Low Dimensional Adaptive Texture Feature Vectors From Class Distance and Class Difference Matrices

Birgitte Nielsen*, Fritz Albregtsen, and Håvard E. Danielsen

***Abstract*—In many popular texture analysis methods, second or higher order statistics on the relation between pixel gray level values are stored in matrices. A high dimensional vector of predefined, nonadaptive features is then extracted from these matrices. Identifying a few consistently valuable features is important, as it improves classification reliability and enhances our understanding of the phenomena that we are modeling. Whatever sophisticated selection algorithm we use, there is a risk of selecting purely coincidental "good" feature sets, especially if we have a large number of features to choose from and the available data set is limited. In a unified approach to statistical texture feature extraction, we have used class distance and class difference matrices to obtain low dimensional adaptive feature vectors for texture classification. We have applied this approach to four relevant texture analysis methods. The new adaptive features outperformed the classical features when applied to the most difficult set of 45 Brodatz texture pairs. Class distance and difference matrices also clearly illustrated the difference in texture between cell nucleus images from two different prognostic classes of early ovarian cancer. For each of the texture analysis methods, one adaptive feature contained most of the discriminatory power of the method.**

***Index Terms*—Adaptive features, Brodatz textures, class distance matrices, early ovarian cancer, image texture analysis, low dimensionality, pattern classification.**

I. Introduction

WITHIN the field of texture analysis, a number of categories of methods exist [1], and even if we restrict ourselves to statistical methods, a large number of methods are available [2]. In many popular texture analysis methods, second or higher order statistics on the relation between gray level values in pixel pairs or sequences of pixels are stored in matrices (e.g., gray level cooccurrence matrix (GLCM) [3], gray level run length matrix (GLRLM) [4]), or the statistical qualities of e.g., roughness and complexity at different scales are represented by vectors (e.g., fractal signature [5], lacunarity [6], and complexity curve [7]). Features are then extracted that directly describe the probability distribution within the matrix and, therefore, indirectly describe the image texture. The features are inherently *ad hoc*, and they do not adapt to the image material under study. The features may often be seen as very simple weight functions in a weighted summation of the probability matrix elements. The result is *nonadaptive weighting* based on either the matrix element *value* or the *position* within the matrix. In the latter case, the weight function varies along only one axis at a time. Consequently, complex matrix structures are not captured in any single feature, and features do not adapt to problem-specific matrix structures. This feature extraction may be repeated for several settings of some free parameters (e.g., number of gray levels in the image, interpixel distance, orientation), resulting in a relatively high dimensionality of the feature space.

Identifying a few consistently valuable features is important for many applications as it improves classification reliability and enhances our understanding of the phenomena that we are modeling. The usual procedure is to perform a feature selection, based on a training data set, in order to obtain a subset of features to be used in a classifier. If we want to perform an exhaustive search through N features for the optimal subset of the n "best," we would have to test $N!/(n!(N-n)!)$ combinations, provided that we have some way of knowing the optimal value of n. Even for moderate values of N and n, the number of combinations may become prohibitive. Several suboptimal feature selection schemes have been suggested and reviewed [8], [9]. A review of statistical pattern recognition methods is given in [10].

Whatever sophisticated feature selection algorithm we use, there is a risk of selecting purely coincidental "good" feature sets, especially if we have a large number of features to choose from and the available data set is limited. The resulting set of features should be evaluated in terms of its classification performance on a different test data set. Using most of the available data in the training set will give the best possible classifier, while using more data in the test set will give a better estimate of the real error rate. If the performance of the classifier is only estimated by resubstitution, the error rate tends to be too optimistic. Better estimates are obtained by cross-validation, but the most reliable estimate is obtained from separate training and test sets [11]–[13].

In a situation as outlined above, the minimum complexity principle (Occam's Razor) is very relevant. So far, this principle has generally motivated the *search* for reduced feature sets [8], [9]. However, it should also motivate us to *generate* only a few, but powerful features, incorporating as much class discriminating information as possible into as few features as possible, preferably including adaptivity into the computation of the features. Then one can perform very simple—if any—feature se-

Manuscript received February 21, 2003; revised May 20, 2003. This work was supported by the Research Council of Norway. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was A. Manduca. *Asterisk indicates corresponding author.*

*B. Nielsen is with the Department of Informatics, University of Oslo, P.O. Box 1080 Blindern, N-0316 Oslo, Norway and also with the Section of Digital Pathology, the Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway (e-mail: birgitn@ifi.uio.no).

F. Albregtsen is with the Department of Informatics, University of Oslo, Blindern, N-0316 Oslo, Norway.

H. E. Danielsen is with the Section of Digital Pathology, the Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway and also with the Division of Genomic Medicine, University of Sheffield, Sheffield S102TN, U.K.

Digital Object Identifier 10.1109/TMI.2003.819923

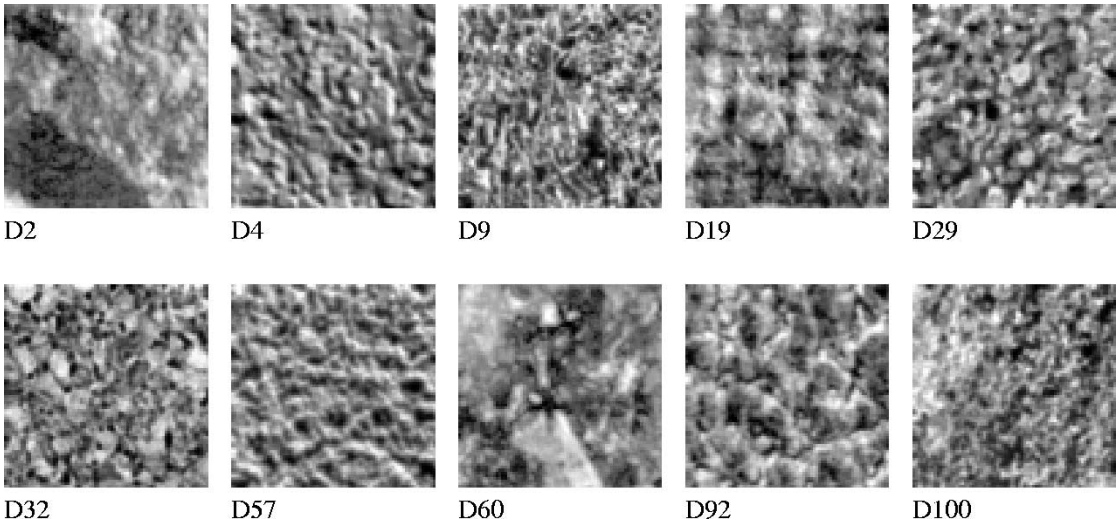

Fig. 1. One 75 × 75 pixels subimage from each of the selected Brodatz textures after gray level histogram normalization (D2: Fieldstone, D4: Pressed cork, D9: Grass lawn, D19: Woolen cloth, D29: Beach sand, D32: Pressed cork, D57: Handmade paper, D60: European marble, D92: Pigskin, D100: Ice crystals).

lection. Perhaps most importantly; the possibility of finding a feature combination that discriminates well by pure chance is reduced.

Based on the work of Walker *et al.* [14] we suggested [15] using *class distance matrices* for the GLRLM texture analysis method [4]. The GLRL class distance matrix contains information on which GLRLM elements that—when taken one matrix element at a time—will provide high class discrimination. For a four-class problem of liver cell nuclei, we found that the class distance matrices between different pairs of classes contained large areas of consistently high values. We could, therefore, combine information from the entries of the normalized run length matrix, based on the class distance matrices, to obtain adaptive features for texture classification.

We have previously [16] introduced the four-dimensional (4-D) cooccurrence of gray level run length matrix (CGLRLM). Adaptive texture features were obtained from the associated two-dimensional (2-D) sum and difference matrix elements, using *class distance matrices* and *class difference matrices*. In [5] we used class distance and class difference matrices to extract adaptive features from fractal signature vectors and lacunarity matrices. Then in [17] we applied this low dimensionality approach to obtain adaptive fractal and GLCM features from cell nucleus images from early ovarian cancer. The class difference matrix for a given texture method contains the difference between the average frequency matrix (i.e., the average GLCM computed from the training set images) of each of two classes.

In the present study, the aim has been to construct class distance matrices for a small number of relevant texture analysis methods, and to extract a minimum number of adaptive features from each of the methods, using the same fundamental approach. The ten most stochastic, isotropic, homogeneous, and fine-grained Brodatz textures [18] have been used to evaluate the new method of adaptive texture feature extraction, comparing it to the classical predefined texture features. Then we have constructed class distance and difference matrices from cell nucleus images taken from two different prognosis classes of early ovarian cancer, using the new adaptive features to classify the samples.

We have found that the new adaptive features outperformed the classical features when applied to all 45 possible Brodatz texture pairs. Class distance and difference matrices also clearly illustrated the difference in chromatin structure between cell nucleus images from the two different prognostic classes of early ovarian cancer. The best feature pairs discriminated the two classes of ovarian cancer with a classification error less than 25%.

II. Materials and Methods

A. Brodatz Textures

From the 112 Brodatz textures [18] we have selected the 10 most stochastic, isotropic, homogeneous and fine-grained textures, see Fig. 1. The Brodatz textures were selected as the most difficult Brodatz data set, but also to give an indication of which texture analysis methods and features that have the best chances of discriminating the stochastic, fine-grained textures of cell nucleus images in the setting of a low dimensional Bayesian classification.

Each Brodatz texture image was partitioned into 48 nonoverlapping 75 × 75 pixels subimages. The 48 subimages were divided randomly but equally into a training set and a test set. Given the 10 Brodatz textures, we have 45 texture pairs. For each of the Brodatz texture pairs, we have two separate sets of 48 subimages for training and test, respectively.

B. Ovarian Cancer Cell Nuclei

Forty cases of ovarian cancer classified as International Federation of Gynecology and Obstetrics (FIGO) stage 1 were included in the analysis [19]. Twenty cases had a good prognosis, which means that they survived the follow-up period without a relapse. The minimum length of follow-up for patients alive without a relapse was ten years. The 20 cases included in the bad prognosis group died of a cancer-related disease or relapsed during the first five years of the follow-up period.

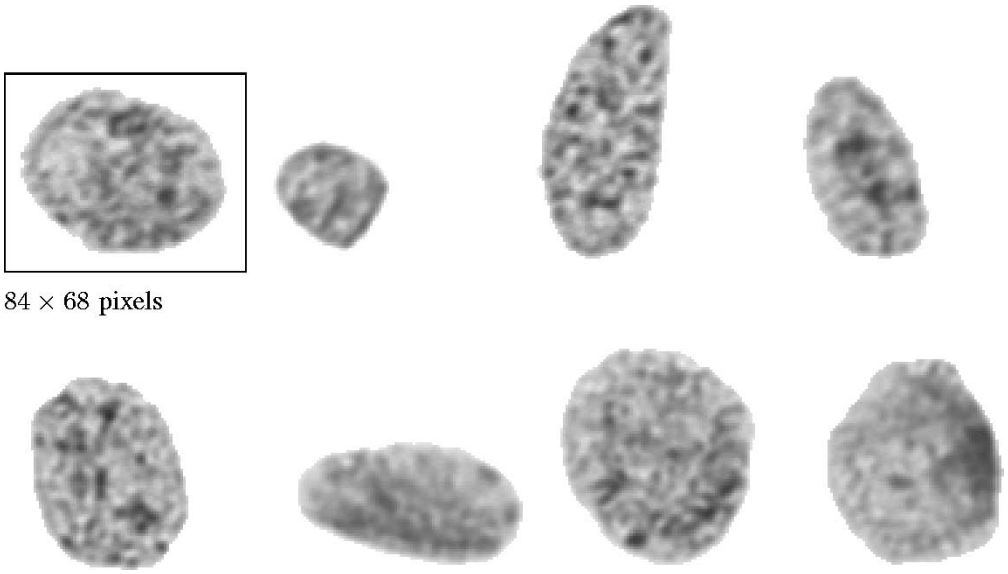

Fig. 2. Four monolayer cell nuclei from a good prognosis sample (upper) and four nuclei from a bad prognosis sample (lower).

Paraffin-embedded tissue samples fixed in 4% buffered formaldehyde were sectioned (2×50 μm) and enzymatically digested (SIGMA protease, type XXIV, Sigma Chemical C., St. Louis, MO) for the preparation of isolated nuclei (monolayers) [20]. The nuclei were Feulgen-Schiff stained according to an established protocol [21]. The tumour tissue to be used for the preparation were selected by the pathologist.

The Fairfield DNA Ploidy System (Fairfield Imaging LTD, Kent, U.K.), which consisted of a Zeiss Axioplan microscope equipped with a 40/0.75 objective lens (Zeiss), a 546-nm green filter and a black and white high-resolution digital camera (C4742-95, Hamamatsu Photonics K.K., Hamamatsu, Japan) was used. A shade correction was performed for each image field and the image was stored with a gray level resolution of 10 bits/pixel. Trained personnel performed a screening of the cells at the microscope and selected tumour cells for the analysis. Stromal cells, necrotic cells, doublets or cut cells were disregarded. The nuclei were segmented from the background by using a global threshold. The mean number of measured tumour nuclei/sample was 256. Fig. 2 shows examples of cell nuclei from each of the two groups.

C. Preprocessing

Each 8-bit Brodatz subimage was normalized to the same mean gray level value ($\mu = 127.5$) and standard deviation ($\sigma = 50.0$). In the same manner, the histograms of all 10-bit cell nucleus images were normalized to a preset mean value (650.0) and standard deviation (120.0).

In order to get separate estimates of texture features in the periphery and center of each 2-D image of the cell nucleus, an 8-neighbor backtracking bug follower [22] was used as a spiral scanning algorithm [6], [17]. Starting with the 2-D image of the segmented cell nucleus, we follow the (outer) contour of the nucleus, and spiral inwards as we peel off pixels, layer by layer, from the nucleus, forming a discrete one–dimensional (1-D) gray level signal.

The 1-D gray level signal resulting from the "peel-off scanning" of each cell nucleus image was divided into a peripheral (representing 30% of the total area of the nucleus) and a central (representing 70% of the area) segment [17]. Separate texture features were extracted from these segments.

D. Adaptive Feature Extraction

For a given texture analysis method, we compute the probability matrices (e.g., GLCMs) $P_n(i,j\,|\,\omega_c), n = 1,2,\ldots,N(\omega_c)$ from the $N(\omega_c)$ training set images of class ω_c. For each element (i,j) in the matrix we then estimate the class conditional probability distribution of the normalized matrix value. Based on these class conditional distributions, we compute the average matrix $\bar{P}(i,j\,|\,\omega_c)$ for each class ω_c, the class variance matrix $\sigma_P^2(i,j\,|\,\omega_c)$, the class difference matrix $\Delta_P(i,j\,|\,\omega_1,\omega_2)$, and finally the Mahalanobis class distance matrix $J_P(i,j\,|\,\omega_1,\omega_2)$ between the two classes ω_1 and ω_2

$$\bar{P}(i,j\,|\,\omega_c) = \frac{1}{N(\omega_c)} \sum_{n=1}^{N(\omega_c)} P_n(i,j\,|\,\omega_c) \tag{1}$$

$$\sigma_P^2(i,j\,|\,\omega_c) = \frac{1}{N(\omega_c)} \sum_{n=1}^{N(\omega_c)} (P_n(i,j\,|\,\omega_c) - \bar{P}(i,j\,|\,\omega_c))^2 \tag{2}$$

$$\Delta_P(i,j\,|\,\omega_1,\omega_2) = \bar{P}(i,j\,|\,\omega_1) - \bar{P}(i,j\,|\,\omega_2) \tag{3}$$

$$J_P(i,j\,|\,\omega_1,\omega_2) = 2\frac{(\bar{P}(i,j\,|\,\omega_1) - \bar{P}(i,j\,|\,\omega_2))^2}{\sigma_P^2(i,j\,|\,\omega_1) + \sigma_P^2(i,j\,|\,\omega_2)}. \tag{4}$$

Fig. 3 illustrates the computation of a GLCM class distance matrix. The GLCM class difference and class distance matrices obtained in the present study for the two different prognostic classes of early ovarian cancer are given in Figs. 4 (gray-scale plots) and 5 (surface plots).

Now the Mahalanobis class distance matrix and the class difference matrix form the basis upon which we construct only two adaptive features from the probability matrix of a given texture method. Our low dimensionality adaptive feature extraction is

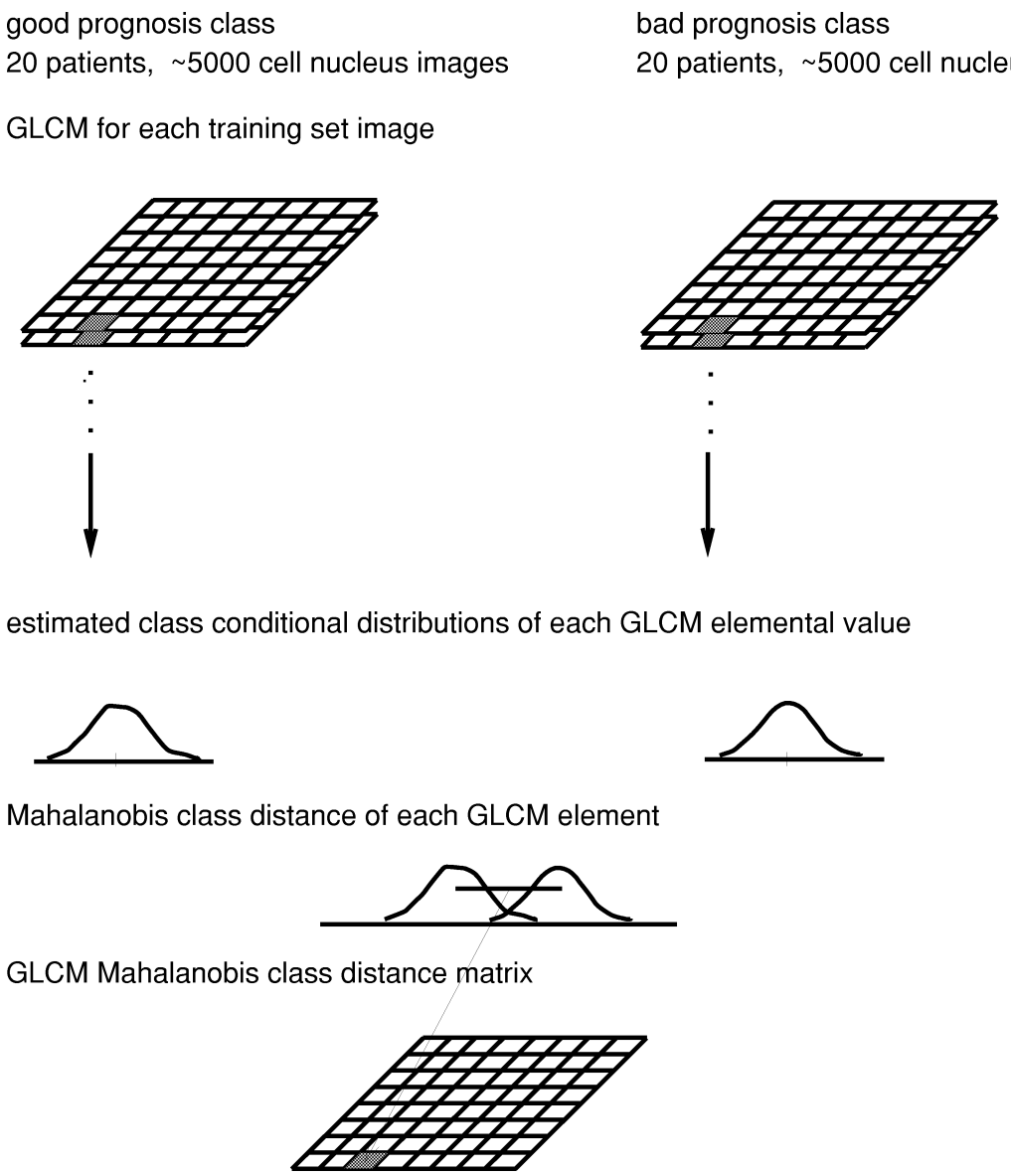

Fig. 3. The computation of a GLCM class distance matrix $J_A(i,j\,|\,\omega_1,\omega_2)$ between two different prognostic classes of early ovarian cancer. The GLCM class distance matrix contains information on which GLCM matrix elements that—when taken one matrix element at a time—will provide high class discrimination.

based on *a priori* information about the existence of large areas of consistently high values within the class distance and class difference matrices. For each texture image, we combine those probability matrix elements (e.g., GLCM elements) that contribute the most to the class separability into two adaptive features, simply by using the squared class distance matrix values as summation weights. We utilize the fact that in the two-class problem, the class difference has a *sign*, depending on whether the first or the second class matrix element contains the highest average probability. We use the two disjoint positive and negative parts of the class difference matrix as the domains of the weighted summation. Thus, an image having a probability matrix $P_k(i,j)$ will give two adaptive feature values

$$F_+ = \sum_{\Delta_P(i,j\,|\,\omega_1,\omega_2)\geq 0} P_k(i,j\,|\,\omega_c)[J_P(i,j\,|\,\omega_1,\omega_2)]^2$$
$$F_- = \sum_{\Delta_P(i,j\,|\,\omega_1,\omega_2)< 0} P_k(i,j\,|\,\omega_c)[J_P(i,j\,|\,\omega_1,\omega_2)]^2. \quad (5)$$

This guarantees that the highest weight is put on the most discriminatory parts of the matrix. We have applied this approach to four different texture analysis methods.

E. Gray Level Cooccurrence Matrix

The GLCM method [3] is one of the most popular methods of extracting second-order statistical texture features. The number of gray levels, G, in the image is often reduced by re-quantization prior to the accumulation of the matrix. Without any pretesting, we have chosen an interpixel distance $d = 3$ and $G = 16$ gray levels, which are commonly used parameters. From each Brodatz image, we have calculated an average (isotropic) matrix out of four matrices, $\theta = 0°, 45°, 90°, 135°$.

The following nine GLCM features were extracted for comparison with the two new adaptive features: angular second moment (ASM), contrast (CON), correlation (COR), variance (VAR), inverse difference moment (IDM), entropy (ENT) [3],

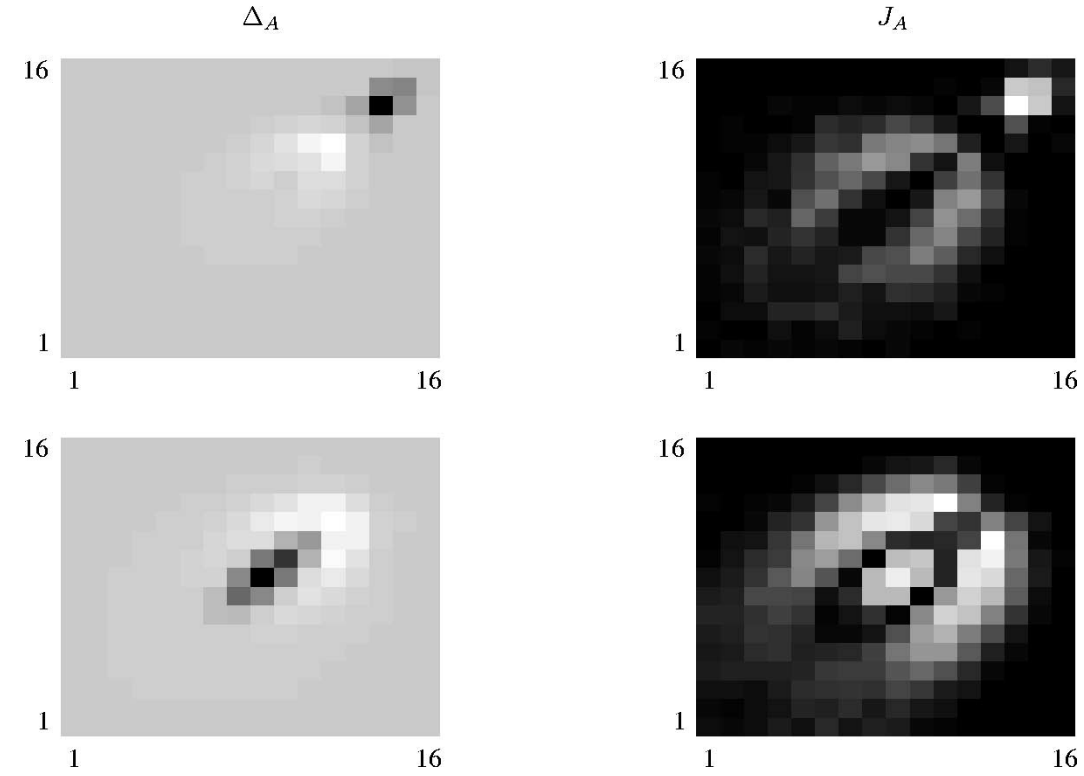

Fig. 4. (left) The GLCM class difference matrix, $\Delta_A(i,j \mid \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) segments extracted from the 2-D cell nucleus images. The lighter areas in the gray-scale plots correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. (right) The Mahalanobis class distance matrix, $J_A(i,j \mid \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the GLCM matrices, computed from the peripheral 30% (upper) and the central 70% (lower) segments. The number of gray levels in each nucleus image was reduced by re-quantization to $G = 16$ prior to the accumulation of the matrices.

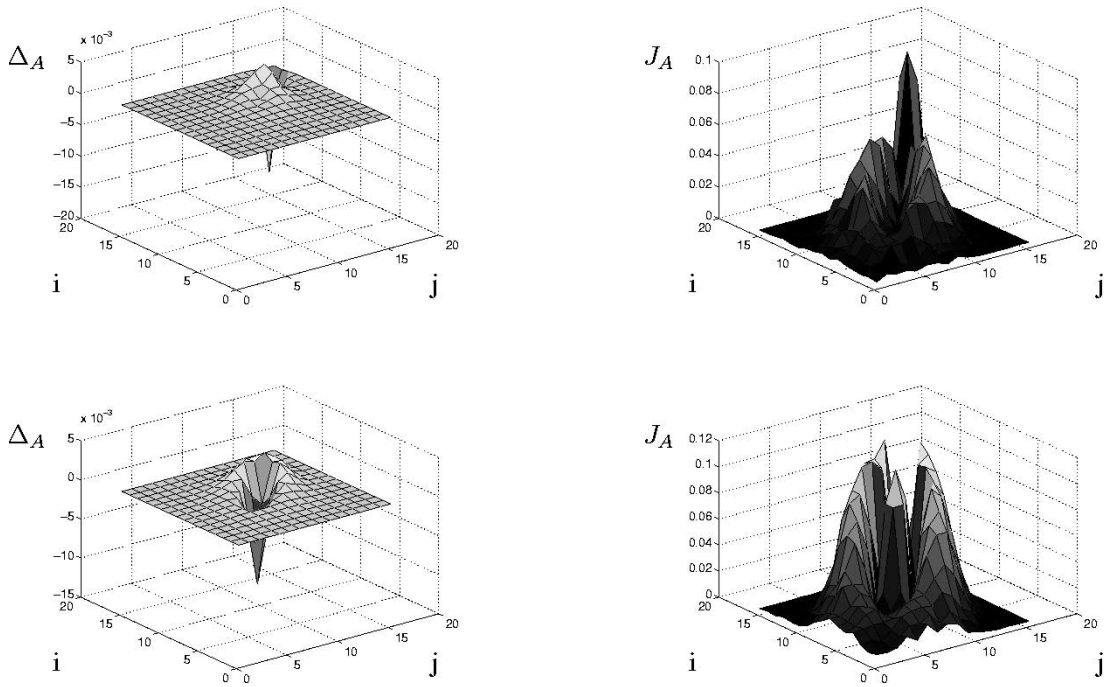

Fig. 5. (left) The GLCM class difference matrix, $\Delta_A(i,j \mid \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) segments extracted from the 2-D cell nucleus images. The positive values in the surface plots correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. (right) The Mahalanobis class distance matrix, $J_A(i,j \mid \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the GLCM matrices, computed from the peripheral 30% (upper) and the central 70% (lower) segments.

cluster shade (SHADE), cluster prominence (PROM) [23], and diagonal moment (DM) [24].

F. Gray Level Run Length Matrix

The GLRLM method extracts higher order statistical texture information [4], [25]. We have used $G = 16$, and the run lengths were partitioned logarithmically into ranges, i.e., 1, 2–3, 4–7, 8–15, 16–31, 32–63, 64–127, and 128–255. From each Brodatz image, we have calculated an average (isotropic) matrix out of two matrices, $\theta = 0°, 90°$.

The following seven classical features were compared with the two adaptive features: short runs emphasis (SRE), long runs emphasis (LRE), gray level nonuniformity (GLN), run length nonuniformity (RLN), run percentage (RP) [4], low gray level runs emphasis (LGRE), and high gray level runs emphasis (HGRE) [25].

As $p(i, j \mid \theta)$ is given as the *number* of occurrences of runs of length j at gray level i in a given direction θ, the original expressions for the features GLN and RLN [4] are not invariant to image size. Thus, they need to be redefined, and we have, therefore, formed the normalized GLRLM, $P(i, j \mid \theta)$, containing the probability of occurrences of runs of length j at gray level i, in a given direction θ [15].

G. Cooccurrence of Gray Level Run Length Matrix

The 4-D normalized CGLRLM [16] $P(i, j, k, l)$ may be seen as a natural extension of the 2-D GLRLM, containing the probability of cooccurrence of two neighboring runs of (gray level, run length) $= (i, j)$ and (k, l). Following the reasoning by Unser [26] for the 2-D GLCM case, two independent runs of (gray level, run length) $= (i, j)$ and (k, l) may be viewed as two random variables with the same variance. Thus, the 4-D normalized cooccurrence probability matrix $P(i, j, k, l)$ may be replaced by its associated 2-D sum $(P_s(\xi, \psi), \xi \in \{2, 2G\}, \psi \in \{2, 2R\})$ and difference $(P_d(\gamma, \delta), \gamma \in \{-G + 1, G - 1\}, \delta \in \{-R+1, R-1\})$ run length matrices, where R is the length of the longest run. So for all neighboring runs, the sum and differences of the two gray levels and the two run lengths were computed, and the entries in the two matrices accumulated. Finally, the sum and difference run length matrices were normalized.

Again we have used $G = 16$ and the run lengths were partitioned logarithmically into ranges, i.e., 1, 2–3, 4–7, 8–15, 16–31, 32–63, 64–127, and 128–255. The adaptive low dimensionality feature extraction as described above was then performed separately on the sum and difference run length matrices [16].

H. Complexity Curve

Baheerathan *et al.* [7] proposed five new texture features to characterize the first order statistics of the complexity curve proposed by Kamata *et al.* [27], thereby giving a second-order description of 2-D image texture: maximum value (MV), average value (AV), sample mean (SM), sample standard deviation (SSD), and entropy (ENT). Here, we have followed the findings of Baheerathan *et al.* [7], that computing five features for the gray levels below the mean pixel value (D) and five for the gray levels above (B) will give improved classification results. The two new adaptive complexity features were also extracted from the complexity curve.

TABLE I
THE MEAN CLASSIFICATION ERROR, $\overline{\mathrm{ERR}}$ (IN %), AVERAGED OVER THE 45 TEXTURE PAIRS FOR: 1) THE PREDEFINED FEATURE THAT GAVE THE LOWEST AVERAGE TRAINING SET ERROR; 2) THE PREDEFINED FEATURE THAT GAVE THE LOWEST TRAINING SET ERROR IN EACH OF THE 45 CLASSIFICATIONS WAS SELECTED, AND THE AVERAGE OF THESE ERROR RATES WAS COMPUTED; 3) THE ADAPTIVE FEATURE THAT GAVE THE LOWEST TRAINING SET ERROR IN EACH OF THE 45 CLASSIFICATIONS WAS SELECTED, AND THE AVERAGE OF THESE ERROR RATES WAS COMPUTED. THE MEAN BHATTACHARYYA DISTANCES $\overline{J_B}$ (AVERAGED OVER THE 45 TRAINING SETS) ARE ALSO GIVEN

Method	$\overline{ERR}$, training	$\overline{ERR}$, test	$\overline{J_B}$
"best" GLCM average (CON)	9.4	9.6	2.9
average of "best" GLCM	5.7	6.5	3.1
adaptive GLCM	2.6	6.9	3.8
"best" GLRLM average (RLN)	10.8	10.2	1.9
average of "best" GLRLM	7.8	8.3	2.0
adaptive GLRLM	5.6	8.8	2.0
"best" Complexity average (AV(D))	10.1	9.1	2.2
average of "best" Complexity	5.6	4.7	3.2
adaptive complexity	5.7	5.0	3.1
adaptive CGLRLM	2.5	3.7	5.0

I. Classification and Feature Evaluation

Bayesian classification with equal prior probabilities for each class was used as the rule for classification. The feature distribution within each class was assumed to be multivariate normal and the within-class covariance matrices were assumed equal. The Bayesian classification rule then becomes a linear discriminant function. Linear discriminant functions have a variety of pleasant properties from an analytical point of view. Even when they are not optimal, one might be willing to sacrifice some performance to gain the advantage of simplicity [28].

For each of the 45 Brodatz training sets, we have used the leave-one-out method to estimate the misclassification rates of single features and feature pairs. The statistical *Bhattacharyya distance*, $J_B(\omega_1, \omega_2)$, between classes ω_1 and ω_2 [29] was used as a second evaluation criteria when more than one feature were ranked as the "best" according to classification error. Again, we have assumed normal distribution of features within each class and equal *a priori* probabilities. For each Brodatz pair, the designed classifiers (based on the "best" single features and the "best" feature pairs) were then tested on the test set.

Initially, the value of each texture feature used to classify each patient of the ovarian data set was the mean value of the distribution of feature values extracted from about 250 cell nuclei/sample (patient). However, the corresponding standard deviation, and the 10th and 90th percentiles of the distributions have also been tried. Because of the small number of ovarian samples available, we have used the leave-one-out method to estimate the misclassification rates. However, for each cycle of the leave-one-out, new class difference and class distance matrices were obtained. If all data had been used to compute these matrices prior to the leave-one-out error estimation, a bias would have been introduced [30], [31].

III. RESULTS

We have constructed class distance matrices for four relevant texture analysis methods. For each method we have extracted a

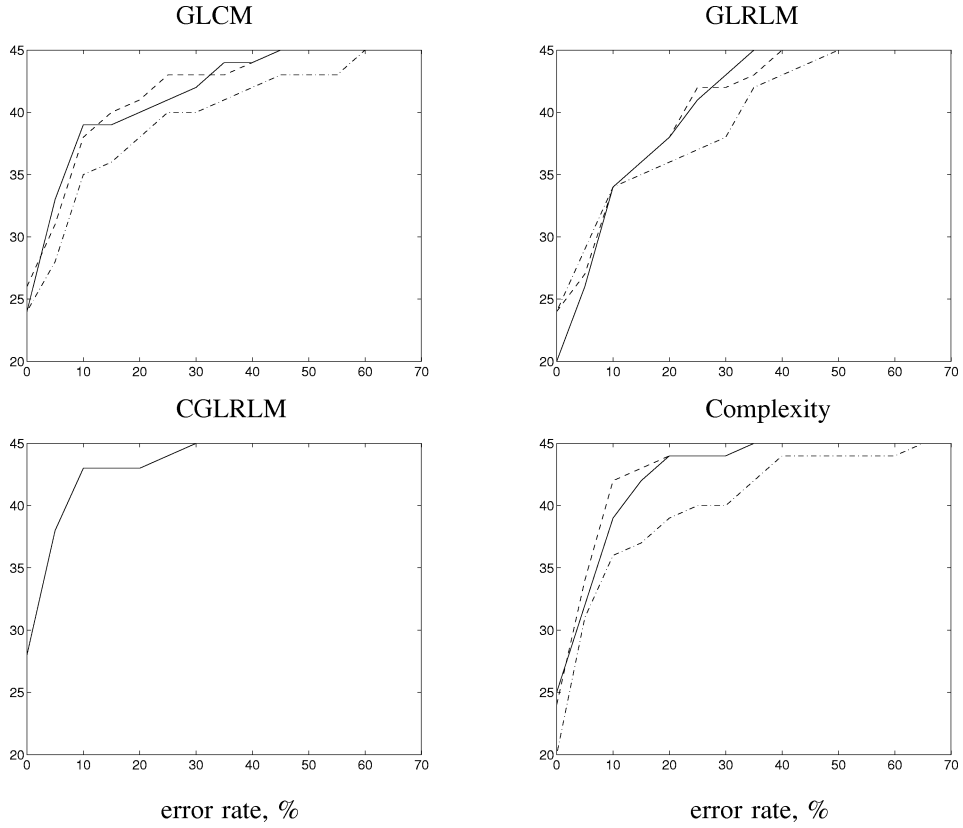

Fig. 6. The cumulative histogram of the error rates obtained from the 45 Brodatz texture test sets for: 1) the "best" average predefined feature (-.); 2) the average of the "best" predefined features (- -); and 3) the adaptive features (-) for each of the four texture methods.

small number of adaptive features, and evaluated them against a number of classical predefined features. The evaluation was performed on the most challenging set of Brodatz texture pairs as well as on 10 000 cell nucleus images taken from 40 patients of two different prognostic classes of early ovarian cancer.

A. Brodatz Textures

Based on the 45 training sets, we have selected the "best" single features and the "best" feature pairs. For each texture analysis method, Table I gives the mean classification error and the mean Bhattacharyya distance averaged over the 45 texture pairs for three cases: 1) the predefined feature that gave the lowest average training set error rate [i.e., contrast (CON), run length nonuniformity (RLN), average value (AV(D))], 2) the predefined feature that gave the lowest training set error in each of the 45 classifications was selected, and the average of these error rates was computed, 3) the adaptive feature that gave the lowest training set error in each of the 45 classifications was selected, and the average of these error rates was computed. For CGLRLM, only the results of the adaptive features are given, as no predefined features have been defined for this method. Fig. 6 shows the distribution of the test set error rates of the 45 classifications for the three cases above. The test set errors for the three cases for the different Brodatz pairs are given in Table II. The Brodatz pairs (D2D4, D2D9, D2D29, D2D57, D4D19, D4D60, D4D92, D9D19, D9D60, D19D57, D29D60, D32D60, D57D60) that gave a test set error rate of 0.0% for all three cases are not included in the table. The table gives the possibility to compare the performance of the adaptive and classical features, and also to compare the different texture analysis methods for each Brodatz pair.

In Table III the mean classification errors computed from the combinations of the two predefined and the two adaptive features that gave the lowest training set error rates in each of the 45 Brodatz texture classifications are given.

B. Ovarian Cancer

The class difference and class distance matrices for each of the texture methods are given in Figs. 4, 5, and 7–10. The matrices for the GLCM method are given both as gray-scale plots (Fig. 4) and surface plots (Fig. 5).

The mean feature value extracted from all cells from each patient gave a 35–40% classification error for all adaptive features, regardless of texture analysis method. However, the error was reduced if the standard deviation and the 10th and 90th percentile of the distribution of the feature over the approximately 250 cells/patient were used as features to classify each patient, and two such features were combined. The combination of two CGLRLM features gave an error rate of 25.0% and the combination of two complexity features gave an error rate of 22.5%.

IV. Discussion

A. Brodatz Textures

The predefined feature which gives the lowest classification error when averaged over all the 45 Brodatz texture pairs, is outperformed by the new adaptive features (Table I). This is true for all three methods having static predefined features.

TABLE II
THE CLASSIFICATION ERRORS, ERR (IN %), OF THE TEST SETS OF SOME OF THE 45 BRODATZ TEXTURE PAIRS FOR: 1) THE PREDEFINED FEATURE THAT GAVE THE LOWEST AVERAGE TRAINING SET ERROR RATE OVER THE 45 CLASSIFICATIONS; 2) THE PREDEFINED FEATURE THAT GAVE THE LOWEST TRAINING SET ERROR IN EACH OF THE 45 CLASSIFICATIONS; 3) THE ADAPTIVE FEATURE THAT GAVE THE LOWEST TRAINING SET ERROR IN EACH OF THE CLASSIFICATIONS. THE BRODATZ PAIRS THAT GAVE AN ERROR RATE OF 0.0% FOR ALL THREE CASES ARE NOT INCLUDED IN THE TABLE

	GLCM			GLRLM			CGLRLM	Complexity		
Brodatz pairs	1) CON	2)	3)	1) RLN	2)	3)	3)	1) AV(D)	2)	3)
D2D19	8.33	14.58	6.25	6.25	8.33	6.25	6.25	4.17	8.33	12.50
D2D32	0.00	0.00	0.00	0.00	0.00	2.08	0.00	0.00	0.00	0.00
D2D60	58.33	27.08	35.42	41.67	41.67	37.50	31.25	66.67	33.33	35.42
D2D92	2.08	2.08	2.08	0.00	0.00	0.00	0.00	0.00	4.17	0.00
D2D100	22.92	25.00	12.50	10.42	8.33	8.33	4.17	12.50	12.50	12.50
D4D9	10.42	12.50	4.17	37.5	27.08	25.00	25.00	39.58	12.50	14.58
D4D29	4.17	12.50	4.17	4.17	4.17	4.17	0.00	6.25	6.25	10.42
D4D32	16.67	10.42	18.75	10.42	18.75	10.42	6.25	18.75	10.42	18.75
D4D57	45.83	14.58	33.33	43.75	41.67	29.17	8.33	10.42	6.25	10.42
D4D100	0.00	0.00	2.08	12.50	12.50	6.25	6.25	4.17	4.17	4.17
D9D29	58.33	45.83	29.17	2.08	0.00	4.17	0.00	4.17	4.17	2.08
D9D32	35.42	8.33	10.42	4.17	16.67	8.33	8.33	37.50	4.17	8.33
D9D57	10.42	6.25	8.33	31.25	20.83	31.25	0.00	4.17	2.08	4.17
D9D92	4.17	2.08	4.17	2.08	0.00	0.00	0.00	2.08	2.08	0.00
D9D100	4.17	4.17	4.17	2.08	2.08	4.17	6.25	4.17	4.17	4.17
D19D29	0.00	0.00	0.00	10.42	10.42	18.75	0.00	4.17	0.00	0.00
D19D32	0.00	0.00	0.00	0.00	0.00	14.58	0.00	0.00	0.00	0.00
D19D60	8.33	2.08	10.42	4.17	4.17	8.33	4.17	12.50	10.42	8.33
D19D92	10.42	12.50	6.25	8.33	8.33	22.92	2.08	12.50	6.25	8.33
D19D100	41.67	39.58	41.67	33.33	10.42	14.58	12.50	41.67	20.83	18.75
D29D32	22.92	4.17	2.08	22.92	10.42	8.33	0.00	0.00	0.00	0.00
D29D57	8.33	4.17	6.25	0.00	0.00	2.08	0.00	33.33	4.17	6.25
D29D92	2.08	0.00	4.17	52.08	22.92	37.50	4.17	27.08	8.33	14.58
D29D100	4.17	4.17	4.17	37.50	33.33	27.08	6.25	6.25	6.25	6.25
D32D57	18.75	10.42	12.50	6.25	4.17	12.50	12.50	0.00	0.00	0.00
D32D92	0.00	0.00	0.00	16.67	27.08	10.42	0.00	0.00	0.00	0.00
D32D100	0.00	0.00	2.08	20.83	14.58	12.50	4.17	4.17	2.08	4.17
D57D92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.42	0.00	0.00
D57D100	0.00	0.00	0.00	2.08	2.08	6.25	6.25	4.17	12.50	4.17
D60D92	2.08	2.08	4.17	0.00	0.00	0.00	0.00	4.17	0.00	0.00
D60D100	20.83	18.75	25.00	0.00	0.00	2.08	2.08	20.83	8.33	2.08
D92D100	10.42	10.42	12.50	37.50	22.92	20.83	8.33	14.58	16.67	14.58

TABLE III
THE MEAN CLASSIFICATION ERROR, $\overline{\mathrm{ERR}}$ (IN %), COMPUTED FROM THE CLASSIFICATION ERRORS OF THE "BEST" COMBINATIONS OF TWO PREDEFINED AND TWO ADAPTIVE FEATURES IN EACH OF THE 45 CLASSIFICATIONS. THE MEAN BHATTACHARYYA DISTANCE $\overline{J_B}$ (AVERAGED OVER THE 45 TRAINING SETS) FOR THE SAME COMBINATIONS ARE ALSO GIVEN

Method	$\overline{ERR}$, training	$\overline{ERR}$, test	$\overline{J_B}$
average of "best" GLCM	3.7	4.9	5.5
adaptive GLCM	2.6	6.6	4.2
average of "best" GLRLM	4.9	6.2	4.4
adaptive GLRLM	5.4	7.1	2.4
average of "best" complexity	4.0	5.0	4.3
adaptive complexity	4.6	4.2	4.3
adaptive CGLRLM	1.9	3.2	6.6

Using the new adaptive features, we get the same classification performance as the average performance of the "best" predefined feature from each of the 45 classifications. This mixture of predefined features may contain 7–10 different features. The CGLRLM method gives the best results. This is confirmed by the mean Bhattacharyya distance. The reason for this may be that this method on the average is using more pixels in the image/entry in the probability matrices, and thus is destilling more contextual information from the image into a few features. The strong adaptivity of the new features probably causes a lower training error than test set error.

When we consider pairs of features applied to the task of correctly classifying the 45 possible pairs of the 10 Brodatz textures, the "best" pair of predefined features is chosen in each of the 45 situations, and an average is computed. Thus, there are 36, 21, or 45 possible pairs of predefined features involved in the case of GLCM, GLRLM, and complexity, respectively. For the new adaptive features on the other hand, there is only one pair of features in these three methods, while there are six possible pairs of the four features from the sum and difference run length matrices of the CGLRLM method.

When we allow a ranking of the pairs of predefined features according to their performance in the classification of each of the 45 Brodatz texture pairs, and then compute an average performance of this mixture of feature pairs, we get approximately the same results as when we use the only adaptive feature pair available (Table III). So the low dimensional adaptive feature pairs are able to perform on par with a weighted mixture of predefined feature pairs that come from a much higher dimensional feature space. Again, we get better results from the CGLRLM method than from any of the other methods.

B. Ovarian Cancer

In the case of the ovarian cancer, the sampling unit used for classification may either be the cell or the tumour [32]. In accordance with the recommendation of [32], we have chosen to classify the tumour, computing average values and a few other first order statistics to characterize every lesion. The formulation above is based on a single level statistical aggregation of the data from all cells within each class of the training set in order to

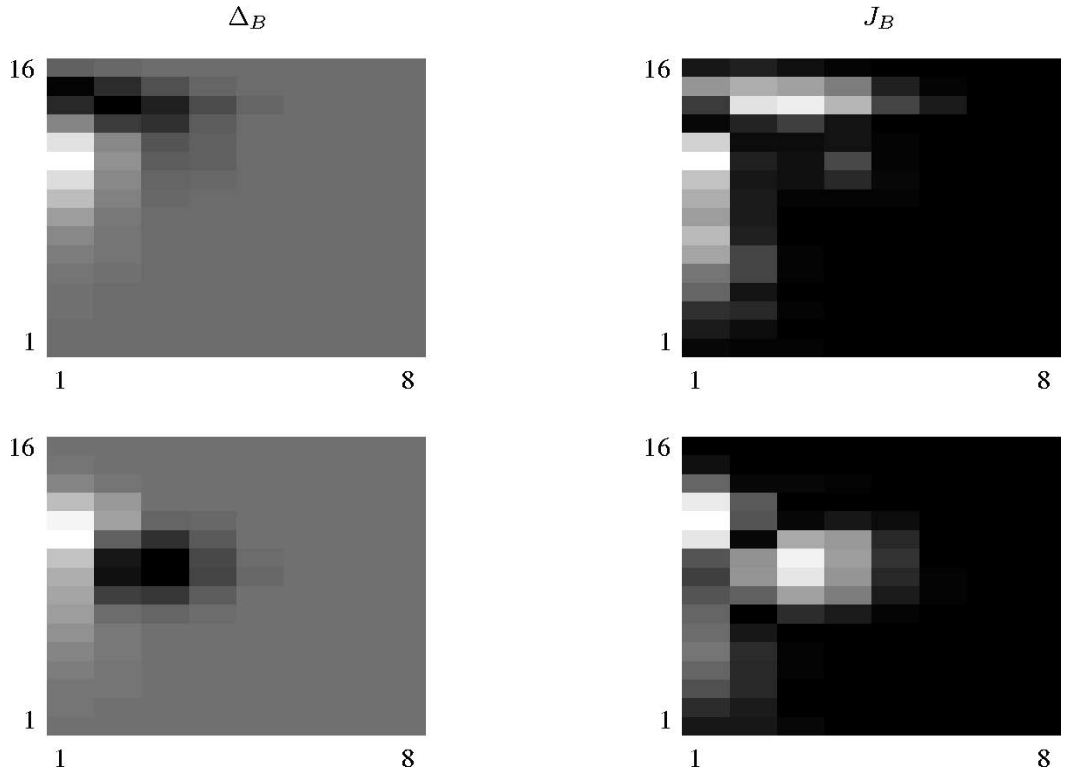

Fig. 7. (left) The GLRLM class difference matrix, $\Delta_B(i, j \mid \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) segments extracted from the 2-D cell nucleus images. The lighter areas in the gray-scale plots correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. (right) The Mahalanobis class distance matrix, $J_B(i, j \mid \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the GLRLM matrices, computed from the peripheral 30% (upper) and the central 70% (lower) segments.

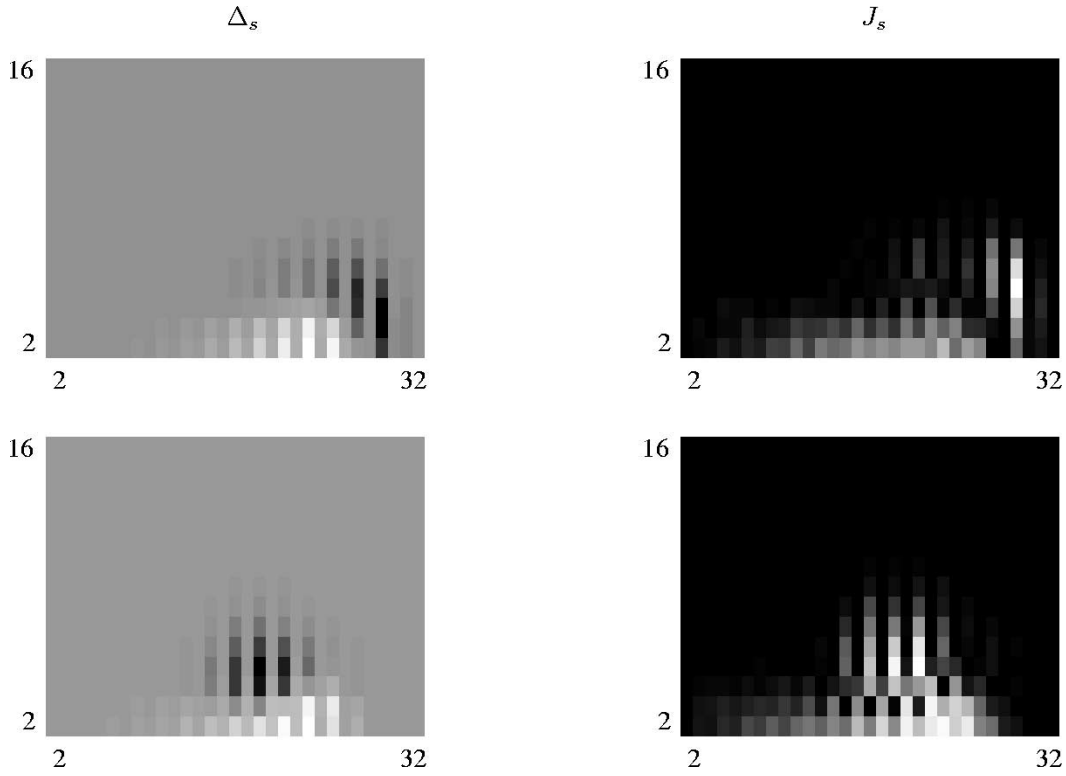

Fig. 8. (left) The sum CGLRLM class difference matrix, $\Delta_s(\xi, \psi \mid \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) segments extracted from the 2-D cell nucleus images. The lighter areas in the gray-scale plots correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. (right) The Mahalanobis class distance matrix, $J_s(\xi, \psi \mid \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the sum CGLRLM matrices, computed from the peripheral 30% (upper) and the central 70% (lower) segments.

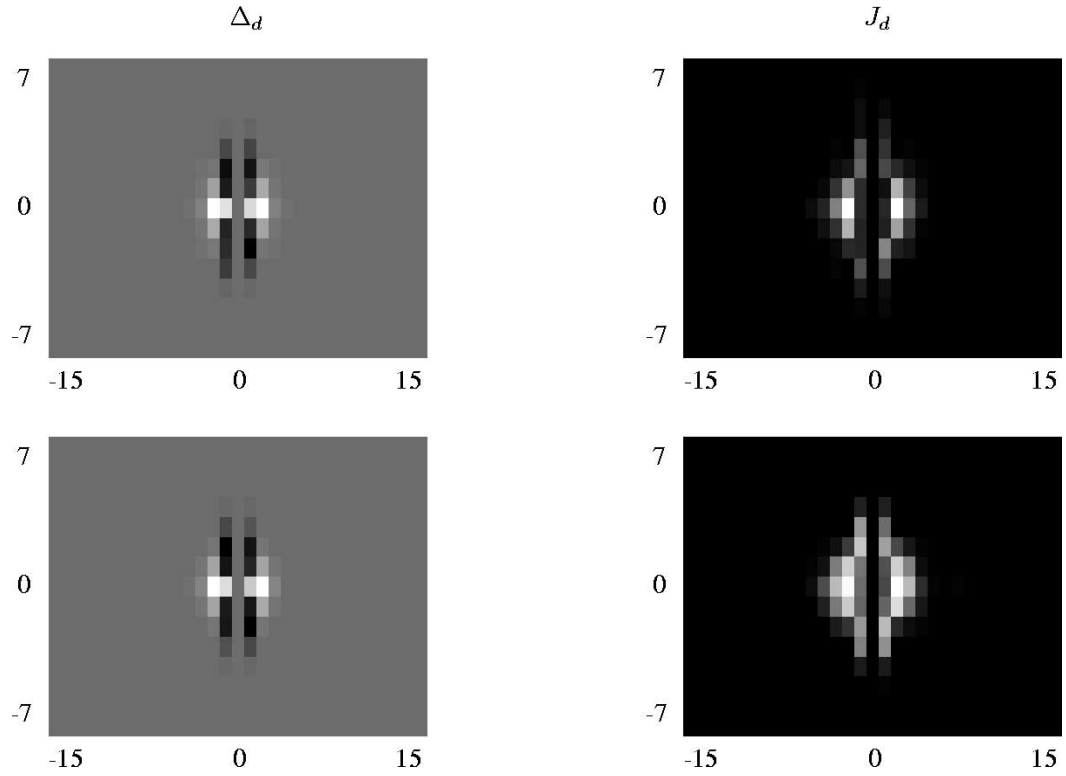

Fig. 9. (left) The difference CGLRLM class difference matrix, $\Delta_d(\gamma, \delta \,|\, \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) segments extracted from the 2-D cell nucleus images. The lighter areas in the gray-scale plots correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. (right) The Mahalanobis class distance matrix, $J_d(\gamma, \delta \,|\, \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the difference CGLRLM matrices, computed from the peripheral 30% (upper) and the central 70% (lower) segments.

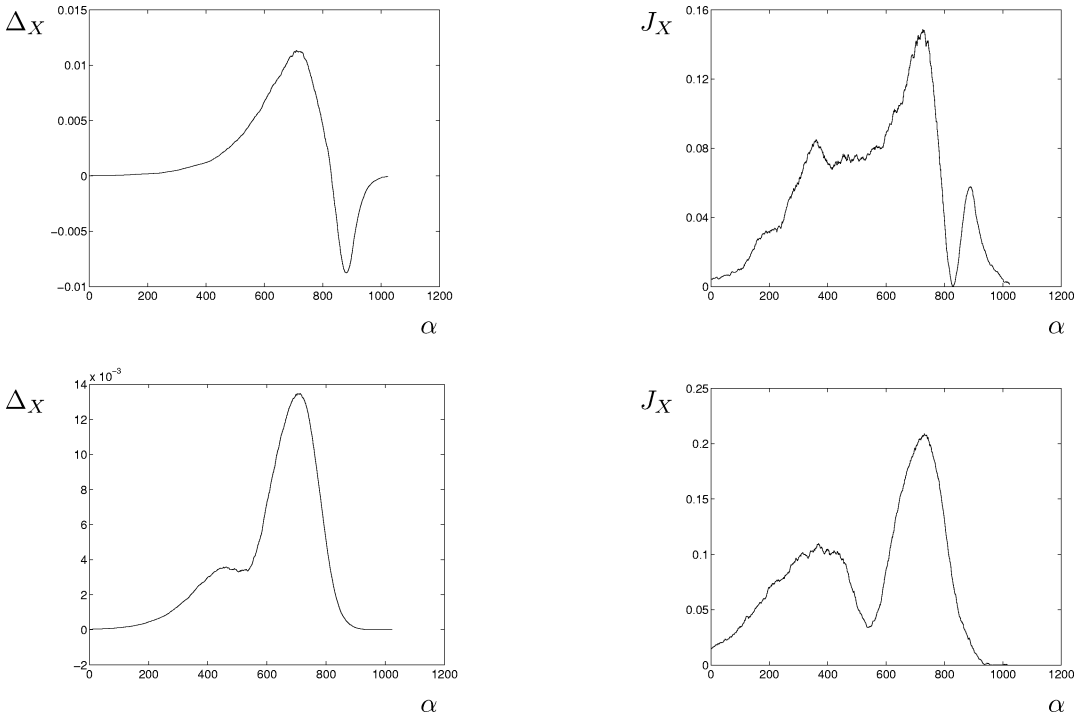

Fig. 10. (left) The complexity class difference vector, $\Delta_X(\alpha \,|\, \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) segments extracted from the 2-D cell nucleus images. The positive values correspond to vector elements that are more probable for the good prognosis class than for the bad prognosis class. (right) The Mahalanobis class distance vector, $J_X(\alpha \,|\, \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the complexity curves, computed from the peripheral 30% (upper) and the central 70% (lower) segments.

establish the squared class distance matrix $J^2(i, j \,|\, \omega_1, \omega_2)$ and the class difference matrix $\Delta(i, j \,|\, \omega_1, \omega_2)$.

The class distance matrices contain localized areas of consistently high values, and these correspond to separate areas of

positive or negative values in the difference matrices (Figs. 4, and 7–10). Earlier [5], [15], [16], we have found that the class distance matrices between different pairs of mouse liver nuclei classes contained large areas of consistently high values, and that these areas corresponded to separate areas of consistently positive or negative class difference values.

In the case of GLCM (Fig. 4), GLRLM (Fig. 7), CGLRL sum matrices (Fig. 8), and complexity (Fig. 10) there is a marked difference between the matrices computed from the peripheral 30% and the central 70% segments of the cell nuclei. Using a partitioning of the cell nuclei into a peripheral and a central part, we have repeatedly seen that the classification based on chromatin texture will benefit from a radial dichotomy of the texture feature extraction [5], [6], [15], [16], [33].

A gray level difference of one between neighboring runs is less probable in cell nuclei from the good prognosis class than for the bad prognosis, while larger gray level differences (from two to five) are more probable in the good prognosis case (Fig. 9). This indicates a difference in texel size between the two classes. This also explains the higher probability of odd number sums of gray levels in the bad prognosis class (Fig. 8).

The class difference vector in the complexity case is not a difference between normalized probability distributions, as in the other methods. Therefore, the overall sum of the difference vector is not necessarily zero. The result is that one may have two features, as in the peripheral part of the cell nuclei (Fig. 10, upper), or only one feature, as for the central part of the cell nuclei (Fig. 10, lower), where the class difference vector is always nonnegative.

Using the two percentiles and the standard deviation in addition to the mean will obviously increase the total number of classification features by a factor of four. The reduction in feature space dimensionality accomplished by the new adaptive texture features is, therefore, an important point. This is particularly valuable in the case of the ovarian cancer material, where we extract the texture features from two separate segments (center and periphery) for each cell nucleus. In the case of GLCM, if we had started out with only the nine most popular predefined GLCM features, using only one (d, G) parameter setting, we would have had a total of 72 possible features when we use the mean, the standard deviation and the two percentiles (10th and 90th), and a staggering 2556 possible pairs of features. This in contrast to the 16 adaptive features obtained when we combine two matrix features with two cell segments and four distribution parameters. The number of possible feature pairs is now only 120, reduced by a factor of 21.3, substantially reducing the risk of selecting good feature pairs by pure coincidence.

C. Summary

In the present study we have eliminated the classical step of predefined feature extraction, and instead computed a minimum number of adaptive features by using squared Mahalanobis class distance matrices as weights in a summation over matrix elements, using the disjoint positive and negative parts of the class difference matrices as the domains of the weighted summation.

When using the low dimensional feature extraction scheme presented here, the features will automatically adapt if the location of the most class discriminating matrix elements is shifted, either because of different parameter settings, changes in the physical conditions of the image material, or changes from one data set to a completely different one.

In conclusion, the new adaptive features outperformed the predefined features when applied to the 45 most challenging Brodatz texture pairs. Class distance and difference matrices clearly illustrated the difference in chromatin texture between cell nucleus images from two different prognostic classes of early ovarian cancer. For each of the texture analysis methods, we found that one adaptive feature contained most of the discriminatory power of the method.

ACKNOWLEDGMENT

The authors would like to thank W. Hansen, W. Kildal, and R. Puntervold of the Section of Digital Pathology, the Norwegian Radium Hospital, for their skillful technical assistance. They would also like to thank Prof. K. Liestøl of the Department of Informatics, University of Oslo, for useful discussions.

REFERENCES

[1] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds, Singapore: World Scientific, 1993, pp. 235–276.

[2] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[3] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610–621, 1973.

[4] R. M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graphic. Image Processing*, vol. 4, pp. 172–179, 1975.

[5] B. Nielsen, F. Albregtsen, and H. E. Danielsen, "Fractal signature vectors and lacunarity class distance matrices to extract new adaptive features from cell nuclei," in *Fractals in Biology and Medicine*, G. A. Losa, D. Merlini, T. F. Nonnenmacher, and E. R. Weibel, Eds. Basel, Switzerland: Birkhäuser-Verlag, 2002, vol. 3, pp. 55–65.

[6] ——, "The use of fractal features from the periphery of cell nuclei as a classification tool," *Analytical Cellular Pathol.*, vol. 19, pp. 21–37, 1999.

[7] S. Baheerathan, F. Albregtsen, and H. E. Danielsen, "New texture features based on the complexity curve," *Pattern Recogn.*, vol. 32, no. 4, pp. 605–618, 1999.

[8] D. Zongker and A. Jain, "Algorithms for feature selection: An evaluation," in *Proc. 13th Int. Conf. Pattern Recognition*, vol. 2, 1996, pp. 18–22.

[9] A. K. Jain and D. Zonker, "Feature selection: Evaluation, application and small sample performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 153–158, Feb. 1997.

[10] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 4–37, Jan. 2000.

[11] H. Schulerud, G. B. Kristensen, K. Liestøl, L. Vlatkovic, A. Reith, F. Albregtsen, and H. E. Danielsen, "A review of caveats in statistical nuclear image analysis," *Analytical Cellular Pathol.*, vol. 16, no. 2, pp. 63–82, 1998.

[12] H. Schulerud, "Bias of error rates in linear discriminant analysis caused by feature selection and sample size," in *Proc. 15th Int. Conf. Pattern Recognition*, vol. 2, 2000, pp. 372–377.

[13] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 252–264, Mar. 1991.

[14] R. F. Walker, P. Jackway, and I. D. Longstaff, "Recent developments in the use of the co-occurrence matrix for texture recognition," presented at the *13th Int. Conf. Digital Signal Processing*, Santorini, Greece, 1997.

[15] F. Albregtsen, B. Nielsen, and H. E. Danielsen, "Adaptive gray level run length features from class distance matrices," in *Proc. 15th Int. Conf. Pattern Recognition*, vol. 3, 2000, pp. 746–749.

[16] F. Albregtsen and B. Nielsen, "Texture classification based on cooccurrence of gray level run length matrices," *Aust. J. Intelligent Information Processing Syst.*, vol. 6, no. 1, pp. 38–45, 2000.

[17] B. Nielsen, F. Albregtsen, W. Kildal, and H. E. Danielsen, "Prognostic classification of early ovarian cancer based on very low dimensionality adaptive texture feature vectors from cell nuclei from monolayers and histological sections," *Analytical Cellular Pathol.*, vol. 23, pp. 75–88, 2001.

[18] P. Brodatz, *Textures—A Photographic Album for Artists & Designers*. New York: Dover, 1966.

[19] G. B. Kristensen, W. Kildal, V. M. Abeler, E. O. Pettersen, J. Kaern, C. G. Tropé, I. Vergote, and H. E. Danielsen, "High-resolution image analysis—A better method for predicting prognosis in early ovarian cancer," *Gyn. Oncol.*, vol. 76, no. 2, p. 236, 2000.

[20] D. W. Hedley, "DNA analysis from paraffin-embedded blocks," *Methods Cell Biol.*, vol. 41, pp. 231–240, 1994.

[21] H. J. Tanke and E. M. van Ingen, "A reliable Feulgen-acriflavine-SO_2 staining procedure for quantitative DNA measurements," *J. Histochem. Cytochem.*, vol. 28, pp. 1007–1013, 1980.

[22] W. K. Pratt, *Digital Image Processing*, 2nd ed. New York: Wiley, 1991, pp. 623–625.

[23] R. W. Conners, M. M. Trivedi, and C. A. Harlow, "Segmentation of a high-resolution urban scene using texture operators," *Comput. Vis. Graphic. Image Processing*, vol. 25, pp. 273–310, 1984.

[24] J. M. Carstensen, "Description and Simulation of Visual Texture," Tech. Univ. Denmark, Lyngby, Denmark, Ph.D. dissertation, 1992.

[25] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recogn. Lett.*, vol. 11, pp. 415–420, 1990.

[26] M. Unser, "Sum and difference histograms for texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 118–125, 1986.

[27] S. Kamata, R. O. Eason, and K. Kawaguchi, "Complexity curves versus histograms and their application to image segmentation," in *Proc. 7th Scand. Conf. Image Analysis*, 1991, pp. 1070–1077.

[28] R. O. Duda, P. E. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.

[29] D. J. Hand, *Discrimination and Classification*. New York: Wiley, 1981.

[30] S. M. Snapinn and J. D. Knoke, "Estimation of error rates in discriminant analysis with selection of variables," *Biometrics*, vol. 45, no. 1, pp. 289–299, 1989.

[31] H. Schulerud and F. Albregtsen, "Effects of many feature candidates in feature selection and classification," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2002, vol. 2396, pp. 480–487.

[32] O. Tsybrovskyy and A. Berghold, "Primary unit for statistical analysis in morphometry: Patient or cell?," *Analytical Cellular Pathol.*, vol. 18, pp. 191–202, 1999.

[33] B. Nielsen, F. Albregtsen, S. Baheerathan, and H. E. Danielsen, "Peel-off-scanning to obtain radial differentiation of fractal and complexity features in cell nuclei," presented at the *Vision Interface 2000*, Montreal, Canada, 2000.