Multimedia Tools and Applications (2022) 81:36473–36488
https://doi.org/10.1007/s11042-022-13531-6

1213: COMPUTATIONAL OPTIMIZATION AND APPLICATIONS FOR HETEROGENEOUS MULTIMEDIA DATA

A new image classification method using interval texture feature and improved Bayesian classifier

Ngoc Lethikim[1] · Thao Nguyentrang[2,3] · Tai Vovan[4]

Received: 17 December 2020 / Revised: 26 November 2021 / Accepted: 13 July 2022 /
Published online: 5 August 2022

Abstract

In this paper, a novel technique for image classification is proposed with the three main contributions. First, we give the texture extraction technique for each image to have the two-dimensional interval based on the Grey Level Co-occurrence matrices. Second, the automatic fuzzy clustering algorithm for interval data to determine the prior probability for the classification problem by Bayesian method is created. Finally, the new principle to classify for image is established. Combining the above three improvements, we have the effective method to classify the images. In addition, the proposed method can be performed rapidly for the real data by the established Matlab procedure. Four image data sets with the different characters are used to illustrate the proposed method, and to compare to the well-known algorithms like Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Fisher method, Naive Bayes, Multi-Supported Vector Machine (Multi-SVM), Convolutional Neural Networks (CNN), and VGG-19. The results show that the proposed method has the good and stable empirical error, and give the outstanding result about time cost.

Keywords Classification · Extraction · Image · Interval data · Overlap distance

✉ Tai Vovan
vvtai@ctu.edu.vn

Ngoc Lethikim
ngoc.ltk@vlu.edu.vn

Thao Nguyentrang
nguyentrangthao@tdtu.edu.vn

[1] Faculty of Engineering, Van Lang University, Ho Chi Minh City, Vietnam

[2] Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[3] Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[4] College of Natural Science, Can Tho University, Can Tho City, Vietnam

1 Introduction

Image classification refers to the procedure of classifying images into several categories, based on their similarities. This procedure aims to recognize the observation content of the image. A typical image classification process consists of two major phases: feature extraction (i) and classification (ii).

In the feature extraction process (i), we need to find the most useful features by which the image categories can be well-discriminated. Color, texture, and shape information, among several kinds of features, are the primary features in content-based image classification systems [4, 13]. Depending on the image classification problem being solved, different types of features can be used. This paper concentrates on the image classification problem based on the texture feature. In this approach, we first calculate the Grey level occurrence matrix (GLCM) and then calculate the statistical texture features based on the obtained GLCM [16, 17, 19, 20, 32, 40]. Haralick [11] defined 14 statistical features where contrast, correlation, homogeneity, and energy are the most common variables. The texture features contain significant information about the basic arrangement of the surface. They also can define the surface with environment relationship, and can describe the physical composition of the surface. Therefore, texture-based image classification has been widely studied in many research areas so far [2, 6, 14, 18, 34, 37]. Previous methods extract the texture as a discrete point that is relatively sensitive to noise, background cluster, and image quality. In other words, the existing texture-based image classification method is no longer suitable for the big data area in which the uncertainty property needs to be considered. Therefore, the first motivation of this paper is to present a novel method for texture extraction based on GLCM, called the interval texture feature. Particularly, for each image, an interval is rather than a discrete point that is extracted by the state-of-the-art method. Because the interval data allow considering variability, and uncertainty, the extracted texture intervals are expected to be more comprehensive than the traditional methods. The experimental results showed that the use of the proposed technique for interval featured extraction is a good way of describing the image content. Moreover, if we use the proposed method to extract feature images, dimensional input will decrease. As a result, this work can reduce related processing time. In this study, we believe that the texture featured interval is the special type of features extraction and the important contribution to the classification result.

In classification process (ii), Linear Discriminant Analysis (LDA), Quadratic discriminant analysis (QDA), Logistic regression, Fisher method, Bayesian classifier, Multi-Supported vector machine (Multi-SVM), and Convolutional Neural Network (CNN) are some of most well-known methods [1, 5, 8–10, 12, 15, 21, 23, 24, 26, 33, 36, 38]. Compared to other methods, the Bayesian classifier has some advantages because it is not constrained by assumptions of normal distribution and equal variance of classes [35]. Nevertheless, the existing Bayesian classifier as well as the other methods can only be used to classify the traditional data, which is the discrete element. In other words, none of them is built for classifying the interval data, i.e., the existing classification algorithms cannot be used for classifying the texture intervals extracted in Phase (i). Therefore, the second motivation of this paper is to propose a new algorithm for classifying the interval data, based on the Bayesian classifier. For performing the Bayesian classifier, we need to determine the prior probabilities (ii-a) as well as the probability density functions (ii-b). For (ii-a), the prior probabilities are often determined based on expert knowledge, or a previous statistical conclusion to a research problem. Thus, they are not generally applicable to data sets of different disciplines. For this reason, we propose a new approach for determining the

prior probabilities for each interval. Particularly, a new fuzzy clustering technique for interval data is utilized. This algorithm can estimate the membership degrees of each interval to the given classes. In other words, for determining prior probabilities, it considers both the information about the new object (interval) and the information of training data. In this respect, this algorithm can be expected to contain more information than traditional methods thereby reducing the empirical error. In this study, the empirical error is normally measured on testing set only, and also known as the overall error rate of classification. This is the proportion of cases where the prediction is wrong. For (ii-b), due to the lack of a technique for estimating the probability function of interval data, the overlap distance normalized from range 0 to 1 is proposed as an alternative measure. Using the overlap distance has the two following advantages. First, due to the high computational cost of direct calculating the probability density function, this technique can produce a more efficient algorithm. Second, the overlap distance has demonstrated its superiority over the other distances in calculating the dissimilarity between two intervals, especially in the case of complex objects, such as images [25, 34]. Therefore, we use the normalized overlap distance rather than the probability density function of intervals in the proposed method. The contribution of this paper can be highlighted as follow.

- Proposing a new method that can extract image texture interval. This new feature is expected to be more appropriate than the existing texture features when classifying images with uncertainty.
- Proposing a new method for calculating the prior probabilities of intervals.
- Proposing a new method that can avoid calculating the probability function of interval data, which is so complicated in practice.

The remainder of the paper is organized as follows. In Section 2, some distance measurements are defined for multi-dimension cases. In addition, clustering evaluation criteria are also defined in this section. The proposed technique is given in Section 3. In section 4, three examples are investigated to compare the proposed method with some existing methods. Section 5 is the conclusion.

2 Measures for divergence of intervals

2.1 The popular distances of two intervals

Given two p-dimensional intervals, $p \geq 1$:

$$a = \left(a^1, a^2, \ldots, a^p\right) = \left(\left[a_1, \hat{a}_1\right], \left[a_2, \hat{a}_2\right], \ldots, \left[a_p, \hat{a}_p\right]\right),$$

$$b = \left(b^1, b^2, \ldots, b^p\right) = \left(\left[b_1, \hat{b}_1\right], \left[b_2, \hat{b}_2\right], \ldots, \left[b_p, \hat{b}_p\right]\right).$$

The following common distances are used in clustering for interval data:
Hausdorff distance:

$$d_H(a, b) = \sum_{i=1}^{p} \left(\max\{|a_i - b_i|, |\hat{a}_i - \hat{b}_i|\}\right). \tag{1}$$

City-block distance:

$$d_C(a,b) = \sum_{i=1}^{p} (|a_i - b_i| + |\hat{a}_i - \hat{b}_i|). \tag{2}$$

Euclidean distance:

$$d_E(a,b) = \sqrt{\sum_{i=1}^{p} [(a_i - b_i)^2 + (\hat{a}_i - \hat{b}_i)^2]}. \tag{3}$$

The above distances have been designed for a relatively simple way where only the lower and upper bounds are considered. Those distances have not considered the overlapping degree between two intervals, thereby providing unsuitable results in some cases.

2.2 The overlap distance

Let $a = [a_1, \hat{a}_1]$ and $b = [b_1, \hat{b}_1]$ be two p-dimensional intervals. The overlap distance between them is given by

$$d_O(a,b) = D(a,b).\left(1 - \frac{O(a,b)}{2r_a + 1}\right), \tag{4}$$

where $r_a = \frac{1}{p}\sum_{i=1}^{p} |a_i - \hat{a}_i|$,

$O(a,b)$ is the overlapped area between a and b,

$D(a,b)$ is a distance originated from Hausdorff distance by considering all points inside the interval:

$$D(a,b) = \max_{a' \in [a]} \{\min_{b \in [b]} \{d_E(a', b')\}\}, \tag{5}$$

with $d_E(a,b)$ is the Euclidean distance.

In case of the univariate intervals, set $c_a = \frac{a_1 + \hat{a}_1}{2}$, $r_a = \frac{a_1 - \hat{a}_1}{2}$, $c_b = \frac{b_1 + \hat{b}_1}{2}$, and $r_b = \frac{b_1 - \hat{b}_1}{2}$. Formula (5) is presented as follows:

$$d_O(a,b) = \begin{cases} 0 & \text{if (i)} \\ (|c_a - c_b| + r_a - r_b)\left(1 - \frac{2r_b}{2r_a+1}\right) & \text{if (ii)} \\ |c_a - c_b| & \text{if (iii)} \\ (|c_a - c_b| + r_a - r_b)\left(1 - \frac{r_a + r_b - |c_a - c_b|}{2r_a+1}\right) & \text{if (iv)} \\ (|c_a - c_b| + r_a - r_b)\left(1 + \frac{|c_a - c_b| - (r_a + r_b)}{2r_a+1}\right) & \text{if (v).} \end{cases} \tag{6}$$

(i) the interval a is completely falling inside of interval $b : \|c_a - c_b\| \leq r_b - r_a$,
(ii) the interval b is completely falling into interval $a : \|c_a - c_b\| \leq r_a - r_b$,
(iii) the interval b overlaps with a on the left side of $a : r_a = r_b = 0$,
(iv) the interval b overlaps with a on the right side of $a : |r_a - r_b\| < \|c_a - c_b\| < r_a + r_b$,
(v) the interval b is not overlapping with a, and b is on the left side of a or b is on the right side of $a : |c_a - c_b| \geq r_a + r_b$.

In case of p-dimensions ($p > 2$), the overlap distance is defined as Formula (6), where

$$c_a = \frac{1}{p}\sum_{i=1}^{p}(a_i + \hat{a}_i), r_a = \frac{1}{p}\sum_{i=1}^{p}\left|a_i - \hat{a}_i\right|,$$
$$c_b = \frac{1}{p}\sum_{l=1}^{p}(b_i + \hat{b}_i), r_b = \frac{1}{p}\sum_{i=1}^{p}\left|b_i - \hat{b}_i\right|.$$

It can be seen that d_O evaluates the similarity of two intervals based on useful information about the center and the overlap area. As a result, the(d_O) can be a more suitable choice than the Euclidean distance (d_E), the city-block distance (d_C), and the Hausdorff distance (d_H). Based on the practical experience, we see that, the d_O can overcome the drawback of other measures in evaluating the similarity of the intervals for many cases. For example, given 8 intervals $a_1 = [3, 5]$, $a_2 = [3, 6]$, $a_3 = [5, 6]$, $a_4 = [2, 7]$, $a_5 = [3, 7]$, $a_6 = [2, 6]$, $a_7 = [5, 8]$, and $b = [0, 4]$, the similarities between b and $a_1, a_2, a_3, a_4, a_5, a_6, a_7$ are respectively evaluated by d_C, d_H, d_E, and d_O, and are presented in Table 1.

It can be observed that the City-block distance and the Euclidean distance cannot capture the difference between (a_2, b) and (a_4, b); the Hausdorff distance cannot capture the differences among (a_1, b), (a_2, b), (a_4, b) and (a_5, b). Meanwhile, all differences can be captured using the overlap distance, d_O, as shown in Table 1. The overlap distance is therefore suitable to measure the dissimilarity between two intervals.

3 Methods to extract image data

3.1 Extraction of featured interval method based GLCM

The Grey-level co-occurrence matrices (GLCM) is a matrix P size of $G \times G$. Each element $p(i, j)$ presents the number of occurrences of intensity i and intensity j at fixed distance d and orientation θ. Generally, $p(i, j)$ is calculated by Formula (7) and showed in Fig. 1.

$$p_{d\theta}(i, j) = \#\{ \left((r, c), \left(r', c'\right)\right) \in M \times N | d = \left|\left|(r, c), \left(r', c'\right)\right|\right|, \theta = \oplus\left((r, c), \left(r', c'\right)\right), I(r, c) = i, I\left(r', c'\right) = j\}. \quad (7)$$

Based on the GLCM, the texture featured interval is estimated as follow.

$$[\mu_x - r_1/2, \mu_x + r_1/2], [\mu_y - r_2/2, \mu_y + r_2/2], \quad (8)$$

where

r_1 and r_2 are random numbers with uniform distribution over [0; 1],

$$\mu_x = \frac{1}{N_y}\sum_{j}^{N_y}\left(\frac{1}{N_x}\sum_{i}^{N_x}(i)p_{d\theta}(i, j)\right); \mu_y = \frac{1}{N_y}\sum_{i}^{N_x}\left(\frac{1}{N_x}\sum_{j}^{N_y}(j)p_{d\theta}(i, j)\right), \quad (9)$$

Table 1 The similarity between a_i and b

Distance	a_1	a_2	a_3	a_4	a_5	a_6	a_7
d_C	4.00	5.00	7.00	5.00	6.00	4.00	9.00
d_H	3.00	3.00	5.00	3.00	3.00	2.00	5.00
d_E	3.16	3.61	5.39	3.61	4.24	2.83	2.83
d_O	0.67	1.50	3.00	2.00	2.40	1.20	5.00

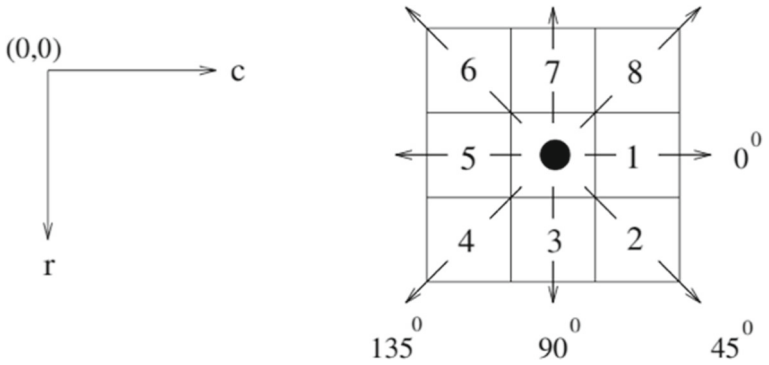

Fig. 1 Arranged space of pixels

N_x and N_y are the size of GLCM, and $p_{d\theta}(i, j)$ is determined by Formula (7).

3.2 The proposed algorithm

Given k images groups $w_1, w_2, ..., w_k$ consisting of N images. The n new images are classified into one of k given groups using the following algorithm.

1. Extract the interval features of $N + n$ training and testing images using Formula (7), (8), and (9). Let a_j be the interval representing for the image j, $j = 1, 2, \ldots, N + n$.
2. Calculate the pairwise distances d_{ij} between each group i and each new image j by Formula (10).

$$d_{ij} = \max_{a \in w_i} d_O(a, a_j), \tag{10}$$

   where $i = 1, 2, \ldots, k$; $j = N + 1, N + 2, \ldots, N + n$. After that, normalize the set of d_{ij} into the range [0 1].
3. Calculate the prior probabilities using the following steps.
   - Set $t = 0$, and establish the initial partition matrix $\mathbf{U}^{(t)} = \left[\mu_{ij}\right]_{k\times(N+n)}$, where the first N columns are extracted from known training data with $\mu_{ij} = 1$ if the jth image belongs to the w_i and $\mu_{ij} = 0$ for the opposite. The remain columns are chosen by uniform distribution.
   - Find the prototype interval v_i for each group by Formula (11).

$$v_i^{(t)} = \frac{\sum_{j=1}^{N+n} \mu_{ij}^m a_j}{\sum_{j=1}^{N} \mu_{ij}^m}, \tag{11}$$

     where $m \in [1, \infty)$.
   - Update the new partition matrix by the following Formula:

$$\mu_{ij}^{(t+1)} = \frac{1}{\sum_{l=1}^{k} \left(d_O(v_i, a_j)\right) / \left(d_O(v_l, a_j)\right)^{2/(m-1)}}. \tag{12}$$

   - Repeat the two above steps until $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$, where $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| = \max_{ij} |\mu_{ij}^{(t+1)} - \mu_{ij}^{(t)}|$.

   At the end of the above three steps, we receive the prior probabilities q_{ij} of the new n images through the last n column of partition matrix.

4. For each image, compute the quasi-posterior probability density functions using Formula (13).

$$g_{ij} = q_{ij}(1 - d_i). \tag{13}$$

5. Classify the new image j by the following rule. If $\max\{q_{ij}(1 - d_i)\} = q_{cj}(1 - d_c)$, the image j is classified as belonging to the class w_c, $c = 1, 2, \ldots, k$.

The flowchart of the proposed method is shown by Fig. 2.

In Step 3, the proposed algorithm will be finished if $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\|$ is less than ε. The bigger ε is, the fewer iterations are taken while the quantity of group can be unsuitable. In this article, we choose $\varepsilon = 10^{-4}$ for all numerical examples (see [22, 35]).

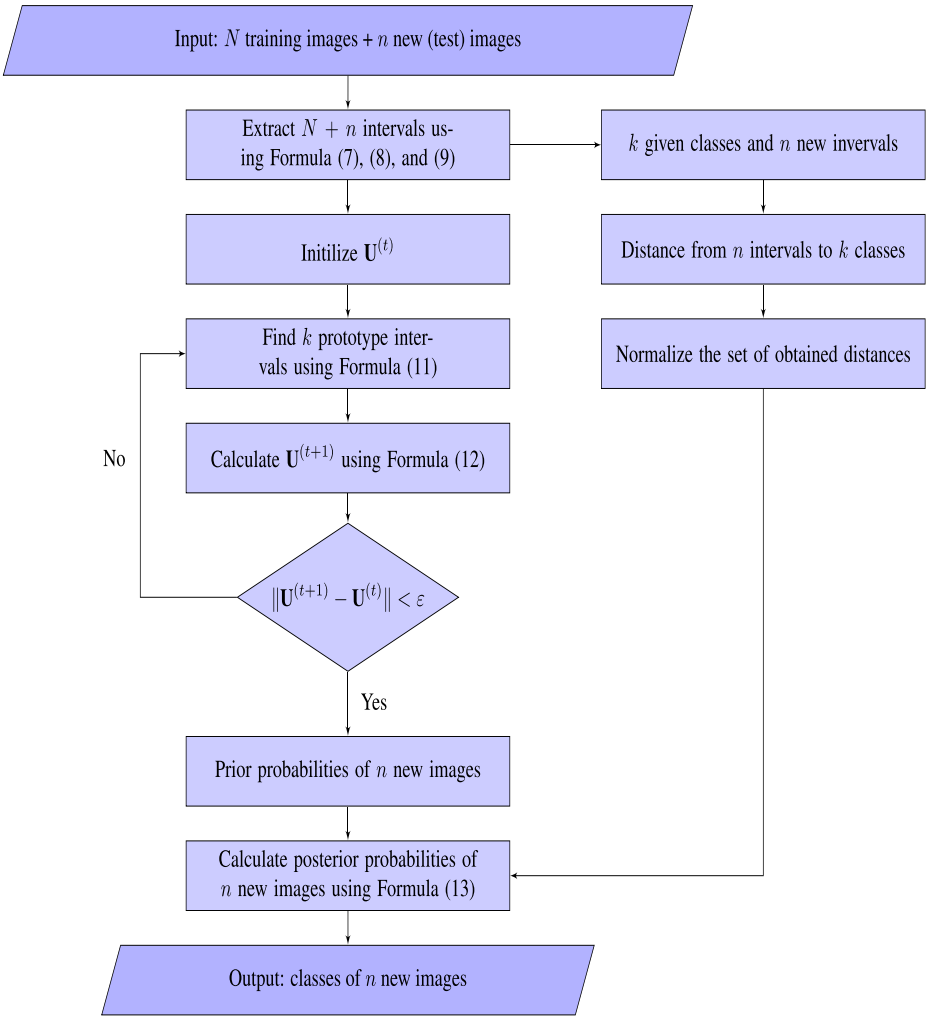

Fig. 2 Flowchart for the proposed method

4 Experiments

To illustrate the potential of the proposed method in the image classification problem, three data sets downloaded from [7, 29] are considered. Example 1 is a simple binary classification problem where we need to classify each image into "motorbike" or "car" categories. The purpose of this example is to illustrate step-by-step the proposed algorithm. Example 2 is more complicated problems in which skin cancer detection problem is solved. In addition, a much more challenging dataset have been used to derive classification accuracy in Example 3 that consist of 1440 images in 20 categories.

The proposed method is compared to traditional classification algorithms such as LDA, QDA, Fisher method, Multi-SVM, Naive Bayes, CNN, and VGG-19. Table 2 shows the self-training CNN parameters for all examples, while others are set to the suggestion [9, 12, 23, 24].

We use a transfer learning strategy for the VGG-19 model, which preserves the network structure and weights in the initial layers. We employ layers in the following order in the final layers: Fully Connected Layer (64), Dropout Layer (0.1), and Fully Connected Layer (k), where k is the number of classes in each problem. Other settings include: learning rate of 0.001, minimum batch size of 30, number of epochs = 25, Adam as the optimizer, and so on.

Table 2 The parameter of the CNN method for Examples

Name	Type	Activations	Learnables
Imageinput	Image input	$128 \times 128 \times 1$	-
Conv_1	Convolution	$128 \times 128 \times 8$	Weights $5 \times 5 \times 1 \times 8$
			Bias $1 \times 1 \times 8$
Batchnorm_1	Batch Normalization	$128 \times 128 \times 8$	Offset $1 \times 1 \times 8$
			Scale $1 \times 1 \times 8$
Relu_1	ReLU	$128 \times 128 \times 8$	-
Maxpool_1	Max Pooling	$64 \times 64 \times 8$	-
Conv_2	Convolution	$64 \times 64 \times 16$	Weights $5 \times 5 \times 8 \times 16$
			Bias $1 \times 1 \times 16$
Batchnorm_2	Batch Normalization	$64 \times 64 \times 16$	Offset $1 \times 1 \times 16$
			Scale $1 \times 1 \times 16$
Relu_2	ReLU	$64 \times 64 \times 16$	-
Maxpool_2	Max Pooling	$32 \times 32 \times 16$	-
Conv_3	Convolution	$32 \times 32 \times 32$	Weights $5 \times 5 \times 16 \times 32$
			Bias $1 \times 1 \times 32$
Batchnorm_3	Batch Normalization	Offset $1 \times 1 \times 32$	
			Scale $1 \times 1 \times 32$
Relu_3	ReLU	$32 \times 32 \times 32$	-
fc	Fully Connected	$1 \times 1 \times k$	Offset 2×32768
			Bias 2×1
Softmax_1	Softmax	$1 \times 1 \times k$	-
Classoutput	Classification output	-	-

To test the generalization capacity of prediction models and avoid overfitting, we utilize the 5-folds method, which divides the data into 5 equal sets. The first four sets act as a training set, while the last four function as a test set. In each data set, we use the empirical error, iteration, and computational time as criteria to compare together.

Example 1 The first data set utilized in this example is composed of two different groups, namely Motorbike and Car, with 250 images and 280 images, respectively. Some image samples are given in Fig. 3. The categorization procedure is illustrated step-by-step with a first fold as follows, using 5-folds cross validation.

1. Extract the interval features. Based on Formula (7), (8), and (9), we extract the interval feature for 530 images including in 424 training images divided into two groups and 106 test images.
2. Calculate the pairwise distances. Using Formula (10), we calculate the normalized overlap distances between the considered intervals and the two classes. The obtained result is represented by $\mathbf{D}$, which is a matrix of size 2×530. Based on the last 106 columns in $\mathbf{D}$, the corresponding alternative measures for the probability density functions of the 106 test images, namely $(1 - d_i)$, are also defined.

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 0.9359 & ... & 0.8763 & 1 & 1 \\ 0.9541 & 0.9734 & 1 & ... & 1 & 0.9415 & 0.9800 \end{bmatrix}.$$

3. Calculate the prior probabilities.

– Build the initial partition matrix $U^{(0)}$. Particularly, for $i = 1, 2, \ldots, 424$, $j = 1, 2$, $u_{ij} = 1$ if the image i belongs to class j, and vice versa. For $i = 424, 425, \ldots, 530$, $j = 1, 2$ (the test set), we initialize $u_{ij} = 0.5$.

$$U^{(0)}_{2,530} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & ... & 0.5 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0 & 1 & 1 & ... & 0.5 & 0.5 & 0.5 \end{bmatrix}.$$

– We then compute the prototype interval using the Formula (11) and obtain the following result.

$$v = \begin{pmatrix} [5.9114, 6.3211]; [5.8453, 6.3899] \\ [3.8385, 4.3548]; [3.8945, 4.3232] \end{pmatrix}.$$

– Using Formula (12), we update the partition matrix and obtain the following $U^{(1)}$:

$$U^{(1)}_{2,530} = \begin{bmatrix} 0.0462 & 0.0024 & ... & 0.9800 & 0.9712 \\ 0.9538 & 0.9976 & ... & 0.0200 & 0.0288 \end{bmatrix}.$$

(a) (b)

Fig. 3 Samples of two groups

– Repeat the two above steps until $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$. After after 15 iterations, we obtain the final partition matrix $U^{(15)}$ in which the last nine columns, $q_{2,106}^{(15)}$, are used to determine the prior probabilities of test images.

$$q_{2,106}^{(15)} = \begin{bmatrix} \mathbf{0.9931} & \mathbf{0.9926} & \ldots & 0.0118 & 0.0343 \\ 0.0069 & 0.0074 & \ldots & \mathbf{0.9882} & \mathbf{0.9657} \end{bmatrix}$$

The iterations to find the prior probability is shown by Fig. 4.

Figure 4 illustrates the convergence in finding the prior probability of the 530 images. It can be observed that the prior probabilities values have been significantly updated at some first iterations, and quickly converges at some final iterations.

4. Classify the new image j. If $max\{q_{ij}(1-d_i)\} = q_{cj}(1-d_c)$, the image j is classified to w_c, with $c = 1, 2$. In the mentioned rule, q_{ij} indicates the prior probability determined in Step 3, and d_i indicates the overlap distance between the test image and the class i, calculated in Step 2. The ultimate conclusion is that 56 photos are assigned to the Motor bike group, while the rest are assigned to the Car group.

As shown in Table 3, whether the proposed technique likewise produces high accuracy, around 96 percent, the VGG-19 method provides the best result with an empirical error of 0. However, in terms of computing cost, the suggested method outperforms the VGG-19, which spends a lot of time on classification (Time cost of the proposed method is only more 1 second while VGG-19 is about 100 minutes). The results show that the proposed method is a viable solution to the image categorization problem.

Example 2 Cancer is today's most fragile and serious disease due to its difficult treatment method. As a result, finding a way to identify cancer at an early stage is essential. Moles

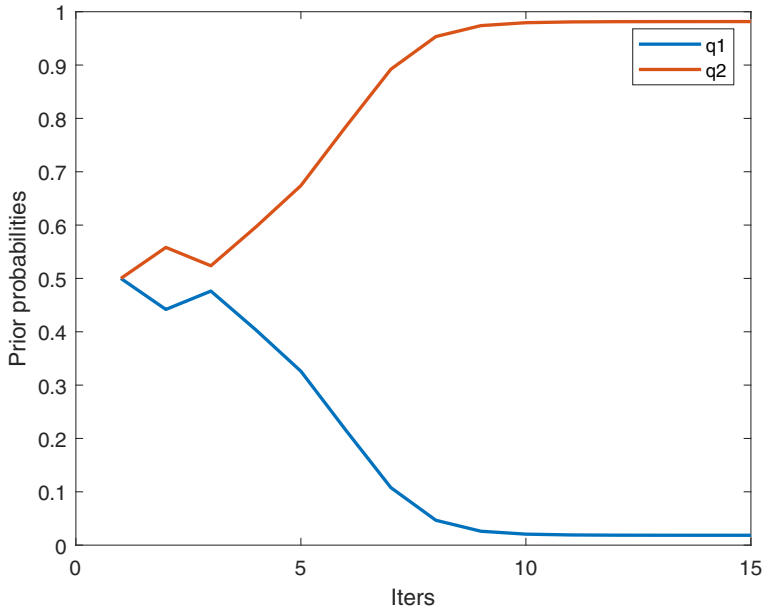

Fig. 4 The prior probability of a test image after 15 iterations

Table 3 The empirical error of the methods for Example 1

Method	Inter	Time (s)	Empirical Error
Linear Discriminant Analysis (LDA)	-	502.711	0.0962
Quadratic discriminant analysis (QDA)	-	401.337	0.0623
Naive Bayes	-	111.129	0.1698
Fisher method	-	573.764	0.1226
Multi-SVM	-	338.648	0.0568
CNN	-	492.6	0.0509
VGG-19	-	6185	0.0000
Proposed - C	20	1.7088	0.0819
Proposed - E	22	1.7276	0.0623
Proposed - H	25	1.0025	0.0774
Proposed	15	1.1809	0.0393

that gradually emerge on the skin in cancer groups are a direct external manifestation of skin cancer. These rashes are infectious rashes. It's difficult to distinguish between a disease and a benign mole simply by glancing at them. This example uses a Kaggle.com dataset with two types of tumors: benign and malignant. Each set contains 160 shots, with 90 images in the other. The image samples in Fig. 5 show a significant disparity between the two groups.

We have 200 training images and 50 test images based on the 5-folds approach. The featured interval extraction approach is then used to extract the corresponding training and test intervals. After 15 iterations, we get the following categorization result using Phase (ii) of the proposed technique:

– Prior probabilities: 0.8717, 0.8922, 0.9481, ..., 0.9036, 0.9531.
– Empirical error: 0.0974

Comparing the proposed method and others, we obtain Table 4.

The proposed technique yields an empirical error of 0.0974, which is the least error, as shown in Table 4. This indicates that the recommended method is superior to the others. One possible explanation for this result is that CNN and VGG-19 must train to find the suitable

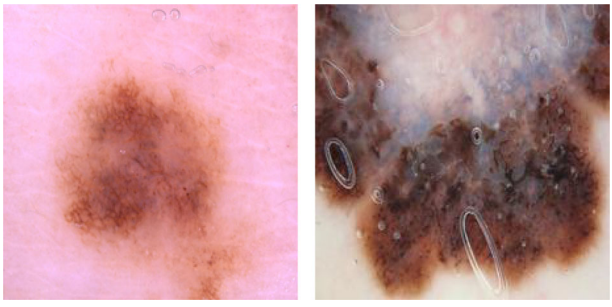

Fig. 5 Sample images of 2 groups: Begign tumor and Melanoma

Table 4 The empirical error of the methods for Example 3

Method	Inter	Time (s)	Empirical Error
Linear Discriminant Analysis (LDA)	-	517.534	0.1746
Quadratic discriminant analysis (QDA)	-	349.276	0.1746
Naive Bayes	-	428.1140	0.1986
Fisher method	-	332.1910	0.1783
Multi-SVM	-	153.7810	0.1709
CNN	-	172.400	0.1600
VGG-19	-	3095.00	0.1200
Proposed - C	45	45.2161	0.1600
Proposed - E	49	37.5994	0.1857
Proposed - H	44	56.3471	0.1857
Proposed	34	43.7360	0.0974

features, however, the color and texture features, which we extracted in the preprocessing step, appear to be the most suitable features. In the other words, the feature found by CNN may be not as good as the original color and textures features. Therefore, in those examples, the proposed method achieves the better performance compared to CNN and VGG-19.

Example 3 We use a dataset that is larger than the previous ones to demonstrate the efficacy of the suggested strategy. It's made up of 1440 texture images from the Columbia Object Image Library (COIL-20) collection (http://deeplearning.net/datasets). There are 20 groups, each with 72 images as a sample size. Figure 6 shows a few image examples.

We also utilized 1160 training images and 280 test images ($T_1, T_2, ...T_{280}$) for classification. We carry out the classification phases in the same way as in the previous cases using the proposed technique. We get the following classification result after 225 iterations:

– Prior probabilities: 0.9023, 0.9571, ..., 0.8820, 0.9652.
– Empirical error: 0.0625

After comparing the data in Table 5, we discovered that the proposed method has a 94 percent better classification accuracy than the others, with the exception of CNN and VGG-19. The CNN and VGG-19 models, on the other hand, take more time to train, reaching at 1742s and 17103s, respectively, meanwhile the proposed methods consume the least time, nearly 200s seconds.

In general, three examples have proved the benefits of the proposed method in comparing others. It has the outstanding advantage about the number of iterations to perform and the time cost. With the large image data sets and divided into many groups, this advantage is even more evident. Regarding experimental error, the proposed method overcomes many popular methods such as LDA, QDA, Naive Bayes, Fisher, and Multi-SVM, and is equivalent to CNN and VGG-19 methods. For the simple images such as Example 1, the proposed method is better than CNN but not as good as VGG-19. With Example 3, the proposed method is better than method CNN but not as good as method VGG-19. The proposed method shows its worth in Example 2, where it outperforms CNN and VGG-19 in classification. There are three primary reasons why the proposed method performs better than others in terms of categorization.

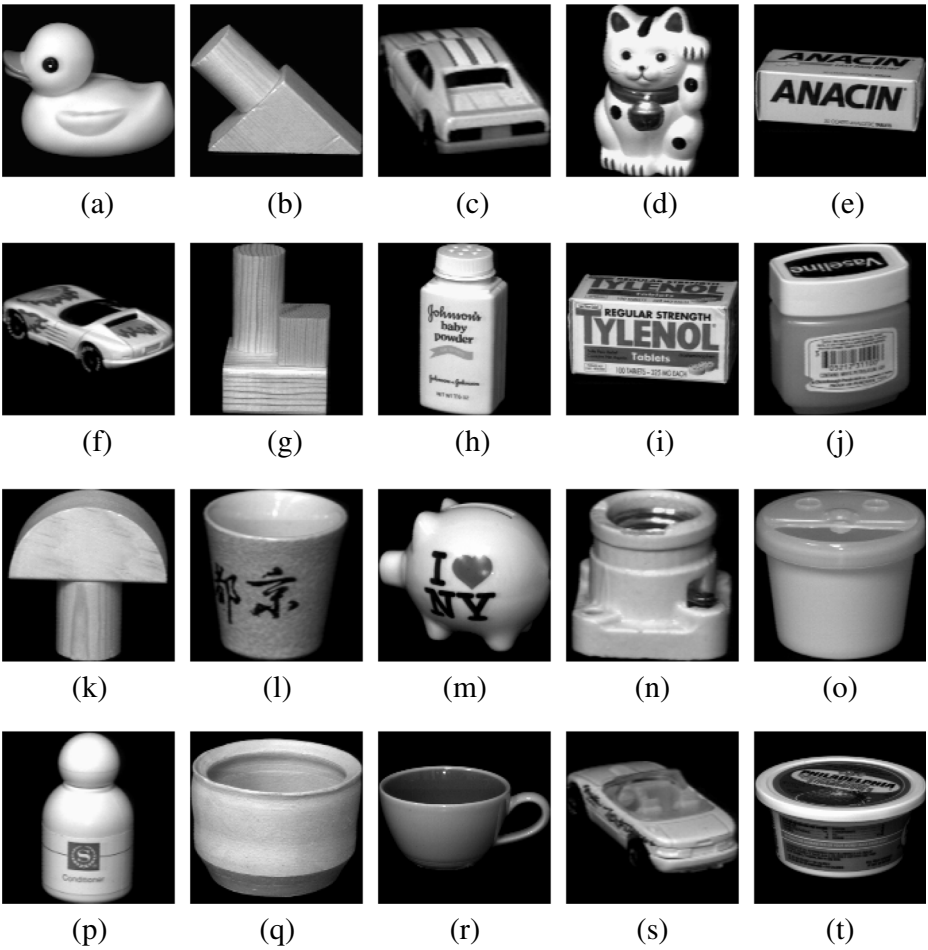

Fig. 6 Sample images of 20 groups

– Firstly, interval texture feature extraction based the proposed method is a great approach to reduce dimensional input. As a result, the proposed method achieved better classification result than others (LDA, QDA, Naive Bayes, Fisher method, and Multi-SVM) that used traditional extraction as discrete points. Conversely, CNN and VGG-19 gain good output because method can automatically extract training sample features through multi-layer convolution. That puts the input images using a set of convolution filters, each of which activates certain features from the images. However, from examples of illustration, CNN and VGG-19 method is more time-consuming than the proposed ones. The increased computational cost can be attributed to the increased width (numbers of filters) [3, 27, 39], depth (number of layers) [28, 31], smaller strides [27, 28, 39], and their combinations.
– Moreover, overlap distance is a suitable measure for interval data that can greatly affect the accuracy of the final classification as well as time cost. Thus, the total performance

Table 5 The empirical error of the methods for Example 4

Method	Inter	Time (s)	Empirical Error
Linear Discriminant Analysis (LDA)	-	387.193	0.1426
Quadratic discriminant analysis (QDA)	-	479.211	0.1137
Naive Bayes	-	321.616	0.1715
Fisher method	-	265.871	0.1285
Multi-SVM	-	315.785	0.1075
CNN	-	1742.00	0.0607
VGG-19	-	17103.00	0.0064
Proposed - C	288	249.448	0.0656
Proposed - E	248	228.205	0.0832
Proposed - H	214	101.145	0.0729
Proposed	225	204.200	0.0625

of the proposed method will be improved if we use the overlap distance rather than the City-Block, the Euclidean, and the Hausdorff distances.

– Finally, the proposed method use fuzzy clustering algorithm to find the prior probability for each test image that achieves more effectiveness than a prior assumptions used by other methods. This does not guarantee the global minimum [30]. For instance, Naive Bayesian is a simple probabilistic classifier based on the Bayesian theorem with the (naive) independence assumption.

5 Conclusion

This study considers the classification problem with a proposed method. In this study, the new method for extracting interval features from an image is given. In addition, the prior probabilities for upgrading the Bayesian classifier are discovered by the automatic algorithm that their values depend on the fuzzy relationship between the classified image with the populations. Moreover, using the suitable measure for intervals, the classification rule is created. The numerical results show that the proposed method outperforms the existing ones including LDA, QDA, Naive Bayes, Fisher method, Multi-SVM and CNN ones. For the VGG-19, a very powerful method at present, our method is considered equivalent about empirical error but has the outstanding advantages about time cost. The proposed model's application in image categorization is a significant contribution. This article is the combination of several improvements made at various stages. As a result, we think that the proposed method is a significant contribution, and believe that it can be applied to the variety of practical image classification challenges, particularly in the medical system.

Acknowledgements This research is funded by Ministry of Education and Training in Vietnam under grant number B2022–TCT–03.

Declarations

Conflict of Interests No potential conflict of interest is reported by the authors.

References

1. Ali L, Wajahat I, Golilarz NA et al (2020) Lda–ga–svm: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine. Neural Computing and Applications, pp 1–10
2. Armi L, Fekri-Ershad S (2019) Texture image classification based on improved local quinary patterns. Multimed Tools Applic 78(14):18995–19018
3. Chatfield K, Simonyan K, Vedaldi A, Zisserman A (2014) Return of the devil in the details: Delving deep into convolutional nets. In: BMVC
4. Chen J, Shan S, He C et al (2009) Wld: Arobustlocalimage descriptor. IEEE Trans Pattern Anal Mach Intell 32(9):1705–1720
5. Cortes C, Vapnik V (1995) Support-vector networks. Machin elearning 20(3):273–297
6. Csevik U, Karakullukcsu E, Berber T et al (2019) Automatic classification of skin burn colour images using texture based feature extraction. IET Image Process 13(11):2018–2028
7. Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In: 2004 conference on computer vision and pattern recognition workshop. pp 178–178. IEEE
8. Fisher RA (1938) The statistical utilization of multiple measurements. Ann Eugen 8(4):376–386
9. Fisher RA (1992) Statistical methods for research workers. In: Breakthroughs in statistics, pp 66–70. Springer
10. Ha CN, Thao NT, Tuan NB et al (2020) A new approach for face detection using the maximum function of probability density functions. Annals of Operations Research 1–21
11. Haralick RM (1979) Statistical and structural approaches to texture. Proc IEEE 67(5):786–804
12. Hearst MA, Dumais ST, Osuna E et al (1998) Support vector machines. IEEE Intell Syst Applic 13(4):18–28
13. Hiremath P, Pujari J (2007) Content based image retrieval based on color, texture and shape features using image and its complement. Int J Comput Sci Secur 1(4):25–35
14. Hoang ND (2019) Automatic detection of asphalt pavement raveling using image texture based feature extraction and stochastic gradient descent logistic regression. Autom Constr 105:102843
15. Isa NM, Amir A, Ilyas M et al (2019) Motor imagery classification in brain computer interface (bci) based on eeg signal by using machine learning technique. Bull Electr Eng Inform 8(1):269–275
16. Jardine M, Miller J, Becker M (2018) Coupled x-ray computed tomography and grey level co-occurrence matrices as a method for quantification of mineralogy and texture in 3d. Comput & Geosci 111:105–117
17. Khaldi B, Aiadi O, Kherfi ML (2019) Combining colour and grey-level co-occurrence matrix features: a comparative study. IET Image Process 13(9):1401–1410
18. Khan MN, Ahmed MM (2019) Snow detection using in-vehicle video camera witht exture-based image features utilizing k-nearest neighbor, support vector machine, and random forest. Trans Res Rec 2673(8):221–232
19. Lloyd K, Rosin PL, Marshall D et al (2017) Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (glcm) based texture measures. Mach Vis Applic 28(3-4):361–371
20. Mohebian R, Riahi MA, Yousefi O (2018) Detection of channel by seismic texture analysis using grey level co-occurrence matrix based attributes. J of Geophysics Eng 15(5):1953–1962
21. Murphy KP et al (2006) Naive bayes classifiers. University of British Columbia 18:60
22. Ngoc L, Tuan L, Tai V (2021) Automatic clustering algorithm for interval data based on overlap distance. Communications in Statistics - Simulation and Computation. https://doi.org/10.1080/03610918.2021.1900248
23. Nhu VH et al (2020) Comparison of support vector machine, bayesian logistic regression, and alternating decision tree algorithms for shallow landslide susceptibility mapping along a mountainous road in the west of iran. Appl Sci 10(15):5047
24. Pham BT, Prakash I (2019) Evaluation and comparison of logitboost ensemble, fisher's linear discriminant analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. Geocarto Int 34(3):316–333
25. Ren Y, Liu YH, Rong J et al (2009) Clustering interval-valued data using an overlapped interval divergence. In: Proceedings of the eighth australasian data mining conference, vol 101, pp 35–42
26. Scott DW (2015) Multivariate density estimation: theory, practice, and visualization. Wiley, NJ
27. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y (2014) Overfeat: Integrated recognition, localization and detection using convolutional networks
28. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

29. Spacek L (1996) Description of libor spacek's collection of facial images
30. Sultan KS, Selim SZ (1993) Global algorithm for fuzzy clustering problem. Pattern Recognit 26:1357–1361
31. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S et al (2014) Going deeper with convolutions. arXiv:1409.4842, 2014
32. TAI VV (2018) Some results of classification problem by bayesian method and application incredit-operation. Stat Theory Related Fields 2(2):150–157
33. Tai VV, Ha CN, Thao NT (2017) Textural features selection for image classification by bayesian method. In: 2017 13th International Conference on Natural Computation. Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). pp 733–139. IEEE
34. Tai VV, Thao TN (2018) Similar coefficient of cluster for discrete elements. Sankhya B 80(1):19–36
35. Vovan T, Phamtoan D, Tuan LH et al (2020) An automatic clustering for interval data using the genetic algorithm. Annals of Operations Research, pp 1–22
36. Wang PW, Lin CJ (2014) Iteration complexity of feasible descent methods for convex optimization. J Mach Learn Res 15(1):1523–1548
37. Wang Y, Shi F, Cao L et al (2019) Morphological segmentation analysis and texture-based support vector machines classification on mice liver fibrosis microscopic images. Curr Bioinform 14(4):282–294
38. Yulita I, Novita D, Sholahuddin A et al (2020) Electroencephalography based emotion recognition using fisher's linear discriminant analysis on support vector machine. In: Journal of physics: Conference series 1577, 012004. IOP Publishing
39. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional neural networks. In: ECCV
40. Zhang X, Cui J, Wang W et al (2017) A study for texture feature extraction of high resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. Sensors 17(7):1474