Reduced Multidimensional Co-Occurrence Histograms in Texture Classification

Kimmo Valkealahti and Erkki Oja, *Senior Member, IEEE*

Abstract—Textures are frequently described using co-occurrence histograms of gray levels at two pixels in a given relative position. Analysis of several co-occurring pixel values may benefit texture description but is impeded by the exponential growth of histogram size. To make use of multidimensional histograms, we have developed methods for their reduction. The method described here uses linear compression, dimension optimization, and vector quantization. Experiments with natural textures showed that multidimensional histograms reduced with the new method provided higher classification accuracies than the channel histograms and the wavelet packet signatures. The new method was significantly faster than our previous one.

Index Terms—Texture classification, multidimensional histograms, vector quantization, self-organizing map, feature selection.

1 Introduction

MONOCHROME textures are conventionally described by one-dimensional difference histograms or two-dimensional co-occurrence histograms of gray levels. Histograms may be used as such for texture description [1], [2] but, typically, the description is based on various statistics computed from the histogram [3]. Histograms with more than two dimensions have only occasionally been applied to monochrome texture description [4], [5]. Our interest in the utilization of multidimensional histograms was evoked by our previous experiments suggesting that an increase in the co-occurrence dimensionality, which improves the description of spatial relationships, benefits both monochrome and color texture classification [6]. In the present study, we explore ways to speed up the collection of multidimensional histograms. We compare our method with the channel histogram method of Unser [7] which is a well-justified improvement of the co-occurrence matrix method, and with the wavelet packet signature method of Laine and Fan [8] exemplifying recent developments in texture analysis without histogram basis. Moreover, a comparison is made with multidimensional channel histograms which are somewhat related to the method of He and Wang [4]. All comparisons are performed using Brodatz textures, and to increase the generalizability of the results, the textures are also rotated and scaled.

A K-dimensional histogram of G gray levels consists of G^K bins. In practice, such full-scale multidimensional histograms can only be used at very low values of G and K. In our approach, a substantial increase in co-occurrence dimensionality, without restrictions on pixel quantization, is made possible by vector quantization. Using a vector quantizer, the centers and intervals of histogram bins are determined by a relatively small set of code vectors $\mathbf{r}_n = (r_{n,1}, r_{n,2}, \ldots, r_{n,K})$, $n = 1, 2, \ldots, N$, and a reduced multidimensional histogram is formed as

$$h_q = \#\left\{\mathbf{s} \,\middle|\, q = \arg\min_n \left\|\mathbf{s} - \mathbf{r}_n\right\|\right\}, \tag{1}$$

in which # denotes the number of elements in a set and vector $\mathbf{s}$ comprises the co-occurring pixel gray levels. Bin h_q shows the number of times $\mathbf{r}_q$ has been selected as the best-matching code vector for co-occurrence vectors $\mathbf{s}$ from all locations in a texture block. Thus, a vector quantizer is not used as a texture classifier [9] but as a low-resolution sampling grid for the collection of texture histograms which are then classified.

In our previous study with co-occurrence maps [6], multidimensional histograms reduced with the tree-structured self-organizing map [10], [11] provided higher classification accuracies than co-occurrence matrices or multidimensional full-scale co-occurrence histograms. The tree structure of the self-organizing map enables a fast but suboptimal search for the best-matching code vector. In a trained map, the similarity of neighboring code vectors makes it possible to improve quantization by limited lateral searches in the tree structure. This makes the map somewhat slower than a traditional tree-search vector quantizer with an identical structure. In the present study, we explore whether the texture classification could be speeded up without loss in accuracy by substituting the quadtree vector quantizer [12], [13] for the tree-structured self-organizing map. To obtain effective dimensionality reduction during the selection of co-occurrence components, discrete cosine transform coefficients are used instead of gray levels. A genetic algorithm, especially suited to minimize noisy, discontinuous, and multimodal functions such as the present one, is applied to the optimization of feature selection for all classifiers studied.

2 Reduced Multidimensional Histograms

2.1 Texture Preprocessing

The 32 Brodatz textures used in the study (Fig. 1) [14] were equalized to 256×256 pixels and 256 gray levels. The images were selected according to visual judgment so that 64×64 subimages captured the essential substructures. Otherwise, the selection was independent of the classifiers used. Each image was divided into 16 disjoint 64×64 blocks, and each block was independently histogram-equalized to abolish luminance differences among textures.

Each texture block was transformed into three additional blocks: a block rotated by 90 degrees, a 64×64 scaled block obtained from 45×45 pixels in the middle, and a block which was both rotated and scaled. The 32 texture categories included 2,048 blocks altogether.

Construction and testing of the classifiers were carried out with disjoint sets of blocks: Eight blocks in each texture image, together with the corresponding 24 transformed blocks, were randomly selected into a design set and the other 8 + 24 blocks were used for the evaluation of classifier performance. The classifier performance was evaluated statistically with 10 different randomly selected design and test sets.

2.2 Co-Occurrence Vectors

Two kinds of co-occurrences were analyzed: mean-subtracted gray levels and variance-equalized discrete cosine transform coefficients. Let matrix $\mathbf{G}$ represent the gray levels in a 4×4 neighborhood at some location in a texture block. A 16-dimensional gray-level vector $\mathbf{g}$ is formed by collecting entries $G_{i,j}$ of matrix $\mathbf{G}$ into vector $\mathbf{g} = (g_1, g_2, \ldots, g_{16}) = (G_{1,1}, G_{1,2}, \ldots, G_{4,4})$. The components of the mean-subtracted gray-level co-occurrence vector $\mathbf{s} = (s_1, s_2, \ldots, s_{16})$ are defined as

$$s_i = b_i\left(g_i - \frac{\sum_{j=1}^{16} b_j\, g_j}{\sum_{j=1}^{16} b_j}\right), \tag{2}$$

• *The authors are with the Laboratory of Computer and Information Science, Helsinki University of Technology, P.O. Box 2200, FIN-02015 HUT, Espoo, Finland. E-mail: kimmo.valkealahti@hut.fi; erkki.oja@hut.fi.*

Manuscript received 26 Sept. 1996; revised 10 Nov. 1997. Recommended for acceptance by S. Peleg.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106043.

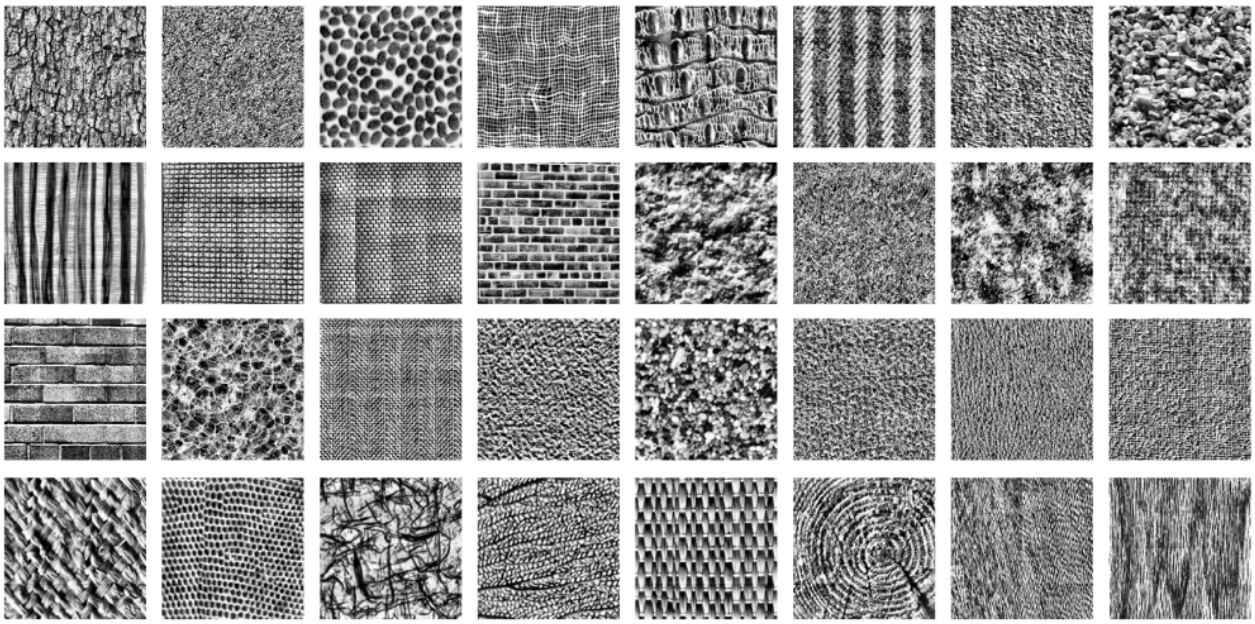

Fig. 1. Thirty-two histogram-equalized Brodatz textures used in the study.

in which binary coefficients b_i determine the active components of **s**. The co-occurring gray levels are selected from adjacent positions because then their mutual dependence is strongest. During feature selection, Section 2.6, the coefficients are adjusted to find the subset of vector components which minimize the classification error rate. Setting one or more coefficients b_i to zero annuls the corresponding elements in the co-occurrence vectors. Therefore, the effective dimensionality of the co-occurrence vectors is $K = \sum_{i=1}^{16} b_i$.

As an alternative representation, the discrete cosine transform (DCT) was used to decorrelate the gray-level co-occurrences [7]. Let matrix **C** be the DCT matrix with unnormalized column vectors,

$$\mathbf{C} = \begin{bmatrix} 1 & \cos\frac{\pi}{8} & \cos\frac{2\pi}{8} & \cos\frac{3\pi}{8} \\ 1 & \cos\frac{3\pi}{8} & \cos\frac{6\pi}{8} & \cos\frac{9\pi}{8} \\ 1 & \cos\frac{5\pi}{8} & \cos\frac{10\pi}{8} & \cos\frac{15\pi}{8} \\ 1 & \cos\frac{7\pi}{8} & \cos\frac{14\pi}{8} & \cos\frac{21\pi}{8} \end{bmatrix}. \tag{3}$$

The DCT-coefficient vector is collected from entries $F_{i,j}$ of matrix $\mathbf{F} = \mathbf{C}^T\mathbf{G}\mathbf{C}$ into vector $\mathbf{f} = (f_1, f_2, \ldots, f_{16}) = (F_{1,1}, F_{1,2}, \ldots, F_{4,4})$. Vector **f** contains the same information as vector **g**, but the components of **f** are almost uncorrelated. The components of the variance-equalized DCT co-occurrence vector **s** are defined by

$$s_i = b_i \frac{f_i}{\sqrt{Var\{f_i\}}}, \tag{4}$$

in which the denominator is the estimated standard deviation of the numerator. The equalization of component variances means that the DCT coefficients with small variance are assumed to possess similar discriminatory power as the coefficients with large variance.

2.3 Quadtree Vector Quantizer

In a quadtree vector quantizer, each nonterminal node branches to four nodes at each succeeding level. In the present case, the tree-structured codebook had six levels $l = 1, 2, \ldots, 6$, each with 4^l vectors. Each codebook with $N = 4{,}096$ code words was trained with 41,000 vectors **s** which were randomly sampled from a design set of all 32 textures. At the beginning of training, the first level is initialized with four distinct vectors close to the mean of all training vectors. The code vectors are trained by repeating the Lloyd iteration [12] until the decrease of quantization error in succeeding iterations is less than 5 percent. During a Lloyd iteration, each code vector is replaced with the mean of the sample vectors closer to it than to any other code vector. After one level has been trained, vector values are fixed and each value is copied to the descendant code vectors, which are then made unequal by changing their values slightly. The search for the best-matching code vector always starts from the first level and proceeds to the subtree descending from the code vector which is closest to the sample vector. Decisions among branches are made until the level under training is reached. During vector quantization with a trained quadtree, only $6 \cdot 4 = 24$ code vectors, instead of 4,096, are matched to a sample.

2.4 Tree-Structured Self-Organizing Map

The quadtree-structured self-organized map [11], [6] consisted of six levels in which the nodes were arranged in squares. The levels are trained using repeated Lloyd iterations as the quadtree vector quantizers. After training of one level, each code vector value is copied to the four descendant code vectors and the succeeding level is initialized by giving each code vector the mean value of its four-connected neighbor nodes. The search for the best-matching code vector always starts from the first level, and when proceeding to the succeeding levels it is limited to the descendant nodes of the current best-matching code vector and its four-connected neighbors. The similarity of neighboring code vectors emerges during the training from the initialization of tree levels and limited lateral searches. In a trained map, about 90 code vectors are matched to a sample during the quantization.

2.5 Histogram Classification

Using the vector quantizer, sample histograms were computed for each texture block separately. The histograms were represented as sequences $\{h_{ijq}\} = \{h_{ij1}, \ldots, h_{ijN}\}$, in which $i = 1, \ldots, 32$ indexes the different texture categories, $j = 1, \ldots, 32$ indexes the 64×64 blocks in the design set of the ith texture, and N is the histogram size equaling the number of code words. A texture model histogram

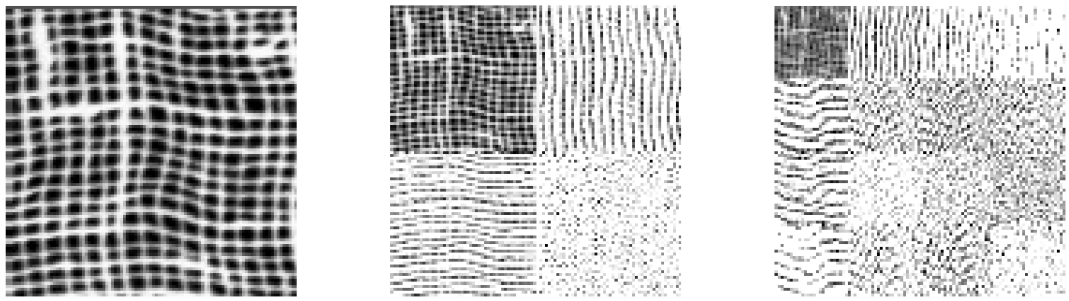

Fig. 2. Demonstration of the wavelet packet transform. Original image (left), the first-level transform (middle), and the second-level transform (right). The high-pass subimages were enhanced by rescaling pixel values.

$\{m_{iq}\} = \{\sum_{j=1}^{32} h_{ijq}\}$ was obtained by summing up the histograms computed from each block of the modeled texture. A new histogram $\{h_q\}$ was classified according to the model histogram that maximized the log-likelihood function

$$L(\{h_q\}, \{m_{iq}\}) = \sum_{q=1}^{N} h_q \ln \frac{m_{iq}}{M_i}, \tag{5}$$

in which $M_i = \sum_{q=1}^{N} m_{iq}$ [1], [2]. Another plausible measure for the histogram matching, the χ^2-statistic, was found inferior to the log-likelihood measure in the present case.

2.6 Selection of Co-Occurrence Components

The classifier was adjusted to the current task by selecting the co-occurrence vector components which minimized the classification error rate of a design set. The error rate was determined with the leave-one-out method: When a sample histogram for a texture block was compared with the model of the same texture, shares of the histogram itself and the histograms collected from all other rotated and scaled variants of the same texture block were subtracted from the model histogram (changing M_i accordingly) prior to the likelihood computation of (5).

A genetic algorithm was used to find the values of coefficients b_i in (2) and (4). Vector $\mathbf{b} = (b_1, b_2, \ldots, b_{16})$ which minimized the error rate determined the component selection. The genetic algorithm was implemented according to the instructions given by Davis [15]. The optimization was carried out using a population of 50 vectors with random initial values. A new unique vector was reproduced from two parent vectors which were selected with the roulette wheel method. Each component was randomly assigned the value of either of the parents. The value was inverted with probability 0.04. If the classification error rate of the reproduced vector was lower than that of the worst population member then a replacement was made. The number of evaluated vectors was 500.

3 Channel Histograms

Channel histograms estimate one-dimensional marginal densities of a decorrelated feature distribution [7]. Thus, a K-dimensional co-occurrence distribution is represented by K one-dimensional histograms. The components of co-occurrence vectors $\mathbf{s}$ in (4) are approximately decorrelated by the discrete cosine transform. Therefore, the ith channel histogram is collected from the values of component s_i. The component values were scalar-quantized to $N_1 = 2, 4, 8, \ldots, 256$ levels so that each bin of the ith histogram had about the same share of the distribution of s_i. This means that the channel histograms over the whole texture data are flat. Each texture is modeled with a combination of $K = \sum_{i=1}^{16} b_i$ channel histograms.

Multidimensional channel histograms were also used as texture models. The component values were scalar-quantized to a small number of levels, $N_2 = 2, 3, \ldots, 8$, and were collected into K-dimensional histograms.

The optimization of channel histogram classifiers included selection of the number of quantization levels N_1 or N_2. For the selection, component b_{17} is added to vector $\mathbf{b}$. Components $b_1, b_2, \ldots, b_{17}$ were constrained so that histogram size $N = KN_1$ or $N = N_2^K$ ensuing from their values did not exceed 4,096 bins. During the selection, component b_{17} was given a random value within its range with probability 0.04, and it was increased or decreased by one step with probability 0.3. The number of evaluated vectors was 2,000.

4 Wavelet Packets

The histogram methods were compared with the wavelet packet method of Laine and Fan [8]. The eight-tap Daubechies wavelet filters are used to transform a texture block into a wavelet packet representation. The transform is described by one-dimensional low-pass filter $\mathbf{u} = (u_1, u_2, \ldots, u_8)$ and high-pass filter $\mathbf{v} = (v_1, v_2, \ldots, v_8)$. Daubechies filter coefficients of different length are given in [16, p. 260], for instance. Four two-dimensional convolution masks, corresponding to one low-pass and three high-pass filters, are obtained from different outer products of the base vectors:

$$\mathbf{D}_1 = \mathbf{u} \otimes \mathbf{u},\ \mathbf{D}_2 = \mathbf{u} \otimes \mathbf{v},\ \mathbf{D}_3 = \mathbf{v} \otimes \mathbf{u},\ \mathbf{D}_4 = \mathbf{v} \otimes \mathbf{v}. \tag{6}$$

The wavelet packet decomposition begins with transforming a 64×64 texture block into one low-pass and three high-pass images of 32×32 pixels, see Fig. 2. The transform is repeated for each of the four subimages yielding sixteen 16×16 images and then for each of those yielding 64 8×8 images. Each $n \times n$ image is convolved circularly, that is, the image is extended periodically.

The classified feature vectors consist of energies, that is, sums of squares, computed over the values of the 32×32, 16×16, and 8×8 transform images. A feature vector is thereby 84-dimensional, containing three-resolution information of a texture block. Let this vector be

$$\mathbf{t} = (t_1, t_2, \ldots, t_{84}). \tag{7}$$

The class conditional probability density functions of $\mathbf{t}$ for the 32 classes are assumed to be multivariate Gaussian with mean vectors μ_i and covariance matrices Σ_i, $i = 1, 2, \ldots, 32$. Feature vector $\mathbf{t}$ is assigned to the class whose distribution parameters minimize the distance

$$(\mathbf{t} - \mu_i)(\mathbf{t} - \mu_i)^T + \ln \det \Sigma_i^{-1}. \tag{8}$$

The components of $\mathbf{t}$ were selected to minimize the leave-one-out classification error of a design set. The selection with the genetic

TABLE 1
AVERAGE CLASSIFICATION RESULTS IN 10 EXPERIMENTS

Method	Accuracy % (SD)		P value
	Design set	Test set	
Reduced multidimensional histograms			
of DCT coefficients with TSOM	93.3 (0.8)	93.9 (0.7)	–
of DCT coefficients with QVQ	93.3 (0.8)	93.4 (0.8)	n.s.
of mean-removed gray levels with TSOM	92.5 (0.8)	92.8 (0.9)	n.s.
Multidimensional channel histograms	89.9 (0.8)	90.4 (0.8)	$p < 0.001$
Wavelet packets	86.4 (0.9)	85.1 (1.1)	$p < 0.001$
One-dimensional channel histograms	78.8 (0.8)	78.2 (1.6)	$p < 0.001$

SD: sample standard deviation.
DCT: discrete cosine transform.
TSOM: tree-structured self-organizing map.
QVQ: quadtree vector quantizer.

algorithm was similar to that described in Section 2.6. A population of 200 vectors was used, and during the optimization, altogether 30,000 different combinations of the 84 vector components were evaluated.

5 RESULTS AND DISCUSSION

5.1 Reduced Multidimensional Histograms

The reduced multidimensional histograms provided significantly higher classification accuracies than the channel histograms or wavelet packets. The results of different methods are shown in Table 1, where they are in descending order of classification accuracy; each accuracy is an average of ten experiments with different design and test sets. Each P value in the table shows the significance of the difference between two consecutive test set classification accuracies (analysis of variance with Tukey test).

As shown in Table 1, the two vector quantizers provided similar classification accuracies. There was a significant difference ($p < 0.002$; paired sign test) between the quantization error per dimension of the tree-structured self-organizing map (median 0.362) and the quadtree vector quantizer (median 0.418). This shows that the quantization accuracy was not critical for the classification performance.

Table 1 also shows that the DCT-coefficient and gray-level vectors performed equally well. During the optimization, the reduction of the mean-removed gray-level components was modest: The median number of selected components was 13.5 (lower quartile 12, upper 15). The high number of selected components demonstrates that our method can represent quite high-dimensional co-occurrences. The reduction of DCT coefficients was more pronounced: The median number of selected components was 8 (8, 9) with both vector quantizers. The DCT coefficients with the highest variance preceding the normalization in (4) were most frequently selected and those with the lowest variance were most frequently rejected. The optimization of DCT-coefficients thus resembled principal-component-type feature reduction: The "principal" DCT coefficients captured most of the discriminatory information, whereas the "minor" DCT coefficients could be regarded as noise. As a separable two-dimensional transform, the DCT is faster to compute than the principal component transform.

Suppose that it takes 100 time units to compute a multidimensional histogram of the optimized mean-removed gray-level vectors with the tree-structured self-organizing map. The time was decreased to 72 units by the use of the optimized DCT coefficients. Replacing the tree-structured self-organizing map with the quadtree vector quantizer decreased the time further to 32 units.

5.2 Channel Histograms and Wavelet Packets

As shown by Table 1, multidimensional channel histograms performed much better than one-dimensional channel histograms suggesting that there were significant dependencies among the channel values. The presence of dependencies corroborates the use of multidimensional analysis. The median number of selected channels for the collection of one-dimensional channel histograms was 8.5 (8, 9) and the median number of quantization levels was $N_1 = 256$ (256, 256). The median number of channels for the collection of multidimensional channel histograms was 11 (11, 11) and the number of quantization levels was always $N_2 = 2$. The performance of wavelet packets was better than that of one-dimensional channel histograms, Table 1. The median number of selected energy features for the wavelet packet method was 10 (8, 10). The level-wise medians were two first-level, three second-level, and 4.5 third-level features.

The classification accuracies obtained with the one-dimensional channel histograms and the wavelet packets were around 80 percent. The authors of the methods have reported accuracies of 100 percent [7], [8]. Such differences originate from experimental setups. For instance, the authors of wavelet packets used four times larger blocks than we did. Moreover, both for wavelet packets and channel histograms the texture descriptors were computed from overlapping texture blocks and thus the design and test sets may not have been quite separate. The performance of a classifier depends on the representativeness of the textures from which the models are computed. In general, the comparison of studies is difficult because of differences in the character and number of textures, in histogram equalization, in the size of the block from which the texture descriptor is computed, and in the testing of classifiers.

5.3 Minimization of Leave-One-Out Error

Table 1 shows no significant differences between the design and test set accuracies for any method. Thus, the minimization of the leave-one-out error during the optimization did not result in significant overfitting of the texture models. Owing to the leave-one-out method, a lower number of texture blocks were used to compute the models of the design than the test sets. This may have somewhat favored the test set accuracies. The suitability of genetic algorithms for complex tasks, such as the optimization of classifiers, is well documented [17] but it is not excluded that simpler search methods [18] might perform equally well in the present case.

6 CONCLUSION

In comparison with the other methods, the reduced multidimensional histograms provided the highest classification accuracies. In our previous study [6], the tree-structured self-organizing map was chosen instead of the traditional tree-search vector

quantizer because it was suggested to provide lower quantization error. The present results verified this suggestion, but they also showed that the significant difference in the quantization error was not reflected in the classification performance. In comparison to the previous study, the use of traditional tree-search vector quantizer and optimized linear compression significantly speeded up the classification.

The results encourage the development of methods making use of multidimensional feature distributions. The bottleneck of the present method is the search for the best-matching code vector. The codebook search could be eliminated by finding suitable transformations of local pixel values which could be scalar-quantized to a few levels and collected directly into multidimensional histograms of practical size. If the transformations are linear, as with multidimensional channel histograms, the set of scalar quantizers is, in fact, a lattice-codebook vector quantizer [13]. Application of scalar quantization to the present task would speed up the classification. The accuracy of the present method could, however, not be exceeded.

ACKNOWLEDGMENT

The study was funded by the Academy of Finland as a part of a project on intelligent processing and analysis of images and speech.

REFERENCES

[1] A.L. Vickers and J.W. Modestino, "A Maximum Likelihood Approach to Texture Classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 4, no. 1, pp. 61–68, Jan. 1982.
[2] M. Unser, "Sum and Difference Histograms for Texture Classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 118–125, Jan. 1986.
[3] R.M. Haralick, "Statistical and Structural Approaches to Texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.
[4] D.-C. He and L. Wang, "Unsupervised Textural Classification of Images Using the Texture Spectrum," *Pattern Recognition*, vol. 25, no. 3, pp. 247–255, 1992.
[5] V. Kovalev and M. Petrou, "Multidimensional Co-Occurrence Matrices for Object Recognition and Matching," *Graphical Models and Image Processing*, vol. 58, no. 3, pp. 187–197, May 1996.
[6] E. Oja and K. Valkealahti, "Co-Occurrence Map: Quantizing Multidimensional Texture Histograms," *Pattern Recognition Letters*, vol. 17, no. 7, pp. 723–730, June 1996.
[7] M. Unser, "Local Linear Transforms for Texture Measurements," *Signal Processing*, vol. 11, no. 1, pp. 61–79, July 1986.
[8] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1,186–1,191, Nov. 1993.
[9] G.F. McLean, "Vector Quantization for Texture Classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 637–649, May/June 1993.
[10] T. Kohonen, *Self-Organizing Maps*. Berlin: Springer-Verlag, 1995.
[11] P. Koikkalainen, "Fast Deterministic Self-Organizing Maps," *Proc. Int'l Conf. Artificial Neural Networks*, vol. 2, pp. 63–68, Paris, 9–13 Oct. 1995.
[12] R.M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4–29, Apr. 1984.
[13] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Publishers, 1992.
[14] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York: Dover Publications, 1966.
[15] L. Davis, ed., *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.
[16] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, N.J.: Prentice Hall, 1995.
[17] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Mass.: Addison-Wesley Publishing Company, 1989.
[18] P.A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. London: Prentice Hall International, 1982.