

## Color Image Quantization and Dithering Method based on HVS Characteristics

Yeong-Ho Ha

School of Electronic and Electrical Engineering, Kyungpook National University

E-mail:yhha@ee.kyungpook.ac.kr

### Abstract

New methods for both color palette design and dithering based on human visual system (HVS) characteristics are proposed. Color quantization for palette design uses the relative visual sensitivity and spatial masking effect of HVS. The dithering operation for printing uses nonlinear quantization, which considers the overlapping phenomena among neighbor printing dots, and then a modified dot-diffusion algorithm is followed to compensate the degradation produced in the quantization process. The proposed techniques can produce high quality image in the low-bit color devices.

Keywords: color quantization, distortion measure, color dithering, and dot diffusion

### I. Introduction

Recently, the use of color image data has been growing fast in the area of image processing[1-7]. To express natural color in conventional low-cost color devices, the color image should be quantized for monitoring and dithered for printing.

A video monitor displays a color image by modulating the intensity of three primary colors(red, green, and blue) at each pixel of the color image. In a digitized color image, each primary color is usually quantized with 8 bits of resolution to eliminate distinguishable quantization steps in trichromatic specification (luminance, hue, and saturation). Thus, full-color digital display systems use 24 bits to specify the color of each pixel on the screen. However, the cost of high-speed memory needed to support such a display on a high-resolution monitor makes many applications impractical. An alternative approach in currently available displays is to provide a limited number of bits, such as 8 bits, for specifying the color at each pixel. Each of these  $2^8$  values is then used as an index of user-defined color table, i.e. color palette. Each entry in the table contains a 24-bit value that specifies each primary component of the color image. In this way, the user is allowed to select a small subset of color palette from the full range of  $2^{24}$  colors. The drawback of this scheme is that it restricts the number of colors which may be simultaneously displayed[1-7].

Because natural images typically contain a large number of distinguishable colors, displaying such images with a limited palette is difficult. Several techniques exist for color quantization, some of which are based on a more general class of vector quantization (VQ) techniques. One approach involves the iterative refinement of an initially selected palette. Variations on this idea include the manner in which the initial palette is chosen and the color space in which the quantization is performed. The refinement algorithm, commonly known as the K-means or Linde-Buzo-Gray (LBG) algorithm[1], is a vector extension of the Lloyd quantizer for scalars. The algorithm seeks to reduce the total squared error (TSE) between the original and the quantized image at each iteration until a (local) minimum is found. This method yields high-quality images and, with a properly chosen initial palette, will result in the lowest TSE for a given palette size. The method is, however, computationally intensive, and its performance is sensitive to the choice of the initial palette. Also there are splitting algorithms[2,3,6] that divide the color space into disjoint regions and pick a representative color from each region as a palette color. The algorithms vary according to the methods used to split the color space. As an example of such an algorithm, the median-cut algorithm[2] invented by Heckbert was undertaken as an alternative to the popularity algorithm. The median-cut algorithm repeatedly subdivides the color space into smaller rectangular boxes until the desired number of boxes is generated. The split point is the median point, i.e., the plane that divides the box into two halves so that an equal number of colors are on each side. The main advantage of the split algorithm is lower computational time cost and memory space for the spatial storage scheme, mostly because it is simple to compute the split point. However, a number of problems are associated with this method. One problem is that partitioning a box by a plane passing through the median point does not necessarily lead to a lower quantization error.

As explained above, the conventional color quantization algorithms usually use the TSE as the distortion measure[1-7]. This measure is, however, perceptually insufficient when accurately estimating the perceptual difference between an original image and its quantized representation[5-7]. The TSE does not take into account

the spatial correlation linked with perceptually adjacent pixels. With such a measure we have no means of knowing whether the observed degradation is the result of several particularly noticeable degradation. The measure must be reconsidered, this time by duly integrating the notion of locally observable errors. Thus, we propose to use a new distortion measure that takes into account the spatial activity in local region of the input image. The activity is computed as the mean of the sensitivity-weighted difference between the input and the local mean color in a  $4 \times 4$  region. Then, using the distortion measure, a hierarchical quantization algorithm is proposed by considering the spatial masking effect[8,9] a characteristic of the HVS. The algorithm consists of initial and subdivision steps both to reduce the computational time and to minimize the measured distortion.

Digital dithering is the process for generating a pattern of dots with a limited number of gray levels to reproduce a continuous tone image on the paper media for hard copy. Display of continuous tone images on the paper media is impossible without the dithering process. Thus, many dithering techniques can be found in the printing algorithms[10-13]. Printers are able to produce marked dots on a part of a Cartesian grid with horizontal and vertical spacing. Conventional digital dithering methods use linear quantization in which input gray levels are equally divided by the printer resolution, dots per inch (dpi). Errors produced by the quantization are diffused to the neighborhood to compensate for the local gray level[14-15]. In this process, accurate gray levels are not represented because of pure printer resolution, in generally, and also because diffusion of the contrast in the edge region is not considered. Therefore, new method using linear quantization and diffusion is required to reduce the degradation of images produced by the conventional dithering method.

For printing the monitor-displayed image on the paper media, we propose nonlinear quantization that considers the overlapping of printing dots and modified dot-diffusion method. For the nonlinear quantization, each quantization step is adjusted in proportion to the overlapping area that is subtracted from total area of each neighboring printing dot. To compensate for errors produced in the quantization, the modified dot-diffusion adjusts the quantization error to be diffused according to the characteristics of the input image. Filtering of the quantization error by a low pass filter is carried out before dot diffusion so that the quantization error has to be diffused only in the smooth region. Also, to apply the proposed method to color printing, intensity and color change must be considered simultaneously. Thus quantization error is diffused in the Hue, Saturation, and Intensity (HIS) color coordinate that represents the HVS color coordinate system. The low pass filtering must be considered to prevent color degradation due to the diffusion of quantization error in the color edge region.

This report is organized as follows. In the next

section, which starts by defining the color space in which the quantization algorithm operates, the proposed distortion measure and hierarchical quantization algorithm are described. The succeeding section describes the printer model, the proposed nonlinear quantization, and the modified dot-diffusion method for color printing. The section that follows reports the simulation results of the proposed techniques. Finally we presents out conclusion.

## II. Proposed Color Quantization Algorithm

The color quantization for palette design is described as selecting the prescribed number of colors to display an image with almost no noticeable perceived difference. The process is usually performed by treating three color components (red, green, and blue) independently in the RGB color space. Although three color components can be decorrelated by transforming the color space to YIQ, Lab, or some other color spaces, independent quantization in these spaces is inefficient because certain proportion of these spaces lies outside the RGB color space. In any event, color transformations are of little use in quantization for display; their proper place is the image compression systems. Thus, we quantize the colors of original image into  $K$  (usually  $K=256$  or less) colors, called color palette, in the RGB color space.

The color image is assumed to be on the rectangular grid of  $M \times M$  pixels. The set of all grid points is denoted by  $F$  and its members  $f \in F$  may be explicitly written as  $f = (i, j)$ , where  $i$  and  $j$  are the row and column indices ( $0 \leq i, j \leq M-1$ ). The color value of the pixel at grid point  $f$  is denoted by  $\vec{c}_f = [r_f, g_f, b_f]^T$ , where the components are the red, green, and blue tristimulus values for the pixel in the RGB color space, and superscript  $T$  means transpose. As explained above, the color quantization is selecting  $K$  colors to be displayed on the monitor. The  $K$  colors are usually composed of the centroids of  $K$  color clusters. In this report, the  $k$ -th cluster is denoted by  $\Omega_k (1 \leq k \leq K)$  and the centroids of the clusters are denoted by  $\vec{\mu}_k = [r_k, g_k, b_k]^T$ , ( $1 \leq k \leq K$ ), each of which composes the color palette. Thus, the input image colors are mapped to the centroid colors after the quantization, and the mapped colors are displayed on the monitor simultaneously. Therefore, the color quantization algorithm finds the optimal color cluster and its centroid with almost no difference between the original image and the reconstructed image.

## III. Weighted-Distortion Measure of HVS Color Activity

In this paper, a new distortion measure based on the spatial masking effect of the HVS is proposed. The spatial masking effect means that human vision is more

sensitive to quantization errors in the smooth region than to those in the edge region. To use this effect, the activity of color in the local region must be computed to judge whether the color is in the smooth region or in the edge region. Because human vision is sensitive to each color component differently, the proposed color activity is computed as the mean of sensitivity-weighted differences between input color in the local region of  $16(4 \times 4)$  pixels and mean color ( $\vec{c}_m = [\bar{r}_m, \bar{g}_m, \bar{b}_m]^T$ ) of all input colors in the region, as follows.

$$\begin{aligned} A(\vec{c}) &= 1/16 \sum_{n=1}^{16} (\vec{c}_n, \vec{c}_m) \\ &= 1/16 \sum_{n=1}^{16} \{ \alpha(r_n - \bar{r}_m)^2 + \beta(g_n - \bar{g}_m)^2 + \gamma(b_n - \bar{b}_m)^2 \}^{\frac{1}{2}} \end{aligned} \quad (1)$$

where  $A(\vec{c}_n, \vec{c}_m)$  is the sensitivity-weighted difference and  $\vec{c}_n = [\bar{r}_n, \bar{g}_n, \bar{b}_n]^T$  is the  $n$ -th input color in the local region. The coefficients ( $\alpha, \beta$  and  $\gamma$ ) are the relative visual sensitivities of the HVS in the CIE standard, and they are the coefficients of the luminance equation. This computed activity is similarly assigned to each color included in the local region, that is, the same activity value to each color in the same local region. This is performed for all local regions of the image. Then we can obtain the activity ( $A(\vec{c}_f)$ ) of a pixel  $f$  as the activity of the local region which the pixel is included. According to the masking effect, the higher activity means that the color is in the edge region and its quantization error is less sensitive in human vision, whereas the lower activity means that the color is in the smooth region and its quantization error is more sensitive. Therefore, we can see that the activity is inversely proportional to the error sensitivity of the HVS.

Using this property, the new distortion measure is proposed to determine which color should be more finely quantized to decrease the distortion based on the masking effect. The total distortion is given by

$$\begin{aligned} D &= \sum_{k=1}^K d_k \\ &= \sum_{k=1}^K \sum_{c_f \in D_k} \frac{1}{A(\vec{c}_f)^2} \|\vec{c}_f - \vec{\mu}_k\|^2 \\ &= \sum_{k=1}^K \sum_{c_f \in D_k} \frac{1}{A(\vec{c}_f)^2} E_q \end{aligned} \quad (2)$$

where  $d_k$  is the distortion of  $k$ -th color cluster and  $E_q$  is the quantization error. Because the input color is included in  $k$ -th cluster, the quantization error is inversely weighted by the activity of the input color. If so, the distortion error is decreased by the activity. Therefore, if the input color is in the edge region, the activity is much higher and the quantization error of the color is relatively much decreased, which means that the distortion of the cluster is decreased. If the input color is in the smooth region, the activity is much lower and the quantization error of the color is relatively less decreased,

which means that the distortion of the cluster is increased. Thus, if a color cluster of maximum distortion is more finely quantized, the total distortion based on the masking effect can be much less.

The proposed quantization algorithm has a hierarchical structure, so that the color of lower activity should be more finely quantized.

#### IV. Proposed Hierarchical Color Quantization Algorithm

Based on the activity and the distortion measure, the proposed color quantization algorithm can be performed in two hierarchically steps. In the first step, the input colors are divided into eight initial color clusters by a thresholding method using intercluster variances[16]. To compute the intercluster variance, an activity histogram ( $x_l$ ,  $x = R, G$ , or  $B$  component) with respect to gray level ( $l, 0 \leq l \leq 255$ ) of each color component is obtained. The activity histogram of the red component of Peppers image used in the experiment is shown in Fig. 1(a).

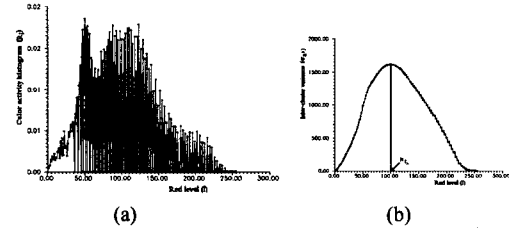


Figure 1. Activity histogram and its inter-cluster variance in the red component of the Peppers image: (a) activity histogram and (b) intercluster variance.

According to the activity histogram, the intercluster variance for each component is calculated as follows:

$$\begin{aligned} \sigma_x^2 &= P_1 \left( \frac{1}{T_h + 1} \sum_{l=0}^{T_h} x_l - \bar{x} \right)^2 \\ &\quad + P_2 \left( \frac{1}{255 - T_h} \sum_{l=T_h+1}^{255} x_l - \bar{x} \right)^2, \quad l \leq T_h \leq 255 \\ P_1 &= (x_0 + x_1 + \dots + x_{T_h}) / X \\ P_2 &= (x_{T_h+1} + x_{T_h+2} + \dots + x_{255}) / X = 1 - P_1 \end{aligned} \quad (3)$$

where  $X = \sum_{l=0}^{255} x_l$  and  $\bar{x} = X/256$ . The value  $T_h$  signifies threshold, and it is the gray value of maximum intercluster variance. The calculated variance of Fig. 1(a) is shown in Fig. 1(b). The threshold values ( $R_{T_h}$ ,  $G_{T_h}$ , and  $B_{T_h}$ ) for red, green, and blue can be selected as  $T_h$ .

In the second step, the color cluster of maximum distortion is again divided into eight color clusters, using the proposed distortion measure and thresholding method based on the masking effect. This process is repeatedly applied until  $K$  ( $=256$  or less) color clusters are obtained.

Then the color palette is composed of the centroids of the obtained color clusters, and, finally, the input colors are mapped to the color palette. The proposed algorithm is shown in Fig. 2.

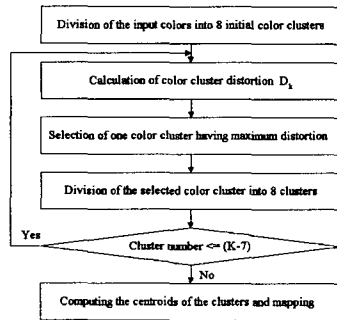


Figure 2. The proposed hierarchical color quantization algorithm.

## V. Proposed Color Dithering Method Printer Model and Nonlinear Quantization

Most conventional dithering methods use linear quantization that equally divides input gray level according to printer resolution. This algorithm has simple processing and less computational time. However, because the algorithm does not consider hardware characteristic of printer, a gray-level difference between the printed image and the displayed image on the monitor is produced.

As shown in Fig. 3, printers produce circular black dots, rather than square ones. The most elementary degradation introduced by most printers is illustrated as following: their dots are larger than the minimal covering size, as if "ink spreading" is occurred. Other degradations are caused by heat finishing, reflections of light within the paper, etc. Also the inks are not perfect. As a result, the gray level produced by the printer in the vicinity of pixel (i, j) depends in some complicated way and neighboring dots. However, due to the close spacing of dots and the limited spatial resolution of the eye, the gray level can be modeled as having a constant value within the area of the ideal circular pixel (i, j).

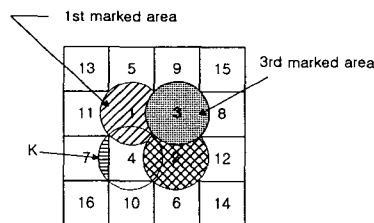


Figure 3. The marking area of dots in the dither matrix.

Therefore, we develop a quantization method that improves the quality of printed images. The method decreases the quantization error and distributes the remaining quantization error more effectively, both spatially and across the intensity range.

We propose here a nonlinear quantization that accounts for the dot overlap degradation due to hardware characteristics. This approach is adequate to illustrate quantization errors produced in the multilevel quantization due to printer resolution.

In black-and-white printing, all the overlapping areas have the same gray level as the dots. So that the reproduced gray level is proportional to the marked area. The marked area in the printing is changed according to marking position and dither matrix. If the difference in the marking area is not considered, the printed image is not equal in gray level to the image displayed on the monitor.

The proposed nonlinear quantization for dithering can improve this drawback. The quantization step is adjusted in proportion to the overlapping area, which is subtracted from total area of each printed neighbor dot. The total area of each pixel is the area of the  $4 \times 4$  dither matrix in the 300-dpi printer, and the area of each dot is the area of a circle. Therefore, the reconstructed gray level due to actual marked area is calculated as follows.

$$r = \left\{ 1 - N \times \left( \frac{\pi}{32} \right) + P_N \times K' \right\} \times 255 \quad (4)$$

where  $r$  is reconstructed gray level,  $N$  is number of the marked dots,  $P_N$  is the sum of overlapped area and half of dots marked in outer region of the  $4 \times 4$  dither matrix among the marked dots,  $K'$  is  $2 \times K$ , and the area of  $K$  is  $((\pi - 2)/32)/4$ . In the printed image, this nonlinear quantization reduces the degradation which is generated due to hardware characteristic.

## VI. Modified Dot Diffusion for Color Printing

In this section we present models for color printers and modified dot diffusion for color printing. Such printers are capable of producing colored dots on a piece of paper, at any and all sites of a Cartesian grid with horizontal and vertical spacing. The reciprocal of spacing is the printer resolution in dots per inch (dpi). Color printers use cyan (C), magenta (M), and yellow (Y) inks to produce color dots. These colors form the basis for the subtractive system of colors. The relationship to the additive colors (RGB) is followed.

$$\begin{aligned} R &= 1 - C \\ G &= 1 - M \\ B &= 1 - Y \end{aligned} \quad (5)$$

Cyan ink absorbs red light, magenta absorbs green,

and yellow absorbs blue. Dots of different inks can be superimposed to produce red, green, blue, and black dots. In practice, however, the inks are not perfect, i.e., there are unwanted absorptions. Thus, many printers use a separate black ink (K) to produce better black dots. In the remainder of this discussion we will assume that the printer uses all four types of ink. The printer is controlled by an array of four-dimensional vectors with binary components.

$$b_{i,j} = (b_{i,j}^C, b_{i,j}^M, b_{i,j}^Y, b_{i,j}^K) \quad (6)$$

where  $b_{i,j}^C=1$  indicates that a cyan dot is to be placed at pixel  $(i, j)$  and  $b_{i,j}^C=0$  indicates that no cyan dot is to be placed at the site. The magenta  $b_{i,j}^M$ , yellow  $b_{i,j}^Y$ , and black  $b_{i,j}^K$  component are defined similarly. When all components are zero, the site is the remain white. When more than one component is equal to 1, different inks are printed on top of each other. In principle, we can specify  $2^4=16$  different colors for each dots, but nine of these colors are variations of black. Usually, the black ink is used only in combination with the other three inks to produce solid black dots, thus reducing the number of colors that each dot can include to  $2^3=8$ .

As we can see in black-and-white printing, printers produce circular, rather than square, dots, and the quantization error produced in the quantization process has to be compensated. To apply the proposed algorithm to color printing, color change, as well as intensity, must be considered. Thus, quantization error is diffused according to the characteristics of the input image in the HSI color-coordinate system, which represents HVS color-sensing properties. In the dithering process, dot diffusion is modified by adjusting the amount of hue and saturation diffusion according to the intensity and saturation of the input image in the HSI color-coordinate system. The HSI coordinate system has higher sensitivity in the highest and lowest intensities and lower saturation in the mid-range intensities. To compensate the nonuniform property in the HSI coordinate system, the amount of diffusion is given by

$$\begin{aligned} \Delta H' &= \begin{cases} \Delta H \times S \times \frac{I}{127}, & I \leq 127 \\ \Delta H \times S \times \frac{(255-I)}{127}, & \text{otherwise} \end{cases} \\ \Delta S' &= \begin{cases} \Delta S \times S \times \frac{I}{127}, & I \leq 127 \\ \Delta S \times S \times \frac{(255-I)}{127}, & \text{otherwise} \end{cases} \end{aligned} \quad (7)$$

where  $\Delta H'$  and  $\Delta S'$  are the hue and saturation to be diffused,  $\Delta H$  and  $\Delta S$  are the hue and saturation errors that occurred in the quantization process,  $S$  is the saturation value, and  $I$  is the intensity value.

In color printing using the modified dot-diffusion, we suggest filtering the quantization error by a low-pass

filter before dot diffusion, because the quantization error must be diffused only in the smooth region. Therefore, the printed image can prevent color degradation through the diffusion of quantization error between neighboring regions of different color.

## VII. Experimental results

In the experiment, we used a PC with VGA board of 256 colors and a HP Deskjet 560K of 300-dpi ink-jet printer. Color quantization operations were performed separately on each color component: R, G, and B. The images used in the experiment were  $256 \times 256$  Girl, Lena, Peppers, and Zelda. Figure 5 shows the results of the color quantization with the Girl image. Here, Fig. 5(a) is the original image, 5(b) is the result of LBG algorithm, 5(c) is the result of Heckbert algorithm, and 5(d) is the result of the proposed algorithm. Figure 6 compares magnified images of Fig. 5. The displayed image on the monitor using the proposed color quantization method shows almost no noticeable difference from the result of the LBG method and from the original image. Figure 7 shows printed images using (a) linear quantization and ordered dithering, (b) linear and dot diffusion, and (c) the proposed method, along with the original image. The printed result using the proposed color dithering method is high quality and shows less color degradation than that of the control image produced by a conventional printing method.

## VIII. Summary and Conclusion

New methods for displaying and printing high-resolution full-color images on limited color output devices are proposed. The color quantization for palette design consists of a hierarchical method based on HVS characteristics. This technique uses the activity-weighted distortion measure based on the relative visual sensitivity and spatial masking effect of the HVS. In color printing, dithering of nonlinear quantization is proposed to reduce color degradation by considering the overlapping phenomena of printing dots. In the dithering process, modified dot diffusion is also proposed. The proposed techniques enables the limited-color-output devices to display and print high-quality color images.

## References

- [1] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Tran. on Commun.*, vol. COM-28, (1), pp.84-95, 1980.
- [2] P. S. Heckbert, "Color image quantization for frame buffer display," *Comput. Graph.*, 16 (3) pp.297-307, 1982.
- [3] N. Goldbert, "Color image quantization for high resolution graphics display," *Image and vision computing*, 9 (5), 1991.
- [4] M. T. Orchard and C. A. Bouman, "Color Quantization of

- Images," *IEEE Trans. on Signal Processing*, **39** (12), 1991.
- [5] A. Tremeau, M. Calonnier, and B. Laget, "Color Quantization Error In Terms Of Perceived Image Quality," in *1994 ICASSP*.
- [6] R. Balasubramanian, "Color-image quantization with use of a fast binary splitting techniques," *J. Opt. Soc. Am. A*, **11** (11), 1994.
- [7] T. S. Liu and L.-W. Chang, Fast color image quantization with error diffusion and morphological operations, *Signal Processing* **43**, pp.293-303, 1995.
- [8] A. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. of IEEE* **65**, pp.536-548, 1977.
- [9] J. D. Mollon and L. T. Sharpe, *COLOR VISION: Physiology and Psychophysics*, Academic Press Inc., 1983.
- [10] R. Eschbach and K. T. Knox, "Error Diffusion Algorithm with Edge Enhancement," *J. Opt. Soc. Am. A*, **8** (12), pp.1844-1850, 1991.
- [11] R. Ulichney, *Digital Halftoning*, The MIT Press Cambridge, Massachusetts, 1987.
- [12] R. S. Gentile, E. Walowit, and J. P. Allebach, "Quantization and Multilevel Halftoning of Color Images for Near-Original Image Quality," *J. Opt. Soc. Am. A* **7**, pp.1019-1026, 1990.
- [13] S. C. Wells, G. J. Williamson, and Susan E. Carrie, "Dithering for 12-Bit True-Color Graphics," *IEEE Computer Graphics & Appl.*, pp.18-29, 1991.
- [14] D. E. Knuth, "Digital Halftones by Dot Diffusion," *ACM Trans. on Graphics*, **6** (4): pp. 245-273, 1987.
- [15] F. Fettuer and O. Bryngdahl, "Quantization Noise and the Error Diffusion Algorithm," *Journal of Electronic Imaging* **3**(1), pp.37-44, 1994.
- [16] N. Otsu, "A Threshold Selection Method for Gray Level Histograms," *IEEE Trans. on Sys. Man and Cybern.*, SMC-9: pp.62-66, 1979.

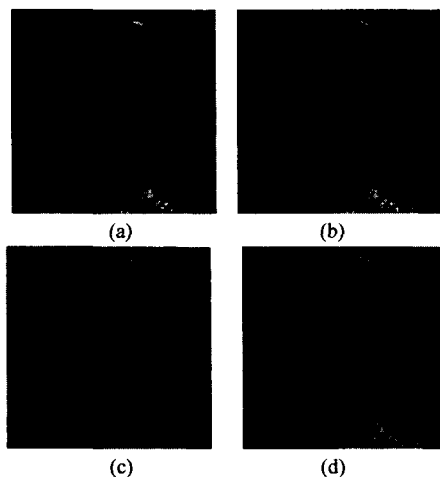


Figure 5. Original and result image of various quantization algorithm: (a) original image, (b) result of LBG algorithm, (c) result of Heckbert algorithm, and (d) result of the proposed algorithm.

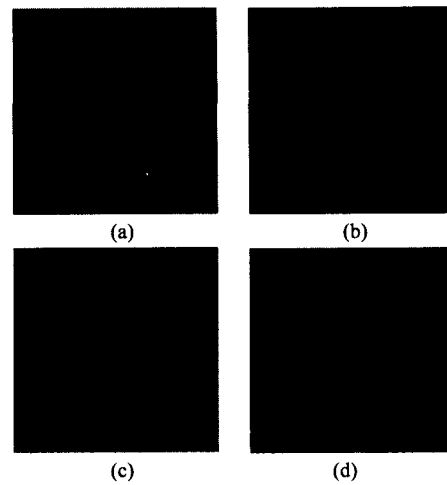


Figure 6. Magnified images of Fig. 5: (a) Magnified image of original, (b) magnified image of LBG algorithm, (c) magnified image of Heckbert algorithm, and (d) magnified image of the proposed algorithm.

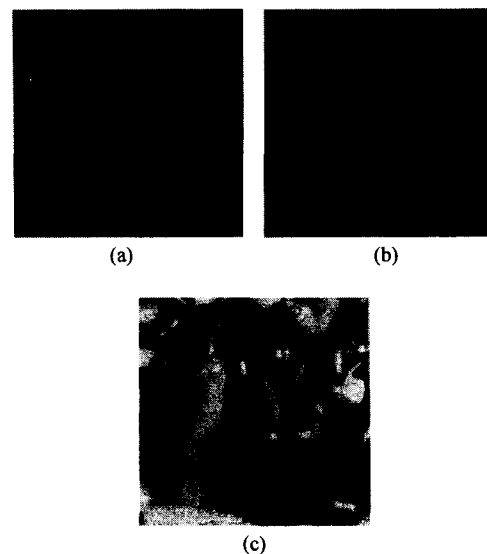


Figure 7. Printed image using various methods: (a) Printed image using linear quantization and ordered dithering, (b) printed image using linear and dot diffusion, and (c) printed image using the proposed method.