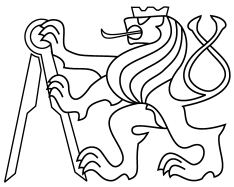
CENTER FOR
MACHINE PERCEPTION



CZECH TECHNICAL
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

Discriminable and Stable Feature Selection for Computer-aided Diagnosis from Sonographic Images – PhD Thesis Proposal

Martin Švec

xsvecm@cmp.felk.cvut.cz

CTU–CMP–2005–05

February 28, 2005

Available at
ftp://cmp.felk.cvut.cz/pub/cmp/articles/svec/Svec-TR-2005-05.pdf

Supervisor: Radim Šára

This work has been supported by the Grant Agency of the Czech Academy of Sciences under project 1ET101050403 and by the Czech Ministry of Health under projects NO/7742-3 and NB 5472-3.

Research Reports of CMP, Czech Technical University in Prague, No. 5, 2005

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Discriminable and Stable Feature Selection for Computer-aided Diagnosis from Sonographic Images – PhD Thesis Proposal

Martin Švec

February 28, 2005

Abstract

This paper summarises state-of-the-art in feature selection and classification for medical purposes. The main focus is put on sonographic imaging, especially images of the thyroid gland. State-of-the-art covers related works dealing with feature selection and classification (i) in general, (ii) for medical purposes, and (iii) for thyroid gland diseases. It also summarizes works that deal with the task of different scanners and digitizer setting, which may lead to different classification results. The second part of this paper describes our methods and experiences with feature selection and classification applied on sonographic images of thyroid gland. It deals with generating features from sonographic textural images and reproducibility under different sonograph settings. The last part involves tentative plan for future work to conclude PhD thesis.

Contents

1 Introduction

Sonographic imaging is cheap non-invasive technique that is widely used for treating and for diagnostic purposes. Sonographic image arises during process that consists of several steps:

(i) Ultrasound waves are sent by transducer into the tissue.

(ii) Interferences of waves and tissue make ultrasound echoes.

(iii) Echoes returned from different tissue depths are detected by the same transducer.

(iv) Detected signals are processed and transformed into image called B-mode.

B-mode image is two-dimensional grey-scale image in which vertical dimension is related to the depth in the tissue and the horizontal dimension belongs to ultrasound beam position.

Resulted B-mode image has textural character. Visual changes in texture lead an expert while making decision about appropriate diagnosis, based on their knowledge and experiences. Besides this qualitative assessment, texture can also be described and classified independently on the observer, using quantitative description. Quantifiers for this description are called features. There can be computed many features for an image. Methods that are used to find the most important ones are called feature selection methods. Since general feature selection is not the aim of this paper, just several works are overviewed in Sec. 2.1. Texture features should bear the main characteristics of texture. Examples of works on texture features are given in Sec. 2.2. Features describing textural character of sonographic image were used for tissue classification in many applications. These works are described in Sec. 2.3.

We deal with classification of sonographic images of thyroid gland (position of the thyroid gland and the transducer can be seen in Fig. 1). There are few works that deal with similar tasks and are mentioned in Sec. 2.4. Most of these works examined images acquired under one specific setting of the sonographic tool. Classification results achieved under changed setting should ideally be the same as for the initial setting. This property is called reproducibility. So far, general reproducibility for such classification was not achieved. Works that deal with reproducibility are described in Sec. 2.5. As has been stated above, we deal with sonographic images of the thyroid gland. Our goal is to find the most discriminable features to classify between normal tissue and one particular disease called Hashimoto's lymphocytic thyroiditis.

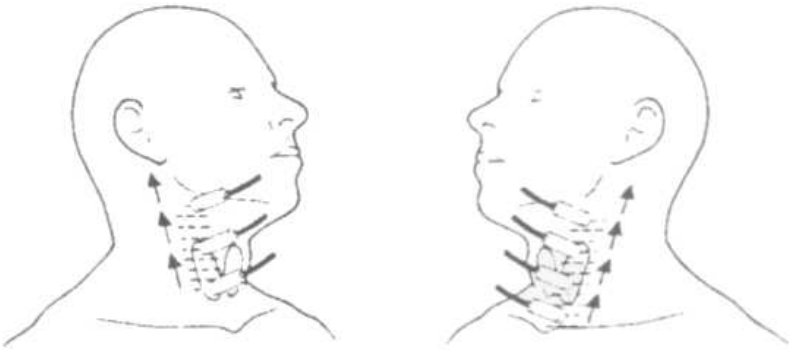


Figure 1: Scanning of the thyroid gland.

Such features should also be the most stable ones, it means the best for reproducibility. Our methods and results are given in Sec. 3. Conclusions and tentative plan to conclude PhD thesis is outlined in Sec. 4.

2 State of the Art

2.1 Feature Selection

Basic principles for general feature selection are described in several works. General points of feature selection and future possible improvements are outlined in [1]. An overview of feature subset selection algorithms and experiments with floating methods and branch and bound method are given in [2]. Feature subset selection by Estimation of Bayesian Network Algorithm is presented in [3]. Feature selection methods are also summarized in [4]: sequential forward selection, sequential backward selection, sequential floating search (forward or backward) and both wrapper and filter approaches are explained here. Somol introduced improvements for branch and bound algorithm for feature selection in [5]. Definition of the objective function, feature construction, feature ranking, multivariate feature selection, efficient search methods, and feature validity assessment methods are overviewed in [6]. Bressan discriminates features into two statistically dependent and statistically independent in [7]. For independent, features with the highest between-class divergence are chosen, for dependent, Independent Component Analysis is used prior to feature selection. Law used mixture models for simultaneous feature selection and clustering in [8]. The sum of kernel functions with controlling parameters that weight importance of

training examples and importance of features is presented in [9]. Feature selection was used for practical purpose for example in [10] for magnetic resonance images of human knees and in [11] for computer tomography lung images.

2.2 Feature Selection for Texture Recognition

Texture features should bear the main characteristics of texture. However, texture cannot be easily described. There exist several definitions, the one used by Turceyan and Jain [12] is stated here:

> "The notion of texture appears to depend upon three ingredients: (i) some local 'order' is repeated over a region which is large in comparison to the order's size, (ii) the order consists in the nonrandom arrangement of elementary parts, and (iii) the parts are roughly uniform entities having approximately the same dimensions everywhere within the textured region"

Davies [13] denotes this order as texture elements that are replicated over a region of the image – *texels*. He characterized texture in following ways:

(i) The texels have various sizes and degrees of uniformity.
(ii) The texels are oriented in various directions.
(iii) The texels are spaced at varying distances in different directions.
(iv) The contrast has various magnitudes and variations.
(v) Various amounts of background may be visible between texels.
(vi) The variations composing the texture may each have varying degrees of regularity and randomness.

Many approaches exist to characterize texture, some of them are overviewed by Haralick and Turceyan [12, 14, 15]. Early works in image texture analysis aimed to discover features based on the use of Fourier analysis. This approach was tested mainly on periodic textures, but the results were not always encouraging. Autocorrelation is another approach to texture analysis, but it is not a very good discriminator of isotropy of natural textures. Hence researchers widely used the co-occurrence matrix approach introduced by Haralick in 1973. It became the "standard" approach to texture analysis. This approach will be mentioned later in this thesis.

Shirazi [16] used texture classification methodology that was based on stochastic modelling of textures in the wavelet domain. Chang and Kuo [17] and Laine and Fan [18] also dealt with a multiresolution approach based on wavelet transform. Shen [19] computed features from different image

resolutions and extracted feature frequency matrices. He used weighted distance between feature vectors instead of Euclidean distance. Pitas and Kotropoulos [20] dealt with segmentation of seismic images by Hilbert transform, minimum entropy learning technique, and by region growing algorithm. Sullins [21] tackled the problem that while most features are useful in some situations, none are totally effective in all of them. He used distributed learning system to learn relevant texture descriptors from a set of first and second-order grey-level statistics.

Some works deal with fractal dimension. The fractal dimension was used as a measure of the characteristics of texture. Kakemura [22] avoided the problem that some textures can easily be discriminated as different textures by human vision, but cannot be discriminated based on their fractal dimensions (white-noise texture and Brownian-noise texture). However, fractal dimension is not sufficient to capture all textural properties.

Biologically motivated nonlinear texture operator, introduced by Kruizinga [23], the grating cell operator was compared to co-occurrence features. It was pointed out by using Fisher linear discriminant on the problem of texture segmentation that the grating cell operator responses only to texture whereas co-occurrence features response also to edges.

2.3 Applications for Sonographic Imaging

Textures of unique organs differ and can be changed by specific diseases in many different ways. There is no general approach and solutions of these classification tasks can differ for each application.

Chen [24] used fractal dimension to discriminate between images with normal livers and abnormal livers. They used an estimation of fractal dimension as a feature. Since the fractal dimension of medical images changes with the scale, they estimated dimension for 27 different scales. Limitation of this approach is the choice of the sample from image. Classification is successful only on samples that contain the least number of blood-vessels. Several authors dealt with textural analysis of ultrasonic images of the liver. Bleck [25] used autoregressive periodic random field models to distinguish between patients without and with microfocal lesions of the liver. Sujana [26] achieved classification accuracy of 100% on the set of images of normal, hemangioma, and malignant livers. It was performed using a multilayered back-propagation neural network. Horng [27] applied a novel approach, called texture feature coding method for texture classification of normal liver, hepatitis, and cirrhosis with correct classification rate of 83.3%. Mojsilovic [28] used wavelet decomposition to detect liver cirrhosis in its early stage. They achieved accuracy of 92%. Jeong [29] improved single feature classification of normal

livers, fat livers and cirrhosis by selecting multiple texture feature vector and its classification using Bayes rule. Yeh [30] used features from co-occurrence matrix and wavelet transform to classify 6 grades of liver fibrosis.

Sutton and Hall [31] dealt with automated screening of chest images for the detection of textural type abnormalities. The disease processes known as interstitial pulmonary fibrosis were considered. Features were based on the statistical properties of the spatial distribution of image pixels. Classification accuracy was 84% for the test set of 24 patients. The classification results using measurements obtained from the Fourier transform domain were disappointing despite of general expectation in the early days of texture analysis. It is because of the existence of no optimum method for feature selection for all types of texture. Segyeong [32] detected solid breast nodules using artificial neural networks. He used 5 features that describe apparent properties of the nodule (e.g. shape, intensity and frequency properties). Garra [33] used co-occurrence matrix features to distinguish sonographic images of breast lesions: cancers, cysts, fibroadenomata and fibrocystic nodules.

Pohle [34] tackled the task of skeletal muscle sonography by computing large set of features: features of run-length matrix, first and second order statistic features, frequency spectrum, and fractal features. He then focused on feature selection, i.e. choosing the most appropriate subset of features for given task. Muzzolini used similar method in [35].

Pujol [36] used features computed from co-occurrence matrix and cumulative moments to find vessel borders in intravascular ultrasound images. Zhang [37] characterised plaque composition from intravascular ultrasound images. He selected features among features computed from gray-level-based measures, co-occurrence matrices, run length measures, and fractal-based measures. The correctness of classification approached 90%. Intravascular ultrasound images were also within the scope of Nailon. He used features derived from co-occurrence matrix and features from fractal texture analysis to (i) distinguish among three different plaque groups: loose fibrotic tissue, dense fibrotic tissue and calcium [38], (ii) characterize disease from samples of: foam cells, foam cells containing lipid and loose fibrotic tissue [39], and (iii) distinguish three types of intracoronary thrombus: red, white and plasma [40]. Coronary plaque in intravascular images was also within the scope of Dixon [41]. He used Haralick's texture features derived from co-occurrence matrix to distinguish adventitia, calcified, fibrous, media and necrotic core samples.

Amin [43] tried to characterize beef quality by assessing fat percentage. Features derived from spatial gray-level dependence matrices and gray-level run-length matrices were used.

Krass [42] derived features from the frequency power spectrum of coronary images to distinguish different types of thrombi.

2.4 Applications for Sonographic Imaging of Thyroid Gland

Few studies focused on the thyroid gland. Hirning [44] analyzed computerized B-mode images. Two features from a set of 109 features proved to be the most efficient statistical parameters: the upper decile of grey level distribution (it allowed to classify normal tissue and cyst from other diagnosis with 100% success), and the entropy (distinguished cyst from the rest with 100% success). Diagnostic classes were normal tissue, carcinoma, adenoma, struma nodosa, cyst, autonomous adenoma, thyroiditis, and Graves' disease. Thyroiditis was classified with success of 87%, 13% was misclassified and denoted as carcinoma. According to the small number of test set (15 patients), we can not deduce ultimate solutions about thyroiditis classification. Mailloux [45] used histogram for segmentation of normal tissue and Hashimoto's disease. Cluster analysis was applied by using K-means algorithm, supposing four classes: background in sonographic image, surrounding tissue, normal thyroid tissue, and thyroid tissue with Hashimoto's disease. Texture of diseased parenchyma seemed to be of two different types, some of them different from normal tissue, some closer to normal tissue. These two types were denoted as parallel to its histologic development. This conclusion was made on 10 patients, which does not seem to be a sufficiently large set. Schiemann [46] used grey scale histogram analysis to show that tissue with Graves' disease has significantly lower echogenicity than normal tissue. Youssef [47] tried to distinguished adenomas and malignancies of thyroid gland using three features: the average scatterer separation, 10% percentile, and entropy (computed from the co-occurrence matrix).

2.5 Reproducibility of the Sonographic Image Classification

There is one important thing in ultrasound image classification. Most of previously mentioned works do not take into account different parameters of the sonographic tool. Results are achieved on a fixed setting of the sonograph, so they are not reproducible on the images taken under different conditions. The reproducibility issue is a long-standing problem in similar quantitative methods [33]. Many relevant works are limited to short statements without closer details, it means that e.g. parameter settings were adjusted for optimal

visualization ([45]), fixed to have standardized conditions ([34, 44]), or kept at values normally used in clinical practice ([41]). Chan [48] tried to tackle this problem by changing the gain setting during the experiment and capturing for each gain at least five images of the object. Mojsilovic [28] removed the mean of each image in order to eliminate effects of unequal ultrasound gain settings.

Gain is the main parameter that influences resulted sonographic image. It can be changed manually or automatically in order to obtain image in which the least amount of relevant information will be lost. Pye [49] examined 17 algorithms for adaptive time gain compensation. These algorithms are used to automatically change gain in real time for given image. However, although automatic gain setting can be in many cases better than manual change, the relevant question is the following: having classifying system learned on images that are scanned under one specific gain, can we obtain the same results under different gains?

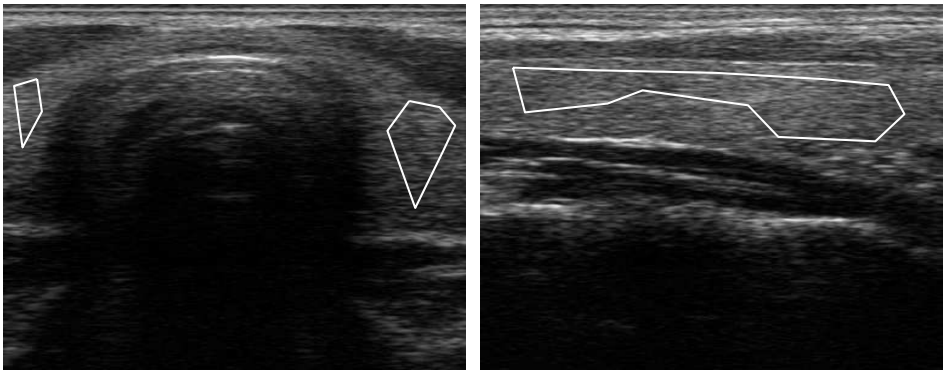
Jirak [50] tried to find texture features independent on used scanning tools. He found no texture parameters that can be reliably reproducible for all examined imagers. Although this was done for magnetic resonance imaging, we can deduce that this will be the same for sonographic imaging.

3 Previous Work and Results

3.1 Introduction

Hashimoto's lymphocytic thyroiditis (LT), one of the most frequent thyroid disorders, is a chronic inflammation of the thyroid gland. The incidence rate of LT is higher in women (0.35%) than in men. This disease changes the structure of the tissue. Changes are diffuse (they affect the entire gland) and can be detected by sonographic imaging. Images from the sonographic tool are used by physicians to determine diagnosis of the patient. Not all information present in the sonographic image is accessible to the naked eye. Information extracted from images by computers may provide additional support for diagnostic hypothesis. Automatic recognition of LT has been attempted based on textural image features [51, 52]. Classification was done with optimal features selected by a search procedure out of 129 features. The optimal features achieved sensitivity and specificity [1] of 100% in a cross-validation experiment on an independent set of 18 subjects [52].

[1] Sensitivity is the proportion of subjects with disease who have a positive test result, specificity is the proportion of subjects without disease who have negative test result.



(a) Transversal scan (b) Longitudinal scan

Figure 2: Segmentation by a different physician of the same images as in Fig. 3.

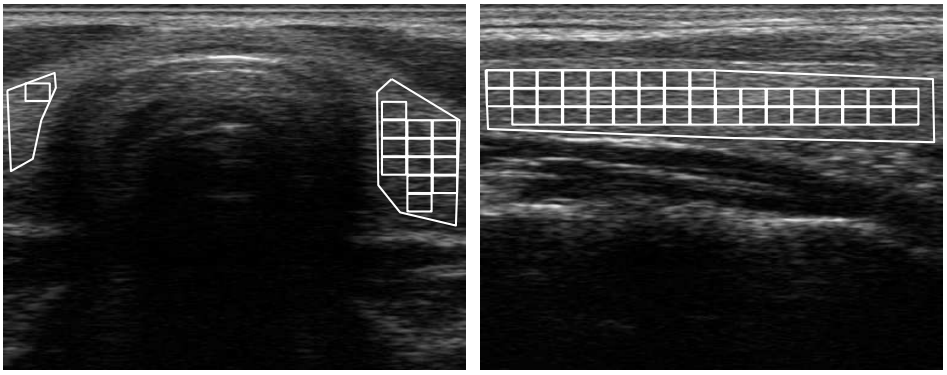
Although high success rate was achieved, the results were limited to one particular setting of the sonographic tool. This has been recognized as the most important obstacle to bringing the method to online clinical practice.

Our goal is to *quantify reproducibility* of optimal features used previously [52]. Reproducibility is the possibility to achieve the same classification results under different sonograph setting, different gland delineations in the manual segmentation step (depending on physician's knowledge and experience, see Fig. 2), and different scan type (longitudinal or transversal). The proposed analysis is general enough to be applied to other data interpretation problems involving complex and non-linear dependencies on variables that cannot be controlled.

Texture features are described and sensitivity analysis method is proposed in Sec. 3.2. Sec. 3.3 describes a feature sensitivity experiment and results. Discussion follows in Sec. 3.4.

3.2 Methods

Given a sonographic image, a two-class classification problem is considered in the previous work [51, 52]: distinguishing healthy tissue (denoted here as N) from tissue changed due to Hashimoto's Lymphocytic Thyroiditis (denoted as LT). The classification is done on textural features computed from a set of fixed-size rectangular regions referred to as texture samples, as shown in Fig. 3. The non-overlapping samples are obtained from a manually segmented thyroid gland. Automatic segmentation of thyroid gland in sonographic im-



(a) Transversal scan (b) Longitudinal scan

Figure 3: Sonographic images with manually segmented thyroid gland and covered by automatically selected rectangular texture samples.

ages is difficult and was not subject of this work. Optimally performing features from the set of 129 candidates including Haralick's texture features and Muzzolini's spatial features [35] were automatically searched for. Haralick texture features are given in Table 1, where C is the co-occurrence matrix, i denotes row in the matrix C, j denotes column in the matrix C, $m \times n$ is the texture window,

$$\mu_i = \sum_{i=1}^{n}\sum_{j=1}^{m} iC_{ij} \qquad \mu_j = \sum_{i=1}^{n}\sum_{j=1}^{m} jC_{ij} \qquad C_i = \sum_{j=1}^{m} C_{ij}$$

$$\mathrm{var}(i) = \sum_{i=1}^{n}\sum_{j=1}^{m}(i-\mu_i)^2 C_{ij} \qquad \mathrm{var}(j) = \sum_{i=1}^{n}\sum_{j=1}^{m}(j-\mu_j)^2 C_{ij}.$$

Muzzolini spatial features are based on gray-level transformations (denoted as t_i) obtained from each of texture samples. These transformations are gradient magnitude (t_1), difference from sample mean (t_2), horizontal curvature (t_3), vertical curvature (t_4) and original pixel gray levels (t_5). Histogramming and Kolmogorov-Smirnov distance between histograms of transformations $t_1, ..., t_5$ are used to compute features $F_1, ..., F_5$. The other features are computed as their combinations, e.g. features F_6 and F_7 are Euclidean distances from $F_1, ..., F_5$ to their mean and median (for more details see [52]).

Their performance was measured as Bayes classifier error. The classifier was learned on a training set of 81 patients and classification error was evaluated on an independent test set of 18 subjects. Three one-dimensional

Table 1: Haralick features.

Num.	**Name**	**Equation**
1	texture cluster tendency	$\sum_{ij}(i-\mu_i+j-\mu_j)^2 C_{ij}$
2	texture entropy	$-\sum_{ij} C_{ij}\log C_{ij}$
3	texture contrast	$\sum_{ij}\lvert i-j\rvert C_{ij}$
4	texture correlation	$\frac{\sum_{ij}(i-\mu_i)(j-\mu_j)C_{ij}}{\sqrt{\mathrm{var}(i)\mathrm{var}(j)}}$
5	texture homogeneity	$\sum_{ij}\frac{C_{ij}}{1+\lvert i-j\rvert}$
6	texture inverse difference moment	$\sum_{ij,i\neq j}\frac{C_{ij}}{\lvert i-j\rvert}$
7	maximum texture probability	$\max_{ij} C_{ij}$
8	texture probability of run length of 2	$\sum_{i}\frac{(C_i-C_{ii})^2 C_{ii}}{C_i^2}$
9	uniformity of texture energy	$\sum_{ij} C_{ij}^2$

optimal features turned up, each for a different optimal texture sample size, i.e. F2 for 41×41, F6 for 31×31 and F7 for 21×21 texture samples (the size is given in pixels). The optimal features achieved sensitivity and specificity of 100% in a cross-validation experiment on the independent set of 18 subjects. The principal parameters of the sonograph[2] were fixed in the study: the gain of 92, medium sensitivity by depth, maximal acoustic power, frequency of 8MHz, repetition rate of 19Hz, and maximum spatial resolution of 4cm. All details concerning data acquisition and processing are given in [52].

Features used in the current sensitivity analysis are the optimal features F2, F6 and F7, as described above. Feature probability distributions for each class were estimated by histogramming. Optimal (the least bias and variance) histogram resolution according to Scott's rule was used [53]. The idea of the sensitivity analysis is to quantify the changes of these histograms under various modifications of the data acquisition and processing pipeline that produced them.

A suitable statistic for this purpose is a divergence measure between two

[2]Toshiba ECCO-CEE, console model SSA-340A, transducer model PLF-805ST.

feature probability distributions. For this purpose, Kullback-Leibler distance (KL) is often used. Let X be the range of a discrete random variable and let p_1 and p_2 be two probability distributions over X. Kullback-Leibler distance is then defined as

$$KL(p_1, p_2) = \sum_{x \in X} p_1(x) \log \frac{p_1(x)}{p_2(x)}. \tag{1}$$

The KL distance is semi-definite, additive but not symmetric [54]. It is undefined if $p_2(x) = 0$ and $p_1(x) \neq 0$, hence to compute this divergence it is recommended to exclude such occurrences from the sum [55]. Divergence measure that does not require this fix is called Jensen-Shannon divergence (JS). Jensen-Shannon divergence is symmetric. These are the reasons we will use it in this work. The JS is defined in terms of discrete Shannon entropy $H(p)$ of a pdf p as

$$JS(p_1, p_2) = H(\frac{1}{2}(p_1 + p_2)) - \frac{H(p_1) + H(p_2)}{2}. \tag{2}$$

A detailed description is given in [56] and recommendations for practical usage in [57]. Comparison of KL and JS measures is given in [58].

The feature probability distribution (FPD) required in computing JS is estimated by the optimal histogramming. Using JS the sensitivity is measured by comparing the inter-class difference d_I in FPD and the within-class difference $d_{W|N}$ or $d_{W|LT}$ between feature probability distribution and the same distribution changed under a different 1) sonograph gain setting; 2) thyroid gland segmentation; and 3) scan type according to the following diagram

$$\begin{array}{ccc} & \overset{d_{W|N}}{\longleftrightarrow} & \\ F|N & & F'|N \\ d_I \updownarrow & & \\ & \underset{d_{W|LT}}{\longleftrightarrow} & \\ F|LT & & F'|LT \end{array} \tag{3}$$

where $F|N$ stands for 'probability distribution of feature F given class N' and $F'|N$ stands for the same under changed conditions. The inter-class distance d_I was measured on the full dataset (99 subjects) under the standard gain and the standard segmentation. The within-class distances $d_{W|N}$ were measured between FPD of the class N and FPD obtained from a set of class N images processed under modified conditions. Similarly for $d_{W|LT}$. The FPDs were computed from all longitudinal and transversal scans combined.

3.3 Experiments and Results

Sensitivity of the optimal features to the sonograph gain setting was determined by computing the distance $d_{W|N}$ between the N class FPDs under the standard gain of 92 and the respective distributions obtained from a set of images from one subject (class N) under two other gain settings (90, 94). The result was compared to the inter-class distance d_I.

A similar method was used to assess the sensitivity of optimal features to the thyroid gland segmentation. Three different boundary delineations for one subject of class N were drawn by another physician in addition to the one used in optimal feature selection.

Finally, the influence of the scan type on the optimal feature vectors was assessed. Given the optimal feature vector the distance between the longitudinal and transversal scans was measured in each class denoted here as $d^s_{W|N}$, $d^s_{W|LT}$, respectively. The larger of the two values $\max(d^s_{W|N}, d^s_{W|LT})$ was compared to the smaller of the two inter-class distances $\min(d_{I|\text{long}}, d_{I|\text{trans}})$ computed for each scan type separately according to the following diagram

$$\begin{array}{ccc} & \overset{d^s_{W|N}}{\longleftrightarrow} & \\ F|N,\text{long} & & F'|N,\text{trans} \\ d_{I|\text{long}} \updownarrow & & \updownarrow d_{I|\text{trans}} \\ & \underset{d^s_{W|LT}}{\longleftrightarrow} & \\ F|LT,\text{long} & & F'|LT,\text{trans} \end{array} \tag{4}$$

For instance, the $d_{I|\text{trans}}$ is the distance between N and LT class in transversal scans.

The results of the sensitivity analysis for *JS* are shown in Tab. 2. In 21×21 texture samples the differences due to varying gain setting are consistently smaller than the inter-class difference d_I. In 31×31 and 41×41 samples the differences shown in bold are already comparable to d_I.

Tab. 3 shows that changes due to the different segmentations s_1, s_2, s_3 are bigger than the inter-class difference d_I for all three features. Again, the least influenced are the 21×21 texture samples.

In Tab. 4 results for scan type (longitudinal, transversal) are shown. The inter-class distances (last two rows) for 31×31 and 41×41 samples are consistently greater than the inter-scan distances (first two rows). Only in the 21×21 samples the two values (in bold) are comparable.

The feature probability distributions for N and LT tissue used in this analysis are shown in Fig. 4, Fig. 5 and Fig. 6. We can see the *JS* distances capture the relative differences between the histograms well.

Table 2: The *JS* distances between optimal features under the gains of 90 and 94 and the same features under the standard gain of 92 according to Eq. (3). The last row shows inter-class distances.

	F7, 21×21	F6, 31×31	F2, 41×41
$d_{W\|N,\text{ gain=90}}$	0.031	**0.573**	**0.542**
$d_{W\|N,\text{ gain=94}}$	0.158	0.409	0.475
d_I	0.225	0.532	0.510

Table 3: The *JS* distances between optimal features computed under different thyroid gland segmentations s_1, s_2, s_3 and the same features under the standard segmentation according to Eq. (3).

	F7, 21×21	F6, 31×31	F2, 41×41
$d_{W\|N,s_1}$	0.208	**0.701**	**0.850**
$d_{W\|N,s_2}$	0.151	**0.859**	**0.888**
$d_{W\|N,s_3}$	**0.390**	**0.904**	**0.844**
d_I	0.225	0.532	0.510

Table 4: The *JS* distances between optimal features from individual scan types (first and second row) as compared to the inter-class distances in individual scan types (third and fourth row) according to Eq. (4).

	F7, 21×21	F6, 31×31	F2, 41×41
$d^s_{W\|N}$	0.033	0.042	0.021
$d^s_{W\|LT}$	**0.252**	0.092	0.014
$d_{I\|\text{trans}}$	0.478	0.276	0.409
$d_{I\|\text{long}}$	**0.184**	0.575	0.389

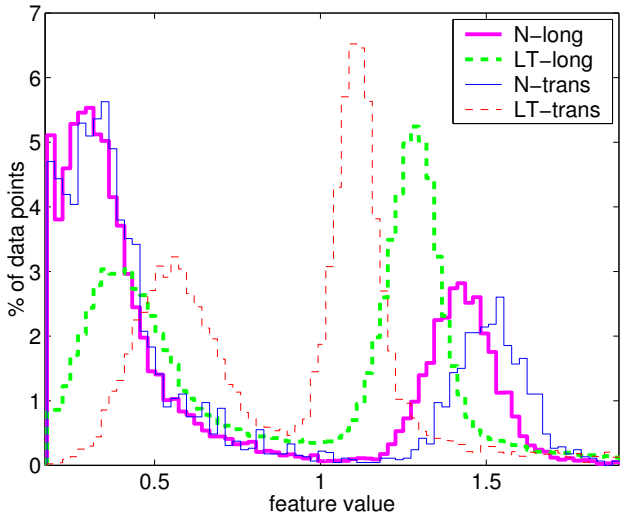


Figure 4: Histograms for N and LT tissue (feature F7, 21×21 texture samples). We can see histograms for longitudinal scans are more similar than for transversal scans, see also Tab. 4.

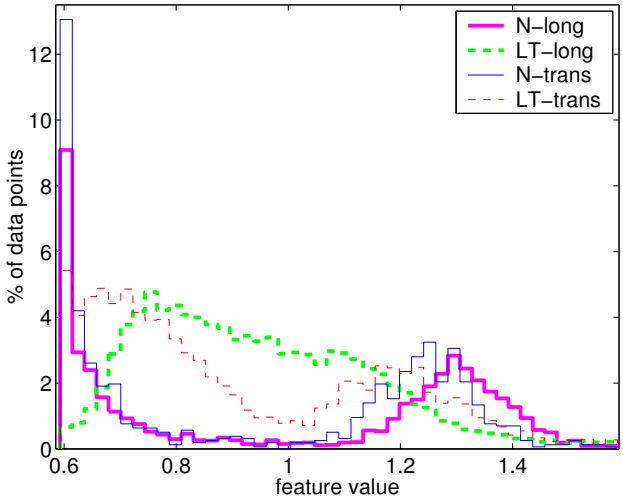


Figure 5: Histograms for N and LT tissue (feature F6, 31×31 texture samples).

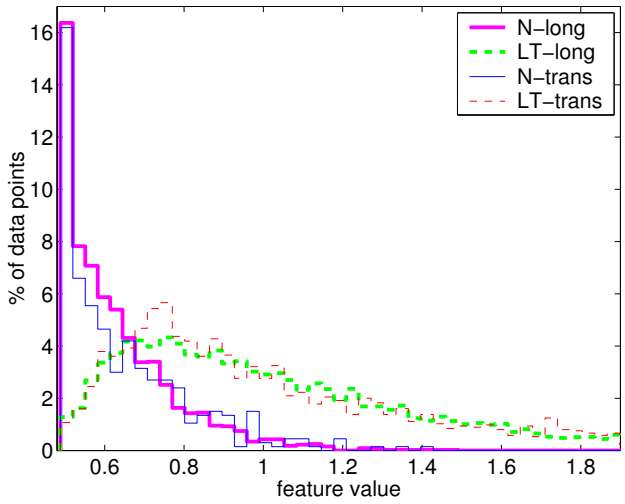


Figure 6: Histograms for N and LT tissue (feature F2, 41×41 texture samples).

Since gain setting is the parameter that has the greatest influence on visual appearance of sonographic image, features were extracted for a wider gain range (82–100). Measurements were done on a grey-scale phantom[3]. The scatter plots of feature $F7_{92}$ under gain 92 and $F7_g$ under several other gains g are shown in Fig. 7. One data point corresponds to one texture sample (21×21). It can be seen that points make clusters and the mapping induced by the gain change is too complex to be mathematically described. Next we considered simple feature F0 defined as the mean value over the whole rectangular texture sample. Analogical plots to those in Fig. 7 using F0 are shown in Fig. 8.

3.4 Discussion

Note that the image area of displayed thyroid tissue is much smaller in transversal scans than in longitudinal ones (see Fig. 3). This means that a smaller number of texture samples fit within the boundary of thyroid gland in a transversal scan as compared to a longitudinal scan. Larger texture samples do not cover the area of the transversal scans well. Therefore, substantial part of available information can be lost for feature construction process from transversal scans. Longitudinal scans provide greater amount of image data from a larger contiguous area of the gland tissue, therefore they should be more useful for automatic texture analysis. However, as can

[3]Precision Small Parts Grey Scale Phantom Gammex 404GS LE

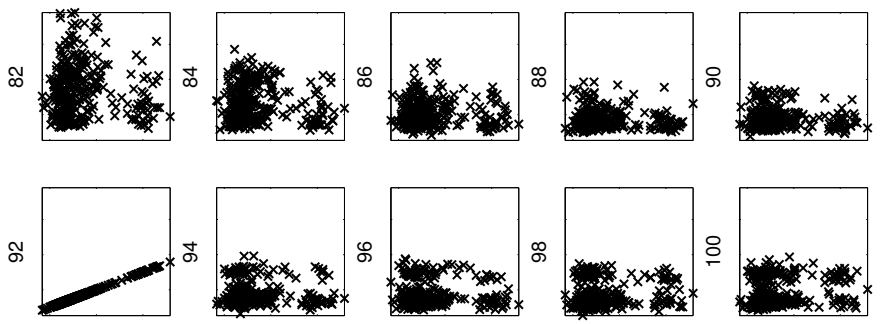


Figure 7: Comparison of feature F7 under standard gain with F7 under gains of 82 to 100. Coordinates of each point are feature values for standard gain (92, x-axis) and under gain change (82-100, y-axis). 21×21 texture samples are used

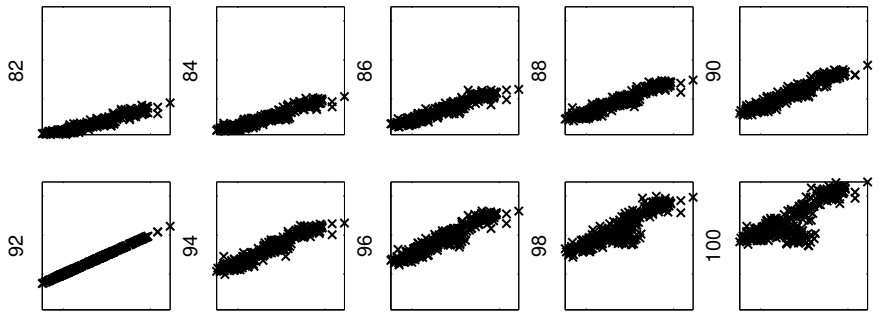


Figure 8: Comparison of feature F0 under standard gain with F0 under gains of 82 to 100. Coordinates of each point are feature values for standard gain (92, x-axis) and under gain change (82-100, y-axis). 21×21 texture samples are used

be seen from the last two rows of Tab. 4, distance between N and LT tissue is not always bigger for longitudinal scans than for transversal scans. This can be due to longitudinal artifacts in surrounding and examined tissue, e.g. muscle fibres or vessels. On the other hand we saw in Tab. 4 that inter-class distance is large in transversal scans when F7, 21×21 features are used and in longitudinal scans when F6, 31×31 features are used. Hence, the results could be improved by taking into account longitudinal and transversal images individually, e.g. by combining two classifiers, one using F7 on transversal scans and another using F6 on longitudinal scans.

There is high sensitivity to thyroid gland segmentation according to *JS* distance in larger samples (see Tab. 3). This can be related to sample placement method that leaves small areas along the boundaries uncovered by

samples.

To guarantee reproducibility of results under different gain settings, transformation to recalculate features from arbitrary gain to standard gain should be found. From the results shown in Figs. 7,8 it follows that direct transformation of complex features is unfeasible but re-mapping of the raw image values prior to feature computation seems feasible. The results on F0 (see Fig. 8) reveal two components: one approximately linear and another random (see the cluster just below the linear cluster). The origin of this cluster is not known. Further analysis is necessary.

An initial calibration (e.g., using a gray-scale phantom) and subsequent customization of the recognition tool could be another approach for solving the problem of reproducibility. The phantom would need to be specially designed to reproduce the statistical distribution of those features that were found to be optimal for the LT/N classification task. Whether this is feasible remains to be ascertained.

4 Summary, Tentative Plan for Future Work

So far we have examined different feature discriminability according to Fisher linear discriminant and classifier stability ([59], [60], [51], [61], [62]), feature compactness, separability criterion and minimal classification error evaluated on an independent validation set ([63], [52]), and deal with reproducibility of achieved results ([64], [65], [66]).

In present work, we concluded, that theoretically well-defined Haralick texture features are not the best discriminable and stable features for our classification task. Contrary to this, quite complicated Muzzolini's features resulted in 100% success in classification under one specific sonograph conditions.

The sensitivity analysis shows that the results for 31×31 texture samples and 41×41 texture samples are sensitive to small changes in sonograph setting. Both are also sensitive to different gland segmentations. They are stable under transversal and longitudinal scans.

The 21×21 pixel samples are insensitive to different gain settings and their sensitivity to different gland segmentations is small. They can also distinguish scan orientation, since there is a significant difference between inter-class distances of longitudinal and transversal scans. Distance between N and LT tissue is bigger for transversal than for longitudinal scans.

For greater difference in sonograph parameter setting it will be necessary to re-map raw image values by a corrective transformation. We believe that if features of small sensitivity are used subsequently, the classification results

will be reproducible. The corrective transformation is a topic for ongoing work.

To conclude PhD thesis, we would like to reach following:

- There are features that are inappropriate for our specific task (Haralick texture features). On the other hand, there are features that can discriminate substantial characteristics of given images (Muzzolini spatial features). Hence, selection of appropriate features in the sense of discriminability and classification success is the core for this and similar applications in pattern recognition. For our task and future work, it is important to understand characteristics that are described by these successful features and automatically generate a large class of similar simple ones.

- Feature selection defined in previous statement is the core of the pattern recognition view on the task. From the medical point of view, those features should be selected which will not change with the changed conditions of the image acquisition. In medical image applications, this is the frequent problem that prevents general solutions of similar tasks. Hence, beside feature selection we also have to find features insensitive to the similar changes, i.e. features for achieving reproducibility of the classification.

- There can be many features that can be beneficial for classification. However, texture in each image can slightly differ (e.g. due to small movements in transducer position) and each of the features can describe these differences diversely. From this point of view, the task is also to evaluate these features separately and combine individual feature classification to get better classifier. For this purpose, principle of some classifier combination schemes (e.g. AdaBoost) can be used ([67, 68, 69, 70, 71]). This would assure better classification results on previously unseen images, i.e. images that were not used to train classifier.

- Selected features and their classifier combination results can be combined with some additional information that need not to be related to the textural character of the image, however bears another important characteristic, e.g. depth of the ultrasound echo.

Solving previously formulated task would be beneficial in both, pattern recognition and medicine.

References

[1] P. Langley. Selection of relevant features in machine learning. In *AAAI Fall Symposium on Relevance*, pages 1–5, New Orleans, 1994. AAAI Press.

[2] A. Jain and D. Zongker. Feature selection: Evaluation, application and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, February 1997.

[3] I. Inza, P. Larrañaga, R. Etxeberria, and B. Sierra. Feature subset selection by Bayesian network-based optimization. *Artificial Intelligence*, 123:157–184, 2000.

[4] L. Portinale and L. Saitta. Feature selection. Technical report, Dipartimento di Informatica, Universita del Piemonte Orientale, Spalto Marengo 33, Alessandria 15100, Italy, April 2002.

[5] P. Somol, P. Pudil, and J. Kittler. Fast branch & bound algorithms for optimal feature selection. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 26(7):900–912, July 2004.

[6] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(7-8):1157–1182, November 2003.

[7] M. Bressan and J. Vitria. On the selection and classification of independent features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1312–1317, October 2003.

[8] M.H.C. Law, M.A.T. Figueiredo, and A.K. Jain. Simultaneous feature selection and clustering using mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1154–1166, September 2004.

[9] B. Krishnapuram, A.J. Harternink, L. Carin, and M.A.T. Figueiredo. A Bayesian approach to joint feature selection and classifier design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1105–1111, September 2004.

[10] C.C. Reyes-Aldasoro and A. Bhalerao. Volumetric texture description and discriminant feature selection for MRI. Simple feature selection is used: For each feature, Battacharyya distances betwenn every couple of

classes is computed. Features are ordered according to the sum of these distances., November 2002.

[11] J.G. Dy, C.E. Brodley, A. Kak, L.S. Broderick, and A.L. Aisen. Unsupervised feature selection applied to content-based reatrieval of lung images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):373–378, March 2003.

[12] M. Turceyan and A. K. Jain. *Handbook of Pattern Recognition and Computer Vision*, chapter Texture Analysis, pages 235–276. World Scientific Publishing Company, 1993.

[13] E. R. Davies. *Machine Vision: Theory, Algorithms, Practicalities*, chapter Texture, pages 561–581. Academic Press, 2nd edition, 1997.

[14] R.M. Haralick. Statistical and structural approaches to texture. In *Proceedings of the IEEE*, volume 67, pages 786–804, May 1979.

[15] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley Publishing Company, 1992.

[16] M. N. Shirazi, H. Noda, and N. Takao. Texture classification based on markov modeling in wavelet feature space. *Image and Vision Computing*, 18(12):967–973, September 2000.

[17] C.-C.J. Chang, T.; Kuo. Texture analysis and classification with tree-structured wavelet transform. In *IEEE Transactions on Image Processing*, volume 2, pages 429–441, Oct 1993.

[18] J. Laine, A.; Fan. Texture classification by wavelet packet signatures. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 15, pages 1186–1191, Nov 1993.

[19] H. C. Shen, C. Y. C. Bie, and D. K. Y. Chiu. A texture-based distance measure for classification. *Pattern Recognition*, 26(9):1429–1437, September 1993.

[20] I. Pitas and C. Kotropoulos. A texture-based approach to the segmentation of seismic images. *Pattern Recognition*, 25(9):929–945, September 1992.

[21] J. R. Sullins. Distributed learning of texture classification. In O. Faugeras, editor, *First European Conference on Computer Vision Proceedings*, pages 349–358. Springer-Verlag, April 1990.

[22] Higadhi T-Koichi I. Kakemura, A. Texture characteristic variables based on virtual volume. *Systems and Computers in Japan*, 29(6):38–48, June 1998.

[23] P. Kruizinga and N. Petkov. Nonlinear operator for oriented texture. *IEEE Transactions on Image Processing*, 8(10):1395–1407, October 1999.

[24] C. C. Chen, J. S. Daponte, and M. D. Fox. Fractal feature analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging*, 8(2):133–142, June 1989.

[25] J. S. Bleck, U. Ranft, M. Gebel, H. Hecker, M. Westhoff-Bleck, C. Thiesemann, S. Wagner, and M. Manns. Random field models in the textural analysis of ultrasonic images of the liver. *IEEE Transactions on Medical Imaging*, 15(6):796–801, December 1996.

[26] H. Sujana, S. Swarnamani, and S. Suresh. Application of artificial neural networks for the classification of liver lesions by image texture parameters. *Ultrasound in Medicine and Biology*, 22(9):1177–1181, 1996.

[27] M.-H. Horng, Y.-N. Sun, and X.-Z. Lin. Texture feature coding method for classification of liver sonography. In B. Buxton and R. Cipolla, editors, *Proceedings of Fourth European Conference on Computer Vision. ECCV '96*, volume 1, pages 209–218, Berlin, Germany, April 1996. Springer-Verlag.

[28] A. Mojsilovic, M. Popovic, and D. Sevic. Classification of the ultrasound liver images with the $2N$ multiplied by 1-D wavelet transform. In *Proceedings of the 1996 IEEE International Conference on Image Processing, ICIP'96*, volume 1, pages 367–370, Los Alamitos, CA, USA, September 1996.

[29] J.W. Jeong and D. Kim. The design of multi texture feature vector classifiers for the diagnosis of ultrasound liver images. In *Proceedings of the 1996 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 1147–1148, Piscataway, NJ, USA, October/November 1996. IEEE.

[30] W.-Ch. Yeh, S.-W. Huang, and P.-Ch. Li. Liver fibrosis grade classification with B-mode ultrasound. *Ultrasound in Medicine and Biology*, 29(9):1229–35, September 2003.

[31] R. Sutton and E. L. Hall. Texture meassures for automatic classification of pulmonary disease. *IEEE Transactions on Computers*, 21(7):667–676, July 1972.

[32] J. Segyeong, S.Y. Yoon, K.M. Woo, and Ch.K. Hee. Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. *EEE Transactions on Medical Imaging*, 23(10):1292–1300, October 2004.

[33] B. S. Garra, B. H. Krasner, S. C. Horii, S. Ascher, S. K. Mun, and R. K. Zeman. Improving the distinction between benign and malignant breast-lesions: The value of sonographic texture analysis. *Ultrasonic Imaging*, 15(4):267–285, October 1993.

[34] R. Pohle, L. von Rohden, and D. Fisher. Skeletal muscle sonography with texture analysis. In *Medical Imaging 1997: Image Processing*, volume 3034 of *Proceedings of the SPIE – The International Society for Optical Engineering*, pages 772–778, Newport Beach, CA, USA, February 1997. SPIE.

[35] R. Muzzolini, Y.-H. Yang, and R. Pierson. Texture characterization using robust statistics. *Pattern Recognition*, 27(1):119–134, 1994.

[36] O. Pujol, M. Rosales, P. Radeva, and E. Nofrerias-fernandez. Intravascular ultrasound images vessel characterization using adaboost. *Functional Imaging and Modeling of the Heart*, 2674:242–251, June 2003.

[37] X. Zhang, S.C. DeJong, C.R. McKay, S.M. Collins, and M. Sonka. Automated characterization of plaque composition from intravascular ultrasound images. In A. Murray and R. Arzbaecher, editors, *Computers in Cardiology 1996*, pages 649–652, New York, NY, USA, September 1996. IEEE.

[38] W.H. Nailon, S. McLaughlin, T. Spencer, and M.P. Ramo. Intravascular ultrasound image interpretation. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 3, pages 503–507, Comput. Soc. Press , Los Alamitos, CA, USA, August 1996. IEEE.

[39] W. H. Nailon, S. McLaughlin, T. Spencer, P. Ramo, D. M. Salter, G. R. Sutherland, K. A. Fox, and W. Norman McDicken. Fractal texture analysis: An aid to tissue characterization with intravascular ultrasound. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 534–537, Chicago, IL, USA, 1997. IEEE.

[40] W.H. Nailon, S. McLaughlin, T. Spencer, and M.P. Ramo. Comparative study of textural analysis techniques to characterise tissue from intravascular ultrasound. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 303–306, New York, NY, USA, September 1996. IEEE.

[41] K.J. Dixon, D.G. Vince, R.M. Cothren, and J.F. Cornhill. Characterization of coronary plaque in intravascular ultrasound using histological correlation. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering'*, volume 2, pages 530–533, IEEE , Piscataway, NJ, USA, October/November 1997. IEEE.

[42] S. Krass, R. Brennecke, T. Voigtlaender, R. Kuprat, P. Staehr, H.J. Rupprecht, A. Fisch, H. Darius, and J. Meyer. Tissue classification by texture and spectral analysis of intracoronary ultrasound radio-frequency data. In *Computers in Cardiology 1995*, pages 641–643, New York, NY, USA, 1995. IEEE.

[43] V. R. Amin, D. E. Wilson, R. Roberts, and G. Rouse. Tissue characterization for beef grading using texture analysis of ultrasonic images. In M. Levy and B.R. McAvoy, editors, *Proceedings of IEEE Ultrasonics Symposium*, volume 2, pages 969–972, New York, NY, USA, October/November 1993. IEEE.

[44] T. Hirning, I. Zuna, D. Schlaps, D. Lorenz, H. Meybier, C. Tschahargane, and G. van Kaick. Quantification and classification of echographics findings in the thyroid gland by computerized B-mode texture analysis. *European Journal of Radiology*, 9(4):244–247, November 1989.

[45] G. Mailloux, M. Bertrand, R. Stampfler, and S. Ethier. Computer analysis of echographic textures in Hashimoto disease of the thyroid. *Journal of Clinical Ultrasound*, 14(7):521–527, September 1986.

[46] U. Schiemann, R. Gellner, B. Riemann, G. Schierbaum, J. Menzel, W. Domschke, and K. Hengst. Standardized grey scale ultrasonography in Graves' disease: correlation to autoimmune activity. *Eur J Endocrinol*, 141(4):332–336, October 1999.

[47] A. B. M. Youssef, A. I. Badran, A. A. R. Sharawi, and A. G. Ahmad. Thyroid tissue characterization using computerized ultrasound b-mode. In *Biomedical Engineering Perspectives: Health Care Technologies for*

the 1990's and Beyond Proceedings of the Annual Conference on Engineering in Medicine and Biology, pages 352–353, IEEE Service Center, Piscataway, NJ, USA, 1990. IEEE.

[48] K.L. Chan. Adaptation of ultrasound image texture characterization parameters. In H.K. Chang and Y.T. Zhang, editors, *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 804–807, Piscataway, NJ, USA, October/November 1998. IEEE.

[49] S.D. Pye, S.R. Wild, and W.N. McDicken. Adaptive time gain compensation for ultrasonic imaging. *Ultrasound in Medicine and Biology*, 18(2):205–212, 1992.

[50] D. Jirak, M. Dezortova, and M. Hajek. Phantoms for texture analysis of MR images. long-term and multi-center study. *Medical Physics*, 31(3):616–622, February 2004.

[51] R. Šára, M. Švec, D. Smutek, P. Sucharda, and Š. Svačina. Texture analysis of sonographic images for diffusion processes classification in thyroid gland parenchyma. In Jiří Jan, Jiří Kozumplík, Ivo Provazník, and Zoltán Szabó, editors, *Proceedings Conference Analysis of Biomedical Signals and Images*, pages 210–212, Brno, Czech Republic, June 2000. VUTIUM.

[52] D. Smutek, R. Šára, P. Sucharda, T. Tjahjadi, and M. Švec. Image texture analysis of sonograms in chronic inflammations of thyroid gland. *Ultrasound in Medicine and Biology*, 29(11):1531–1543, November 2003.

[53] D.W. Scott. *Multivariate Density Estimation*. John Wiley, 1992.

[54] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Willey, 2001.

[55] A. Korhonen and Y. Krymolowski. On the robustness of entropy-based similarity measures in evaluation of subcategorization acquisition systems. In *Proc Conf Natural Language Learning*, 2002.

[56] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, Jan 1991.

[57] I. Dhillon, S. Manella, and R. Kumar. Information theoretic feature clustering for text classification. Technical Report TR–02–17, Dept of CS, U of Texas at Austin, TX USA, 2002.

[58] L. Lee. Measures of distributional similarity. In *Proceedings of the 37th Annual Meeting of the ACL*, pages 25–32, 1999.

[59] R. Švec, M.and Šára. Analýza textury sonografických obrazů difúzních procesů parenchymu štítné žlázy. Research Report CTU-CMP-1999-12, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, November 1999.

[60] R. Šára, M. Švec, D. Smutek, P. Sucharda, and Š. Svačina. Diffusion process classification in thyroid gland parenchyma based on texture analysis of sonographic images: Preliminary results. pages 45–47, 2000.

[61] D. Smutek, R. Šára, M. Švec, P. Sucharda, and Š. Svačina. Chronic inflammatory processes in thyroid gland: Texture analysis of sonographic images. In A. Hasman, B. Blobel, J. Dudeck, R. Engelbrecht, G. Gell, and H. U. Prokosch, editors, *Telematics in Health Care – Medical Infobahn for Europe, Proc eedings of the MIE2000/GMDS2000 Congress*, Berlin, Germany, August 2000. Quintessenz Verlag.

[62] D. Smutek, T. Tjahjadi, R. Šára, M. Švec, P. Sucharda, and Š. Svačina. Kvalitativní ukazatele ultrazvukového vyšetření štítné žlázy. *Diabetologie, Metabolismus, Endokrinologie, Výživa*, 3(2):16, December 2000.

[63] D. Smutek, T. Tjahjadi, R. Šára, M. Švec, P. Sucharda, and Š. Svačina. Image texture analysis of sonograms in chronic inflammations of thyroid gland. Research Report CTU–CMP–2001–15, Center for Machine Perception, K333 FEE Czech Technical University, Prague, Czech Republic, April 2001.

[64] M. Švec, R. Šára, and D. Smutek. Sensitivity analysis for reproducibility of ultrasound image classification. In Daniel Skočaj, editor, *Proceedings of the Computer Vision Winter Workshop 2004 (CVWW'04)*, pages 89–98, Ljubljana, Slovenia, February 2004. University of Ljubljana, Faculty of Computer and Information Science, Slovenian Pattern Recognition Society, Ljubljana.

[65] M. Švec and R. Šára. Citlivostní analýza přenositelnosti výsledků klasifikace ultrazvukových snímků. In Milan Chmelař, editor, *Medifórum 3.11.2004, sborník abstrakt*, pages 42–42, Kongresové centrum Brno, Výstaviště 1, 647 00, Brno, Czech Republic, November 2004.

[66] M. Švec, R. Šára, and D. Smutek. On reproducibility of ultrasound image classification. In *Proc. of the 2nd Iberian Conference on Pattern*

Recognition and Image Analysis. Spanish and Portuguese associations for pattern recognition, June 2005. Accepted.

[67] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518. IEEE Comput. Soc , Los Alamitos, CA, USA, December 2001.

[68] T. G. Dietterich. Ensemble methods in machine learning. In J. Kittler and F. Roli, editors, *First International Workshop on Multiple Classifier Systems*, Lecture Notes in Computer Science, pages 1–15, New York, 2000. Springer Verlag.

[69] Y. Freund and R.E. Schapire. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, September 1999.

[70] R.E. Schapire. The boosting approach to machine learning: An overview. In *MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[71] S.Z. Li and Z.Q. Zhang. FloatBoost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1112–1123, September 2004.