

Improving the maximum-likelihood co-occurrence classifier: a study on classification of inhomogeneous rock images

P.Paclík, S.Verzakov, R.P.W.Duin

Information and Communication Theory Group
Delft University of Technology, The Netherlands
{P.Paclik,S.Verzakov,R.P.W.Duin}@ewi.tudelft.nl
<http://www-ict.et.tudelft.nl/~pavel>

Abstract. An industrial rock classification system is constructed and studied. The local texture information in many image patches is extracted and classified. The decisions made at the local level are fused to form the high-level decision on the image/rock as a whole. The main difficulties of this application lay in significant variability and inhomogeneity of local textures caused by uneven rock surfaces and intrusions. Therefore, an emphasis is paid to the derivation of informative representation of local texture and to robust classification algorithms. The study focuses on the co-occurrence representation of texture comparing the two frequently used strategies, namely the approach based on Haralick features and methods utilizing directly the co-occurrence likelihoods. Apart of maximum-likelihood (ML) classifiers also an alternative method is studied considering the likelihoods to prototypes as feature of a new space. Unlike the ML methods, a classifier built in this space may leverage all training examples. It is experimentally illustrated, that in the rock classification setup the methods directly using the co-occurrence estimates outperform the feature-based techniques.

1 Introduction

This study is concerned with classification of rocks in an industrial application in which a rock is placed beneath a light source and imaged by a color camera. Based on a high-resolution color image, the rock is as a whole assigned to a class of interest. In our application, the association of a rock specimen to a class depends on local texture properties of the rock surface. The nature of industrial rock classification poses several challenges on the design of a pattern recognition system. Because the rock surfaces are uneven the local texture information may exhibit considerable variability over a single image. Moreover, the rocks contain intrusions, which further increase the multi-modal nature of the class description.

In this study, we partition the design of the rock classification system into three steps: the derivation of an informative representation of local texture, the training of a local patch classifier and, eventually, the combination of per-patch decisions into a single decision on the entire image/rock. We focus on an

investigation of the first two steps and fix the combination strategy by using the majority vote over the classification results in a set of local image patches. We do not consider the use of color information in this study and work solely with gray-level textures.

There exists abundant literature on the representation of texture for the sake of classification [9, 4, 15]. In our study, we have focused on one of the most commonly used strategies based on co-occurrence matrices [5]. A co-occurrence matrix (CM) estimates the gray-level dependencies in a local neighborhood for a given displacement step and angle. Two major approaches for building classifiers using the gray-level co-occurrences are used. The majority of studies assume that the rich texture description present in the co-occurrence matrix must be first reduced to a set of features because the original co-occurrence distribution is too large to be used directly for classification [5, 1]. Typically only a subset of the original features, proposed by Haralick in [5], is used [9, 6]. Principally different approach was adopted by Vickers and Modestino [18] who consider the co-occurrence entries directly as features. The decision is derived by a maximum-likelihood classifier (ML), operating on co-occurrence matrices estimated per-class. Instead of using the class prototypes¹, Ojala et.a. [10] derived separate prototype from each image patch in the training set. Therefore, the resulting ML classifier mimics rather the nearest neighbor approach.

It is an open question which of the two major strategies is beneficial in the rock classification problem. Focusing more on the likelihood-based methods, we observe that the ML-based classifiers effectively derive their decisions only from the stored prototypes of the class or local co-occurrences. The existing abundant collections of training image patches are used only for the estimation of the class co-occurrences, or even entirely discarded apart of local prototypes in the nearest-neighbor sense. In order to fully leverage the existing training sets, we also adopt a recently developed strategy for building classifiers on (dis)similarity representations (Duin et.al. [2, 14, 12]). Here the likelihoods to prototypes are considered as dimensions of a new space, which is populated by all available training examples. A general-purpose classifier, such as the Fisher linear discriminant (FLD), built in this space may thereby exploit the correlations between the likelihoods to prototypes. Additional advantage of this approach is that apart of likelihoods also other dissimilarity measures for probability distributions may be used, such as the Kullback-Leibler divergence. We illustrate, that this classification strategy is beneficial for multi-modal rock classification problems as it facilitates derivation of non-linear classifiers.

In the next section, we introduce the rock classification system. Section 3 explains how the discussed data representation and classifiers may be built. In Section 4 we describe a set of experiments on a dataset of rock images. Finally, in the last section, we conclude our findings.

¹ In order to emphasize that the co-occurrence models are estimated from the training data, we adopt the pattern recognition terminology in this paper and refer to prototypes.

2 The rock classification system

The rock classification system is trained using a set of labeled images. Because the classification is performed on the basis of local textures, the set of local image patches must be first extracted from the training images. Each image patch is accompanied with the information on the class of rock it represents. The set of all labeled image patches is processed so that the local texture information captured in each patch is properly represented. In this paper, we represent the local textures by the co-occurrence estimates. For the sake of classification, this intermediate texture representation must be transformed accordingly. It is either further reduced to a set of feature values or represented by a set of likelihoods to prototype co-occurrences. In this representation, the classifier is finally built using the training set of labeled local patches.

When processing of a new image by the trained rock classification system, a set of image regions is first extracted. The texture within each patch is represented by the co-occurrence matrix and the classifier-aware representation is derived similarly to the training stage. The trained patch classifier is invoked on each of the image patches and its decisions are fused by the majority voting combiner to a decision on the level of the complete image/rock.

3 Likelihood-based classification of local textures

The co-occurrence matrix \mathbf{P} , estimated from the image patch r given the displacements Δx and Δy is defined as:

$$P(g_1, g_2) = \{pairs(g_1, g_2) | r(x, y) = g_1 \text{ and } r(x + \Delta x, y + \Delta y) = g_2\}, \quad (1)$$

where g_1 and g_2 denote the gray-levels, $r(x, y)$ represents the gray-level at coordinates x and y in the image patch r and the functional *pairs* returns the number of situations.

The likelihood of a co-occurrence estimate \mathbf{P} with respect to the prototype co-occurrence \mathbf{Q} may be expressed as:

$$L(\mathbf{P}, \mathbf{Q}) = \sum_{g_1, g_2} P(g_1, g_2) \ln \frac{Q(g_1, g_2)}{\sum_{g_1, g_2} Q(g_1, g_2)} \quad (2)$$

The C -class maximum-likelihood classifier assigns a new observation \mathbf{P} to the class of the closest prototype:

$$\omega(\mathbf{P}) = \arg \max_c \max_k \{L(\mathbf{P}, \mathbf{Q}_k^c)\}, \quad (3)$$

where $\omega(\mathbf{P})$ denotes the class of the observation \mathbf{P} and \mathbf{Q}_k^c represents the k -th prototype of the class ω_c , $c = 1, \dots, C$.

Traditionally, the ML classifier represents each class by a single prototype. This is estimated by averaging the training population of the local co-occurrences for each class [18]. This approach is analogous to the nearest mean classification. The ML classifier may also utilize more prototypes per class each of them directly selected from the training set [10]. This strategy resembles the nearest-neighbor classifier.

3.1 Training the classifier on the likelihood representation

The alternative classification scheme considers the likelihoods computed with respect to a set of M prototypes as dimensions of a new M -dimensional feature space where a co-occurrence estimate \mathbf{P} may be represented by the vector:

$$\mathbf{x}^L = \{L(\mathbf{P}, \mathbf{Q}_m)\}_{m=1}^M, \quad \mathbf{x}^L \in R^M. \quad (4)$$

In order to train a classifier in this representation, the space is populated using a set of likelihoods corresponding to N_p training image patches. The training set will, therefore, contain N_p labeled feature vectors \mathbf{x}^L in the M -dimensional space. We propose to train the Fisher linear discriminant (FLD) in this new space. For the two-class situation it becomes:

$$\omega(\mathbf{P}) = \text{sign} \left(\sum_m^M w_k L(\mathbf{P}, \mathbf{Q}_m) + b \right). \quad (5)$$

In the multi-class situation the minimum least square solution may be used, based on regression [3]. Note that the class membership of prototypes is not directly used by this classification rule, on contrary to the ML classifier (3).

The Equations (2) and (5) may be rearranged to show that the linear discriminant built in the space spanned by likelihoods delivers linear solution with respect to the original space defined by the bins of the co-occurrence distribution. The proposed linear discriminant yields, in fact, a *weighted-likelihood solution* as opposed to the maximum-likelihood classifier (3). A non-linear classifier may be constructed by leveraging a different distance measure for probability distributions, such as the Kullback-Leibler divergence (KL) [16]:

$$d_{KL}(\mathbf{P}, \mathbf{Q}) = \sum_{g_1, g_2} P(g_1, g_2) \ln \frac{P(g_1, g_2)}{Q(g_1, g_2)} + \sum_{g_1, g_2} Q(g_1, g_2) \ln \frac{Q(g_1, g_2)}{P(g_1, g_2)}. \quad (6)$$

By comparing the Equations (6) and (2) we can observe the close relation between both measures. Each of the sums in the Kullback-Leibler divergence (6) may be decomposed into entropy and likelihood terms.

4 Experiments

4.1 Dataset description

The dataset used in our experiments contains 72 color images from two classes of rocks, each represented by 36 images with resolution of 1392 times 1040 pixels. Images were acquired under controlled lighting conditions. Apart from cleaning, the rocks remain untreated with no cutting or polishing applied. Each image, therefore, views a multi-faceted rock surface. The Figure 1 shows several patches extracted from the images. Patches in each row were extracted from images of the same class. Note that the patches exhibit significant inhomogeneity and intensity variation caused by uneven rock surfaces and frequent intrusions.

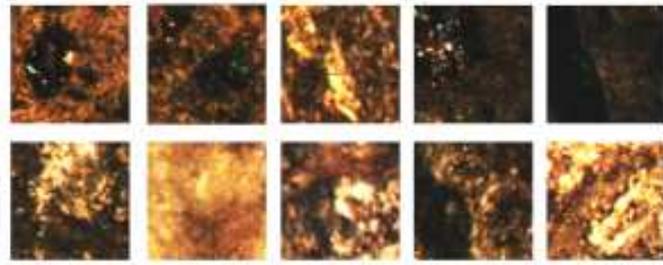


Fig. 1. Examples of image patches for each of the two classes. Each row corresponds to examples of one class.

4.2 Texture descriptions

We consider the following texture descriptions based on co-occurrence:

- Co-occurrence bins are directly considered as features.
- 14 Haralick features computed from the co-occurrence matrix.
- Likelihoods with respect to co-occurrence prototypes considering either mean class prototypes or local prototypes as discussed in Section 3.
- Kullback-Leibler (KL) divergences to local co-occurrence prototypes.

For the sake of comparison, we also include two base methods widely used for texture classification:

Local Binary Pattern features (LBP) describe a distribution of binary patterns constructed by thresholding a micro-region and accumulating this information over an entire image patch [11]. The LBP features were found to be considerably more robust to varying illumination than the techniques using the gray-levels directly. Three types of LBP features are computed with 50 features in total.

Gabor filters A bank of 24 Gabor filters is designed according to [8]. A total set of 48 features is formed by means and standard deviations of the filter responses computed over an image patch.

4.3 Data representation and classification

Prototype selection For the sake of simplicity, we employ the random selection of prototypes. This allows us to estimate the classifier performance for a growing number of prototypes.

Regularization of the similarity measures While the class co-occurrence prototypes are derived from thousands of observations, each of the local co-occurrences is estimated from a single image patch. Due to the limited amount of data, some bins of the local co-occurrence histogram may remain empty. As it can be seen from Equation (2), the likelihood of any observation to the prototype with an empty bin becomes infinitely small. In order to avoid situations where a test object cannot be classified because the likelihoods to all prototypes are infinitely small, we introduce the regularized likelihood measure using prototypes:

$$\mathbf{Q}_{\text{reg}} = \mathbf{Q}(1 - \delta) + \delta G^{-1} \sum_{g_1, g_2} \mathbf{Q}(g_1, g_2), \quad (7)$$

where G denotes the total number of co-occurrence bins and δ stands for the regularization parameter. For the Kullback-Leibler divergence (6), both the observed and prototype co-occurrences \mathbf{P} and \mathbf{Q} are regularized. The regularization parameter was fixed to $\delta = 10^{-6}$ in all the experiments.

Principal Component Analysis In several experiments, the supervised version of PCA was employed which estimates the pooled covariance matrix using the per-class covariances [3]. In all cases 0.99 of variance was preserved.

Classifiers In the experiments, standard classifiers were used such as the nearest neighbor classifier (1-NN), Fisher linear discriminant (FLD) or quadratic classifier assuming normal densities (QDC) [3]. We refer to the maximum-likelihood classifier with class prototypes as NMC (nearest mean classifier).

4.4 Experimental setup

Cross-validation In order to estimate the performance of the rock classification system utilizing different texture characterization and classifiers we have adopted a 10-fold cross-validation procedure over images. Thereby, all the local patches originating from one image, appear always together either in training set or in the test set. All the steps required for training of the image classifier i.e. the estimation of the co-occurrences, computation of likelihoods or dissimilarities, extraction of texture features and the training of the patch classifiers is carried only on the images in the training set.

Preprocessing Vickers [18] and Valkealahti [17] propose to equalize the histograms of local image patches. This procedure assumes that the first-order statistics of the local histogram are not informative and thus may be removed. Experiments on our dataset show that the mean and standard deviation of the gray-level computed over local neighborhoods carry important discriminatory information. We have, therefore, decided not to perform any histogram equalization and rather maintain as constant illumination as possible during the acquisition process.

The original color images were converted to gray-level. All the methods used image patches of 64×64 pixels. The patches are extracted without mutual overlap so that each image is represented by cca 330 image regions. In each fold of the cross-validation procedure, about 20 000 local patches are used for training and about 2 600 image patches during testing. Prior to the estimation of co-occurrence matrices, the histograms of the local regions are reduced from 256 to 8 levels. All co-occurrence matrices are estimated for one pixel displacement in the vertical direction [5].

4.5 Results and discussion

The Table 1 summarizes our rock classification experiments. For each method, three error measure are given, namely the classification error over image patches, the error over images obtained by majority voting and the total number of test images (e), misclassified during the cross-validation.

<i>method</i>	<i>texture description</i>	<i>classifier</i>	<i>error (patches)</i> $\hat{\mu} (\hat{\sigma}_{\mu})$	<i>error (images)</i> $\hat{\mu} (\hat{\sigma}_{\mu})$	<i>e</i>
1	CM, ML, class proto.	NMC	0.327 (0.023)	0.258 (0.040)	19
2	CM, ML, regul., 10 proto.	1-NN	0.416 (0.035)	0.371 (0.044)	27
3	CM, ML, regul., 50 proto.	1-NN	0.380 (0.020)	0.238 (0.052)	17
4	CM, ML, regul., 200 proto.	1-NN	0.362 (0.012)	0.188 (0.045)	14
5	CM, likelihood, class proto.	FLD	0.328 (0.022)	0.258 (0.032)	19
6	CM, likelihood, regul., 10 proto.	FLD	0.314 (0.021)	0.217 (0.043)	16
7	CM, likelihood, regul., 50 proto.	FLD	0.306 (0.020)	0.229 (0.049)	17
8	CM, likelihood, regul., 200 proto.	FLD	0.305 (0.020)	0.217 (0.043)	16
9	CM, KL, regul., 200 proto.	1-NN	0.354 (0.012)	0.167 (0.029)	12
10	CM, KL, regul., 10 proto.	FLD	0.304 (0.022)	0.167 (0.039)	12
11	CM, KL, regul., 50 proto.	FLD	0.269 (0.018)	0.125 (0.039)	9
12	CM, KL, regul., 200 proto.	FLD	0.271 (0.017)	0.125 (0.039)	9
13	CM, directly used as features	FLD	0.305 (0.021)	0.229 (0.049)	17
14	CM, 14 Haralick features	FLD	0.316 (0.024)	0.221 (0.045)	16
15	CM, 14 Haralick features, PCA	QDC	0.378 (0.018)	0.354 (0.032)	26
16	LBP, 50 features	FLD	0.371 (0.021)	0.196 (0.047)	14
17	LBP, 50 features, PCA 0.99	QDC	0.363 (0.018)	0.225 (0.047)	16
18	Gabor filers, 48 features	FLD	0.345 (0.029)	0.283 (0.053)	21
19	Gabor filers, 48 features, PCA 0.99	QDC	0.355 (0.024)	0.271 (0.052)	20

Table 1. Results of the rock classification experiment with 72 images. Estimated mean errors over image patches and over images are given with the respective standard deviations of the mean estimators. The last column e denotes the total number of erroneously classified images in the cross-validation experiment.

First part of the table corresponds to the likelihood-based methods. The per-image performance of the traditionally-used ML classifier with class prototypes (row 1) can be improved by utilizing the local co-occurrence prototypes in the nearest neighbor fashion (rows 2-4). Note, that while the method 4 reaches lower per-image error than the classifier 1, its error over patches remains higher.

In order to understand this behaviour, one needs to focus on the voting combiner based on the ratio of correctly classified patches within an test image. The distribution of this ratio computed for all the test images is depicted in Figure 2 for the ML classifier 1 (on the left) and the nearest neighbor rule 4 (on the right). Although exhibiting lower error over patches, the model-based ML classifier 1 suffers (on the image level) from the problem heterogeneity i.e. from the existence of images that are very different from the model but still belong to the same class. On the contrary, the nearest neighbor rule 4 is able to cope with the problem multi-modality.

This point is illustrated again on the performance of the FLD classifiers trained on the likelihood representation (methods 6-8). Compared to the nearest neighbor rules based on the same prototypes (rows 2-4) the classifiers 6-8 exhibit significantly better per-patch performances. This is understandable as the methods 6-8 utilize all available 20 000 training examples in the 200D spaces while the methods 2-4 use only the 200 training examples. The per-image error of the nearest-neighbor classifiers 2-4, however, decreases significantly faster and, eventually, the method 4 even outperforms the FLD classifier 8. As mentioned in Section 3.1, the FLD classifier based on likelihoods is still a linear discriminant,

similarly to the ML classifier 1. Therefore, we attribute the better performance of the weaker nearest neighbor classifier 4 to its non-linear nature.

The rows 9-12 in Table 1 refer to classifiers based on the Kullback-Leibler divergence. Based on the identical sets of prototypes as the likelihood-based classifier 4, the nearest neighbor rule 9 yields a better performance. Further improvements are possible by training the FLD classifiers on the Kullback-Leibler divergences as all the available training examples are now exploited. The algorithms in rows 11 and 12 yield the best overall results in our study (12.5% of error over images). We conclude that this improvement over the likelihood-based methods 6-8 is caused by the non-linearity introduced by the Kullback-Leibler divergence.

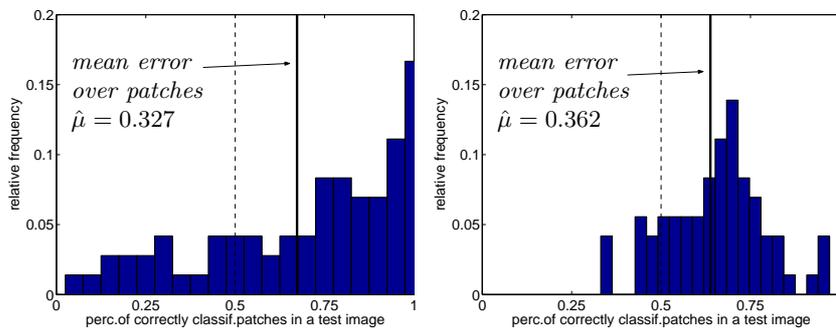


Fig. 2. The effect of the voting combiner on two likelihood-based classifiers: the linear ML classifier 1 (left) and the nearest neighbor rule 4 (right). For each of the test images in cross-validation, the ratio of correctly classified patches is computed. The figures depict the histogram of this ratio over all 72 images. Although the mean error over all patches is lower for the classifier on the left, the fraction of images misclassified by voting (left of the dashed line) is higher.

The rows 13-15 refer to the feature-based methods utilizing directly the co-occurrence matrices (13) or Haralick features (14,15). It is interesting to note that the co-occurrence bins directly used as features (13) yield better result than the traditionally used ML classifier 1. Employing the Haralick features results in an additional minor performance improvement. From the results, we conclude that the use of quadratic classifier in a PCA-reduced representation (15) is not beneficial.

The rows 16-19 represent methods based on different principles than co-occurrence, namely the Local Binary Pattern features (Ojala et.al. [11]) and the features derived from the bank of Gabor filters. While the Gabor features deliver only mediocre performance, the LBP features yield better or comparable results than the best likelihood-based approaches. The LBP features are designed to be more resilient to variable illumination than the co-occurrence matrices. Based on the similar performance of both approaches we conclude that the illumination in our dataset is rather constant.

5 Conclusions

The rock classification employs a local texture description in order to classify the high-resolution images into a set of pre-defined classes. Unlike the existing studies which deal with images of inhomogeneous but polished and cut rocks [13, 7], the images used in this study depict rocks with uneven surfaces and intrusions. The local textures extracted from a single image therefore exhibit significant variability. The aim of this paper was to understand which texture representations and what types of classifiers are robust and well-performing for this type of problem. We have focused on the family of texture representations based on the co-occurrence matrix investigating two distinct approaches to texture characterization. The first derives the Haralick features from the co-occurrence matrices and employs a conventional classifier. The second approach leverages the co-occurrence estimates directly and is traditionally bundled with the maximum-likelihood classifier.

The maximum-likelihood classifier operating in the nearest neighbor fashion on the local co-occurrence estimates appears to be beneficial to the traditionally employed ML classifier utilizing the class-specific co-occurrences. Although weaker in classifying individual patches, the nearest neighbor classifier yields a non-linear class-separation boundary. The traditional maximum-likelihood classifier operates on the nearest mean principle and, therefore, cannot cope well with the multi-modal rock-classification problem.

The maximum-likelihood classifiers use the available training set only for estimation of class prototypes (the nearest mean scenario) or even entirely discard all the training examples apart of prototypes (the nearest neighbor approach). The proposed alternative classifier is derived by training the Fisher linear discriminant on likelihoods to prototypes. It uses all the available training examples and leverages correlations between the likelihoods (or dissimilarities) to prototypes. While this method still yields a linear classifier when applied to likelihoods, the use of other distance measures such as the Kullback-Leibler divergence results in a non-linear decision rule. The dissimilarity-based classifiers built using the Kullback-Leibler divergence also deliver the highest performance of all studied approaches.

Surprisingly, the linear classifier built on Haralick features outperforms the state-of-the-art maximum-likelihood classifier operating directly on co-occurrences. However, the more sophisticated treatment of dissimilarity representations yields systems, performing significantly better than any of the feature-based classifiers. We conclude that while the Haralick features are simple and provide good accuracy, the dissimilarity-based classifiers offer higher flexibility and eventually better performance.

6 Acknowledgements

The authors would like to thank to David Tax, Carmen Lai and Thomas Landgrebe for support and stimulating discussions on the manuscript. This research is/was supported by the Technology Foundation STW, applied science division of NWO and the technology program of the Ministry of Economic Affairs.

References

1. F.S. Cohen, Z. Fan, and M.A. Patel. Classification of rotated and scaled textured images using gaussian markov random field models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(2):192–202, 1991.
2. Robert P W Duin, Dick de Ridder, and David M J Tax. Experiments with object based discriminant functions; a featureless approach to pattern recognition. *Pattern Recognition Letters*, 18(11-13):1159–1166, 1997.
3. R.P.W. Duin, P. Juszczak, D. de Ridder, P. Paclík, E. Pekalska, and D. M. J. Tax. PR-Tools 4.0, a Matlab toolbox for pattern recognition. Technical report, ICT Group, TU Delft, The Netherlands, January 2004. <http://www.prtools.org>.
4. R.M. Haralick. Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 67:786–804, 1979.
5. R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *IEEE Trans. Sys.Man and Cybernetics*, 3(6):610–621, November 1973.
6. Lepistö., I. Kunttu, J. Autio, and A. Visa. Classification of non-homogenous textures by combining classifiers. In *Proc.of IEEE International Conference on Image Processing, Barcelona, Spain, September 14.-17.*, volume 1, pages 981–984, 2003.
7. Lepistö., I. Kunttu, J. Autio, and A. Visa. Rock image classification using non-homogenous textures and spectral imaging. In *WSCG proc., WSCG'2003, Plzen, Czech Republic, Feb. 3.-7.*, 2003.
8. B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.
9. P.P. Ohanian and R.C. Dubes. Performance Evaluation for Four Classes of Textural Features. *Pattern Recognition*, 25:819–833, 1992.
10. T Ojala, M Pietikäinen, and J Kyllönen. Gray level cooccurrence histograms via learning vector quantization. In *Proc. 11th SCIA, Kangerlussuaq, Greenland*, pages 103–108, 1999.
11. T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
12. Pavel Paclík and Robert P.W. Duin. Dissimilarity-based classification of spectra: computational issues. *Real Time Imaging*, 9(4):237–244, 2003.
13. M. Partio, B. Cramariuc, M. Gabbouj, and A. Visa. Rock texture retrieval using gray-level co-occurrence matrix. In *ORSIG-2002, 5th Nordic Signal Processing Symposium, On Board Hurtigruten M/S Trollfjord, Norway, October 4-7*, 2002.
14. E. Pekalska, P. Paclík, and R.P.W. Duin. A generalized kernel approach to dissimilarity based classification. *Journal of Machine Learning Research*, 1(2):175–211, 2001. Special Issue "New Perspectives on Kernel Based Learning Methods".
15. T. Randen and J.H. Husøy. Filtering for texture classification: A comparative study. *IEEE Trans. Patt.Anal. and Mach.Int.*, 21(4):291–310, 1999.
16. S. Theodoridis and K. Koutroubas. *Pattern Recognition*. Academic Press, 1998. ISBN 0-12-686140-4.
17. K. Valkealahti and E. Oja. Reduced multidimensional co-occurrence histograms in texture classification. *IEEE Trans. Patt.Anal. and Mach.Int.*, 20(1):90–94, 1998.
18. A.L. Vickers and J.W. Modestino. A maximum likelihood approach to texture classification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4(1), 1982.