Hindawi Publishing Corporation
Advances in Multimedia
Volume 2016, Article ID 4985313, 10 pages
http://dx.doi.org/10.1155/2016/4985313

Research Article

Classification of Error-Diffused Halftone Images Based on Spectral Regression Kernel Discriminant Analysis

Zhigao Zeng,[1,2] Zhiqiang Wen,[1] Shengqiu Yi,[1] Sanyou Zeng,[3] Yanhui Zhu,[1] Qiang Liu,[1] and Qi Tong[1]

[1]*College of Computer and Communication, Hunan University of Technology, Hunan 412007, China*
[2]*Intelligent Information Perception and Processing Technology, Hunan Province Key Laboratory, Hunan 412007, China*
[3]*Department of Computer Science, China University of Geosciences, Wuhan, Hubei 430074, China*

Correspondence should be addressed to Zhigao Zeng; zzgzzg99@163.com

Received 21 January 2016; Revised 22 March 2016; Accepted 18 April 2016

Academic Editor: Stefanos Kollias

This paper proposes a novel algorithm to solve the challenging problem of classifying error-diffused halftone images. We firstly design the class feature matrices, after extracting the image patches according to their statistics characteristics, to classify the error-diffused halftone images. Then, the spectral regression kernel discriminant analysis is used for feature dimension reduction. The error-diffused halftone images are finally classified using an idea similar to the nearest centroids classifier. As demonstrated by the experimental results, our method is fast and can achieve a high classification accuracy rate with an added benefit of robustness in tackling noise.

1. Introduction

As a popular image processing technology, digital halftoning [1] has found wide applications in converting a continuous tone image into a binary halftone image for a better display on binary devices, such as printers and computer screens. Usually, binary halftone images can only be obtained in the process of printing, image scanning, and fax, from which the original continuous tone images need to be reconstructed [2, 3], using an inverse halftoning algorithm [4], for image processing, for example, image classification, image compression, image enhancement, and image zooming. However, it is difficult for inverse halftoning algorithms to obtain the optimal reconstruction quality due to unknown halftoning patterns in practical applications. Furthermore, a basic drawback of the existing inverse halftone algorithms is that they do not distinguish the types of halftone images or can only coarsely divide halftone images into two major categories of error-diffused halftone images and orderly dithered halftone images. This inability of exploiting a prior knowledge on the halftone images largely weakens the flexibility, adaptability, and effectiveness of the inverse halftoning techniques, making the study on the classification of halftone images imperative for not only optimizing the existing inverse halftoning schemes, but also guiding the establishment of adaptive schemes on halftone image compression, halftone image watermarking, and so forth.

Motivated by observing the significance of classifying halftone images, several halftone image classification methods have been proposed. In 1998, Chang and Yu [5] classified halftone images into four types using an enhanced one-dimensional correlation function and a backpropagation (BP) neural network, for which the data sets in the experiments are limited to the halftone images produced by clustered-dot ordered dithering, dispersed-dot ordered dithering, constrained average, and error diffusion. Kong et al. [6, 7] used an enhanced one-dimensional correlation function and a gray level cooccurrence matrix to extract features from halftone images, based on which the halftone images are divided into nine categories using a decision tree algorithm. Liu et al. [8] combined support region and least mean square (LMS) algorithm to divide halftone images into four categories. Subsequently, they [9] used LMS to extract features from Fourier spectrum in nine categories of

halftone images and classify these halftone images using naive Bayes. Although these methods work well in classifying some specific halftone images, their performance largely decreases when classifying error-diffused halftone images produced by Floyd-Steinberg filter, Stucki filter, Sierra filter, Burkers filter, Jarvis filter, and Stevenson filter, respectively. They are described as follows.

Different Error Diffusion Filters. Consider the following:

(a) Floyd-Steinberg filter:

$$\left(\frac{1}{16}\right)\begin{array}{ccc} & \bullet & 7 \\ 3 & 5 & 1 \end{array} \tag{1}$$

(b) Sierra filter:

$$\left(\frac{1}{32}\right)\begin{array}{ccccc} & & \bullet & 5 & 3 \\ 2 & 4 & 5 & 4 & 2 \\ & 0 & 2 & 3 & 2 \end{array} \tag{2}$$

(c) Burkers filter:

$$\left(\frac{1}{32}\right)\begin{array}{ccccc} & & \bullet & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \end{array} \tag{3}$$

(d) Jarvis filter:

$$\left(\frac{1}{48}\right)\begin{array}{ccccccc} & & & \bullet & 7 & 5 \\ 3 & 5 & 7 & 5 & 3 \\ & 1 & 3 & 5 & 3 & 1 \end{array} \tag{4}$$

(e) Stevenson filter:

$$\left(\frac{1}{200}\right)\begin{array}{ccccccccc} & & & & \bullet & & 32 & & \\ 12 & & 26 & & 30 & & 16 & & \\ & 12 & & 26 & & 12 & & & \\ 5 & & 12 & & 12 & & 5 & & \end{array} \tag{5}$$

The Error Diffusion of Stucki

(a) Error kernel of Stucki filter is

$$\left(\frac{1}{42}\right)\begin{array}{ccccc} & & \bullet & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{array} \tag{6}$$

(b) Matrix of coefficients template is

$$\begin{array}{|c|c|c|c|c|} \hline 0 & 0 & \underline{0} & 8 & 4 \\ \hline 2 & 4 & 8 & 4 & 2 \\ \hline 1 & 2 & 4 & 2 & 1 \\ \hline \end{array} \tag{7}$$

(c) O denotes the pixel being processed; A, B, C, and D indicate the four neighborhood pixels:

$$\left(\frac{1}{42}\right)\begin{array}{ccccc} & & O & A & C \\ \hline 0 & 0 & \underline{0} & 8 & 4 \\ \hline 2 & 4 & 8 & 4 & 2 \\ \hline 1 & 2 & 4 & 2 & 1 \\ \hline \end{array}\begin{array}{l} \\ \\ \\ \\ \rightarrow B \\ \\ \downarrow D \end{array} \tag{8}$$

based on different error diffusion kernels, as summarized in [10–12]. Moreover, these literatures did not consider all types of error diffusion halftone images. For example, only three error diffusion filters are included in [6, 7, 9] and only one is involved in [5, 8]. The idea of halftoning for the six error diffusion filters is quite similar, with the only difference lying in the templates used (shown in different error diffusion filters and the error diffusion of Stucki described above; the templates are shown at the right-hand side in each equation). It is difficult to classify the error-diffused halftone images because of the almost inconspicuous differences among various halftone features extracted from, using these six error diffusion filters, the error-diffused halftone images. However, as a scalable algorithm, the error diffusion has gradually become one of the most popular techniques, due to its ability to provide a solution of good quality at a reasonable cost [13]. This asks for an urgent requirement to study the classification mechanism for various error diffusion algorithms, with the hope to promote the existing inverse halftone techniques widely used in different application fields of graphics processing.

This paper proposes a new algorithm to classify error-diffused halftone images. We first extract the feature matrices of pixel pairs from the error-diffused halftone image patches, according to statistical characteristics of these patches. The class feature matrices are then subsequently obtained, using a gradient descent method, based on the feature matrices of pixel pairs [14]. After applying the spectral regression kernel discriminant analysis to realize the dimension reduction in the class feature matrices, we finally classify the error-diffused halftone images using the idea similar to the nearest centroids classifier [15, 16].

The structure of this paper is as follows. Section 2 presents the method of kernel discriminant analysis. Section 3 describes how to extract the feature extraction of pixel pairs from the error-diffused halftone images. Section 4 describes the proposed classification method for the error-diffused halftone image based on the spectral regression kernel discriminant analysis. Section 5 shows the experimental results. Some concluding remarks and possible future research directions are given in Section 6.

2. An Efficient Kernel Discriminant Analysis Method

It is well known that linear discriminant analysis (LDA) [17, 18] is effective in solving classification problems, but it

fails for nonlinear problems. To deal with this limitation, the approach called kernel discriminant analysis (KDA) [19] has been proposed.

2.1. Overview of Kernel Discriminant Analysis. Suppose that we are given a sample set $\{x_1, x_2, \ldots, x_m\}$ of the error-diffused halftone images with $x_1, x_2, \ldots, x_m$ belonging to different class C_i ($i = 1, 2, \ldots, K$), respectively. Using a nonlinear mapping function ϕ, the samples of the halftone images in the input space R^n can be projected to a high-dimensional separable feature space f; namely, $R^n \to f$, $x \to \phi(x)$. After extracting features, the error-diffused halftone images will be classified along the projection direction along which the within-class scatter is minimal and the between-class scatter is maximal. For a proper ϕ, an inner product described as $\langle \cdot, \cdot \rangle$ can be defined on f to form a reproducing kernel Hilbert space. That is to say, $\langle \phi(x_i), \phi(x_j) \rangle = \kappa(x_i, x_j)$, where $\kappa(\cdot, \cdot)$ is a positive semidefinite kernel function. In the feature space f, let S_b^ϕ be the between-class scatter matrix, let S_w^ϕ be the within-class scatter matrix, and let S_t^ϕ be the total scatter matrix:

$$
\begin{aligned}
S_b^\phi &= \sum_{k=1}^{c} m_k \left(\mu_\phi^k - \mu_\phi\right)\left(\mu_\phi^k - \mu_\phi\right)^T, \\
S_w^\phi &= \sum_{k=1}^{c} \left(\sum_{i=1}^{m_k} \left(\phi\left(x_i^k\right) - \mu_\phi^k\right)\left(\phi\left(x_i^k\right) - \mu_\phi^k\right)^T \right), \\
S_t^\phi &= S_b^\phi + S_w^\phi = \sum_{i=1}^{m} \left(\phi\left(x_i\right) - \mu_\phi\right)\left(\phi\left(x_i\right) - \mu_\phi\right)^T,
\end{aligned}
\tag{9}
$$

where m_k is the number of the samples in the kth class, $\phi(x_i^k)$ is the ith sample of the kth class in the feature space f, $\mu_\phi^k = (1/m_k) \sum_{i=1}^{N_i} \phi(x_i^k)$ is the centroid of the kth class, and $\mu_\phi = (1/N) \sum_{k=1}^{c} \sum_{i=1}^{m_k} \phi(x_i^k)$ is the global centroid. In the feature space, the aim of the discriminant analysis is to seek the best projection direction, namely, the projective function v to maximize the following objective function:

$$
v_{\text{opt}} = \arg\max \frac{v^T S_b^\phi v}{v^T S_w^\phi v}. \tag{10}
$$

Equation (10) can be solved by the eigenproblem $S_b^\phi v = \lambda S_t^\phi v$. According to the theory of reproducing kernel Hilbert space, we know that the eigenvectors are linear combinations of $\phi(x_i)$ in the feature space f: there exist weight coefficients a_i ($i = 1, 2, \ldots, m$) such that $v^\phi = \sum_{i=1}^{m} a_i \phi(x_i)$. Let $a = [a_1, a_2, \ldots, a_m]^T$; then it can be proved that (10) can be rewritten as follows:

$$
a_{\text{opt}} = \arg\max \frac{a^T KWKa}{a^T KKa}. \tag{11}
$$

The optimization problem of (11) is equal to the eigenproblem

$$
KWKa = \lambda KKa, \tag{12}
$$

where K is the kernel matrix and $K_{ij} = \kappa(x_i, x_j)$; W is the weight matrix defined as follows:

$$
w_{ij} = \begin{cases} \dfrac{1}{m_k}, & \text{if } x_i, x_j \text{ belong to the } k\text{th class}, \\ 0, & \text{others}. \end{cases} \tag{13}
$$

For sample x, the projective function in the feature space f can be described as

$$
\begin{aligned}
f^*(x) &= \langle v, \phi(x) \rangle = \sum_{i=1}^{m} a_i \langle \phi(x_i), \phi(x) \rangle \\
&= \sum_{i=1}^{m} a_i \kappa(x_i, x) = a^T \kappa(\cdot, x).
\end{aligned}
\tag{14}
$$

2.2. Kernel Discriminant Analysis via Spectral Regression. To efficiently solve the eigenproblem of the kernel discriminant analysis in (12), the following theorem will be used.

Theorem 1. *Let y be the eigenvector of the eigenproblem $Wy = \lambda y$ with eigenvalue λ. If $K\alpha = y$, then α is the eigenvector of eigenproblem (12) with the same eigenvalue λ.*

According to Theorem 1, the projective function of the kernel discriminant analysis can be obtained according to the following two steps.

Step 1. Obtain y by solving the eigenproblem in (12).

Step 2. Search eigenvector α which satisfies $K\alpha = y$, where K is the positive semidefinite kernel matrix.

As we know, if K is nonsingular, then, for any given y, there exists a unique $\alpha = K^{-1} y$ satisfying the linear equation described in Step 2. If K is singular, then, the linear equation may have infinite solutions or have no solution. In this case, we can approximate α by solving the following equation:

$$
(K + \delta I)\alpha = y, \tag{15}
$$

where $\delta \geq 0$ is a regularization parameter and I is the identity matrix. Combined with the projective function described in (14), we can easily verify that the solution $\alpha^* = (K + \delta I)^{-1} y$ given by (15) is the optimal solution of the following regularized regression problem:

$$
\alpha^* = \min_{f \in F} \sum_{i=1}^{m} \left(f\left(x_i\right) - y_i\right)^2 + \delta \left\| f \right\|_K^2, \tag{16}
$$

where y_i is the ith element of y and F is the reproducing kernel Hilbert space induced from the Mercer kernel K with $\| \cdot \|_K$ being the corresponding norm. Due to the essential combination of the spectral analysis and regression techniques in the above two-step approach, the method is named as spectral regression (SR) kernel discriminant analysis.

3. Feature Extraction of the Error-Diffused Halftone Images

Since its introduction in 1976, the error diffusion algorithm has attracted widespread attention in the field of printing applications. It deals with pixels of halftone images using, instead of point processing algorithms, the neighborhood processing algorithms. Now we will extract the features of the error-diffused halftone images which are produced using the six popular error diffusion filters mentioned in Section 1.

3.1. Statistic Characteristics of the Error-Diffused Halftone Images. Assume that $f_g(x, y)$ is the gray value of the pixel located at position (x, y) in the original image and $F(x, y)$ is the value of the pixel located at position (x, y) in the error-diffused halftone image. For the original image, all the pixels are firstly normalized to the range $[0, 1]$. Then, the pixels of the normalized image are converted to the error-diffused image F line by line; that is to say, if $f_g(x, y) \geq T_0$, $F(x, y)$, which is the value of the pixel located at position (x, y) in error-diffused image F, is 1; otherwise, $F(x, y)$ is 0, where T_0 is the threshold value. The error between $F(x, y)$ and T_0 is diffused ahead to some subsequent pixels not necessary to deal with. Therefore, for some subsequent pixels, the comparison will be implemented between T_0 and the value which is the sum of $f_g(x, y)$ and the diffusion error e. A template matrix can be built using the error diffusion modes and the error diffusion coefficients, as shown in the error diffusion of Stucki described above, for example, (a) the error diffusion filter and (b) the error diffusion coefficients which represent the proportion of the diffusion errors. If the coefficient is zero, then the corresponding pixel does not receive any diffusion errors. According to the error diffusion of Stucki described above, pixel A suffers from more diffusion errors than pixel B; that is to say, O-A has a larger probability to become 1-0 pixel pair than O-B. The reasons are as follows. Suppose that the pixel value of O is $0 \leq p_O \leq 1$, and pixel O has been processed by the thresholding method according to the following equation:

$$q_O = \begin{cases} 1, & p_O \geq T_0, \\ 0, & \text{else.} \end{cases} \tag{17}$$

In general, threshold T_0 is set as 0.5. According to the template shown in the error diffusion of Stucki described above, we can know that the diffusion error is $e = q_O - p_O$, the new value of pixel A is $p_A = f(x_A + y_A) + 8e/42$, and the new value of pixel B is $p_B = f(x_B + y_B) + e/42$, where $f_g(x_A + y_A)$ and $f_g(x_B + y_B)$ are the original values of pixels A and B, respectively.

Since the value of each pixel in the error-diffused halftone image can only be 0 or 1, there are 4 kinds of pixel pairs in the halftone image: 0-1, 0-0, 1-0, and 1-1. Pixel pairs 0-1 and 1-0 are collectively known as 1-0 pixel pairs because of their exchange ability. Therefore, there are only three kinds of pixel pairs essentially: 0-0, 1-0, and 1-1. In this paper, three statistical matrices are used to store the number of different pixel pairs with different neighboring distances and different directions, which are of size $L \times L$ and are referred to as M_{00}, M_{10}, and M_{11}, respectively (L is an odd number satisfying $L = 2R + 1$ and R is the maximum neighboring distance). Suppose that the center entry of the statistical matrix template covers pixel O of the error-diffused halftone image with the size $W * H$, and other entries overlap other neighborhood pixels $F(u, v)$. Then, we can compute three statistics on 1-0, 1-1, and 0-0 pixel pairs within the scope of this statistics matrix template. If the position (i, j) of pixel O changes continually, the matrices M_{00}, M_{10}, and M_{11} with zero being the initial values can be updated according to

$$\begin{aligned} M_{11}(x, y) &= M_{11}(x, y) + 1 \\ &\qquad \text{if } F(i, j) = 1,\ F(u, v) = 1, \\ M_{00}(x, y) &= M_{00}(x, y) + 1 \\ &\qquad \text{if } F(i, j) = 0,\ F(u, v) = 0, \\ M_{10}(x, y) &= M_{00}(x, y) + 1 \\ &\text{if } F(i, j) = 1,\ F(u, v) = 0 \text{ or } F(i, j) = 0,\ F(u, v) = 1, \end{aligned} \tag{18}$$

where $u = i - R + x - 1$, $v = j - R + y - 1$, $1 \leq i, u \leq W$, $1 \leq j, v \leq H$, $1 \leq x \leq L$, and $1 \leq y \leq L$. After normalization, the three statistic matrices can be ultimately obtained as the statistical feature descriptor of the error-diffused halftone images.

3.2. Process of Statistical Feature Extraction of Halftone Images. According to the analysis described above, the process of statistical feature extraction of the error-diffused halftone images can be represented as follows.

Step 1. Input the error-diffused halftone image F, and divide F into several blocks B_i $(i = 1, 2, \ldots, H)$ with the same size $K \times K$.

Step 2. Initialize the statistical feature matrix M (including M_{00}, M_{10}, M_{11}) as the zero matrix, and let $i = 1$.

Step 3. Obtain the statistical matrix M_i of block B_i according to (18), and update M using the equation $M = M + M_i$.

Step 4. Set $i = i + 1$. If $i \leq H$, then return to Step 3. Otherwise, go to Step 5.

Step 5. Normalize M such that it satisfies $\sum_{x=1}^{L} \sum_{x=1}^{L} M(x, y) = 1$ and $M(x, y) \geq 0$, where K satisfies $K \geq \min(T_i)$ and T_i $(1 \leq i \leq 6)$ represents the size of the template matrix of the ith error diffusion filter.

According to the process described above, we know that the statistical features of the error-diffused halftone image F are extracted based on the method that divides F into image patches, which is significantly different with other feature extraction methods based on image patches. For example, in [20], the brightness and contrast of the image patches are normalized by Z-score transformation, and whitening (also called "sphering") is used to rescale the normalized data to remove the correlations between nearby pixels (i.e., low-frequency variations in the images) because

these correlations tend to be very strong even after brightness and contrast normalization. However, in this paper, features of the patches are extracted based on counting statistical measures of different pixel pairs (0/0, 1/0, and 1/1) within a moving statistical matrix template and are optimized using the method described in Section 3.3.

3.3. Extraction of the Class Feature Matrix. The statistics matrices $M_{00}^i, M_{10}^i, M_{11}^i$ $(i = 1, 2, \ldots, N)$, after being extracted, can be used as the input of other algorithms, such as support vector machines and neural networks. However, the curse of dimensionality could occur, due to the high dimension of M_{00}, M_{10}, M_{11}, making the classification effect possibly not significant. Thereby, six class feature matrices $G_1, G_2, \ldots, G_6$ are designed in this paper for the error-diffused halftone images produced by the six error diffusion filters mentioned above. Then, a gradient descent method can be used to optimize these class feature matrices.

$N = 6 \times n$ error-diffused halftone images can be derived from n original images using the six error diffusion filters, respectively. Then, N statistics matrices M_i $(M_{00}^i, M_{10}^i, M_{11}^i)$ $(i = 1, 2, \ldots, N)$ can be extracted as the samples from N error-diffused halftone images using the algorithm mentioned in Section 3.2. Subsequently, we label these matrices as label(M_1), label(M_2), ..., label(M_N) to denote the types of the error diffusion filters used to produce the error-diffused halftone image. Given the ith sample M_i as the input, the target out vector $t_i = [t_{i1}, \ldots, t_{i6}]$ $(i = 1, \ldots, N)$, and the class feature matrices $G_1, G_2, \ldots, G_6$, the square error e_i between the actual output and the target output can be derived according to

$$e_i = \sum_{j=1}^{6} \left(t_{ij} - M_i \bullet G_j\right)^2, \tag{19}$$

where

$$t_{ij} = \begin{cases} 1 & \text{if } j = \text{label}\left(M_i\right), \ j = 1, 2, \ldots, 6 \\ 0 & \text{else.} \end{cases} \tag{20}$$

The derivatives of $G_j(x, y)$ in (19) can be explicitly calculated as

$$\frac{\partial e_i}{\partial G_j(x, y)} = -2\left(v_{ij} - M_i \bullet G_j\right) M_i(x, y), \tag{21}$$

where $1 \le x \le L$, $1 \le y \le L$, and $\bullet$ is the dot product of matrices defined, for any matrices A and B with the same size $C \times D$, as

$$A \bullet B = \sum_{u=1}^{C} \sum_{v=1}^{D} A(u, v) \times B(u, v). \tag{22}$$

The dot product of matrices satisfies the commutative law and associative law; that is to say, $A \bullet B = B \bullet A$ and $(A \bullet B) \bullet C = A \bullet (B \bullet C)$. Then, the iteration equation (23) can be obtained using the gradient descent method:

$$\begin{aligned} G_j^{k+1}(x, y) &= G_j^k(x, y) - \eta \frac{\partial e_i}{\partial G_j^k(x, y)} \\ &= G_j^k(x, y) \\ &\quad + 2\eta\left(v_{ij} - M_i \bullet G_j\right) M_i(x, y), \end{aligned} \tag{23}$$

where η is the learning factor and k means the kth iteration. The purpose of learning is to seek the optimal matrices G_j $(j = 1, 2, \ldots, 6)$ by minimizing the total square error $e = \sum_{i=1}^{N} e_i$, and the process of seeking the optimal matrices G_j can be described as follows.

Step 1. Initialize parameters: initialize the numbers of iterations *inner* and *outer*, the iteration variables $t = 0$ and $i = 1$, the nonnegative thresholds ε_1 and ε_2 used to indicate the end of iterations, the learning factor η, the total number of samples N, and the class feature matrices G_j $(j = 1, 2, \ldots, 6)$.

Step 2. Input the statistics matrices M_i $(i = 1, 2, \ldots, N)$, and let $G_j^0 = G_j$ $(j = 1, \ldots, 6)$ and $k = 0$. The following three substeps are executed.

(1) According to (23), G_j^{k+1} $(j = 1, \ldots, 6)$ can be computed.

(2) Compute e_i^{k+1} and $\Delta e_i^{k+1} = |e_i^{k+1} - e_i^k|$.

(3) If $k >$ inner or $\Delta e_i^{k+1} < \varepsilon_1$, then set $e_i = e_i^{k+1}$ and go to Step 3; otherwise, set $k = k + 1$ and return to (10).

Step 3. Set $G_j = G_j^{k+1}$ $(j = 1, \ldots, 6)$. If $i = N$, then go to Step 4. Otherwise, set $i = i + 1$, and go to Step 2.

Step 4. Compute the total error $e^t = \sum_{i=1}^{N} e_i$. If $|e^t - e^{t-1}| < \varepsilon_2$ or $t =$ *outer*, then end the algorithm. Otherwise, set $t = t + 1$ and go to Step 2.

4. Classification of Error-Diffused Halftone Images Using Nearest Centroids Classifier

This section describes the details on classifying error-diffused halftone images using the spectral regression kernel discriminant analysis as follows.

Step 1. Input N error-diffused halftone images produced by six error-diffused filters, and extract the statistical feature matrices M_i, including M_{00}^i, M_{10}^i, M_{11}^i, $i = 1, 2, \ldots, N$, using the method presented in Section 3.2.

Step 2. According to the steps described in Section 3.3, all the statistical feature matrices M_i $(M_{00}^i, M_{10}^i, M_{11}^i$ and $i = 1, 2, \ldots, N)$ are converted to the class feature matrices G_i, including $G_{00}^i, G_{10}^i, G_{11}^i$ and $i = 1, 2, \ldots, N$, correspondingly. Then, convert $G_{00}^i, G_{10}^i, G_{11}^i$ $(i = 1, 2, \ldots, N)$ into one-dimensional matrices by columns, respectively.

Step 3. All the one-dimensional class feature matrices $G_{00}^i, G_{10}^i, G_{11}^i$ ($i = 1, 2, \ldots, N$) are used to construct the samples feature matrices F_{00}, F_{10}, F_{11} of size $N \times M$ ($M = L \times L$), respectively.

Step 4. A label matrix *information* of the size $1 \times N$ is built to record the type to which the error-diffused halftone images belong.

Step 5. The first m features $G_{00}^1, G_{00}^2, \ldots, G_{00}^m$ of the samples feature matrices F_{00} are taken as the training samples (the first m features of F_{10}, F_{11}, or F_{all} which is the composition of F_{00}, F_{10}, and F_{11} also can be used as the training samples). Reduce the dimension of these training samples using the spectral regression discriminant analysis. The process of dimension reducing can be described by three substeps as follows.

(1) Produce orthogonal vectors. Let

$$y_k = \left[\underbrace{0, 0, \ldots, 0}_{\sum_{i=1}^{k-1} m_i}, \underbrace{1, 1, \ldots, 1}_{m_k}, \underbrace{0, 0, \ldots, 0}_{\sum_{i=k+1}^{c} m_i}\right]^T \quad k = 1, 2, \ldots, c \tag{24}$$

and $y_0 = [1, 1, \ldots, 1]^T$ be the matrix with all elements being 1. Let y_0 be the first vector of the weight matrix W, and use Gram-Schmidt process to orthogonalize the other eigenvectors. Remove the vector y_0, leaving $c-1$ eigenvectors of W denoted as follows:

$$\{\overline{y}_k\}_{k=1}^{c-1}, \quad \left(\overline{y}_i^T y_0 = 0,\ \overline{y}_i^T \overline{y}_j = 0,\ i \neq j\right). \tag{25}$$

(2) Add an element of 1 to the end of each input data (G_{00}^i ($i = 1, 2, \ldots, m$)), and obtain $c-1$ vectors $\{\vec{a}_k\}_{k=1}^{c-1} \in \mathfrak{R}^{n+1}$, where $\vec{a}_k$ is the solution of the following regular least squares problem:

$$\vec{a}_k = \arg\min \left(\sum_{i=1}^{m} \left(\vec{a}^T G_{00}^i - \overline{y}_i^k\right)^2 + \alpha \|\vec{a}\|^2 \right). \tag{26}$$

Here $\overline{y}_i^k$ is the ith element of eigenvectors $\{\overline{y}_k\}_{k=1}^{c-1}$ and α is the contraction parameter.

(3) Let $A = [\overline{a}_1, \ldots, \overline{a}_{c-1}]$ be the transformation matrix with the size $(M+1) \times (c-1)$. Perform dimension reduction on the sample features $G_{00}^1, G_{00}^2, \ldots, G_{00}^m$, by mapping them to $(c-1)$ dimensional subspace as follows:

$$x \longrightarrow z = A^T \begin{bmatrix} x \\ 1 \end{bmatrix}. \tag{27}$$

Step 6. Compute the mean values of samples in different classes according to following equation:

$$\text{aver}_i = \frac{\sum_{s=1}^{m_i} G_{00}^s}{m_i} \quad (i = 1, 2, \ldots, 6), \tag{28}$$

where m_i is the number of samples in the ith class; $\sum_{i=1}^{6} m_i = m$. Let aver_i be the class-centroid of the ith class.

Step 7. The remaining $N - m$ samples are taken as the testing samples, and the dimension reduction is implemented for them using the method described in Step 5.

Step 8. Compute the square of the distance $|G_{00}^s - \text{aver}_i|^2$ ($s = m+1, m+2, \ldots, N$ and $i = 1, 2, \ldots, 6$) between each testing sample G_{00}^s and different class-centroid aver_i, according to the nearest centroids classifier; the sample G_{00}^s is assigned to the class i if $i = \arg\min |G_{00}^s - \text{aver}_i|^2$.

In Step 8, the weak classifier (i.e., the nearest centroid classifier) is used to classify error-diffused halftone images, because this classifier is simple and easy to implement. Simultaneously, in order to prove that these class feature matrices, which are extracted according to the method mentioned in Section 3 and handled by the algorithm of the spectral regression discriminant analysis, are well suited for the classification of error-diffused halftone images, this weak classifier is used in this paper instead of a strong classifier [20], such as support vector machine classifiers and deep neural network classifiers.

5. Experimental Analysis and Results

We implement various experiments to verify the efficiency of our methods in classifying error-diffused halftone images. The computer processor is Intel(R) Pentium(R) CPU G2030 @3.00 GHz, the memory of the computer is 2.0 GB, the operating system is Windows 7, and the experimental simulation software is matlab R2012a. In our experiments, all the original images are downloaded from http://decsai.ugr.es/cvg/dbimagenes/ and http://msp.ee.ntust.edu.tw/. About 4000 original images have been downloaded and they are converted into 24000 error-diffused halftone images produced by six different error-diffused filters.

5.1. Classification Accuracy Rate of the Error-Diffused Halftone Images

5.1.1. Effect of the Number of the Samples. This subsection analyzes the effect of the number on the feature samples on classification. When $L = 11$, and feature matrices F_{00}, F_{10}, F_{11}, F_{all} are taken as the input data, respectively, the accuracy rate of classification under different conditions is shown in Tables 1 and 2. Table 1 shows the classification accuracy rates under different number of training samples, when the total number of samples is 12000. Table 2 shows the classification accuracy rates under different number of training samples, when the total number of samples is 24000. The digits in the first line of each table are the size of the training samples.

According to Tables 1 and 2, the classification accuracy rates under 12000 samples are higher than that under 24000 samples. Moreover, the classification accuracy rates improve with the increase of the proportion of the training samples when the number of the training samples is lower than the 80% of sample size. And it achieves the highest classification accuracy rates when the number of the training samples is

Table 1: Classification accuracy rates (%) (12000 samples).

	2000	3000	4000	5000	6000	7000	8000	9000	10000
F_{00}	97.750	97.500	99.050	99.150	99.150	99.150	99.100	98.950	98.900
F_{10}	99.950	99.950	100.00	100.00	100.00	100.00	100.00	100.00	100.00
F_{11}	99.500	99.250	99.350	99.400	99.200	99.350	99.300	99.300	99.450
F_{all}	99.100	99.100	99.450	99.500	99.650	99.650	99.650	99.650	99.600

Table 2: Classification accuracy rates (%) (24000 samples).

	14000	15000	16000	17000	18000	19000	20000	21000	22000
F_{00}	95.810	95.878	95.413	95.286	94.933	98.580	98.100	98.267	98.650
F_{10}	97.600	97.378	97.150	96.957	96.450	99.680	99.600	99.533	99.300
F_{11}	96.420	96.256	95.925	95.614	95.000	99.040	98.825	98.700	98.400
F_{all}	97.230	97.100	96.888	96.686	96.233	99.680	99.600	99.533	99.350

Table 3: Classification accuracy rates obtained by different algorithms (%).

F_{00} + SR	F_{10} + SR	F_{11} + SR	F_{all} + SR	ECF + BP	LMS + Bayes	G_{10} + ML	G_{11} + ML
98.40	99.53	98.74	99.54	90.04	90.38	98.32	96.15

Table 4: The training and testing time under different sample sizes (in seconds).

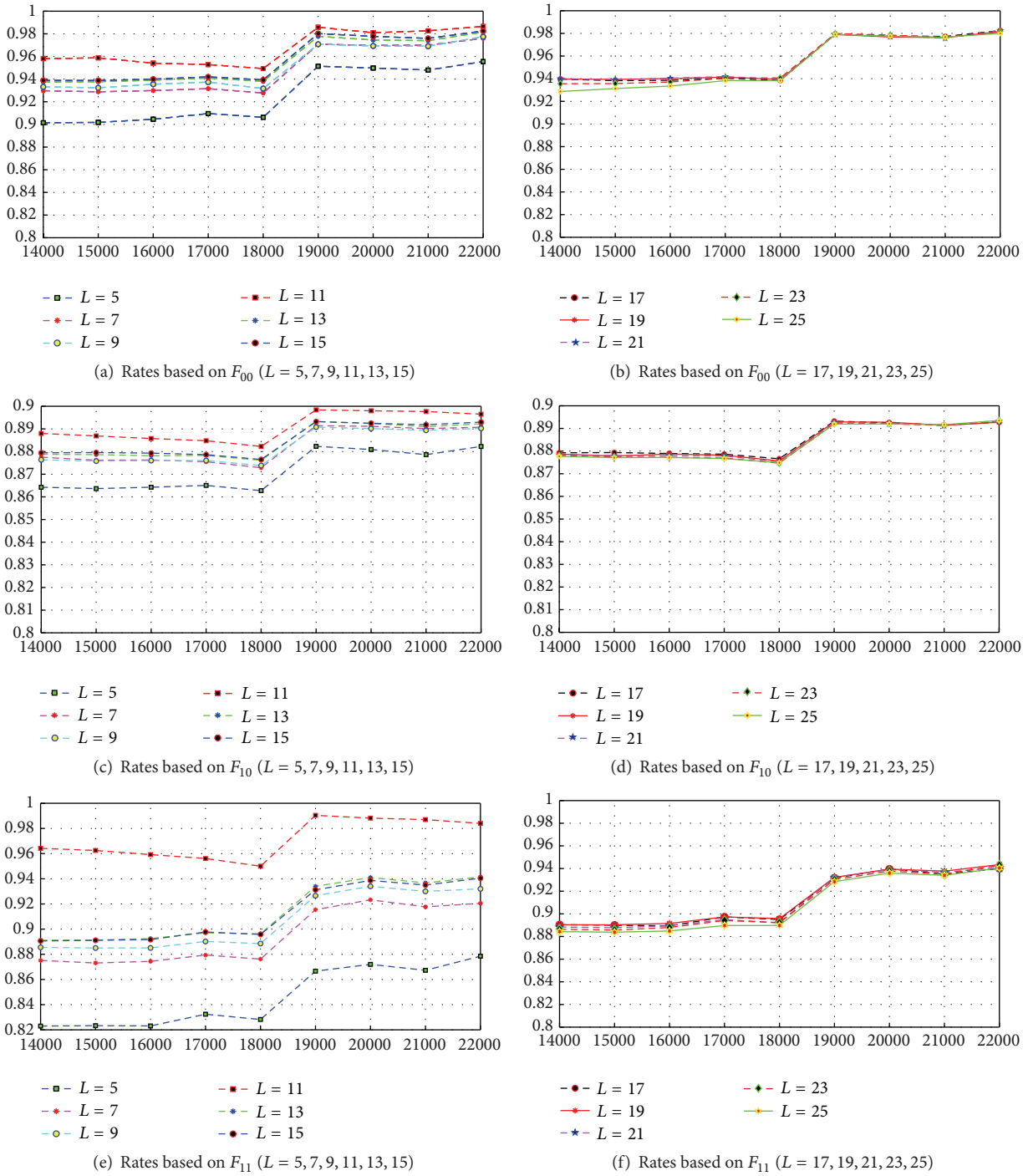
		14000	15000	16000	17000	18000	19000	20000	21000	22000
F_{00}	Training time	65.208	78.317	94.660	110.47	129.32	150.18	178.94	205.11	513.54
	Testing time	6.3742	6.2583	5.8530	5.5251	4.8792	4.4947	3.7872	3.0375	4.3393
F_{10}	Training time	65.851	78.796	95.113	110.70	130.97	150.76	173.98	198.95	240.10
	Testing time	6.3674	6.0804	5.7496	5.503	4.9287	4.3261	3.6775	3.0353	2.9631
F_{11}	Training time	65.662	79.936	95.173	110.69	130.33	150.63	174.41	198.34	239.96
	Testing time	6.5226	6.1817	5.7826	5.3778	4.8531	4.5513	3.7137	2.9735	3.0597
F_{all}	Training time	68.902	82.554	99.189	115.27	135.34	156.71	179.94	204.85	288.36
	Testing time	10.200	9.8393	9.3112	8.7212	8.016	6.902	5.8556	4.9363	3.6921

about 80% of the sample size. In addition, from Tables 1 and 2, we can also see that F_{00}, F_{10}, F_{11} can be used as the input data alone. Simultaneously, they can also be combined into the input data F_{all}, based on which the classification accuracy rates would be high.

5.1.2. Comparison of Classification Accuracy Rate. To analyze the effectiveness of our classification algorithm, the mean values of classification accuracy rate of the four data sets on the right-hand side of each row in Table 2 are computed. The algorithm SR outperforms other baselines in achieving higher classification accuracy rates, when compared with LMS + Bayes (the method composed of least mean square and Bayes method), ECF + BP (the method based on the enhanced correlation function and BP neural network), and ML (the maximum likelihood method). According to Table 3, the mean values of classification accuracy rates obtained using SR and different features F_{00}, F_{10}, F_{11}, and F_{all}, respectively, are higher than the mean values obtained by other algorithms mentioned above.

5.2. Effect of the Size of Statistical Feature Template on Classification Accuracies. Here, different features F_{00}, F_{10}, F_{11} of the 24000 error-diffused halftone images are used to test the effect of the size of statistical feature template. F_{00}, F_{10}, F_{11} are constructed using the corresponding class feature matrices G_{00}, G_{10}, G_{11} with deferent size $L \times L$ ($L = 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25$). Figure 1 shows that the classification accuracy rate achieves the highest value when $L = 11$, no matter which feature is selected for experiments.

5.3. Time-Consumption of the Classification. Now, the experiments are implemented to test the time-consumption of the error-diffused halftone image classification under the condition that the total number of samples is 24000 and $L = 11$. It is well known that the time-consumption of the classification includes the training time and the testing time. From Table 4, we can know that the training time increases with the increase of the number of training samples; on the contrary, the testing time decreases with the increase of the number of training samples.



(a) Rates based on F_{00} ($L = 5, 7, 9, 11, 13, 15$)

(b) Rates based on F_{00} ($L = 17, 19, 21, 23, 25$)

(c) Rates based on F_{10} ($L = 5, 7, 9, 11, 13, 15$)

(d) Rates based on F_{10} ($L = 17, 19, 21, 23, 25$)

(e) Rates based on F_{11} ($L = 5, 7, 9, 11, 13, 15$)

(f) Rates based on F_{11} ($L = 17, 19, 21, 23, 25$)

Figure 1: Classification accuracy rates based on deferent features with deferent sample size.

To compare the time-consumption of the classification method proposed in this paper with other algorithms, such as the backpropagation (BP) neural network, radial basis function (RBF) neural network, and support vector machine (SVM), all the experiments are implemented using F_{10}, which is divided into two parts: 12000 training samples and 12000 testing samples ($L = 11$). From the digits listed in Table 5, we can know that SR achieves the minimal summation of the training and testing time.

In addition, Table 5 implies that the time-consumption of classifiers based on neural networks, such as the classifier based on RBF or BP, are much more than that of other algorithms, especially SR. This is because these neural network algorithms essentially use gradient descent methods to

Table 5: Time-consumption of different algorithms (in second).

	ML	RBF	BP	SVM	SR
Training time	110.42	76.390	2966.7	49.297	43.9337
Testing time	1.2000	276.00	3.6000	62.400	6.3161

Table 6: Classification accuracy rates under different variances (%).

Variance	14000	15000	16000	17000	18000	19000	20000	21000	22000
0.01	95.570	95.422	95.425	95.529	95.217	98.440	98.200	98.033	97.950
0.10	95.370	95.233	95.188	95.157	94.867	98.200	98.000	97.767	98.000
0.50	94.580	94.522	94.600	94.671	94.567	98.320	98.225	98.033	98.550
1.00	93.060	93.067	93.013	93.343	93.200	96.980	97.175	96.800	97.350

Table 7: Classification accuracy rates using other algorithms under different variances.

Variance	ECF + BP	LMS + Bayes	ML	Variance	ECF + BP	LMS + Bayes	ML
0.01	89.5000	88.7577	97.9981	0.50	79.7333	40.2682	82.9169
0.10	88.1333	78.4264	96.8315	1.00	61.5000	33.8547	64.3861

optimize the associated nonconvex problems, which are well known to converge very slowly. However, the classifier based on SR performs the classification task through computing the square of the distance between each testing sample and different class-centroids directly. Hence, the time-consumption of it is very cheap.

5.4. The Experiment of Noise Attack Resistance. In the process of actual operation, the error-defused halftone images are often polluted by noise before the inverse transform. In order to test the ability of SR to resist the attack of noise, different Gaussian noises with mean 0 and different variances are embedded into the error-defused halftone images. Then classification experiments have been done using the algorithm proposed in this paper and the experimental results are listed in Table 6. According to Table 6, the accuracy rates decrease with the increase of the variances. As compared with the accuracy rates listed in Table 7 achieved by other algorithms, such as ECF + BP, LMS + Bayes, and ML, we find that our classification method has obvious advantages in resisting the noise.

6. Conclusion

This paper proposes a novel algorithm to solve the challenging problem of classifying the error-diffused halftone images. We firstly design the class feature matrices, after extracting the image patches according to their statistical characteristics, to classify the error-diffused halftone images. Then, the spectral regression kernel discriminant analysis is used for feature dimension reduction. The error-diffused halftone images are finally classified using an idea similar to the nearest centroids classifier. As demonstrated by the experimental results, our method is fast and can achieve a high classification accuracy rate with an added benefit of robustness in tackling noise. A very interesting direction is to solve the disturbance, possibly introduced by other attacks such as image scaling and rotation, in the process of error-diffused halftone image classification.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (Grants nos. 61170102, 61271140), the Scientific Research Fund of Hunan Provincial Education Department, China (Grant no. 15A049), the Education Department Fund of Hunan Province in China (Grants nos. 15C0402, 15C0395, and 13C036), and the Science and Technology Planning Project of Hunan Province in China (Grant no. 2015GK3024).

References

[1] Y.-M. Kwon, M.-G. Kim, and J.-L. Kim, "Multiscale rank-based ordered dither algorithm for digital halftoning," *Information Systems*, vol. 48, pp. 241–247, 2015.

[2] Y. Jiang and M. Wang, "Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter," *IET Image Processing*, vol. 8, no. 3, pp. 183–190, 2014.

[3] Z. Zhao, L. Cheng, and G. Cheng, "Neighbourhood weighted fuzzy c-means clustering algorithm for image segmentation," *IET Image Processing*, vol. 8, no. 3, pp. 150–161, 2014.

[4] Z.-Q. Wen, Y.-L. Lu, Z.-G. Zeng, W.-Q. Zhu, and J.-H. Ai, "Optimizing template for lookup-table inverse halftoning using elitist genetic algorithm," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 71–75, 2015.

[5] P. C. Chang and C. S. Yu, "Neural net classification and LMS reconstruction to halftone images," in *Visual Communications and Image Processing '98*, vol. 3309 of *Proceedings of SPIE*, pp. 592–602, The International Society for Optical Engineering, January 1998.

[6] Y. Kong, P. Zeng, and Y. Zhang, "Classification and recognition algorithm for the halftone image," *Journal of Xidian University*, vol. 38, no. 5, pp. 62–69, 2011 (Chinese).

[7] Y. Kong, *A study of inverse halftoning and quality assessment schemes [Ph.D. thesis]*, School of Computer Science and Technology, Xidian University, Xian, China, 2008.

[8] Y.-F. Liu, J.-M. Guo, and J.-D. Lee, "Inverse halftoning based on the Bayesian theorem," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1077–1084, 2011.

[9] Y.-F. Liu, J.-M. Guo, and J.-D. Lee, "Halftone image classification using LMS algorithm and naive Bayes," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2837–2847, 2011.

[10] D. L. Lau and G. R. Arce, *Modern Digital Halftoning*, CRC Press, Boca Raton, Fla, USA, 2nd edition, 2008.

[11] *Image Dithering: Eleven Algorithms and Source Code*, http://www.tannerhelland.com/4660/dithering-elevenalgorithms-source-code/.

[12] R. A. Ulichney, "Dithering with blue noise," *Proceedings of the IEEE*, vol. 76, no. 1, pp. 56–79, 1988.

[13] Y.-H. Fung and Y.-H. Chan, "Embedding halftones of different resolutions in a full-scale halftone," *IEEE Signal Processing Letters*, vol. 13, no. 3, pp. 153–156, 2006.

[14] Z.-Q. Wen, Y.-X. Hu, and W.-Q. Zhu, "A novel classification method of halftone image via statistics matrices," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4724–4736, 2014.

[15] D. Cai, X. He, and J. Han, "Efficient kernel discriminant analysis via spectral regression," in *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM '07)*, pp. 427–432, Omaha, Neb, USA, October 2007.

[16] D. Cai, X. He, and J. Han, "Speed up kernel discriminant analysis," *The International Journal on Very Large Data Bases*, vol. 20, no. 1, pp. 21–33, 2011.

[17] M. Zhao, Z. Zhang, T. W. S. Chow, and B. Li, "A general soft label based Linear Discriminant Analysis for semi-supervised dimensionality reduction," *Neural Networks*, vol. 55, pp. 83–97, 2014.

[18] M. Zhao, Z. Zhang, T. W. S. Chow, and B. Li, "Soft label based Linear Discriminant Analysis for image recognition and retrieval," *Computer Vision & Image Understanding*, vol. 121, no. 1, pp. 86–99, 2014.

[19] L. Zhang and F.-C. Tian, "A new kernel discriminant analysis framework for electronic nose recognition," *Analytica Chimica Acta*, vol. 816, pp. 8–17, 2014.

[20] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1403–1416, 2015.