

Textural Features Selection for Image Classification by Bayesian Method

T. Vo-Van*, H. Che-Ngoc†, T. Nguyen-Trang‡†

*Department of Mathematics, Can Tho University, Can Tho, Vietnam

†Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

‡Division of Computational Mathematics and Engineering Institute for Computational Science,

Ton Duc Thang University, Ho Chi Minh City, Vietnam

Email: nguyentrangthao@tdt.edu.vn

Abstract—This article proposes an algorithm to optimize the performance in texture classification by Bayesian method. Specifically, we extract several features from the Grey level co-occurrence matrices (GLCMs) with different distances d and directions θ . We then apply Genetic algorithm to select the suitable features that can minimize the error rate of using the cross validation set. This choice of features continues to be used for classifying test data. Three numerical examples performed with synthetic and real images show the superiority of proposed algorithm over some existing ones. They also present the feasibility and applicability of the proposed method for texture recognition, especially for some practical problems such as material and handwritten digit recognition.

I. INTRODUCTION

In the trend of globalization, we have to receive and process a huge amount of information which need to be archived, handled, summarized and classified in an effective way. In data processing, image recognition is a complicated problem and plays an important role in many areas such as hydrometeorology, environment, physics, etc. Although it has been interested by a lot of researchers in many different areas for a long time, it is quite challenging at present.

In content-based image recognition, we can classify images by three main approaches: color-based approach, texture-based approach and shape-based approach. Each method has its advantages and disadvantages. According to the image properties and analysis purposes, we can choose the suitable method. This paper focuses on the texture-based method that studies about the spatial arrangement of color or intensities in an image. In literature, we can find two main approaches that interested by many researchers: signal processing approach [1]–[3] and statistical approach [4]–[6]. In here comparison to the signal processing approach, statistical approach based on the grey level co-occurrence matrix (GLCM) is easy to use and spend less computational cost. This method, in fact, gives a high classification performance and is applied by many researchers. For example, Clausi [7] combined GLCM and Fisher method to classify natural textures, Bhogle and Patil [8] combined GLCM and Mahalanobis distance to detect oil spill, etc.

In general, the well-known procedure of texture-based image recognition is performed via three steps: (i) compute the GLCMs from each image, (ii) extract the features from

GLCMs and (iii) use the extracted features for classifying by a certain model. For (i), several GLCMs can be created based on different distances d and directions θ . For instance, Ayala [9] used $d = 2$ and $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, Celebi and Alpkocak [6] concluded that using "contrast" with $d = \{1, 2, 3, 4\}$ is the best, another popular choice is $d = 1$ and $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. For (ii), Haralick [4] determined 14 statistical measures that can be extracted from each GLCM, however, there are only 4 features which are contrast, correlation, homogeneity and energy have strongly effects on classification result [6], [10]. For (iii), we can mention a list of classifiers, such as Bayesian classifier [9], [11], Fisher classifier [7], Binary logistic regression [11] and Support-vector machine classifier [12].

Generally, if we have a cases of distance d , b cases of direction θ and c extracted features for each GLCM, we will receive abc Typical Variables for Texture (TVT) of image. However, it is not meaningful to say that the more variables in the model, the better result we will receive. Furthermore, an increase from the number of variables which are uncorrelated to the class of image may create noise and cause a poor performance. In texture-based image recognition, the choices of TVT in [6], [9], [10] and other researches are still based on experience, hence, their suitabilities are not ensured for all cases of image database. Therefore, this paper proposes a method to determine the suitable TVT which are then used in the classification model. Specifically, we use $d = \{1, 2, 3, 4\}$, $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ to create 16 GLCMs. We next extract the contrast, the correlation, the energy and the homogeneity from each GLCM to establish 64 extracted features that lead to 2^{64} ways to choose the TVT. Combining with Bayesian method, the "genetic algorithm for solving integer and mixed integer optimization problems", MI-LXPM [13], which is integrated in Global Optimization Toolbox of Matlab Software, is used for searching reasonable TVT. Specifically, the TVT are chosen so that the error rate in cross-validation set is minimized. Note that, besides the training and test sets, we use an additional cross validation set. The MI-LXPM is used to minimize the error in cross validation set instead of training set to avoid the overfitting problem. The selected TVT are continued to measure the performance via an image dataset for testing. Actually, in the proposed approach, well-known techniques including GLCM, GA and Bayesian classifier are

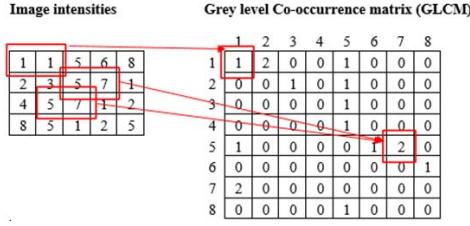


Fig. 1. The illustration for computing GLCM with $d = 1$ and $\theta = 0$

performed, however, the evolutionary approach is utilized to select textural features for the supervised learning, for the first time. The advantage is that the well-known techniques are always easier to read and understand for the first proposal. Certainly, any other metaheuristic and machine learning algorithm could have been used as well. That means the proposed method not only improves the current performance but also suggests a future promising research direction. Three numerical examples performed with synthetic and real images will present the superiority of proposed algorithm over other choices of TTVT in Bayesian classification.

The remainder of the article is arranged as follows. Section 2 summarizes some issues related to establishing GLCM, extracting features from GLCM and classifying by Bayesian method. The Genetic algorithm called MI-LXPM and the proposed method are presented in Section 3. The numerical examples are presented in Section 4 and the final section is destined for the conclusion of the article.

II. SOME RELATED ISSUES

A. Grey level co-occurrence matrix

The Grey Level Co-occurrence matrix (GLCM) presents the information about intensities of pixels and their neighbours, at fixed distance d and orientation θ . The idea is to scan the image and keep track of grey levels for each of two pixels separated within a fixed distance d and orientation θ . If we have an image of size $M \times N$ (M pixels in X -axis and N pixels in Y -axis) and G is the domain of grey level, then GLCM is a matrix P size of $G \times G$. Each element $p(i, j)$ presents the probability of the occurrence of intensity i and intensity j at fixed distance d and orientation θ . The formula to compute $p(i, j)$ is presented by (1). Figure 1 shows how to calculate several values in the GLCM of the 4-by-5 image I in case of $d = 1$ and $\theta = 0$. Element $(1, 1)$ in the GLCM contains the value 1 or $p_{1,0}(1, 1) = 1$ because there is only one instance in the image where two, horizontally adjacent pixels have the values 1 and 1. Similarly, $p_{1,0}(5, 7) = 2$ because there are two instances in the image where two, horizontally adjacent pixels have the values 5 and 7. We continue this processing to fill in all the values in the GLCM.

$$p_{d,\theta} = \# \{(x, y), (x', y') \in M \times N \mid d = \|(x, y), (x', y')\| \theta = \Theta((x, y), (x', y')), f(x, y) = i, f(x', y') = j\} \quad (1)$$

From GLCM, Haralick ([4]) defined 14 statistical measures that can be extracted. However, there are only 4 features

consisting of contrast, correlation, homogeneity and energy have strongly effects to classification result. These features are presented in Table I.

TABLE I
FORMULA TO EXTRACT 4 FEATURES FROM GLCM

Features	Formula
Energy	$\sum_{i,j} p(i, j)^2$
Contrast	$\sum_{i,j} i - j ^k p^l(i, j)$
Homogeneity	$\sum_{i,j} \frac{p(i, j)}{1 + i - j }$
Correlation	$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\delta_i \delta_j}$

Hence, if we use n GLCMs with different distances d and directions θ and 4 features mentioned above, we have up to $4n$ features and 2^{4n} choices of TTVT.

B. Bayesian classification

Given k populations w_1, w_2, \dots, w_k with $p_i \in (0; 1)$ and $f_i(x)$ are the prior probability and probability density function of i th one, respectively, $i = 1, 2, \dots, k$. According to [14], [15], element x_0 will be assigned to w_i if

$$g_i(x_0) = g_{\max}(x_0), \quad (2)$$

where $g_i(x) = p_i f_i(x)$,

$$g_{\max}(x) = \max\{p_1 f_1(x), p_2 f_2(x), \dots, p_k f_k(x)\}.$$

Note that,

- Normally, in case of non-information, we choose prior probability p_i by uniform distribution. If we have some types of past data or training set, the prior probability is often estimated by Laplace method: $p_i = \frac{N_i+1}{N+n}$ and ratio of sample method: $p_i = \frac{N_i}{N}$ where N_i and N are the number of elements in population i and the number of elements in training set, respectively; n is the number of dimensions. In this article, the prior probability is computed by ratio of sample method.
- The likelihood function

$$f_i(x) = \frac{1}{N_i} \frac{1}{h_1 h_2 \dots h_n} \sum_{r=1}^{N_i} \prod_{s=1}^n K_s \left(\frac{x_s - x_{rs}}{h_s} \right)$$

- is estimated from all observations belonging to w_i by kernel method. In this method, the choice of smoothing parameter h_s and the choice of kernel function K role important effects on the result. The smoothing parameter h_s controls the smoothness of the resulting curve estimate, where larger h_s is equivalent to the smaller bandwidth. Although [14]–[17] have discussed on this problem, the optimal choice has not been still found yet. In this article, the smoothing parameter is chosen by the idea suggested by [16], [18] and the kernel function is the Gaussian one.
- In image texture recognition, p_i is the ratio of images having texture i in training set and $f_i(x)$ is the n -dimensional likelihood function for texture i , with n is the number of TTVT in model.

C. Bayes error

Normally, to test the result of a classification method, we compute the correct proportion of test data. It is used to compare methods together. In Bayesian method, before performing for test data, we can effectively evaluate the misclassification via the overlapping areas of posterior probability density functions. This misclassification is called Bayes error and is given by the following formula:

$$Pe_{1,2,\dots,k}^{(p)} = \sum_{i=1}^k \int_{R^n \setminus R_i^n} p_i f_i dx = 1 - \sum_{i=1}^k \int_{R_i^n} p_i f_i(x) dx, \quad (3)$$

where $R_i^n = \{x | p_i f_i(x) > p_j f_j(x),$

$$\forall i \neq j, i, j = \{1, 2, \dots, k\}, (p) = (p_1, p_2, \dots, p_k)\}.$$

From (3), we can prove the following result:

$$Pe_{1,2,\dots,k}^{(p)} = 1 - \int_{R^n} g_{\max}(x) dx. \quad (4)$$

For $k > 2$, let

$$\begin{aligned} R_1^n &= \{x \in R^n : p_1 f_1(x) > p_j f_j(x), 2 \leq j \leq k\}, \\ R_k^n &= \{x \in R^n : p_k f_k(x) > p_j f_j(x), 1 \leq j \leq k\}, \\ R_l^n &= \{x \in R^n : p_l f_l(x) > p_i f_i(x), 1 \leq i \leq k, \\ &\quad 2 \leq l \leq k-1, i \neq l\}. \end{aligned} \quad (5)$$

From (4) and (5), we have

$$\begin{aligned} Pe_{1,2,\dots,k}^{(p)} &= 1 - \int_{R_1^n} p_1 f_1(x) dx \\ &\quad - \sum_{l=2}^{k-1} \int_{R_l^n} p_l f_l(x) dx - \int_{R_k^n} p_k f_k(x) dx. \end{aligned} \quad (6)$$

For $k = 2$, we have

$$\begin{aligned} Pe_{1,2}^{(p,1-p)} &= \int_{R^n} \min\{p f_1(x), (1-p) f_2(x)\} dx \\ &= \lambda_{1,2}^{(p,1-p)} = \frac{1}{2} (1 - \|p f_1, (1-p) f_2\|_1), \end{aligned} \quad (7)$$

where $\lambda_{1,2}^{(p,1-p)}$ is the overlapping area measure of $p f_1(x)$ and $(1-p) f_2(x)$,

$$\|p f_1, (1-p) f_2\|_1 = \int_{R^n} |p f_1(x) - (1-p) f_2(x)| dx.$$

III. GENETIC ALGORITHM FOR SELECTING EXTRACTED FEATURES

Before applying Genetic Algorithm to minimize the error rate, we have to encode each selection of TTVT by a chromosome. Each chromosome contains 64 genes that correspond to 64 extracted features from the image ($d = \{1, 2, 3, 4\}$, $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, features = {contrast, correlation, homogeneity, energy}). Each gene in the chromosome obtains a value of 1 or 0, with 1 means "using this feature as a variable in classification model" and 0 is the opposite. The Genetic

Algorithm for solving the integer optimization problems [13] called MI-LXPM is presented as follows.

Crossover

The crossover operator used in [13] is the Laplace crossover. Given two individual $x^1 = (x_1^1, x_2^1, \dots, x_n^1)$ and $x^2 = (x_1^2, x_2^2, \dots, x_n^2)$, their offspring $y^1 = (y_1^1, y_2^1, \dots, y_n^1)$ and $y^2 = (y_1^2, y_2^2, \dots, y_n^2)$ are generated by the following formula:

$$y_1^1 = x_1^1 + \beta_i |x_1^1 - x_2^1|, y_2^1 = x_2^2 + \beta_i |x_1^1 - x_2^2|.$$

In the above formula, β_i satisfies the Laplace distribution and is generated as:

$$\beta_i = \begin{cases} a - b \cdot \log(u_i) & \text{if } r_i \leq \frac{1}{2} \\ a + b \cdot \log(u_i) & \text{if } r_i > \frac{1}{2} \end{cases}$$

where a is the location parameter and $b > 0$ is the scaling parameter, $u_i, r_i \in [0; 1]$ are uniform random numbers. Specifically, in texture image classification, $n = 64$ is the number of all features that extracted from all of GLCMs, and $x_i \in [0; 1]$ where $x_i = 1$ when "using i th feature as variable for classification model" and $x_i = 0$ is the opposite.

Mutation

The mutation operator used in MI-LXPM is the Power mutation. According to it, a solution x is created in the vicinity of a parent solution \bar{x} in the following manner.

$$x = \begin{cases} \bar{x} - s(\bar{x} - x^l) & \text{if } t < r \\ \bar{x} + s(x^u - \bar{x}) & \text{if } t \geq r \end{cases}$$

In the above equation, s is a random number having power distribution and calculated by $s = (s_1)^p$ where s_1 is selected randomly in interval $[0; 1]$ and p is called the index of mutation (an integer number); $t = \frac{\bar{x} - x^l}{x^u - \bar{x}}$ where x^l and x^u are the lower and upper bounds on the value of the decision variable (in our problem $x^l = 0$ and $x^u = 1$); r is a random number between 0 and 1.

Truncation procedure for integer restriction

In order to ensure that after crossover and mutation operations have been performed, the integer restrictions are satisfied, the following truncation procedure is applied:

x_i is truncated to integer value \bar{x}_i , $i = 1, 2, \dots, n$ by the rule: If x_i is integer then $\bar{x}_i = x_i$, otherwise \bar{x}_i is equal to $[x_i]$ or $[x_i] + 1$ with the probability is 0.5, where $[x_i]$ is the integer part of \bar{x}_i .

Selection

MI-LXPM that uses the tournament selection is presented by Goldberg and Deb [21].

The above paragraphs have presented the detailed MI-LXPM algorithm. This algorithm, as mentioned before, is applied to select the extracted features from GLCMs so that the error rate of Bayesian classifier in cross validation set is the minimum. Minimizing the error in cross validation set not only propose a suitable choice of TTVT but also avoid the overfitting problem which might occur when minimizing the error rate in training data set.

This framework is now called as GA-Bayes and contains two following sub-algorithms:

Algorithm 2: Selecting TTVT by MI-LXPM.

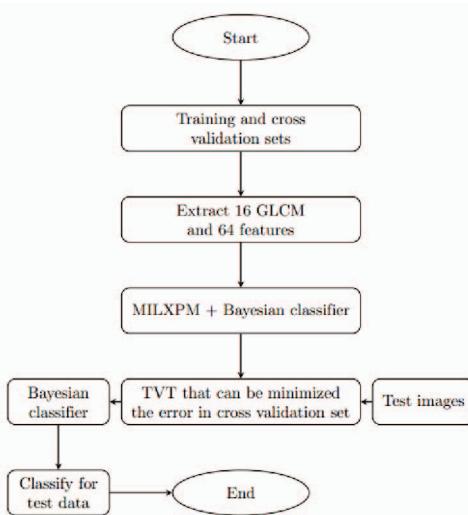


Fig. 2. The frame of GA-Bayesian method

- **Step 1** Select a random number of TVTs where each TVT is presented by a chromosome.
- **Step 2** Perform the genetic operations, such as, crossover, and mutation to introduce new ones.
- **Step 3** Based on training set, compute each prior probability p_i and likelihood function f_i , then we establish the discriminant function $g_i = p_i f_i$ for each choice of TVTs represented by chromosome.
- **Step 4** Evaluate error rate in cross validation set (objective function) for each corresponding chromosome.
- **Step 5** By tournament selection operator, replace the current selections with the new ones that produce smaller objective function values.
- **Step 6** If stopping criterion is met then stop, else go to Step 2.

Algorithm 3: Classifying new objects.

- **Step 1** For each image in training and cross validation sets, compute 16 GLCMs and extract 64 features denoted as S .
- **Step 2** Use Algorithm 1 to select a subset TVT \hat{S} of S so that the error rate in cross validation set is minimized.
- **Step 3** Use the Bayesian classifier, with the set of independent variables \hat{S} , to classify new objects.

Figure 2 shows the flowchart for Algorithm III.

IV. NUMERICAL EXAMPLES

In this section, we denote the proposed method as GA-Bayes, the popular choice of TVTs with $d = 1$ and $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ as D1, the choice of [9] with $d = \{1, 2\}$ and $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ as D12. Similarly, we denote some other methods with different d and $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ as D2, D3, D4, D13, D14, D23, D24, D34, D123, D234, D1234. Furthermore, we examine and denote the method of Celebi and Alpkocak, which uses only the contrast features with $d = \{1, 2, 3, 4\}$ as C1234. In a similar way, we denote C1, C2, C3, C4, C12, C13, C14, C23, C24, C123, C234. Because the

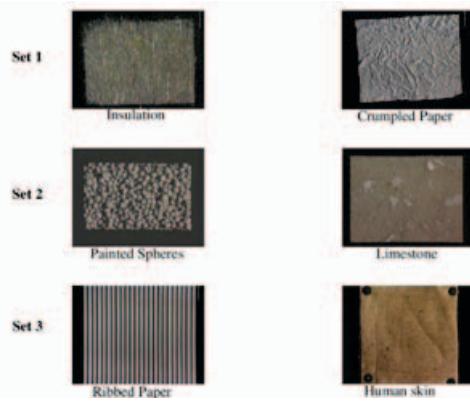


Fig. 3. The sample images of six textures

proposed method is used to optimize the choice of TVT in Bayesian classifier, we only compare the performance of GA-Bayes with other choices of TVT in cases of Bayesian one. The experiments are performed on both synthetic and real images. Specifically, we examine the binary classification in Example 1 that consider the Texture Database, CUReT (<http://www1.cs.columbia.edu/CAVE//software/curet/>). This well-known image database is widely studied in texture classification. If there is a new method related to texture recognition, CUReT is often used as an experiment subject to compare the performance between the new method and existing ones. We also conduct the experiment on classification problem with three classes in Example 2. Moreover, to evaluate the performance of GA-Bayes on practical problems, we conduct an experiment in "handwritten digit recognition", with 4000 images of digits from 0 to 9. In each example, all images are divided to training set, cross validation set and test set with the ratio is 4:3:3, respectively. The proposed method is applied to minimize the error rate in cross validation set and the obtained TVT are continued to perform with test data. Three examples are presented in detail as follows:

Example 1.

We test the proposed algorithm on the CUReT database downloaded from

<http://www1.cs.columbia.edu/CAVE//software/curet/>.

Specifically, 6 different textures separated into 3 subsets are examined:

Subset 1: Insulation (205 images) and Crumpled paper (205 images).

Subset 2: Painted Spheres (204 images) and Limestone (205 images).

Subset 3: Human skin (205 images) and Ribbed paper (410 images).

Figure 3 shows the samples of six textures. The error rates of GA-Bayes and other choices of TVT are presented in Table II.

Table II proves the superiority of GA-Bayes over others when GA-Bayes gives the best results for test images in all subsets, especially the error is really small, about 3%, for the

TABLE II
THE ERROR RATES FOR TEST DATA

Methods	Subset 1	Subset 2	Subset 3
D1	0.1333	0.0811	0.225
D2	0.1585	0.0557	0.2755
D3	0.1992	0.0418	0.3228
D4	0.2602	0.0393	0.3245
D12	0.1545	0.0607	0.2158
D13	0.1472	0.0377	0.2451
D14	0.152	0.0402	0.231
D23	0.165	0.0475	0.3022
D24	0.1707	0.0402	0.2978
D34	0.2228	0.0361	0.331
D123	0.1488	0.05	0.2451
D234	0.1764	0.0426	0.3092
D1234	0.1553	0.0385	0.2641
C1	0.3106	0.2902	0.3424
C2	0.3252	0.1754	0.4011
C3	0.3821	0.1434	0.3821
C4	0.3935	0.1418	0.3734
C12	0.3081	0.2115	0.4022
C13	0.2943	0.1844	0.3679
C13	0.2943	0.1844	0.3679
C14	0.2789	0.1713	0.3418
C23	0.3325	0.1598	0.3951
C24	0.3333	0.1484	0.3815
C34	0.387	0.1418	0.3804
C123	0.3024	0.1844	0.3859
C234	0.3463	0.1459	0.3842
C1234	0.2927	0.1631	0.3625
GABayes	0.1211	0.0352	0.1701

TABLE III
THE p -VALUES OF T-TEST.

Methods	Subset 1	Subset 2	Subset 3
D1	0.2129	0.0000	0.0014
D2	0.0023	0.0041	0.0000
D3	0.0003	0.2353	0.0000
D4	0.0000	0.3629	0.0000
D12	0.0052	0.0049	0.0033
D13	0.0107	0.6849	0.0000
D14	0.0036	0.3572	0.0001
D23	0.0008	0.0617	0.0000
D24	0.0007	0.3133	0.0000
D34	0.0002	0.8402	0.0000
D123	0.0192	0.0239	0.0000
D234	0.0004	0.1708	0.0000
D1234	0.0067	0.6059	0.0000
C1	0.0000	0.0000	0.0000
C2	0.0000	0.0000	0.0000
C3	0.0000	0.0000	0.0000
C4	0.0000	0.0000	0.0000
C12	0.0000	0.0000	0.0000
C13	0.0000	0.0000	0.0000
C14	0.0000	0.0000	0.0000
C23	0.0000	0.0000	0.0000
C24	0.0000	0.0000	0.0000
C34	0.0000	0.0000	0.0000
C123	0.0000	0.0000	0.0000
C234	0.0000	0.0000	0.0000
C1234	0.0000	0.0000	0.0000

TABLE IV
GOOD TVTs FOR THREE SUBSETS.

Sets	TVTs	Bayes error
Subset 1	Correlation with $d = \{1, 2, 3\}$, $\theta = 0$	0.0503
Subset 2	Contrast with $d = 3$, $\theta = \frac{3\pi}{2}$, Energy with $d = \{3, 4\}$, $\theta = \pi$	0.2413
Subset 3	Correlation with $d = \{1, 4\}$, $\theta = 0$, Homogeneity with $d = 1$, $\theta = \frac{3\pi}{2}$	0.1129

Subset 3. To determine whether the differences between the proposed model and other comparative models are statistically significant or not, we apply the t-test for pair samples. As shown in Table III, in almost cases, GA-Bayes outperforms others at the 5% statistical significance level. These results prove that the proposed method is feasible and advantageous for improving the classification performance. Moreover, it presents that GA-Bayes can reach the globally optimal TVT and can be suitable for many cases in texture classification problem.

For the TVTs selection, because the experiments are run 10 times with random training, cross validation and test sets, the TVTs obtained in each time are different. We briefly summary the optimal TVTs in Table 4 where each TVT is consider as a good TVT if it is selected at least 9/10 times. Table 4 also gives Bayes error when we use good TVTs in classifying.

Example 2.

To evaluate the performance of proposed method in case of more than two populations, the Example 1 is extended to the case of three different textures. Particularly, GA-Bayes and others are applied to classify three of texture classes including Human skin (205 images), Ribbed paper (410 images) and Painted Spheres (204 images). Clearly, the current problem becomes more complicated with up to 3 populations and 819 images. The classification and t-test results are presented in Table V whereas the optimal TVTs and Bayes error are briefly summarized in Table VI.

It can be seen that GA-Bayes gives the smallest error (18.2%) and outperforms others at the 5% statistical significance level in almost cases. The results from Examples 1 and

Example 2 demonstrate that GA-Bayes is a feasible and stable algorithm for image classification problem with large data set, multiple populations and high volatilities. We continue to examine the performance of proposed algorithm in a practical problem presented in Example 3.

Example 3.

Handwritten digit recognition is the ability of a computer to receive and interpret intelligible handwritten digit input from sources such as paper documents, photographs, touch-screens and other devices. Due to the current development of information technology, handwritten digit recognition is an active researched topic in many field such as data mining, machine learning, recognized statistics and so on ([22]). In this example, GA-Bayes is continued to apply in this problem to test the classification performance. Specifically, the handwritten digit database, 2k2k ([23]), which includes 4000 handwritten digits between 0 and 9, is examined. Figure 4 illustrates some images that are considered in this example.

In the sort of two above examples, all images are divided to training, cross validation and test set with the ratio is 4:3:3, respectively. Also, the classification result of GA-Bayes is compared with that of other choices of TVT as conducted earlier. Table VII presents the performance of all comparative

TABLE V
THE CLASSIFICATION AND TESTING RESULTS FOR THREE POPULATIONS.

Method	Means of error rates	p-value	Method	Means of error rates	p-value
D1	0.2384	0.0000	C2	0.4424	0.0000
D2	0.2657	0.0000	C3	0.4322	0.0000
D3	0.2878	0.0000	C4	0.4282	0.0000
D4	0.2939	0.0000	C12	0.4527	0.0000
D12	0.2343	0.0000	C13	0.4339	0.0000
D13	0.2306	0.0027	C14	0.4192	0.0000
D14	0.2249	0.0056	C23	0.4420	0.0000
D23	0.2816	0.0001	C24	0.4261	0.0000
D24	0.2755	0.0000	C34	0.4355	0.0000
D34	0.2939	0.0000	C123	0.4433	0.0000
D123	0.2388	0.0006	C234	0.4294	0.0000
D234	0.2788	0.0000	C1234	0.4261	0.0000
D1234	0.2449	0.0015	GABayes	0.1820	
C1	0.4322	0.0000			

TABLE VI
GOOD TVTs FOR EXAMPLE 2.

TVTs	Bayes error
Correlation with $d = 4, \theta = 0$.	0.3868
Homonegity with $d = 1, \theta = \frac{3\pi}{2}$	

methods in detail.

TABLE VII
THE CLASSIFICATION RESULT FOR EXAMPLE 3.

Methods	Error rate	Methods	Error rate	Methods	Error rate
D1	0.5108	D34	0.5175	C13	0.5108
D2	0.5200	D123	0.5067	C14	0.5158
D3	0.5383	D234	0.5042	C23	0.5217
D4	0.5492	D1234	0.4950	C24	0.5242
D12	0.4767	C1	0.5350	C34	0.5600
D13	0.4842	C2	0.5317	C123	0.5183
D14	0.4900	C3	0.5500	C234	0.5342
D23	0.5142	C4	0.6058	C1234	0.5150
D24	0.4900	C12	0.5233	GABayes	0.4317

The results in Table 7 prove the superiority of the proposed method over other choices of TVT when it gives the smallest error at 43.17%. For a practical image database with 10 classes and 4000 observations, this error is acceptable, in the author of view. Moreover, in literature, all of previous methods studied in 2k2k image database have given the error rate in the interval (46.88%, 50.88%) ([23]) whereas GA-Bayes only makes an error at 43.17%. It means the proposed method not only can improve the performance in Bayesian classifier but also can be the best method for 2k2k database in all of machine learning ones. All of them prove that GA-Bayes can be effectively used for further analysis of handwritten digit recognition as well as other practical image classification problems. However, the computational time is always a drawback of Genetic algorithm. In all of experiments, the run-time is about some minutes, even up to over more than an hour for the Example 3. This is a limitation of the proposed method, which is necessary to improve in coming. Furthermore, the study applied GA-Bayes for RGB scale also need to be interested.

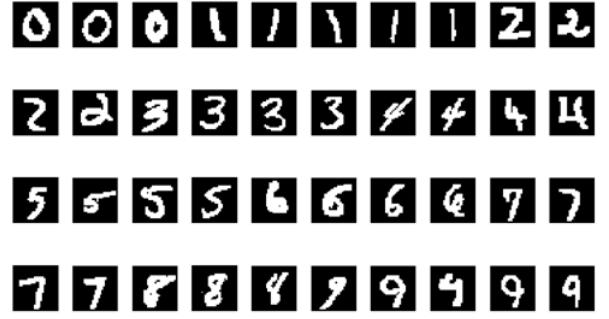


Fig. 4. Some samples in image set 2k2k.

V. CONCLUSION

Based on the features that are extracted from Grey level Co-occurrence matrix, this paper has applied Genetic Algorithm (MI-LXPM) to optimize the choice typical variables for image texture (TVT) in Bayesian classification. The purpose of this method is minimize the error rate in cross validation set. It not only propose a suitable choice of TVT but also avoid the overfitting problem which might occur when minimizing the error rate in training data set. The numerical examples show that the proposed algorithm has good effects on both synthetic images and practical images, for instance handwritten digit recognition. The coming proposed approach will be continued studying to improve the computational cost and to apply for many practical problems associated with image recognition.

REFERENCES

- [1] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, vol. 43, no. 1, pp. 29–44, 2001.
- [2] C. Schmid, "Constructing models for content-based image retrieval," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2. IEEE, 2001, pp. II–39.
- [3] A. Goh, "Unsupervised Riemannian Clustering of Probability Density Functions," pp. 377–392, 2008.
- [4] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.
- [5] L. Shapiro and R. Haralick, "Computer and robot vision," in *Reading: Addison-Wesley*, 1992, vol. 8.
- [6] E. Celebi and A. Alpkocak, "Clustering of texture features for content-based image retrieval," in *Advances in Information Systems*. Springer, 2000, pp. 216–225.
- [7] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of remote sensing*, vol. 28, no. 1, pp. 45–62, 2002.
- [8] P. Bhogle and S. Patil, "Oil spill detection in SAR images using texture entropy algorithm and mahalanobis classifier," *Int. J. Eng. Sci.*, vol. 4, pp. 4823–4826, 2012.
- [9] V. Ayala-ramirez, M. Obara-kepowitz, R. E. Sinchez-ygez, R. Jaime-riavas, and U. D. G. F. M. E. E, "Bayesian texture classification method using a random sampling scheme," in *Systems, Man and Cybernetics (SMC), IEEE International Conference on*, 2003, pp. 2065–2069.
- [10] S. Aksoy and R. M. Haralick, "Content Based Image Database Retrieval Using Variance of Gray Level Spatial Dependencies," in *IAPR International Workshop on Multimedia Information Analysis & Retrieval (MINAR'98)*, Hong Kong, 1998.
- [11] D. Cobzas, N. Birkbeck, M. Schmidt, A. Murtha, and M. Jagersand, "3D Variational Brain Tumor Segmentation using a High Dimensional Feature Set," in *Computer vision, IEEE 11th International Conference on*, 2007, pp. 1–8.

- [12] N. Zhang, S. Ruan, S. Lebonvallet, Q. Liao, and Y. Zhu, "Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation," *Computer Vision and Image Understanding*, no. 2, pp. 256–269.
- [13] K. Deep, K. P. Singh, M. L. Kansal, and C. Mohan, "A real coded genetic algorithm for solving integer and mixed integer optimization problems," *Applied Mathematics and Computation*, vol. 212, no. 2, pp. 505–518, 2009.
- [14] T. Pham-Gia, N. Turkkan, and T. Vovan, "Statistical discrimination analysis using the maximum function," *Communications in Statistics-Simulation and Computation*®, vol. 37, no. 2, pp. 320–336, 2008.
- [15] T. Nguyen-Trang and T. Vo-Van, "A new approach for determining the prior probabilities in the classification problem by Bayesian method," *Advances in Data Analysis and Classification*, pp. 1–15, 2016.
- [16] D. W. Scott, *Multivariate Density Estimation*. Wiley, New York, 1992.
- [17] W. L. Martinez and A. R. Martinez, *Computational statistics handbook with MATLAB*. CRC press, 2007.
- [18] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.
- [19] T. Pham-Gia, N. D. Nhat, and N. V. Phong, "Statistical Classification Using the Maximum Function," *Open Journal of Statistics*, vol. 5, no. 07, p. 665, 2015.
- [20] T. Vo Van and T. Pham-Gia, "Clustering probability distributions," *Journal of Applied Statistics*, vol. 37, no. 11, pp. 1891–1910, 2010.
- [21] D. E. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *Foundations of genetic algorithms*, vol. 1, pp. 69–93, 1991.
- [22] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," *Pattern Recognition*, vol. 36, no. 10, pp. 2271–2285, 2003.
- [23] J. Xu, J. Han, K. Xiong, and F. Nie, "Robust and Sparse Fuzzy K-Means Clustering," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2106, pp. 2224–2230.