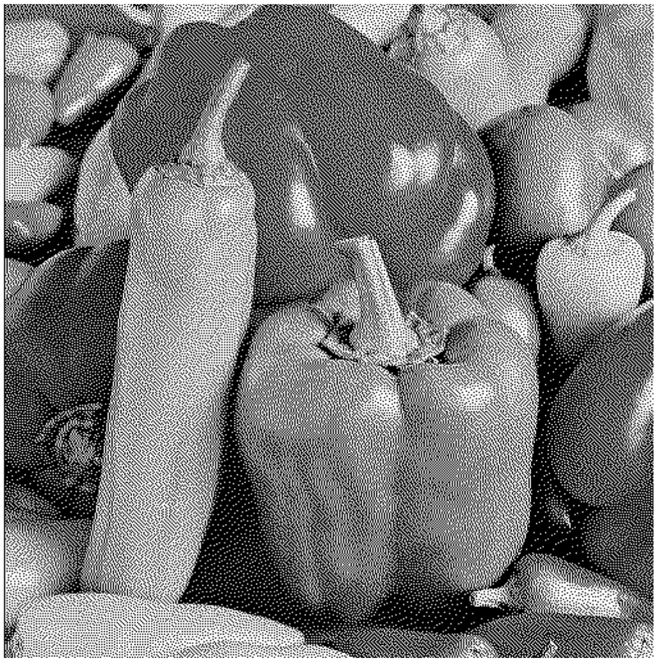

Figure 4. Floyd–Steinberg error diffusion.

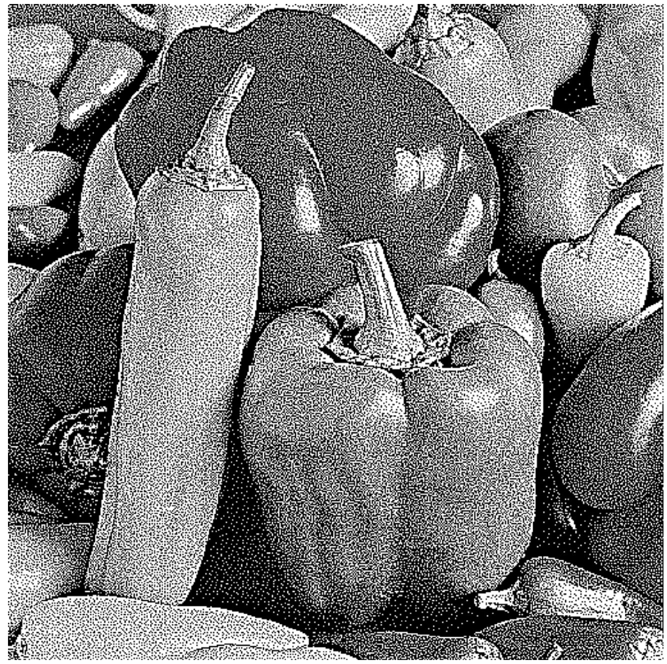

Figure 5. Dot diffusion with HVS optimized 8 × 8 class matrix and enhancement.

bers from 1 to *IJ* such that the related cost is minimized. Because the cost function depends nonlinearly on its parameters, we will use an optimization procedure to get the desired class matrix. The choice of the class matrix that minimizes this cost function was performed using the **pair-wise exchange algorithm**[11] described below.

1. Randomly order the numbers in the class matrix.
2. List all possible exchanges of class numbers.
3. If an exchange does not reduce cost, restore the pair to original positions and proceed to the next pair.
4. If an exchange does reduce cost, keep it and restart the enumeration from the beginning.
5. Stop searching if no further exchanges reduce cost.
6. Repeat the above steps a fixed number of times and deep the best class matrix.†

Optimization Results

In the optimization we have used a gray scale image. First, we have done the optimization for 8 × 8 class matrix. We have optimized this class matrix for $RD = 300$ dpi × 11.5827 in. This class matrix is shown in Table IV. We have summarized the perceived halftoning errors in Table V. In this article, perceived errors are normalized so that perceived error of a gray scale ramp halftoned by the Floyd-Steinberg error diffusion algorithm is unity. As it can be seen from the table, our optimized class matrix achieves 40.04% less PHE than Knuth's class matrix and 51.61% more PHE than error diffusion.

Example: The 512 × 512 continuous tone peppers image was halftoned by using Knuth's class matrix (Fig. 1, *PHE* = 30.77), and by the optimized 8 × 8 class matrix (Fig. 5, *PHE* = 30.35).‡ The images in this article are printed out with $R = 150$ dpi, thus they should be viewed from a distance $D \approx 23$ inches. It is clear that the dot diffusion method with the optimized 8 × 8 class matrix is visually superior to dot diffusion method with Knuth's class matrix. In fact, dot diffusion with the optimized 8 × 8 class matrix offers a quality comparable to Floyd-Steinberg error diffusion method (Fig. 4, *PHE* = 3.86). Note that, we cannot use the PHE values of error diffused images and images obtained by dot diffusion with enhancement to compare the visual quality of these two methods because of the enhancement step. Thus visual inspection is necessary. Error diffused images suffer from worm-like patterns which are not in the original image, whereas dot diffused halftones do not contain these artifacts. Notice that the artificial periodic patterns in Fig. 1 are absent in Fig. 4 and in the dot diffusion with the optimized 8 × 8 class matrix (Fig. 5).

TABLE IV. 8 × 8 Optimized Class Matrix.

37	41	34	14	60	61	7	9
16	12	36	59	46	17	50	24
45	27	33	58	5	3	42	48
29	2	57	30	43	15	20	11
26	18	55	49	4	32	10	54
25	21	53	40	38	6	64	52
8	28	35	13	39	22	63	56
51	44	19	23	31	62	1	47

TABLE V. Perceived Halftoning Errors (PHE) for 8 × 8 Class Matrices.

Halftoning Method	Dot Diffusion Kmuth's Class Matrix	Dot Diffusion Optimized Class Matrix	Error Diffusion (Floyd–Steinberg)
Perceived Halftoning Error	2.53	1.52	1.00

† Note that pair-wise exchange algorithm yields a local minimum of the cost function. We begin the pairwise exchange with a number of random class matrices and take the class matrix having the least local minimum in order to get closer to the global minimum. Global minimum is not guaranteed.

‡ We observed that the enhancement step in dot diffusion is the cause of higher PHE values. In the next section, enhancement step will be removed from dot diffusion algorithm.

TABLE VI. 16 × 16 Class Matrix

202	1	14	18	51	56	45	105	74	98	75	145	150	170	171	173
4	7	24	37	57	52	66	88	146	103	138	159	183	185	198	222
8	15	25	38	68	70	87	6	107	153	144	166	184	193	225	2
16	27	44	54	29	102	116	132	140	137	167	120	196	224	227	5
23	40	53	72	85	104	165	136	158	174	131	200	223	226	228	17
41	86	73	84	114	118	168	134	169	181	201	220	232	229	13	22
48	121	55	106	124	133	147	177	180	203	221	231	246	3	21	42
77	82	128	110	139	135	179	182	207	197	230	245	247	20	43	50
81	100	113	148	143	172	178	204	219	233	244	250	248	34	49	69
109	108	141	151	186	164	208	218	234	243	249	256	19	46	71	80
111	142	89	76	176	206	215	235	242	251	255	39	47	78	117	101
112	149	161	175	205	216	236	241	252	253	254	62	63	94	95	126
152	160	190	191	209	217	237	240	26	32	61	83	93	96	125	115
157	189	192	210	214	238	239	30	33	60	65	92	119	79	129	156
188	195	199	213	10	11	31	36	59	64	91	97	123	130	155	162
194	211	212	9	12	28	35	58	67	90	99	122	127	154	163	187

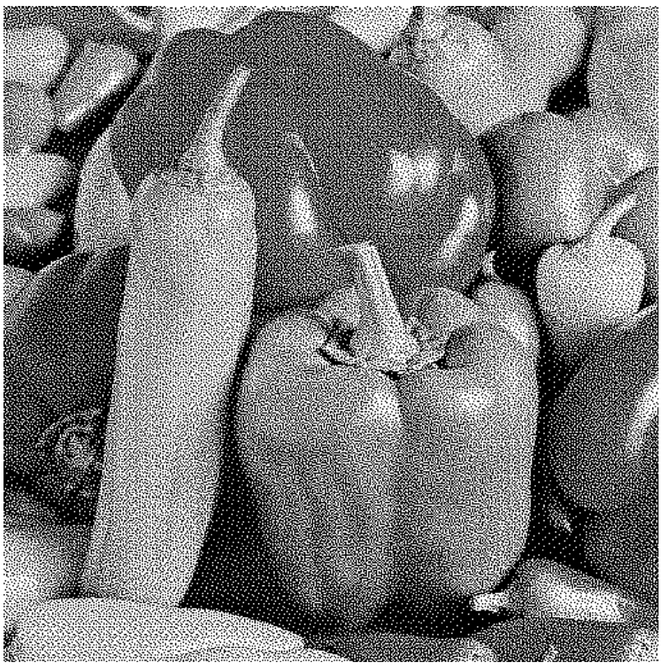

Figure 6. Dot diffusion with HVS optimized 8 × 8 class matrix and no enhancement.

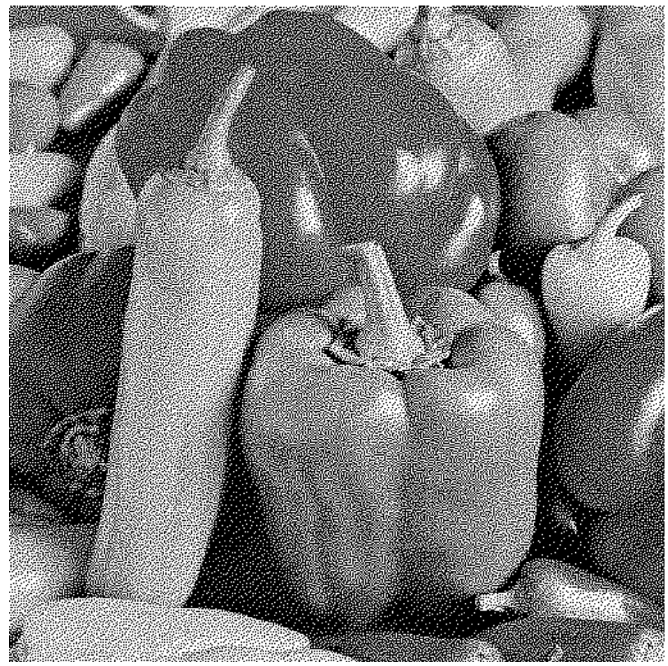

Figure 7. Dot diffusion with HVS optimized 16 × 16 class matrix and no enhancement.

Dot Diffusion Without Enhancement

If we compare the halftone images obtained with enhancement (Fig. 5, *PHE* = 30.35) and without enhancement (Fig. 6, *PHE* = 6.90), we can conclude that the enhancement step reduces the periodic structures in the halftone image, but it might be objectionable in some applications because of its very visible sharpening effect (e.g., see Fig. 5).

It turns out that we can get good halftones without use of the enhancement step provided we make the class matrix larger than the standard 8 × 8 size. The price paid for the larger class matrix is that the parallelism of the algorithm is compromised. However, in practice, even with a 16 × 16 class matrix, we have plenty of parallelism for any desktop printing implementation: Assume that we want to render the image at 600 dpi, and process 16 rasters (lines or rows) of an 8.5 × 11 inch square page simultaneously. Then we will have 600 × 8 × 16)/256 = 300 pixels that can be processed simultaneously.[§] The real disadvantage of increasing the size of the class matrix is the fact that the number of rasters that must be processed at the same time increases.

We found that if a 16 × 16 matrix is used, the halftone images resulting from the optimization of this matrix are very good even without the enhancement step. (For comparison we note here that whenever enhancement is used, the class matrix can be as small as 5 × 5 without creating noticeable periodicity patterns.) Such optimization was carried out using a fray scale ramp as the training image. The HVS function was used in the optimization, and the associated cost was optimized using the pair-wise exchange algorithm. The 16 × 16 optimized class matrix is shown in Table VI. The *PHE* of a gray scale halftoned by dot diffusion with 16 × 16 optimized class matrix is 1.19. We can compare this *PHE* value with the *PHE* values in Table V. The *PHE* for optimized 16 × 16 class matrix is only 19.38% worse than the error diffusion. Also, the PHE for optimized 8 × 8 class matrix is 27.00% worse than the PHE for optimized 16 × 16 class matrix.

The peppers image halftoned with the resulting class matrix is shown in Fig. 7 (PHE = 5.90). There are no

§ The number 8 in the numerator of this expression is based on the assumption of an 8 inch wide active printing area.

periodic artifacts in these results. While the overall visible noise level appears to be higher than for error diffusion, the problematic halftone patterns of error diffusion in the mid gray level are eliminated here. (Examine the body of the middle pepper in Fig. 4). By comparing Figs. 1 and 7 we see that 16 × 16 dot diffusion without enhancement is also superior to 8 × 8 enhanced dot diffusion using Knuth's matrix because there are no noticeable periodic patterns any more, and there are no enhancement artifacts.

Conclusion

Dot diffusion offers more parallelism than error diffusion and the method has been optimized in order to remove the periodic artifacts. The enhancement step prior to dot diffusion was preserved in previous optimizations. Because the enhancement can be objectionable in some cases, the method has been improved by optimizing a larger class matrix. Furthermore, we have shown that error diffusion and binary quantization are special cases of dot diffusion.

Acknowledgment. This work was supported by the National Science Foundation under Grant MIP 0703755 and Microsoft Research, Redmond, VA.

References

1. M. Mese and P. P. Vaidyanathan, Optimized halftoning using dot diffusion and methods for inverse halftoning, *IEEE Transactions on Image Processing* **9,** 4, 691–709 (2000).
2. R. Floyd and L. Steinberg, An adaptive algorithm for spatial greyscale, *Proc. SID*, pp. 75–77, 1976.
3. D. E. Knuth, Digital halftones by dot diffusion, *ACM Tr. on Graphics*, **6,** pp. 245–273 (1987).
4. M. Mese and P. P. Vaidyanathan, Image halftoning using optimized dot diffusion, *Proc. of EUSIPCO*, Rhodes, Greece, 1998.
5. M. Mese and P. P. Vaidyanathan, Image halftoning and inverse halftoning for optimized dot diffusion, *Proc. ICIP*, Chicago, IL, 1998.
6. M. Mese and P. P. Vaidyanathan, A mathematical description of the dot diffusion algorithm in image halftoning, with application in inverse halftoning, *Proc. of ICASSP*, Phoenix, AZ, 1999.
7. M. Mese and P. P. Vaidyanathan, Improved Dot Diffusion for Image Halftoning, *Proceedings of IS&T's Conference on Digital Printing Technologies, NIP14,* IS&T, Springfield, VA, 1999.
8. R. Nasanen, Visibility of halftone dot textures, *IEEE Transactions on Systems, Man and Cybernetics* **14,** No. 6, pp. 920–924, 1984.
9. J. P. Allebach, FM screen deisgn using DBS algorithm, *Proc. of ICIP*, Vol. 1, Lausanne, Switzerland, 1996, pp. 549–552.
10. S. Dally, Subroutine for the generation of a two dimensional human visual contrast sensitivity function, Eastman Kodak Tech. Rep. No. 233203.
11. J. P. Allebach and R. N. Stradling, Computer-aided design of dither signals for binary display of images, *Appl. Opt.* **18,** 2708–2714 (1979).