Algorithm for training the minimum error one-class classifier of images

J. T. Guillen-Bonilla,[1] E. Kurmyshev,[1,*] and E. González[2]

[1]Departamento de Metrología Óptica, Centro de Investigaciones en Óptica A. C., Loma del Bosque No. 115, Col. Lomas del Campestre, C.P. 37150, León, Gto., México

[2]Department of Engineering Design, Universidade de Vigo, ETSII, Campus Universitario, E-36310 Vigo, Spain

*Corresponding author: kev@cio.mx

Received 1 October 2007; revised 10 December 2007; accepted 12 December 2007; posted 19 December 2007 (Doc. ID 88140); published 25 January 2008

We propose a training algorithm for one-class classifiers in order to minimize the classification error. The aim is to choose the optimal value of the slack parameter, which controls the selectiveness of a classifier. The one-class classifier based on the coordinated clusters representation of images is trained and then used for the classification of texture images. As the slack parameter C varies through a range of values, for each C, the misclassification rate is computed using only the training samples. The value of C that yields the minimum misclassification rate, estimated over the training set, is taken as the optimal value, C_{opt}. Finally, the optimized classifier is tested on the extended database of images. Experimental results demonstrate the validity of the proposed method. In our experiments, classification efficiency approaches, or is equal to, 100%, after the optimal training of the classifier.

OCIS codes: 100.3008, 100.5010, 150.3040.

1. Introduction

We are dealing with a one-class classification problem, meaning that one class of objects is well defined, while all other classes are either of no interest or poorly described. In one-class classification a target class is separated from all other objects considered as outliers. Given a test object, we have to decide if the object belongs to the target class or is an outlier. A well-sampled training set of objects of the target class is usually available, while poor (or no) information of the outliers is known. High measurement cost and low frequency of an event are some of the reasons for the poor information about outliers. In multiclass classification a decision boundary is supported from both sides by samples of adjacent classes. In one-class classification (also called outlier rejection) a set of features of a target class is available, and thus only one side of the closed decision boundary is supported. This makes the problem of one-class classification more difficult than multiclass classification.

An example of a one-class classifier problem is given by polished granite (or ceramic) tiles that are widely used as construction elements. In the industry of granite tiles, a usual task is to choose tiles with similar appearance, independent of the origin or of a previous selection. Due to the natural origin of granite, the visual aspect of tiles may differ significantly, even within a particular variety and denomination. The polished granite tiles are usually inspected by a human expert using a chosen master tile as the reference. The qualitative parameters used to evaluate are mainly a dominant color, a texture granularity, and a number of discriminant colors. As a result of this classification, usually called "similarity judgment," the granite tiles of a similar visual aspect are grouped in lots. Such inspection is subjective and qualitative and is not based on formally defined criteria. Application of artificial vision techniques and digital image processing to the classification of ornamental rocks and ceramic tiles looks very promising. Advances in research for methods and systems that perform the visual inspection of granite and ceramic tiles automatically, giving quantitative and reproducible results in real time, are reported in [1–5].

A number of general approaches for solving the one-class problem have been proposed [6–14]. Models of one-class classifiers can be divided into two groups, statistics- (density-based) and domain-based models.

0003-6935/08/040541-07$15.00/0

In the statistical approach, the target class is described by a probability density function in a feature space. A suitable threshold is selected to determine the class (decision) boundary in the feature space. A test object is assigned to the target class if the estimated probability is higher than the given threshold. A statistical approach involves a density estimation of the target class. The estimation of conditional probabilities is a nontrivial problem, especially when a limited amount of data is available.

The domain-based approach determines a boundary around the target class that minimizes the surrounded volume; that is a constrained optimization. The advantage of this approach is that no probability density of a target class has to be estimated; the local density variation is ignored. The decision boundary follows the shape of the target class distribution and does not focus on the high density regions. The training set size is usually smaller for domain-based methods than for density-based approaches. Since classification using a domain-based approach is a constrained optimization, the higher the dimension of the feature space the harder the problem becomes.

In contrast to general approaches, application oriented classifiers usually use a specified (not an abstract) feature space. Recently developed, the coordinated clusters representation (CCR) of images was shown to be efficient in image classification [15–20]. Being a statistical method, the CCR does not estimate the probability densities of texture images explicitly. To describe a class in the CCR feature space we can use only a few training images of the target class; in the limit case, it can be a single image. Because of high dimensionality, the CCR feature space is difficult to use with the conventional statistics- and domain-based methods of classification. These methods are used to determine a closed decision boundary of simply connected domains in a feature space. Low- and high-dimensional feature spaces are treated in the same way (see, for example [12–14]), without paying attention to the ring-shaped probability distribution of feature vectors around the mass center of class, in a high-dimensional feature space [8]. That is the reason for an efficient one-class classifier in the CCR feature space that was proposed in [16]. As shown in the cited work, varying the classifier slack parameter C allows either the recognition of a single image or grouping of similar images in a class. In this paper we propose an algorithm for the calculation of the C value that minimizes the one-class classification error and test this algorithm on a representative image database that includes outliers.

To have a self-contained paper, we give a brief description of one-class image classification in the CCR feature space in Section 2, following [16]. Image feature extraction is described in Subsection 2.A. The criterion for one-class classification is given in Subsection 2.B. An algorithm of optimal training of the classifier is discussed in Subsection 2.C. The classification setup and experimental results are described in Section 3, followed by conclusions in Section 4.

2. Quasi-Statistical Criterion of One-Class Classification

A. Feature Extraction and Training

Suppose that Q images S_q ($q = 1, 2, \dots, Q$) are known images of a class. In the training stage, a random set of P subimages S_q^{α} ($\alpha = 1, 2, \dots, P$) is sampled from each image S_q. The image S_q can be characterized by the three following magnitudes. The CCR prototype distribution function of the qth set of subimages S_q^{α} ($\alpha = 1, 2, \dots, P$) is

$$F_q = \frac{1}{P}\sum_{\alpha=1}^{P} F_q^{\alpha}, \tag{1}$$

where F_q^{α} is the CCR distribution function of the subimage S_q^{α}. Let $d(F_q^{\alpha}, F_q)$ be a distance of F_q^{α} from F_q in the CCR feature space. Since F_q^{α} and F_q are random variables, this distance is also a random variable. Then, the estimated value of distance $d(F_q^{\alpha}, F_q)$ and the variance,

$$\langle d \rangle_q = \frac{1}{P}\sum_{\alpha=1}^{P} d(F_q^{\alpha}, F_q), \tag{2}$$

$$\sigma_q^{2} = \frac{1}{P}\sum_{\alpha=1}^{P} \left(d(F_q^{\alpha}, F_q) - \langle d \rangle_q\right)^2, \tag{3}$$

are also characteristic magnitudes of image S_q.

To characterize the whole class, identified with the known images S_q, we calculate the five magnitudes, F, D, σ, $\tilde{D}$, and $\tilde{\sigma}$ as follows. Since each of the texture images S_q is interpreted as a random field of pixels, and each of Q sets of subimages is randomly sampled from this field, F_q, $\langle d \rangle_q$, and σ_q are random too. Hence, we can calculate the corresponding estimates. The class prototype distribution function is the mean,

$$F = \frac{1}{Q}\sum_{q=1}^{Q} F_q = \frac{1}{PQ}\sum_{q=1}^{Q}\sum_{\alpha=1}^{P} F_q^{\alpha}, \tag{4}$$

that is, the mass center of subimages in the CCR feature space. The mean of distances $\langle d \rangle_q$ and the mean of standard deviations σ_q are, respectively,

$$D = \frac{1}{Q}\sum_{q=1}^{Q} \langle d \rangle_q = \frac{1}{PQ}\sum_{q=1}^{Q}\sum_{\alpha=1}^{P} d(F_q^{\alpha}, F_q), \tag{5}$$

$$\sigma = \frac{1}{Q}\sum_{q=1}^{Q} \sigma_q. \tag{6}$$

D and σ show the distribution of F_q^{α} with respect to the mass center F_q of the qth set of subimages. In addition, we calculate the mean distance $\tilde{D}$ of the qth centers F_q from the center of the class, F, and the variance $\tilde{\sigma}^2$:

$$\tilde{D} = \frac{1}{Q}\sum_{q=1}^{Q} d(F_q, F), \quad (7)$$

$$\tilde{\sigma}^2 = \frac{1}{Q}\sum_{q=1}^{Q} \left(d(F_q, F) - \tilde{D}\right)^2. \quad (8)$$

$\tilde{D}$ and $\tilde{\sigma}$ characterize the distribution of mass centers F_q with respect to the class prototype distribution function F.

The above formulas describe the training phase of the one-class classifier in the CCR feature space. Before proceeding with the recognition phase, we would like to comment on the case when a single image S of a class is known. This can have a practical interest for similarity judgment of granite tiles. To begin training, Q independent random sets of subimages are sampled from the master image S. Each qth set of subimages is considered as the sampling of a virtual image S_q ($q = 1, 2, \ldots, Q$). Then, the training follows Eqs. (1)–(8).

B. One-Class Classification Criterion

The basic idea of the selection criterion is simple. To assign a test image to, or exclude it from, the class of source images, we have to compare the statistics of subimages sampled from the test image. The following five characteristics of the known images of a class are used as references to compare the statistics: the CCR prototype distribution function F, the mean distances $\tilde{D}$ and D, and the standard deviations $\tilde{\sigma}$ and σ.

Given a test image S_{test}, we sample a random set of subimages S^{α} and calculate the CCR distribution functions F^{α}, where $\alpha = 1, 2, \ldots, K$, and K is the number of samples. Then the mass center of functions F^{α} and the mean distance of subimages from the mass center are calculated as follows

$$F_{\text{test}} = \frac{1}{K}\sum_{\alpha=1}^{K} F^{\alpha}, \quad (9)$$

$$D_{\text{test}} = \frac{1}{K}\sum_{\alpha=1}^{K} d(F^{\alpha}, F_{\text{test}}). \quad (10)$$

It is important to realize that, if samples of a random set are drawn from a high-dimensional space (that is the case of CCR distribution functions), most elements will fall down in an N-dimensional ring centered on the mass center, and no samples will fall in the central region, where the value of the density function is the largest. This phenomenon is observed in our computer experiments with texture images and explained in [8]. Thus we can expect random sets of subimages to be similar if they have almost identical distribution rings and the centers of rings are close in the CCR feature space. Following this idea, the test image S_{test} is assigned to the class of source images S ($S_q(q = 1, 2, \ldots, Q)$) if, and only if,

$$\tilde{D} - 2\tilde{\sigma} \le d(F_{\text{test}}, F) \le \tilde{D} + C\tilde{\sigma}, \quad (11)$$

$$D - 2\sigma \le D_{\text{test}} \le D + 2\sigma, \quad (12)$$

where the adjustment parameter C is used to vary the selectiveness of the classification criterion. The one-class classification criterion consists of two conditions. The first one is a constraint to the position of the test prototype F_{test} with respect to the learned prototype distribution function F of the class, while the second determines the mean distance of subimage distribution function F^{α} from the test prototype F_{test}.

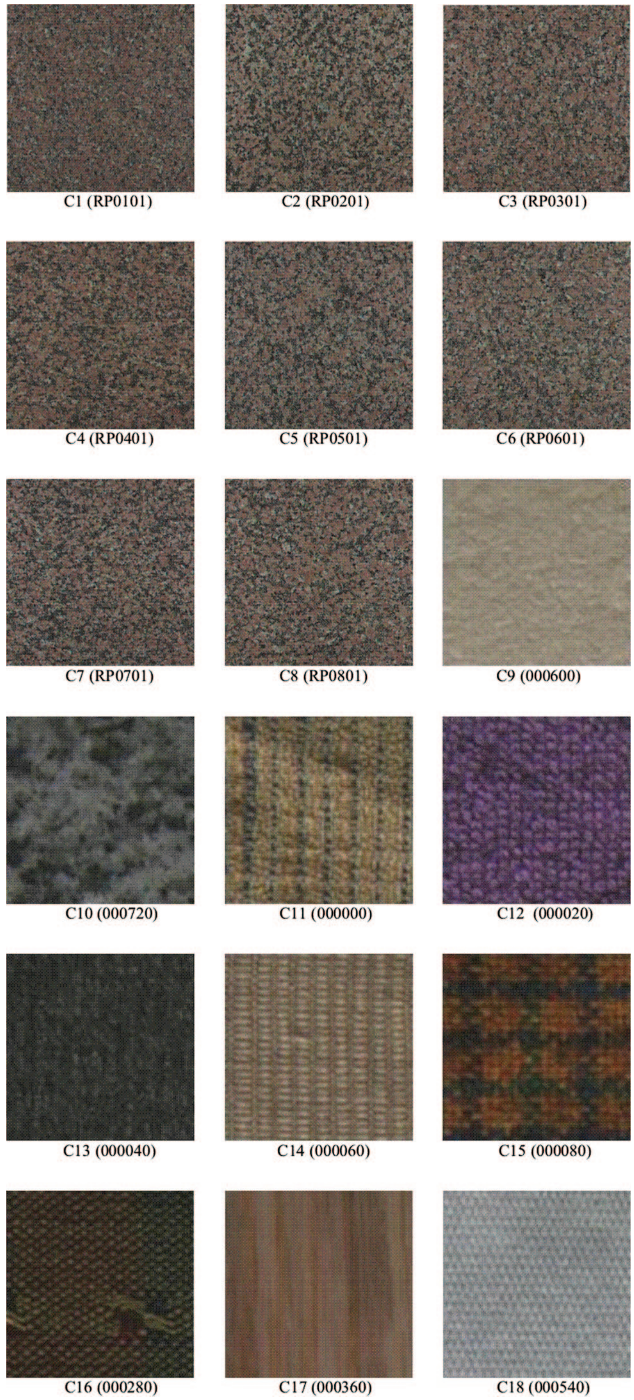

Fig. 1. (Color online) Source images from Rosa Porriño granite and OUTEX catalogs numbered 1–18, in row order.

Table 1. Outliers' Classes Used to Calibrate the Classifier

Target Class	Similar Classes of Outliers		Random Classes of Outliers	
C1	C5	C8	C11	C15
C2	C5	C7	C7	C8
C3	C4	C6	C10	C17
C5	C2	C6	C3	C17
C9	C11	C13	C1	C10
C18	C14	C16	C10	C7

For $C \leq 1$, the one-class classifier recognizes individual images, mainly. No association of images into classes takes place because the criterion is too selective. When C takes larger values, the decision criterion becomes more inclusive, and the association in classes takes place [16]. So, the "purity" of a class is controlled with the parameter C. In Subsection 2.C we solve the inverse problem, showing how to choose the value of the adjustment parameter C to provide a required (low error) selectiveness of the one-class classifier. In other words, we choose the C that minimizes the error for one-class classification into the class that has been defined by means of the training set of images.

C. Minimum Error Classification

Since the selectiveness of the classifier is governed by the adjustment parameter C of Eq. (11), it is important to calibrate this parameter for each particular application. In this subsection we describe an algorithm for calibrating the parameter C that minimizes the prediction error of the classifier. Generalized resampling methods to estimate the prediction error of classifiers are discussed, for example, in [6,21]. These methods imply an estimate of empirical distributions of the learning and test sets to calculate the conditional risk function. To avoid this complication, we use a semiempirical but efficient method for the error estimation in the calibration of the adjustment parameter of the one-class classifier. Note that the calibration needs a training and a validation set of images.

The prediction error of a one-class classifier is expressed as follows

$$Er(\varepsilon_t,\ \varepsilon_o) = \lambda\varepsilon_t + (1-\lambda)\varepsilon_o, \tag{13}$$

where λ is a slack parameter; ε_t and ε_o are the target rejection and the outlier acceptance rates [14]. In one-class classification, only examples of a target class are available, usually, that is, $\lambda = 1$ in Eq. (13). Therefore ε_t can be estimated reliably, but further assumptions are needed to estimate ε_o. When both sets of training images and of outliers are representative, one can take $\lambda = 0.5$.

In the training of the classifier, one might use a simple rule: minimize the total classification error on the validation set. After the five magnitudes, F, D, σ, $\tilde{D}$, and $\tilde{\sigma}$ have been calculated on a training set of images of a target class, we can evaluate the performance of the one-class classifier, Eqs. (11) and (12), by varying the adjustment parameter C in a range of values $C = C_1, C_2, \ldots, C_k$. The validation set, used for this purpose, includes images of both the target class and of the outliers. In classifying an image, one of the four possibilities takes place. The image is: (i) correctly assigned to the class; (ii) rejected correctly; (iii) assigned incorrectly; (iv) rejected incorrectly. Let N_{ac}, N_{rc}, N_{ai}, and N_{ri} be the numbers of images (i) correctly assigned to the class; (ii) rejected correctly; (iii) assigned incorrectly; (iv) rejected incorrectly. The rates of misclassification and correct assignment are defined as

$$M = \left(\frac{N_{\mathrm{ai}} + N_{\mathrm{ri}}}{N}\right) = \varepsilon_o + \varepsilon_t, \tag{14}$$

$$A = \left(\frac{N_{\mathrm{ac}} + N_{\mathrm{rc}}}{N}\right), \tag{15}$$

where N is the number of images used in the performance evaluation of the one-class classifier for a particular C. The equality $A + M = 1$ is satisfied. Searching for the optimal value of C that minimizes the rate of misclassification M, we run the experiment through the range of values $C = C_1$, $C_2, \ldots, C_k$ to obtain the error function $M = f(C)$. The misclassification increases gradually with C, because the criterion becomes "all" inclusive (the number N_{ai} increases). When C gets smaller, the misclassification increases also, because the criterion becomes very exclusive (the number N_{ri} increases). So the function $M = f(C)$ is expected to have a minimum, and the value of C that gives the absolute minimum to M is taken to be the optimal value of the one-class classifier for images of a given class. When the error function minimum is not unique, there is an interval of possible values of $C \in (C_{\min}, C_{\max})$ that provide the function's minimum. Then, we choose the middle point, $C_{\mathrm{opt}} = (C_{\max} + C_{\min})/2$ as the optimal value.

Finally, the algorithm to compute an optimal value of the slack parameter C (or parameters, when a classifier is multiparametric) can be summarized in the following list of instructions:

Table 2. Optimal Values for the Adjustment Parameter *C* Corresponding to Each of the Six Classes[a]

Target Class	C1	C2	C3	C5	C9	C18
Optimal value, C_{opt}	1.4 (1.4)	1.2 (1.2)	1.4 (1.5)	1.3 (1.2)	1.6 (1.7)	2 (2.1)
Classification efficiency (%)	99.62 (98.51)	99.25 (98.88)	99.62 (99.25)	99.25 (98.88)	100 (100)	100 (100)

[a]Calibration results using random classes of outliers are given in parentheses.

1. Train a classifier over a training set of images of a target class.

2. Choose a range and a step of variation for the slack parameter C.

3. Calculate the error function $M = f(C)$ for different values of C, using a validation set of images; this set includes images of the target class and outliers' images.

4. Determine a global minimum for the error function, $M = f(C)$, and take this C as the one-class classifier optimal value for classifying images into the given target class.

3. Experimental Results

A. Classification Setup

The experiment to evaluate the algorithm of optimal training of the one-class classifier was implemented as follows. To obtain 18 classes of training and test images we used 18 source images of 512×512 pixels, shown in Fig. 1. Images 1 through 8 are of Rosa Porriño granite tiles, while images 9 through 18 were taken from the OUTEX database [22]. Note that images 2 and 7 are very similar [19]; human experts consider tiles 2 and 7 to be of the same class, taking into account the visual texture and color. Primary color images were converted into gray level images, since the one-class criterion ignores color features. Then, the source images were reduced to the size of 205×205 pixels, as required by the optimal scale criterion [18]. Each of the 18 classes was obtained by extracting $Q = 30$ random images S_q ($q = 1, 2, \ldots, Q$) of smaller size, 150×150 pixels, from a reduced source image. Fifteen of the 30 images of each class were used in the training of the classifier. The validation set for each given class consisted of 15 training images of the same class and 30 images of two other classes, the outliers. This set was used to calculate the error function $M = f(C)$. Both in the training and the test stage, a set of $P = K = 30$ random subimages S_q^{α} ($\alpha = 1, 2, \ldots, P$) of the size 50×50 pixels was sampled from each image S_q. The fuzzy C-means method with the iterative optimization algorithm was used for the binarization of gray level images. The CCR histograms were calculated with the 3×3 scanning window in all experiments. The Manhattan distance between points in the CCR feature space was used as the measure of distinction between images.

To train the one-class classifier, we calculated the five parameters of a class: the prototype distribution function F, the mean distances $\tilde{D}$ and D, and the standard deviations $\tilde{\sigma}$ and σ as described in Subsection 2.A. Then, the error function $M = f(C)$ was calculated for each of the 18 classes, varying C with the step 0.2 in the interval $C \in [0.2, 10]$. In addition to 15 images of the class, 30 images of two different classes were used as outliers. So, $N = 45$ in Eq. (14). The classes of outliers were selected in two different ways: similar and randomly selected classes. The absolute minimum of the error function $M = f(C)$ is the opti-

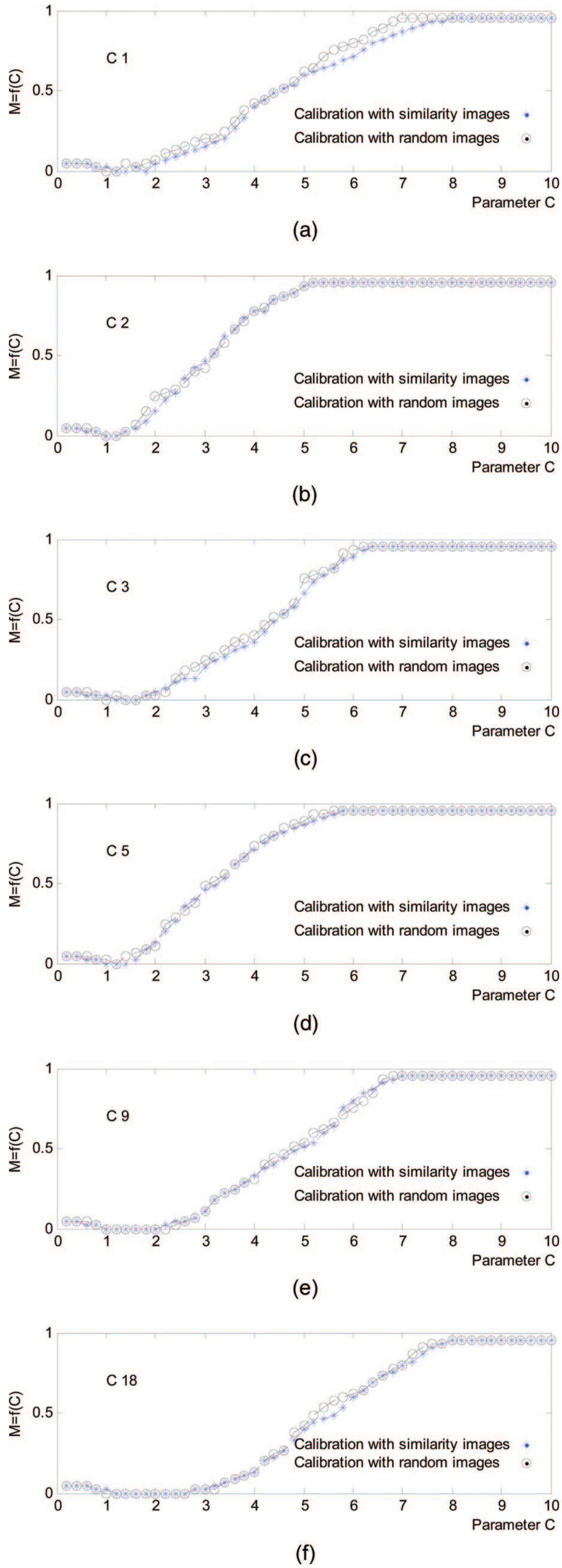

Fig. 2. (Color online) Plots of the error function $M = f(C)$ for target classes C1, C2, C3, C5, C9, and C18 using classes of outliers similar to the target class or outliers of random classes.

Table 3. Performance of the Optimized One-Class Classifier

Target Class	C1	C2	C3	C4	C5	C6	C7	C8	C9
Optimal value, C_{opt}	1.4	1.2	1.4	1.6	1.3	1.4	1.0	1.7	1.6
Classification efficiency (%)	99.62	99.25	99.62	99.62	99.25	99.25	99.62	100	100

Target Class	C10	C11	C12	C13	C14	C15	C16	C17	C18
Optimal value, C_{opt}	1.7	1.9	1.7	1.6	1.9	2.0	2.1	1.9	2.0
Classification efficiency (%)	100	99.62	100	100	100	100	100	100	100

mal value of C for the classifier of images of the given class. When the minimum of the error function is not unique, we take the middle point $C_{opt} = (C_{max} + C_{min})/2$ of the minima interval (C_{min}, C_{max}) as the optimal value of the adjustment parameter. Finally, the classifier's performance with the optimal C_{opt} is evaluated by classifying each of the $15 \times 18 = 270$ images of 150×150 pixels; 15 images being of the target class and others being outliers.

B. Test Results

The outliers used in the calibrating of the one-class classifier can be selected from the available database in two different ways: either outliers are qualitatively similar to the target class images, or they are selected randomly from a database. To see the effect of the outliers selection, we calibrate the classifier for six target classes of images (C1, C2, C3, C5, C9, and C18) in two ways. We then evaluated the classifier efficiency using a 270 images database for 18 classes; 15 images per class. Table 1 shows the target classes along with the outliers selected in both ways. Note that the similarity between classes was estimated qualitatively; we did not use a specific measure of similarity [19]. Results of the classifier calibration for each of the six classes are given in Table 2, along with the classifier efficiency over the 270 images database.

Table 2 shows that the selection of outliers has a small influence on C_{opt} and, hence, on the optimized one-class classifier performance. Nevertheless, the classification efficiency is a little higher when outliers are similar to the target class images, because the uncertainty in the determination of the error function's minimum $M = f(C)$ diminishes, as can be seen from the function's plots (see Fig. 2).

The final classification test was done for each of the 18 classes, using 270 test images in total. Outliers similar to a target class were used in the calibration stage of the one-class classifier. Optimal values of the adjustment parameter, C_{opt}, and the performance of the classifier are shown in Table 3. Classification efficiency with the optimized one-class classifier is shown to be higher than 99%.

4. Conclusions

In this paper we have proposed and studied an algorithm for the optimal training of the one-class classifier, using the coordinated clusters representation of images. In contrast to the approaches based on the constrained optimization, we used a direct calibration of the classifier, varying its adjustment parameter C. For the calibration we used images of the target class and a small number of outliers. The error function's minimum indicates the optimal value for the parameter C. In our experiments, the classification efficiency approaches, or is equal to, 100% when an optimal value of the parameter C is used for the classifier. Overlapping images are used in the training and recognition phases of the classifier (see also [18,19]). Note that the training of a classifier and a definition of a class are related items. Certainly, this influences the classification rate. In our opinion, a definition of a class using overlapping images is more adequate for texture images than that done with nonoverlapping, adjacent images. As a consequence, a higher classification rate is expected, using overlapping images. Unfortunately, the training of classifiers over a database of overlapping images is subestimated and not often used in classification experiments.

In practice the variation of C in the calibration process can be limited to the interval $C \in (0.5, 3)$, which contains the minimum of the misclassification function. In addition, in our experiments we see that any value of the adjustment parameter selected in the interval $C \in (1, 2)$ *a priori*, gives an error for the one-class classification of $<5\%$. When a better performance is required of the one-class classifier, it has to be calibrated using the algorithm described above. The algorithm does not need a special selection of outliers in the classifier's training stage, even though better efficiency is achieved when outliers are similar to the target class images.

E. Kurmyshev thanks the Centro Universitario de los Lagos, Universidad de Guadalajara, México for its hospitality during the sabbatical leave from the Centro de Investigaciones en Óptica (CIO). J. T. Guillen-Bonilla thanks the Consejo Nacional de Ciencia y Tecnologia (CONACYT) of México for the scholarship. This research was partially supported by the Xunta de Galicia under grant PGIDIT04REM303003PR.

References

1. K. Song, J. Kittler, and M. Petrou, "Defect detection in random color textures," Image Vis. Comput. **14,** 667–683 (1996).
2. F. Lumbreras, R. Baldrich, M. Vanrell, J. Serrat, and J. J. Villanueva, "Multiresolution color texture classification of ceramic tiles," in *Recent Research Developments in Optical Engineering* (Research Signpost, 1999), Vol. 2, pp. 213–228.
3. M. L. Smith and R. J. Stamp, "Automated inspection of textured ceramic tiles," Comput. Ind. **43,** 73–82 (2000).

4. G. Paschos, "Fast color texture recognition using chromaticity moments," Pattern Recogn. Lett. **21,** 837–841 (2000).
5. S. Kukkonen, H. Kälviäinen, and J. Parkkinen, "Color features for quality control in ceramic tile industry," Opt. Eng. **40,** 170–177 (2001).
6. R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," 2nd ed. (Wiley, 2001).
7. M. Unser, "A fast texturé classifier based on cross entropy minimization," in *Proceedings of the EUSIPCO'83* (North-Holland, 1983), pp. 261–264.
8. K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd ed. (Academic, 1990).
9. K. Irahama and Y. Furukawa, "Gradient descent learning of nearest neighbor classifiers with outlier rejection," Pattern Recogn. **28,** 761–768 (1995).
10. M. M. Moya and D. R. Hush, "Network constraints and multi-objective optimization for one-class classification," Neural Networks **9,** 463–474 (1996).
11. D. M. J. Tax and R. P. W. Duin, "Support vector domain description," Pattern Recogn. Lett. **20,** 1191–1199 (1999).
12. D. M. J. Tax, "One-class classification: concept-learning in the absence of counterexamples," Ph.D. dissertation (Delft University of Technology, 2001).
13. D. M. J. Tax and R. P. W. Duin, "Uniform object generation for optimizing one-class classifiers," J. Mach. Learn. Res. **2,** 155–173 (2002).
14. P. Juszczak, "Learning to recognize: a study on one-class classification and active learning," Ph.D. dissertation (Delft University of Technology, 2006).
15. R. E. Sánchez-Yáñez, E. V. Kurmyshev, and F. J. Cuevas, "A framework for texture classification using the coordinated clusters representation," Pattern Recogn. Lett. **24,** 21–31 (2003).
16. R. E. Sánchez-Yáñez, E. V. Kurmyshev, and A. Fernández, "One-class texture classifier in the CCR feature space," Pattern Recogn. Lett. **24,** 1503–1511 (2003).
17. E. V. Kurmyshęv and R. E. Sánchez-Yáñez, "Comparative experiment with colour texture classifier using the CCR feature space," Pattern Recogn. Lett. **26,** 1346–1353 (2005).
18. E. V. Kurmyshev, M. Poterasu, and J. T. Guillen-Bonilla, "Image scale determination for optimal texture classification using coordinated clusters representation," Appl. Opt. **46,** 1467–1476 (2007).
19. J. T. Guillen-Bonilla, E. Kurmyshev, and A. Fernández, "Quantifying a similarity of classes of texture images," Appl. Opt. **46,** 5562–5570 (2007).
20. E. V. Kurmyshev, "Classification of texture images using coordinated clusters representation, in *Recent Advances in Optical Metrology* (Research Signpost, 2007), pp. 155–226.
21. A. M. Molinaro, R. Simon, and R. M. Pfeiffer, "Prediction error estimation: a comparison of resampling methods," Bioinformatics **21,** 3301–3307 (2005).
22. OuTex, Texture Database, http://www.outex.oulu.fi/temp/.