CENTER FOR
MACHINE PERCEPTION

CZECH TECHNICAL
UNIVERSITY

Master's Thesis

Analysis of Sonographic Images of Thyroid Gland Based on Texture Classification

Master's Thesis

Martin Švec

xsvecm@cmp.felk.cvut.cz

May 21, 2001

Available at
ftp://cmp.felk.cvut.cz/pub/cmp/users/svec/Svec-MSc01.pdf

Thesis Advisor: Dr. Ing. Radim Šára

This research was supported by the Internal Grant Agency of Ministry of Health of the Czech Republic grant NB 5472-3 and in part by Ministry of Education grant MSM210000012.

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

Acknowledgements

I want to thank my supervisor, Dr. Radim Šára, who initiated me into the project. His willingness, support, and time devoted to me were encouragement throughout working on this project.

I wish also to express my appreciation to MUDr. Daniel Smutek from the 1st Faculty of Medicine at Charles University in Prague. He spent a lot of time preparing dataset used in this thesis.

Prohlášení

Prohlašuji, že jsem na řešení diplomové práce pracoval samostatně s pomocí vedoucího práce a že jsem nepoužil jinou literaturu než je uvedena v seznamu. Zároveň prohlašuji, že nemám námitek proti využití výsledků této práce katedrou kybernetiky fakulty elektrotechnické ČVUT.

Praha, 21.května 2001 Martin Švec

Anotace

Sonografie je v medicíně rozšířena vzhledem ke své rychlosti a možnosti neinvazivní aplikace. Diagnostika ze sonografických snímků je v současné praxi prováděna výhradně expertem. Lidský vizuální systém však nemusí zachytit veškerou informaci důležitou z hlediska rozpoznání onemocnění. Naším cílem je ověřit možnost automatické klasifikace ze sonografických snímků jako pomocné metody při diagnostice štítné žlázy. Uvažujeme dvě třídy: normální tkáň a chronickou lymfocytickou thyroiditidu (Hašimotovu thyroiditidu). Provedli jsme klasifikaci dat reprezentovaných statistickými texturními příznaky: histogramy a Haralickovými příznaky. Použili jsme klasifikátor podle K nejbližších sousedů. Náš závěr je, že struktura dat reprezentovaná Haralickovými příznaky není vhodná k rozlišení zdravé tkáně a Hašimotovy thyroiditidy, na rozdíl od dat reprezentovaných histogramy.

Abstract

Classification from sonographic images of thyroid gland is tackled in semi-automatic way. While making manual diagnosis from images, some relevant information need not to be recognized by human visual system. Quantitative image analysis could be helpful to manual diagnostic process so far done by physician. Two classes are considered: normal tissue and chronic lymphocytic thyroiditis (Hashimoto's Thyroiditis). Data are represented by Haralick features and 1-dimensional histograms. Data structure is analyzed using K-nearest-neighbour classification. Conclusion of this thesis is that unlike the histograms, Haralick features are not appropriate to distinguish between normal tissue and Hashimoto's thyroiditis.

Contents

Chapter 1

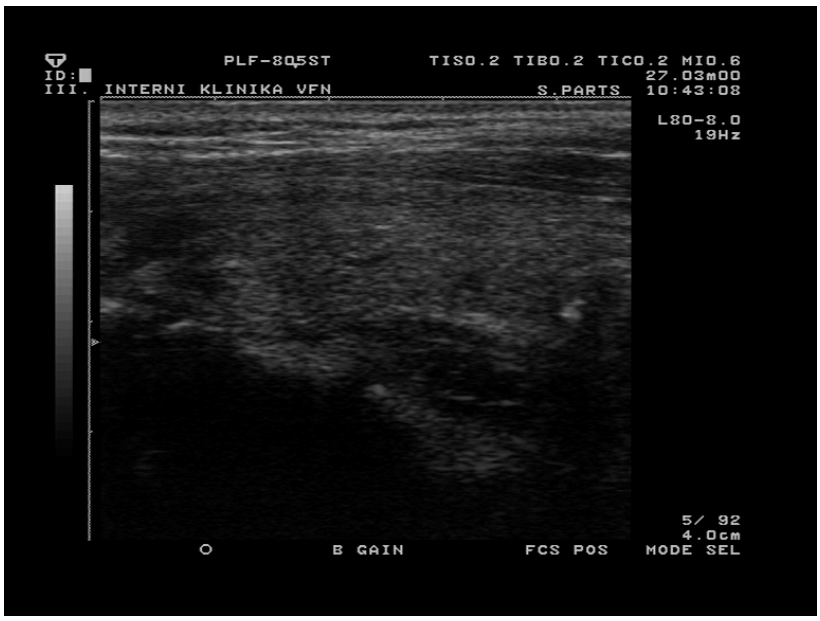
Introduction

This work is a part of the three-year project *Texture Analysis of Sonographic Images for Endocrinopaties and Metabolic Diseases* concluded in collaboration between the Center for Machine Perception at Czech Technical University in Prague and the 1st Faculty of Medicine at Charles University in Prague. I am active participating in this project since its beginning in 1999. Summary of important results achieved in the project is given inside the thesis. New experiments are performed and results discussed.

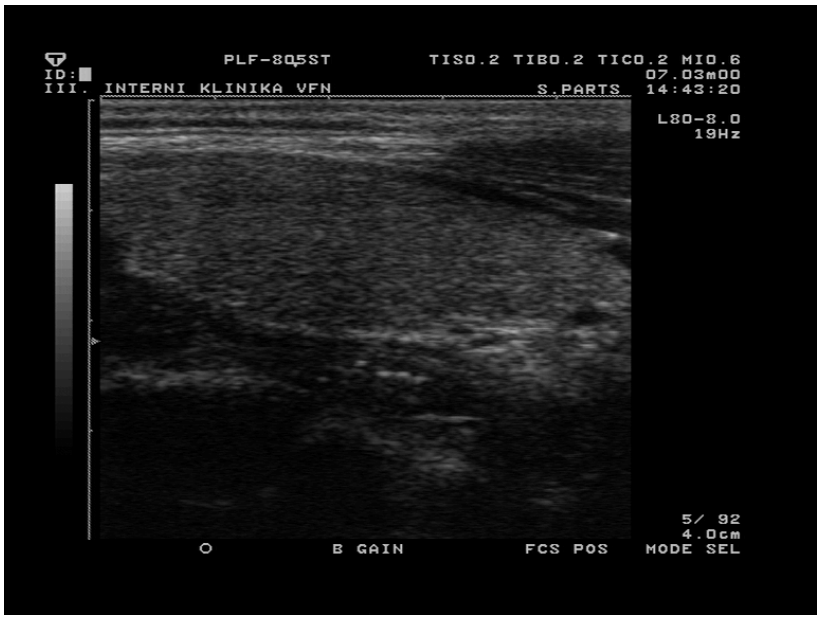
1.1 Goals of the Work

This work deals with computer aided diagnosis of thyroid gland by methods of automatic recognition. Sonographic images with two different tissues of the thyroid gland are processed and analyzed in a semi-automatic way. These two kinds of tissue are normal tissue and diffusely inflamed tissue (chronic lymphocytic thyroiditis – Hashimoto's Thyroiditis, this disease will be in the rest of this thesis called LT) (see Figure 1.1). So far, manual diagnosis is done by a physician. The physician focuses on the textural character of image. This character is influenced by echogenicity and structure of the thyroid parenchyma. Hence texture characteristics carry relevant information about these images. Our task is to classify texture in images where the location of the thyroid gland is segmented out manually.

Image of size $n \times n$ consists of n^2 pixels and can be represented as a point in the n^2-dimensional space. Our aim is to reduce the dimension of this space to the least possible one and make diagnosis in this feature space. Such dimension should fulfill the following requirements: (i) information lost during reduction should be as small as possible, (ii) resulting space should provide sufficient information for certain purpose. This purpose is to recog-



a) Normal tissue.



b) Tissue with lymphocytic thyroiditis (LT).

Figure 1.1: Sonographic images of the thyroid gland.

nize two different diagnoses (normal tissue and inflamed tissue) in this low-dimensional space. These requirements can be met by using formal methods of pattern recognition. Coordinates of the low-dimensional space are called features. For this purpose, methods for feature extraction are used. Methods that provide distinguishing between different kinds of images are called classification methods. More detailed description of methods that provide statistically based results will be given in the subsequent chapters.

1.2 Motivation

Pathology of the thyroid gland has followed generations of people since the earliest period of human existence. It is already mentioned by ancient physicians from China, India, and Egypt several thousand years before Christ. Recently, people are stricken with thyroid gland diseases more frequently than with diabetes (about 900 million people suffer only from diseases caused by lack of iodine). The first important step on the way towards health is a successful diagnosis.

A large number of disease processes influence human body tissue in such a manner as to produce abnormalities. These abnormalities are detectable in images produced by sonography or another imaging technique. Sonography is simple non-invasive diagnostic method and it is one of the most applied imaging techniques. Physician can observe the state of human organs at any time it is necessary. Diagnosis is based on physician's knowledge and experience. The character of sonographic images is textural (it is discussed in Section 1.3) and one can qualitatively characterize texture as having such properties as fineness, coarseness, smoothness, granulation, randomness, lineation or as being mottled or irregular. In case of thyroid gland diseases this qualitative approach is often combined with invasive needle biopsy. It often causes heavy burden and stress for patients. Especially when this process is done repeatedly to evaluate progress of the illness or changes inside a given organ. This motivates us to look for a method that would extract quantitative description in an automatic way. Such method would be able to replace the invasive procedures annoying for patients. Moreover, description obtained by computer need not be observed by human vision and hence can give more information than just visual inspection.

1.2.1 Thyroid Gland

The thyroid (Figure 1.2) is a butterfly-shaped gland which wraps around the front part of the trachea (windpipe) just below the Adam's apple. Location

of the thyroid gland with respect to other human systems can be seen in Figure 1.3.

The thyroid produces hormones that influence essentially every organ, every tissue and every cell in the body. Thyroid hormones regulate the body's metabolism and organ function. The most common thyroid disorder

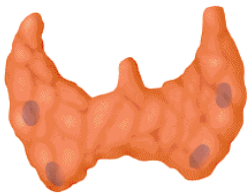
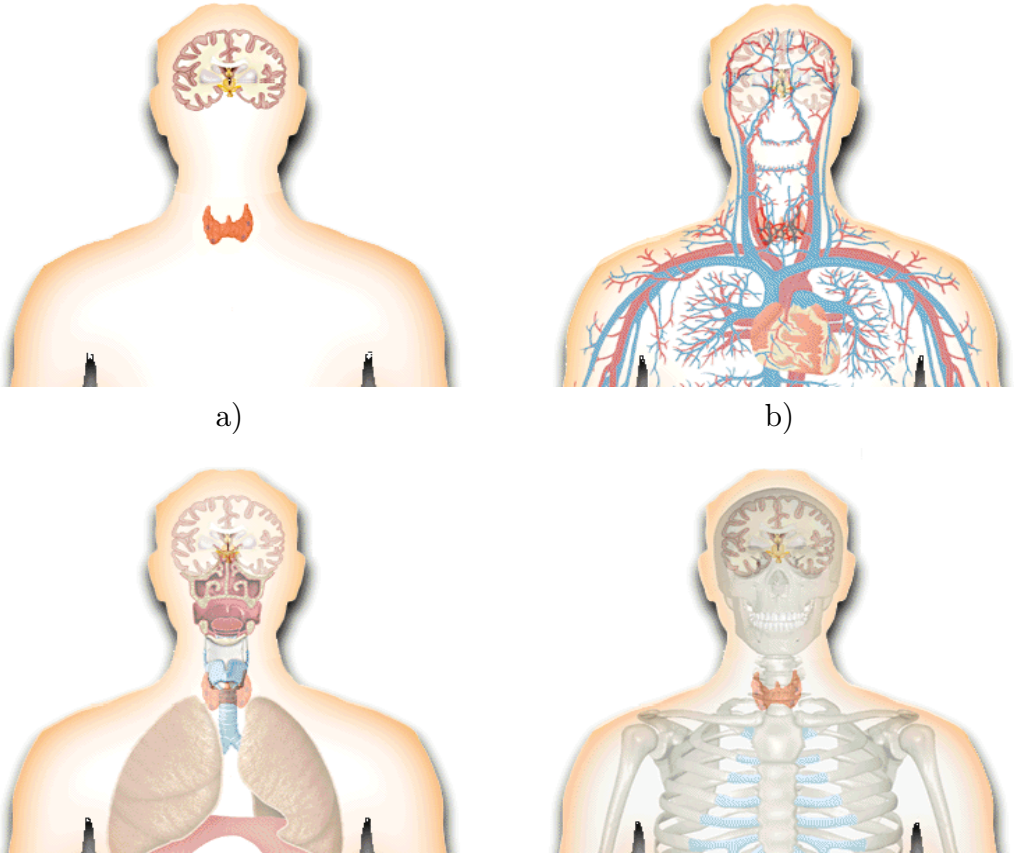


Figure 1.2: Detail of the thyroid gland.



a) b)

c) d)

Figure 1.3: Location of the thyroid gland with respect to different systems.

results from an underactive thyroid gland, or hypothyroidism. This is when the thyroid fails to produce enough hormones. Less frequently, an overactive thyroid condition, or hyperthyroidism, occurs when the thyroid produces more thyroid hormone than is needed.

Hashimoto's disease is the state when function of the thyroid tissue is initially unchanged, but it can not make enough thyroid hormone after certain time. This can result in hypothyroidism. It is named after the Japanese doctor who first described it and it is also called Hashimoto's thyroiditis, chronic autoimmune thyroiditis, or lymphocytic thyroiditis (LT). The disease is five times more common in women than in men. Hypothyroidism caused by Hashimoto's disease results in an overall slowing down of body's functions. A heart may beat more slowly, body temperature may decrease, muscles may weaken, cholesterol level may rise, and one may have difficulty thinking and remembering. In time, this overall slowing down affects most of body's functions, and can seriously affect health. Therefore, it is very important to identify hypothyroidism as early as possible and treat it properly. Human immune system mistakenly identifies stricken thyroid gland as a group of foreign cells and produces antibodies against the thyroid cells. The presence of thyroid antibodies in the blood is in most cases indication of Hashimoto's disease. It can be also recognized by physician from sonographic images. Hypothyroidism is treated with thyroid hormone replacement therapy.

The thyroid gland is small and can be seen in its entirety by some transducers, but it is still evaluated by viewing the lobes individually. Position of the transducer due to the thyroid can be seen in Figure 1.4.

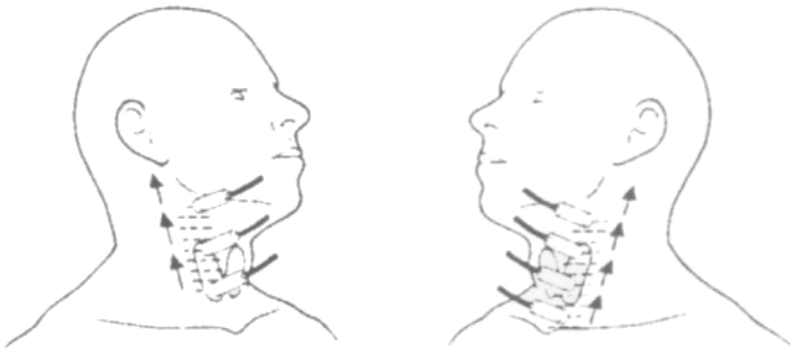


Figure 1.4: Longitudinal section scanning of the thyroid gland.

1.3 Texture Definition

We have shown in [1, 2, 3] that sonographic images of the thyroid gland can be regarded as textures. Texture is an important characteristic for the analysis of medical images. Despite its importance, common precise and satisfactory definition of texture does not exists. Researchers attempted to formulate many different texture definitions. Turceyan and Jain [4] gave examples and we mention here one of them.

> "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 [5] 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.

Generally, image texture can be defined as a function of the spatial variation in pixel intensities.

1.4 State of the Art

An overview of related work is given in this section. First, applications with texture analysis are mentioned in general, then the use of texture analysis for medical purposes is overviewed and after that published works dealing with the thyroid gland diagnosis are given. Previous results of our project are mentioned in the next section (Section 1.5).

Texture For microtextures, the statistical approach seems to work well. These approaches have included autocorrelation functions, optical transforms, digital transforms, textural edgeness, structural element, gray tone co-occurrence, and autoregressive models. For macrotextures, researchers are using histograms and co-occurrence of primitive properties. Note that if an image contains microtexture, the whole image can be considered as a primitive and characterized by its histogram as well. In this way, whether to use statistical or structural approach to texture depends on the point of view of the researcher. We can say that texture discrimination techniques are for the most part ad hoc and many features are constructed by intuition and an *a priori* knowledge about the underlying geometry of the texture.

General Texture Recognition Many approaches exist to characterize texture, some of them are overviewed by Haralick et al. and Turceyan et al. [4, 6, 7]. Early work 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 et al. in 1973. It became the "standard" approach to texture analysis. This approach will be mentioned later in this thesis.

Shirazi et al. [8] used texture classification methodology that was based on stochastic modeling of textures in the wavelet domain. Chang and Kuo [9] and Laine and Fan [10] also dealt with a multiresolution approach based on wavelet transform. Shen et al. [11] 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 [12] dealt with segmentation of seismic images by Hilbert transform, minimum entropy learning technique, and by region growing algorithm. Sullins [13] 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 et al. [14] 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 et al. [15], 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.

Recent studies suggest to combine several approaches. Kittler et al. [16] showed, that combining classifiers is more successful than using only one (under certain conditions). Similarly, it seems to be possible to combine different texture features. Features can be also used with more statistical approach used by Kleinberg's [17] stochastic discrimination. Several methods like bagging and boosting are overviewed by Dietterich in [18].

Texture Recognition in Sonography Texture analysis in medicine is widely used for diagnostic purposes in non-invasive methods. Pohle [19] 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 the given task. Muzzolini used similar method in [20]. Sutton and Hall [21] dealt with automated screening of chest radiographs 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. Uppaluri et al. [22] described a method for evaluating pulmonary parenchyma from computed tomography images. Their method incorporates multiple statistical and fractal texture features. Chen et al. [23] 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 as possible. Several authors dealt with textural analysis of ultrasonic images of the liver. Bleck et al. [24] used autoregressive periodic random field models to distinguish between patients without and with microfocal lesions of the liver. Sujana et al. [25] 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 et al. [26] 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 et al. [27] used wavelet decomposition to detect liver cirrhosis in its early stage. They achieved accuracy of 92%.

Texture Recognition in Thyroid Sonography Few studies focused on the thyroid gland. Hirning [28] analyzed computerized B-mode images. Two features from a set of 109 features proved to be the most efficient statistical parameters:

1. the upper decile of grey level distribution. It allowed to classify normal tissue and cyst from other diagnosis with 100% success,

2. the entropy distinguished cyst from the rest also 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 et al. [29] 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 deduced on 10 patients, which does not seem to be a sufficiently large set. Schiemann [30] used grey scale histogram analysis to show that tissue with Graves' disease has significantly lower echogenicity than normal tissue.

Our results of classifying between normal tissue of the thyroid gland and Hashimoto's thyroiditis were summarized in several articles. Švec and Šára [1, 2] used Haralick texture features and classifier based on the minimum distance from mean value. The most descriptive features were texture entropy, texture correlation, and texture probability of run length of 2. Smutek et al. [3] tried to use subsets of Haralick features combined with features proposed by Muzzolini et al [20]. Šára et al. [31] used systematic feature construction based on the minimization of the conditional entropy of class label. Classification of subjects was done with success between 85% and 96.6% for different features and experiments previously mentioned. Toufik [32] based

classification on histograms with success between 91% and 94.5%. Contributions of our previous work are summarized in the next section.

1.5 Previous Work on the Project

Project "Texture Analysis of Sonographic Images for Endocrinopaties and Metabolic Diseases" has started in 1999. So far we dealt with following tasks:

1. Can sonographic image be characterized by texture? It was shown in [33] that texture distribution is different in different areas in sonographic image. Hence texture can be descriptive for these images.

2. Is lymphocytic thyroiditis (LT) linearly separable from normal tissue? These two classes are partly overlapping and hence not linearly separable [1, 2, 34, 35, 36].

3. The task of feature selection from the set of features was dealt in [2, 3, 36]. For given dataset, three features were denoted as the most descriptive. This will be discussed in subsequent chapters.

4. Systematic feature construction was reported in [31, 37], to generate simplest texture features that are most efficient in distinguishing between normal and LT tissue.

This thesis is based on the results of the work given above. Our aim is to use previous knowledge on extended dataset. The old dataset consisted of 4 patients (subjects). The current dataset has 71 subjects. We also focus on finding appropriate probability density that would fit data represented by features. In our previous work we confirmed that probability density functions on texture features are not separable by simple curves. We therefore need probability density estimates that are as accurate as possible on the class overlap. In this thesis we focus on non-parametric method, the K-nearest-neighbour.

Chapter 2

Data

In this chapter description of the data processing is given. It starts by data capturing with ultrasonographical tool (Section 2.1) through the preprocessing (Section 2.2) to a deeper insight into the data character (Section 1.3).

2.1 Data Acquisition

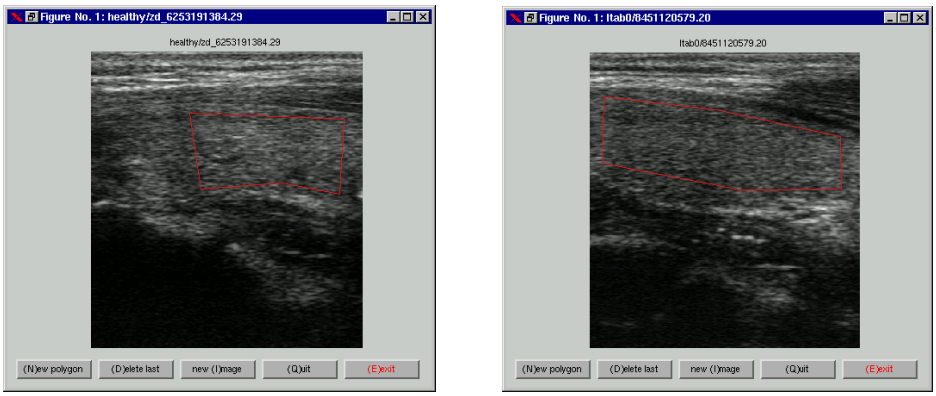
Sonographic images are digitized data from ultrasonographical imaging system Toshiba ECCOCEE (console model SSA-340A, transducer model PLF-805ST at frequency 8MHz). Data are captured in longitudinal cross-sections of both lobes with magnification of 4cm and stored with amplitude resolution of 8 bits (256 grey levels). Examples of these images can be seen in Figure 1.1. There are two kinds of images: image with normal tissue and image with lymphocytic thyroiditis. Real time imaging of the patient's thyroid gland is observed by physician on the monitor of the imaging system while making diagnosis. Since the number of frames observed may play a role in diagnosis, our data consists of approximately 20 images per subject. The overall number of data can be seen in Table 2.1.

Table 2.1: Number of data.

Diagnoses	**N. of subjects**	**N. of images**
normal tissue	33	651
LT tissue	38	754
$\sum$	71	1405

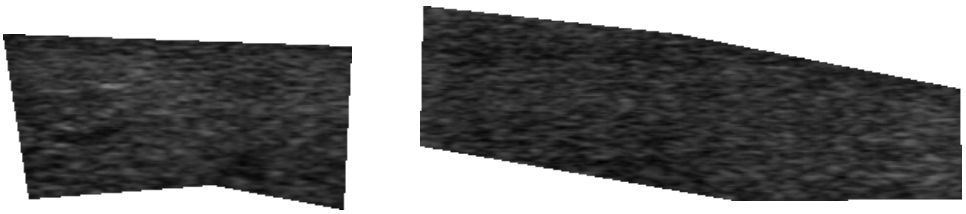
2.2 Data Preprocessing

Images in Figure 1.1 contain tissue of thyroid gland surrounded by other neighbouring tissues. For our purpose we need to have regions with the gland segmented out of the image. Since automatic segmentation is not the aim of this project, the boundary of the thyroid gland is roughly delineated by a physician. For this purpose, simple-to-use interactive tool was implemented (see Figure 2.1). Examples of images with segmented regions of interest (RoI) can be seen in Figure 2.2. In the rest of this thesis, we always consider such segmented images.



a) Normal tissue. b) LT tissue.

Figure 2.1: Expert-drawn boundary of the region of interest.



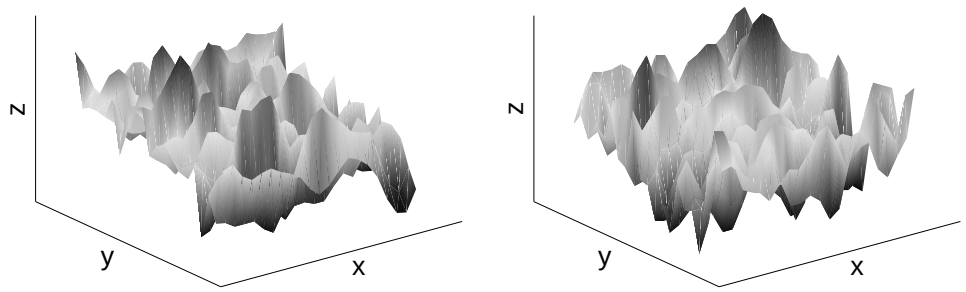
a) Normal tissue. b) LT tissue.

Figure 2.2: Manually segmented regions of interest. All features mentioned in the following chapters will be computed from these RoI's.

2.3 Character of Sonographic Image

For deeper insight into our data, texels (defined in Section 1.3) can be viewed more clearly in 3D space. If we regard the pixel intensity as the height above a plane, the intensity surface of a medical image can be viewed as a rugged surface. Considering image in the coordinate system x, y we represent values of pixels in the z direction. Resulted surface is shown in Figure 2.3. We can say that texel's size in thyroid gland images is microscopic at a level of single pixels or small groups. Arrangement of these texels can be hardly caught by a non-trained eye. Some larger ones can randomly appear in both kinds of sonographic images, so that we can not say whether these texels form the texture and make it distinguishable from another one. Hence this texture can not be analyzed on the structural level since it is not clearly composed of texels. According to that, statistical approach is used.

From the definition of the task, we deal with the texture classification problem. For this purpose, we need to extract statistical texture features (Chapter 3) and to select appropriate classifier (Chapter 4).



a) Normal tissue. b) LT tissue.

Figure 2.3: Texture sample (40×40 pixels) from sonographic image (Figure 2.2) shown in 3D space. Pixel intensity is in z-axis direction.

Chapter 3

Texture Feature Extraction

Sonographical texture of the thyroid gland is analyzed on the statistical level. It means that local features are computed independently at each texture image pixel, and a set of statistics is derived from the distributions of the local features. On this level, features based on statistical distributions of single image pixel values are called first-order statistics. Features that take into account spatial distribution of two or more pixels are called second-order and higher-order statistics. Section 3.1 is devoted to histograms as first-order statistics. Haralick texture features as second-order statistics and their computation is introduced in Section 3.2 and Section 3.3. These features form low-dimensional subspace, which is mentioned in Section 1.1. Measure of quality of such representation can be provided by Fisher linear discriminant (Section 3.4).

3.1 Histograms

Image histogram is a distribution formed by the simplest features: individual pixels. It is obtained simply by dividing the intensity axis into a number of bins (B) and approximating the density at each value of intensity by the fraction of the points which fall inside the corresponding bin. Let n_i ($i = 1, \ldots, B$) be the number of pixels that falls into bin i. N is the total number of pixels. Then distribution

$$h(i) = \frac{n_i}{N}$$

takes on the form of a histogram. The number of bins B (more precisely the bin width) plays a crucial role in image representation. It can be considered as a smoothing parameter. When B is too small, histogram can be very spiky, while if its value is too large, important information can be smoothed out.

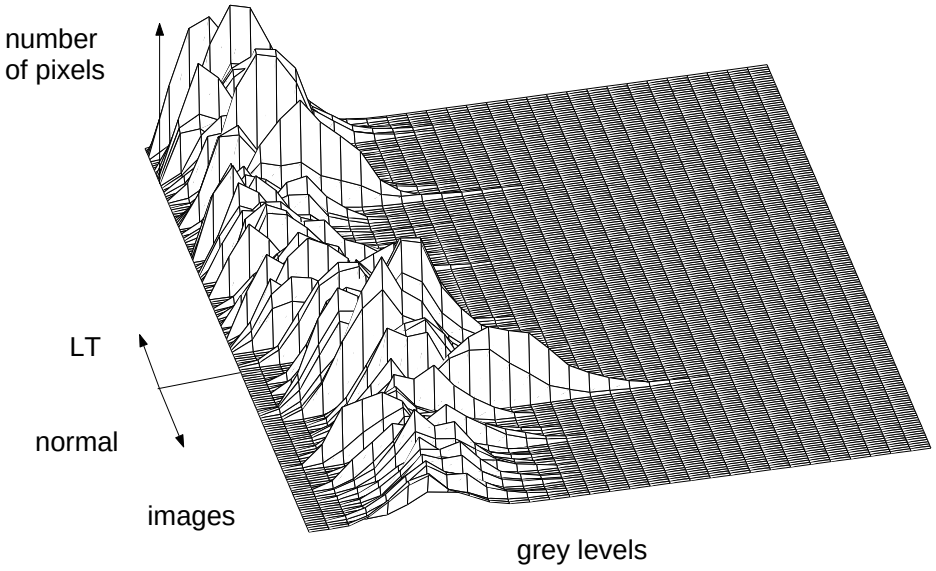


Figure 3.1: Histograms of data shown in 3D.

Our data can be partly overviewed by histograms pictured in 3D as can be seen in Figure 3.1, where we consider 256 grey levels (B).

3.2 Haralick Features

Statistical characteristics can be extracted from textural image by nine Haralick features [7]. Features are computed on co-occurrence matrix. It is a matrix of relative frequencies C_{ij} with which two neighbouring pixels separated by distance vector d occur on the image, one with gray level i and the other with gray level j. Such matrices are a function of the angular relationship between the neighbouring pixels as well as a function of the distance between them. The problem of choosing appropriate distance vector d is overviewed in Section 3.3. Haralick features are given in Table 3.1, 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,

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

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

To achieve data representation in low-dimensional space, subsets from the feature set can be chosen by using Fisher linear discriminant (deeper insight into this method is given in Section 3.4).

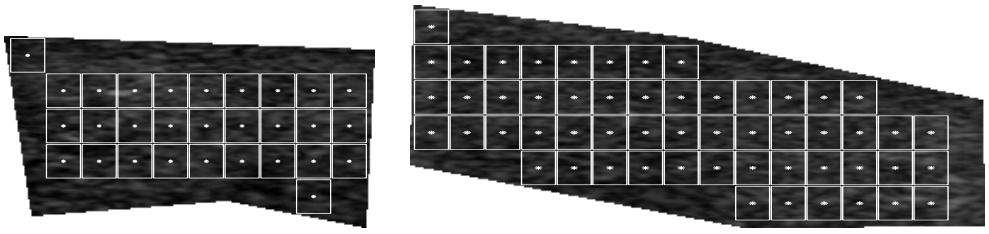
After manual segmentation, texture samples for computing Haralick features were defined as 21×21 rectangular windows inside ROI's. This avoids influence of different shapes and sizes of ROI's on classification. Hence we obtained set of Haralick features for one image. Number of windows (Haralick features) for different classes is given in Table 3.2 (number of histogram is also given, one histogram for each image). Example of covering ROI's by samples can be seen in Figure 3.2.

Table 3.1: Haralick features.

Num.	**Name**	**Equation**
H1	texture cluster tendency	$\sum_{ij}(i-\mu_i+j-\mu_j)^2 C_{ij}$
H2	texture entropy	$-\sum_{ij} C_{ij}\log C_{ij}$
H3	texture contrast	$\sum_{ij} \lvert i-j\rvert C_{ij}$
H4	texture correlation	$\frac{\sum_{ij}(i-\mu_i)(j-\mu_j)C_{ij}}{\sqrt{\mathrm{var}(i)\mathrm{var}(j)}}$
H5	texture homogeneity	$\sum_{ij}\frac{C_{ij}}{1+\lvert i-j\rvert}$
H6	texture inverse difference moment	$\sum_{ij,i\neq j}\frac{C_{ij}}{\lvert i-j\rvert}$
H7	maximum texture probability	$\max_{ij} C_{ij}$
H8	texture probability of run length of 2	$\sum_{i}\frac{(C_i-C_{ii})^2 C_{ii}}{C_i^2}$
H9	uniformity of texture energy	$\sum_{ij} C_{ij}^2$

Table 3.2: Number of Haralick features and histograms.

Diagnoses	N. of subjects	N. of Har. features	N. of hist.
normal tissue	33	18609	651
LT tissue	38	27866	754
$\sum$	71	46475	1405



a) Normal tissue. b) LT tissue.

Figure 3.2: Rectangular windows for computing Haralick features.

3.3 Feature Construction

Systematic feature construction is a procedure suggested by Šára in [31]. It aims at finding a systematic way to generate simplest texture features that are most efficient in distinguishing between normal and LT tissue. The simplest features can be considered as individual image pixels. Let us focus on features done by couples of pixels. The couple of pixels is defined by two pixels separated by distance vector d. Finding appropriate d can be done by computing conditional entropy. Let L be class label variable[1] and $\mathbf{X}$ be a matrix of features created as couples of pixels. Conditional entropy $H(L|\mathbf{X})$ tells us how much information in bits is missing in all data about classes:

$$H(L|\mathbf{X}) = -\sum_{i=1}^{n} p(L, \mathbf{X}) \log p(L|\mathbf{X}), \tag{3.1}$$

where n is the number of features, $p(\cdot)$ is probability and $log(\cdot)$ is the dyadic logarithm. If $H(L|\mathbf{X}) = 0$ classes can be determined by the data, i.e. there exists some (unknown) one-to-one function f such that $L = f(\mathbf{X})$. If $H(L|\mathbf{X}) = H(L)$ the data contain no information about classes. By evaluating conditional entropy for all separation vectors d we can find d for which (3.1) is the smallest. Šára showed that conditional entropy achieves small values for positioning vector d in vertical direction. It is obvious from

[1]Classes may be assigned their numbers, e.g. $0, 1, \ldots$, but this is not strictly necessary for computing entropies.

the fact that it is the principal direction in the sonographic image, the direction in which ultrasonic wave propagates through the tissue. From the searched interval of $0 \dots 30$, distance vector was determined as [11,0]. This is the distance vector d that should be used to compute co-occurrence matrix C for Haralick features.

3.4 Fisher Linear Discriminant

Texture is a complicated entity to measure. The reason is that many parameters (features) are likely to be required to characterize it. In addition, when so many features are involved, it is difficult to decide the ones that are most relevant for recognition. Fisher linear discriminant is a method that provides a measure of information about classes represented by features. The principle of the method can be shown in a feature space. There are clusters of points belonging to different classes. Fisher linear discriminant assumes that each cluster can be represented by its mean value and variance (covariance matrix for more than 1-dimensional feature space). The smaller variance inside clusters and higher distances between them, the more appropriate the features are. Therefore, good features are those for which

$$\frac{\text{inter-class variance}}{\text{intra-class variance}}$$

is higher than for others. Since the variance inside the individual classes can be different we can use following:

$$\frac{\text{variance between classes}}{\text{higher of the intra-class variance}} .$$

Variance inside classes was not computed as covariance of mean values [38], but as covariance of all feature points, which is more precise in our case. Generally, suppose we have data in n classes. Each class is represented by matrix $\mathbf{X}_i$ where columns are feature vectors.
Suppose the following:

$$\mathbf{A} = \mathrm{cov}([\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3 \dots \mathbf{X}_n]) , \tag{3.2}$$

$$\mathbf{B}_i = \mathrm{cov}(\mathbf{X}_i) , \tag{3.3}$$

$$\lambda_i = \max(\mathrm{eig}(\mathbf{B}_i^{-1}\mathbf{A})) . \tag{3.4}$$

Then Fisher linear discriminant is $F = \sqrt{\min_{i=1}^{n} \lambda_i}$. Equation (3.4) is derived in [1].

For deeper insight into values that can be assumed by Fisher linear discriminant, we will do the following. Suppose

$$p = \frac{k_i}{\sum_{j=1}^{N} k_j},$$

where k_i is a number of vectors in matrix $\mathbf{X}_i$,
$\sum_{j=1}^{N} k_j$ is a number of vectors in $[\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3 \ldots \mathbf{X}_n]$.

Values of parameter p can be from interval $< 0, 1 >$.
It was found experimentally that:

1. In case of perfect overlap of classes (they have equal mean values and variance of one class is zero), then $F = p$.

2. For identical classes (equal mean values and variances (covariance matrices)), then $F = 1$.

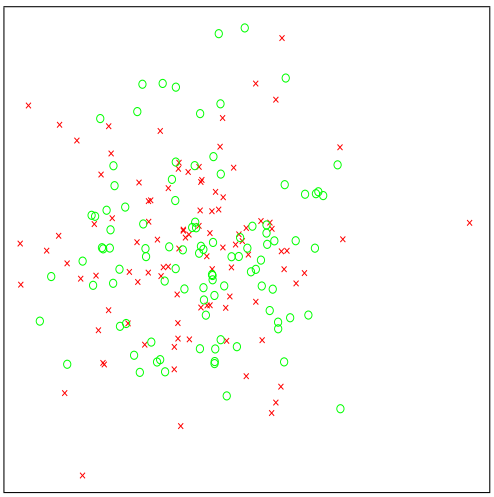
3. If classes are perfectly separable, then $F = \infty$.

4. In case of only partially overlap, then $p < F < 1$.

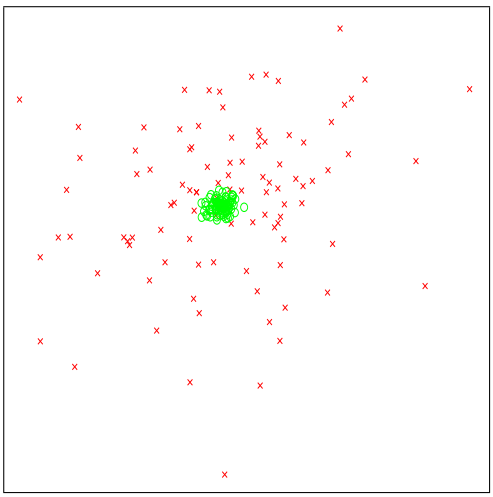
5. If classes are merely separable, then $1 < F < \infty$. The higher F, the better separability.

Examples of two classes in 2-dimensional feature space are given in Figure 3.3.

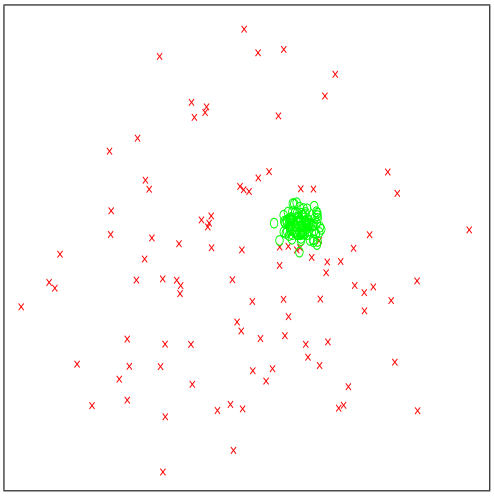
A subset of features appropriate for classification was chosen from 9 Haralick features. Classification (the nearest mean classifier) on different training and test sets for different subsets of features was performed. For each subset, the relative frequency of achieving the best classification rate (called *s*tability) was obtained. We then chose the subset with the highest stability and high Fisher linear discriminant. That yielded in the subset of three features: texture entropy, texture correlation and uniformity of texture energy. They are extracted from co-occurrence matrix with the separation vector [11,0] as mentioned in Section 3.3. They will be used in classifier described next.



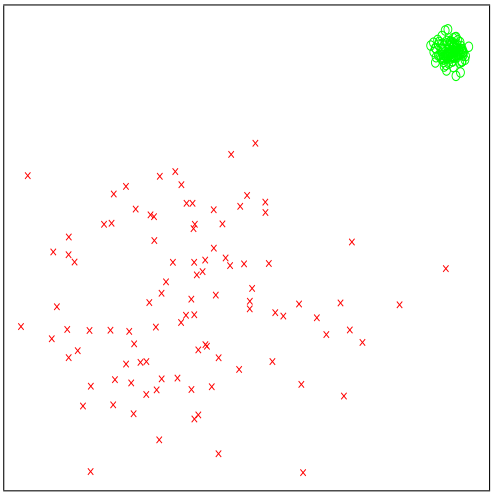
$\mathbf{X}_2 = N(0,1) \rightarrow F = 1.011$



$\mathbf{X}_2 = N(0,0.1) \rightarrow F = 0.505$



$\mathbf{X}_2 = N(0.5,0.1) \rightarrow F = 0.650$



$\mathbf{X}_2 = N(3,0.1) \rightarrow F = 5.520$

Figure 3.3: Basic positions of two classes in 2D space, parameter $p = 0.5$, $k_1 = k_2 = 100$, $\mathbf{X}_1 = N(0,1)$ (it has normal distribution, zero mean value, and variance equals one).

Chapter 4

Classifier Selection

In this chapter we give theoretical background for classifier selection. The task we deal with is supervised learning. If we knew the a priori probabilities and the class-conditional densities, we could have designed an optimal classifier, Bayes classifier. We can obtain probability density estimates by parametric or non-parametric techniques. If we knew the parametric form of the density, we could have derived its parameters by some parametric technique. However, all classic parametric densities are unimodal, whereas many practical problems involve multimodal densities.

In our previous work [1, 2, 32] we evaluated features by classifier that searched for the nearest mean. It is a parametric method for estimating probability distribution under certain simplification, i.e. the assumption of Gaussian distribution, equal covariance matrices for all of the classes, all of the variables statistically independent, and equal a priori probabilities of the classes. The success of classification achieved by this method suggested that automatic classification in thyroid gland diagnostic is possible, but results were still not satisfactory enough. LT diagnosis consists of several sub-units [31] and we can suppose that the variability inside this class can be considerable. Our previous work shows that it can be so (see for example Figure 4.1 in [1]). We also reported (for details look at Section 1.5) that some overlap between LT and normal classes exists in feature space (it can be seen in Figure 4.1 as well). Hence it is necessary to model probability density in the space where these two classes overlapped. Estimation of such probability density should be without any simplification or assumptions. For this purpose non-parametric methods seem to be adequate.

At the beginning of this chapter we overview Bayes decision theory (Section 4.1). After that, methods for estimating the Bayes error are given. We than focus on non-parametric techniques (Section 4.2), mainly the K-nearest-neighbour classification.

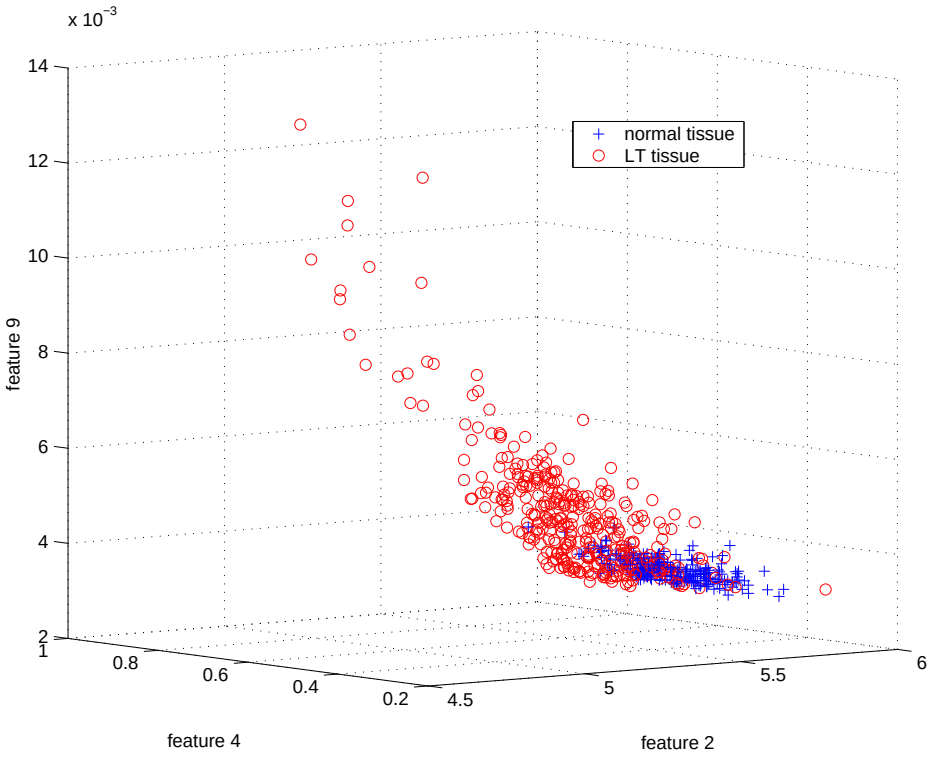


Figure 4.1: Feature space for features H2 H4 H9 from Table 3.1. Notice higher variability inside LT class than inside normal class. Feature space consists of 2 normal and 2 LT subjects [1] (features are derived from co-occurrence matrix given by distance vector $d = [1, 0]$).

4.1 Bayes Decision Theory

Bayes decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on the assumption that all the relevant probabilities are known. Since Bayes theory can be found in every publication devoted to pattern classification [39, 40, 41], we give here merely a brief overview.

Let $P(\omega_i)$ be the a priori probability that an arbitrary feature vector belongs to class ω_i. This reflects our prior knowledge of how likely we are to see one of the classes before feature vector appears. At this moment we know only a priori probabilities, without observing any feature vector X. It is then reasonable to use the following decision rule: Decide ω_i if $P(\omega_k) > P(\omega_i)$ for all $i \neq k$. In most circumstances, there is not so little information. Let $p(X|\omega_i)$ be the conditional density distribution of all feature vectors belonging to ω_i. Suppose that we know both the a priori probabilities

$P(\omega_i)$ and conditional densities $p(X|\omega_i)$ and we measure the next feature vector X.

Then we can use the Bayes theorem

$$P(\omega|X) = \frac{p(X|\omega)P(\omega)}{p(X)}, \tag{4.1}$$

where $p(X)$ is the probability density function of all vectors. $P(\omega_i|X)$ is then a posteriori probability that a specific feature vector X was drawn from class ω_i. The fact that

$$p(X) = \sum_{i=1}^{c} p(X|\omega_i)P(\omega_i) \tag{4.2}$$

assures that

$$\sum_{i=1}^{c} P(\omega_i|X) = 1.$$

The a priori probability $P(\omega_i)$ can be either known from the application or estimated from the training samples. We also need to estimate $p(X|\omega_i)$. This density is estimated in supervised learning stage from a finite number of pre-classified examples. The quality of this training data affect the quality of the classifier approximation. It is often assumed that the density has the form of a normal distribution. In this thesis we will not make any such assumptions.

A given vector X of unknown class is classified to ω_k if $P(\omega_k|X) = \max P(\omega_i|X)$, for all $i \neq k$. Classifier based on this rule is called optimal classifier.

4.1.1 Bounds on the Bayes Error

In general, the classification error is a function of two sets of data, the design and test sets (ρ_D and ρ_T), and may be expressed by

$$\varepsilon(\rho_D, \rho_T), \tag{4.3}$$

where ρ is a set of two densities, as

$$\rho = [\rho_1(X), \rho_2(X)]. \tag{4.4}$$

If the classifier is the Bayes for the given test distributions, the resulting error is minimum. Therefore, we have the following inequality

$$\varepsilon(\rho_T, \rho_T) \leq \varepsilon(\rho_D, \rho_T). \tag{4.5}$$

The Bayes error (error of the Bayes classifier [39]) for the true ρ is $\varepsilon(\rho, \rho)$. However, we never know the true ρ. One way to overcome this difficulty is to find upper and lower bounds of $\varepsilon(\rho, \rho)$ based on its estimate $\hat{\rho} = [\hat{\mathbf{p}}(X), \hat{\mathbf{p}}(X)]$. In order to accomplish this, let us introduce from (4.5) two inequalities as

$$\varepsilon(\rho, \rho) \leq \varepsilon(\hat{\rho}, \rho). \tag{4.6}$$

$$\varepsilon(\hat{\rho}, \hat{\rho}) \leq \varepsilon(\rho, \hat{\rho}). \tag{4.7}$$

Equation (4.6) indicates that ρ is the better design set than $\hat{\rho}$ for testing ρ. likewise, $\hat{\rho}$ is better design set than ρ for testing $\hat{\rho}$. The error estimate is unbiased with respect to test samples [39] . Therefore, the right-hand side of(4.6) can be modified to

$$\varepsilon(\hat{\rho}, \rho) = E_{\hat{\rho}_T}\{\varepsilon(\hat{\rho}, \hat{\rho})\}, \tag{4.8}$$

where $\hat{\rho}_T$ is another set generated from ρ independently of $\hat{\rho}$. Also, after taking the expectation of (4.7), the right-hand side may be replaced by

$$E\{\varepsilon(\rho, \hat{\rho})\} = \varepsilon(\rho, \rho). \tag{4.9}$$

Thus, combining (4.6)-(4.9),

$$E\{\varepsilon(\hat{\rho}, \hat{\rho})\} \leq \varepsilon(\rho, \rho) \leq E_{\hat{\rho}_T}\{\varepsilon(\hat{\rho}, \hat{\rho}_T)\}. \tag{4.10}$$

That is, the Bayes error, $\varepsilon(\rho, \rho)$ is bounded by two sample-based estimates.

The rightmost term $\varepsilon(\hat{\rho}, \hat{\rho}_T)$ is obtained by generating two independent samples sets, $\hat{\rho}$ and $\hat{\rho}_T$, from ρ, and using $\hat{\rho}$ for designing the Bayes classifier and $\hat{\rho}_T$ for testing. The expectation of this error with respect to $\hat{\rho}_T$ gives the upper bound on the Bayes error. Furthermore, taking the expectation of this result with respect to $\hat{\rho}$ does not change this inequality. Therefore, $E_{\hat{\rho}}E_{\hat{\rho}_T}\{\varepsilon(\hat{\rho}), \hat{\rho}_T\}$ also can be used as the upper bound. This procedure is called the Holdout (H) method. On the other hand, $\varepsilon(\hat{\rho}, \hat{\rho})$ is obtained by using $\hat{\rho}$ for designing the Bayes classifier and the same $\hat{\rho}$ for testing. The expectation of this error with respect to $\hat{\rho}$ gives the lower bound on the Bayes error. This procedure is called Resubstitution (RES) method.

The H method gives the upper bound on the Bayes error. It works well if the data sets are generated artificially by a computer. However, in practice, if only one data is available, in order to apply the holdout method, we need to divide sample set into two independent groups. This reduces the number of samples available for designing and testing. Also, how to divide samples is

a serious and nontrivial task. It is necessary to implement a proper dividing algorithm.

A procedure, called the Leave-One-Out (LOO) method, alleviates the above difficulties of the H method. In the LOO method, one sample is excluded, the classifier is designed on the remaining $N-1$ samples, and the excluded sample is tested by the classifier. This operation is repeated N times to test all N samples. Then, the number of misclassified samples is counted to obtain the estimate of the error. Since each test sample is excluded from the design sample set, the independence between the design and test sets is maintained. Also N samples are tested and $N-1$ samples are used for design. Thus the available samples are, in this method, more effectively utilized. Furthermore, we do not need to worry about dissimilarity between the design and test distributions. One of the disadvantages of the LOO is that N classifiers must be designed, one classifier for testing each sample.

The H and LOO methods are supposed to give very similar, if not identical, estimates of the classification error, and both provide upper bounds on the Bayes error.

4.2 Non-parametric Methods for Density Estimation

These methods provide description for probability density functions for which the functional form is not specified in advance. It depends on the data itself, so no assumptions are made about the distribution of the data. Hence, the term non-parametric is apparent.

One of the simplest non-parametric methods is density estimation using histogram. Despite its advantages, such as fast visualization of data in one or two dimensions, it suffers from a number of difficulties which prevent this method to achieve more accurate results. One problem is that the density distribution is not smooth at the boundaries between neighbouring histogram bins. A second problem arises if we divide each variable into M (number of bins) intervals. Then the d-dimensional feature space will be divided into M^d bins. This exponential growth with d is an example of the 'curse of dimensionality' discussed in [40].

Kernel-based methods are more sophisticated than histogram. Instead of using bins defined in advance (taking no account of data), as it is in histogram, the density is estimated by cells of given volume whose locations are determined by the data points [40, 41]. However, the volume (V) is fixed

for all of these points. If V is too large some regions might have high density of points and thus the estimated density is over-smoothed and important spatial variations may be lost. When V is small, many of the volumes will probably be empty and the model density can became noisy. This difficulty is dealt with by K-nearest-neighbour approach, where number of points in the cell is fix (K) and the volume of the cell can vary. This is introduced in Section 4.2.1.

There is also semi-parametric technique called *mixture models* that is not restricted to specific functional forms and where the size of the model only grows with the complexity of the problem being solved and not simply with the size of the data set. In the non-parametric kernel-based approach to density estimation, the density function was represented as a linear superposition of kernel function, with one kernel centered on each data point. Here models are considered in which the density function is again formed from a linear combination of basis functions, but where the number of basis functions is treated as a parameter of the model and is typically much less than the number of data points.

4.2.1 K-nearest-neighbour

K-nearest-neighbour is a non-parametric technique for density estimation and classification. This rule classifies new feature vector X by assigning it the label most frequently represented among the K nearest samples. To explain the principle of classification based on this method sufficiently we use Bayes theorem [40]. It requires computing posterior probabilities from class-conditional densities and a priori probabilities for each class. Suppose our data set contains N_k points in class C_k and N points in total, so that $\sum_k N_k = N$. We then draw a hypersphere (the cell mentioned above) around the point X which encompasses K points irrespective of their class label. Suppose this sphere, of volume V, contains K_k points from class C_k. Then approximations for the class-conditional densities can be given in the form

$$p(X|C_k) = \frac{K_k}{N_k V} \ . \tag{4.11}$$

The unconditional density can be similarly estimated from

$$p(X) = \frac{K}{NV} \tag{4.12}$$

while the a priori probabilities can be estimated using

$$P(C_k) = \frac{N_k}{N} \ . \tag{4.13}$$

We now use Bayes theorem to give

$$P(C_k|X) = \frac{p(X|C_k)P(C_k)}{p(X)} = \frac{\frac{K_k}{N_k V}\frac{N_k}{N}}{\frac{K}{NV}} = \frac{K_k}{K} \ . \tag{4.14}$$

To minimize the probability of misclassifying a new vector X, it should be assigned to the class C_k for which the ratio K_k/K is largest. It means finding a hypersphere around the point X, which contains K points (independent of their class), and then assigning X to the class having the largest number of representatives inside the hypersphere. For the special case of $K = 1$ we simple assign a point X to the same class of which the nearest point from the training set is.

A disadvantage of K-nearest-neighbour method is that the complete set of training samples must be stored and must be searched each time a new feature vector is to be classified. This might result in computational problems while making classification.

Chapter 5

Experiments

Several experiments on the dataset were performed. Data are represented by Haralick features and histograms (see Table 3.2). The choice of histogram resolution is discussed in Section 5.1. To asses measure of information carried by histograms and Haralick features about class labels, we computed Fisher linear discriminant and multi-correlation coefficient (Section 5.2).

Upper and lower bounds on the Bayes error were estimated on K-nearest-neighbour classifier using Resubstitution method and Leave-one-out method. The effect of K on the classification rate can bring insight into the data distribution in the feature space. Classification was done on images and subjects (Section 5.3).

5.1 Histogram Resolution

Histogram is a vector of length t with frequencies of occurring pixels with intensities $1, 2, \ldots, t$. For image with 256 grey values, $t \leq 256$. Resolution t can be smaller without much loss of information. Anděl [42] reported that t can be determined by Sturges rule, so that $t \approx 1 + 3.3 \log n$, where n is number of pixels in image. For our data, computation of t resulted in number 32. Hence histograms experiments consists of 32 bins.

5.2 Feature Evaluation

Haralick features H2,H4,H9 (see Table 3.1) were chosen in [1, 2] as the best subset of 9 Haralick features. Here we can evaluate separability provided by this subset using our new data. We computed Fisher linear discriminant on this subset. This yielded in $F_{har} = 0.996$. For comparison, Fisher linear discriminant was also computed on data represented by histograms. It resulted

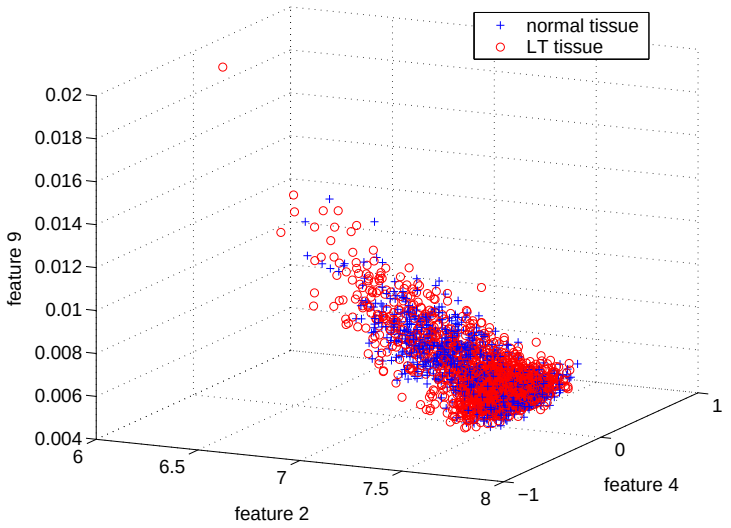


Figure 5.1: Feature space for features H2 H4 H9 from Table 3.1. Feature space consists of 33 normal and 38 LT subjects (features are derived from co-occurrence matrix given by distance vector $d = [11, 0]$).

in $F_{hist} = 15.226$ The resulting space for Haralick features can be easily pictured in 3D space (Figure 5.1). Notice that the variability inside the normal class seems to be similar to the variability of LT tissue, in contrast to data from the old dataset (Figure 5.2).

Another way to compare features is multi-correlation coefficient ϱ. It describes linear dependence between class labels L and $\mathbf{X}$ ($L = \alpha + \beta'\mathbf{X}$). Notation $\varrho_{L,\mathbf{X}}$ is the highest of all correlation coefficients between L and arbitrary nonzero linear function of $\mathbf{X}$. Multi-correlation coefficient can acquire values from interval $\langle 0, 1\rangle$. The higher $\varrho_{L,\mathbf{X}}$, the bigger dependence. Value 1 means that there exist nonzero linear function that unambiguously maps $\mathbf{X}$ onto L. Zero means that such function does not exist. For Haralick features, $\varrho_{L,\mathbf{X}} = 0.006$. For histograms (resolution 32 bins), $\varrho_{L,\mathbf{X}} = 0.559$.

5.3 Classification

K-nearest neighbour classifier was implemented as classifier that makes no assumption about data distribution. In case of histograms as features, subject is classified to be of class normal (N) if majority of its images is of class normal. Data represented by Haralick features consists of set of samples (rectangular windows) for each image. In this case, subject is classified to be of normal class when majority of its samples is classified as normal.

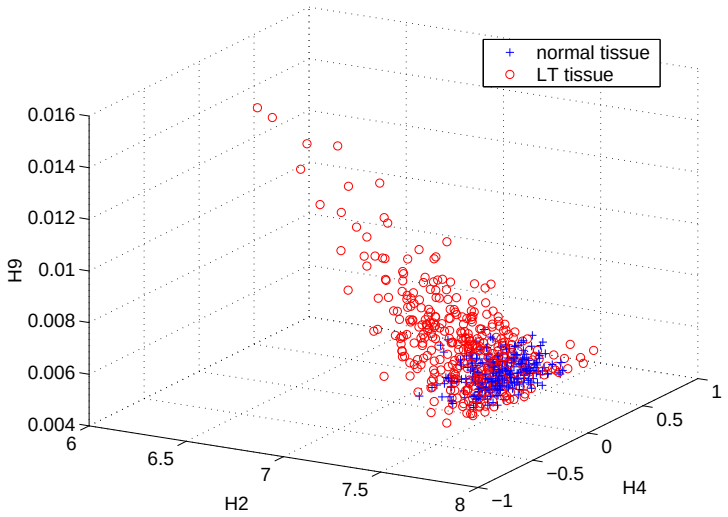


Figure 5.2: Haralick features computed on the old dataset for co-occurrence matrix given by distance vector $d = [11, 0]$.

5.3.1 Leave-one-out

Leave-one-out is a method for estimation of an upper bound on the Bayes error. It consists in dividing feature space into design and test set. The test set consists of features from one subject. The design set is created by features from remaining N-1 subjects. Features from the test set are classified by K-nearest neighbour classifier designed on the design set. This is repeated N times to test features from all N subjects. Subject is then classified by majority vote. After this procedure, leave-one-out error on subjects (LOO subjects) can be computed as number of misclassified subjects divided by the number of all subjects. False negative error for subjects (FN subjects) is

$$\frac{\text{number of LT subjects classified as normal}}{\text{number of all LT subjects}}.$$

False positive error for subjects (FP subjects) is

$$\frac{\text{number of normal subjects classified as LT}}{\text{number of all normal subjects}}.$$

Analogous to that, FN images and FP images can be computed for histograms and FN samples and FP samples for Haralick features. All this process can be repeated over different number of neighbours (K). FN subjects and FP subjects versus K for histograms is shown in Figure 5.3. The same characteristic, but for classification of images can bee seen in Figure 5.4.

The curve of LOO subjects is given for comparison. Similar characteristics were obtained also for Haralick features. They are given in Figure 5.5 and Figure 5.6.

Suppose number of normal subjects be m_N, number of LT subjects m_{LT}, and number of all subjects $M = m_N + m_{LT}$. Relationship between FN, FP, and LOO error is the following:

$$\text{LOO} = \begin{cases} \frac{FN+FP}{2} & \text{if } m_N = m_{LT} = \frac{M}{2}, \\ \frac{FN \cdot m_{LT} + FP \cdot m_N}{M} & \text{otherwise.} \end{cases}$$

5.3.2 Resubstitution

Resubstitution is a method for estimation of a lower bound on the Bayes error. This method is based on the same design and the test set. Characteristic of FN subjects and FP subjects versus K for histograms is shown in Figure 5.7. The same characteristic, but for classification of images can bee seen in Figure 5.8. The curve of LOO subjects is given for comparison. Similar characteristics were obtained also for Haralick features. They are given in Figure 5.9 and Figure 5.10.

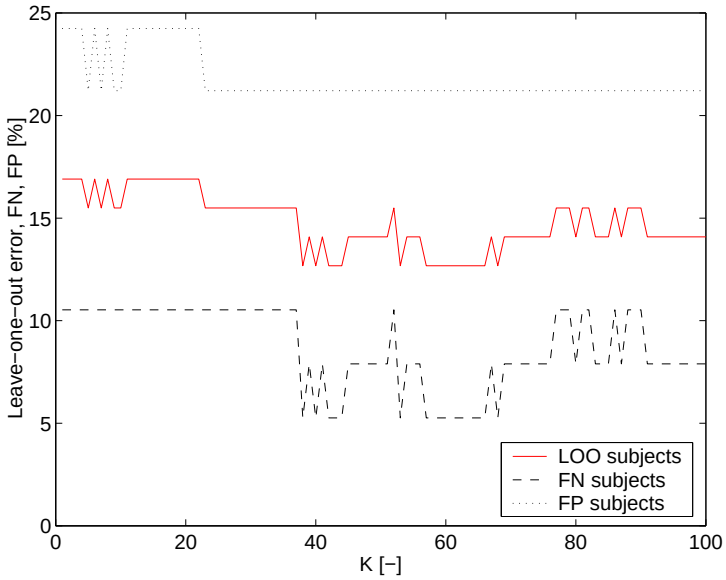


Figure 5.3: **LOO** error of K-NN classifier versus K for **histograms**, classification **on subjects**. FN – false negative error, FP – false positive error.

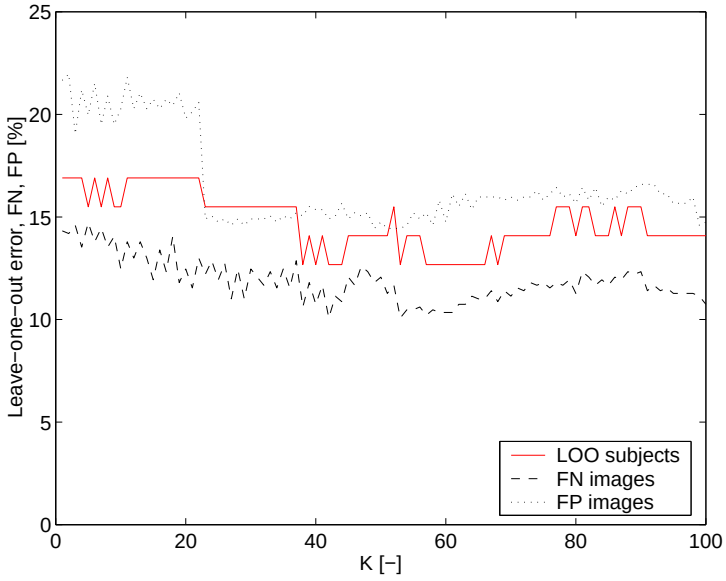


Figure 5.4: False negative (FN) and false positive (FP) error of K-NN classifier versus K for histogram as features, classification **on images**. (LOO error for classification on subjects is given for comparison.)

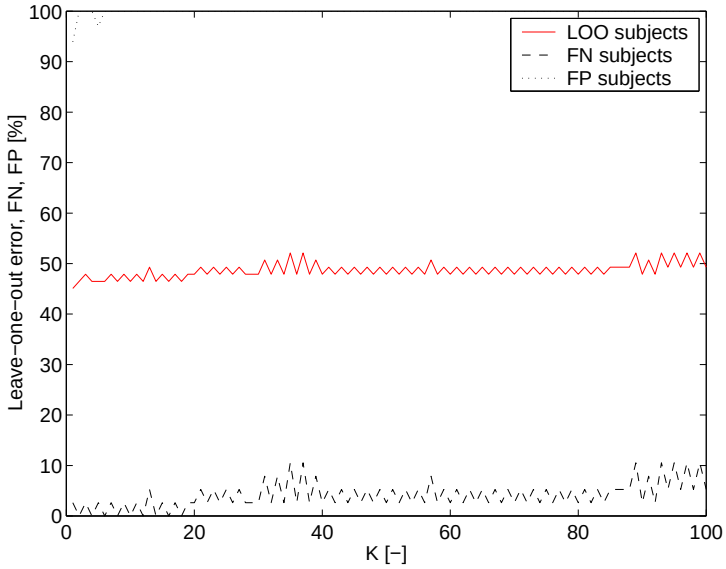


Figure 5.5: **LOO** error of K-NN classifier versus K for **Haralick features**, classification **on subjects**. FN – false negative error, FP – false positive error.

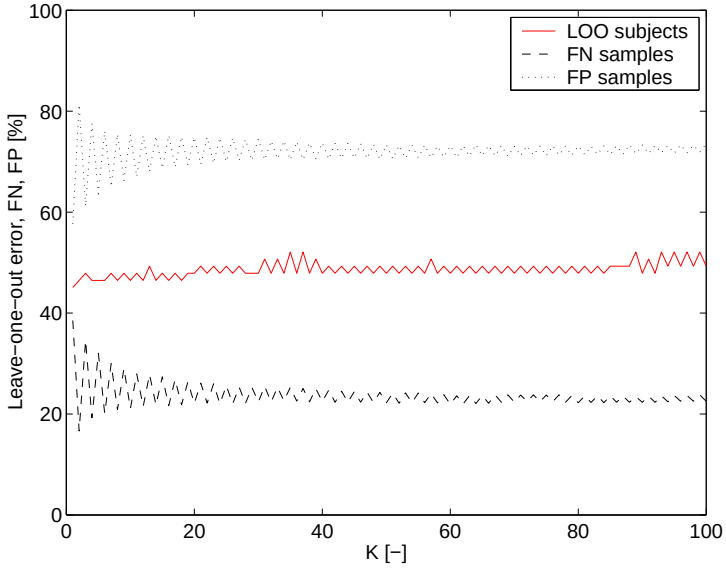


Figure 5.6: False negative (FN) and false positive (FP) error of K-NN classifier versus K for Haralick features, classification **on samples**. (LOO error for classification on subjects is given for comparison.)

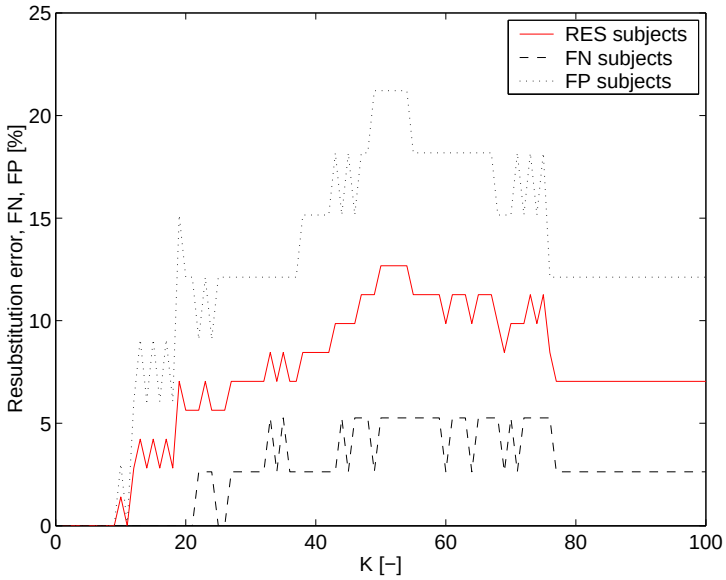


Figure 5.7: **RES** error of K-NN classifier versus K for **histograms**, classification **on subjects**. FN – false negative error, FP – false positive error.

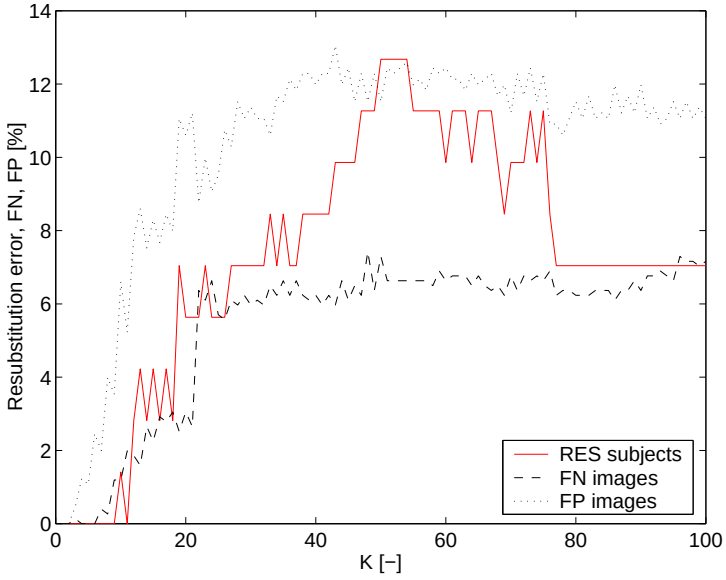


Figure 5.8: False negative (FN) and false positive (FP) error of K-NN classifier versus K for histogram as features, classification **on images**. (RES error for classification on subjects is given for comparison.)

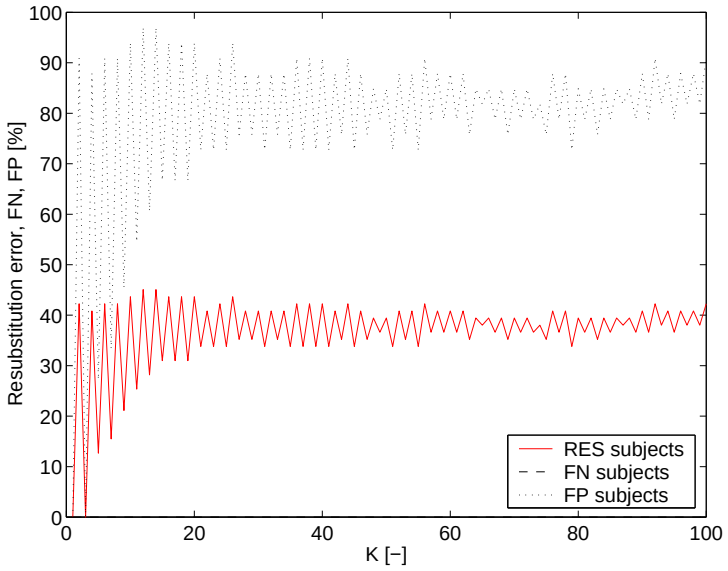


Figure 5.9: **RES** error of K-NN classifier versus K for **Haralick features**, classification **on subjects**. FN – false negative error, FP – false positive error.

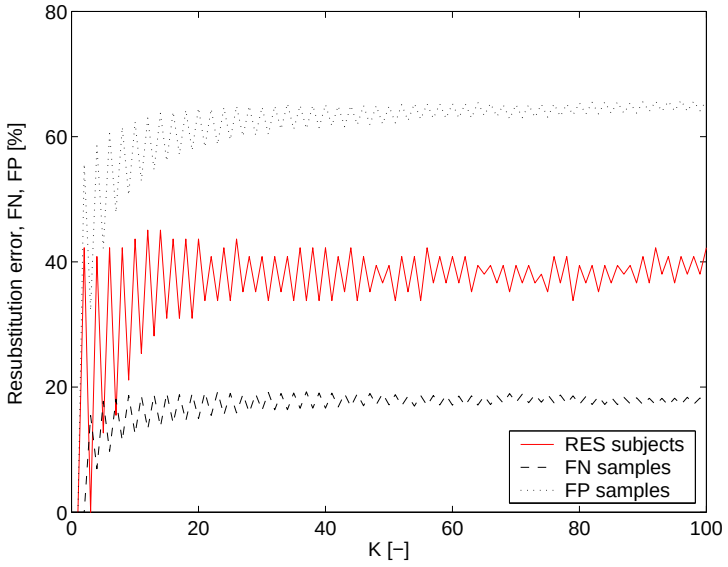


Figure 5.10: False negative (FN) and false positive (FP) error of K-NN classifier versus K for Haralick features, classification **on samples**. (RES error for classification on subjects is given for comparison.)

Chapter 6

Discussion

Fisher linear discriminant supposes that data are from Gaussian distribution, which can be represented by mean value and variance. Multicorrelation coefficient expects linear dependence between class labels and data. K-nearest-neighbour rule recquires no assumptions about data. Nevertheless, the results of all these methods are similar.

Fisher linear discriminat for Haralick features $F_{har} = 0.996$ means that classes have similar mean values and variances. It is when classes are generally overlapped and this is apparent from Figure 5.1. Fisher linear discriminant $F_{hist} = 15.226$ shows that classes represented by histograms of resolution 32 are separable. This separability is not ideal but far better than the one for Haralick features. This fact is confirmed by multicorrelation coefficient. Its value for Haralick features ($\varrho_{L,\mathbf{X}} = 0.006$) is considerably less than for histograms ($\varrho_{L,\mathbf{X}} = 0.559$).

K-nearest-neighbour classification resulted in characteristics that allow to see distribution of the data in feature space (Figure 5.3). LOO error for histograms and subjects classification is less than 17%. The smallest error is achieved for higher K (LOO error = 12.7% for K between 38 and 68). It points out on some partial overlap in feature space when point of one class is surrounded by points of the other class and this whole cluster is again surrounded by points of the first class. Maximal error is for K from 1 to 21. FP error is bigger than FN error, which is good for medical tasks. It is always better to classify normal tissue as inflamed than vice versa. The majority vote causes that FN error is smaller for subjects than for images (maximal FN error for subjects is 10.5%, K smaller than 37, maximal FN error for images is 14.7%, $K = 5$).

RES error for histograms and subject classification is zero for K smaller than 9 since all images of one subject are involved in design set and we can expected that these points form a sub-cluster. Maximal RES error is 12.7%

for K between 50 and 54. FN error for images is higher than for subjects by 3%.

LOO error for Haralick features and subjects classification achieves almost 50%. For K higher than 5 all normal subjects are classified as LT, i.e. the FP error is 100%. It is obvious that separability in feature space is very poor. LOO error is nearly constant with K, it means that overlap of both classes in feature space is homogeneous. Considerable dependence between classification error and number of classes in design set appears in such spaces. It means that for infinite number of feature vectors in feature space, classification error depends only on a priori probability.

RES error for Haralick features and subject classification is zero only for $K = 1$ and 3. Zero for $K = 1$ is typical for resubstitution method since each classified point is included into the design set. Steep increase of FP error for $K = 2$ and K higher than 3 means that almost each normal point has some points of class LT in its neighbourhood. RES error stabilises on 38% for K higher than 15.

The plot curves fluctuation for Haralick features means that neighbourhood of each point contains comparable number of points from the normal and the LT class. The influence of change in number of images is then higher with smaller number of neighbours.

Feature space for Haralick features seems not to provide sufficient representation of our data. It is visible from Figure 5.1 and it is confirmed by the results of Fisher linear discriminant, multicorrelation coefficient and by classification using non-parametric method: K-nearest-neighbour classifier.

This finding differs from results obtained in previous work [2, 3, 36] where Haralick features seemed to be sufficient for classification. It was caused by small dataset. Features computed on co-occurrence matrix with distance vector $d = [11, 0]$ on old dataset can be seen in Figure 5.2. We can see two clusters separable at certain extent. New feature space in Figure 5.1 shows that mainly new normal images caused two clusters to be non-separable in Haralick feature space.

Chapter 7

Conclusions

We aimed to classify sonographical images of thyroid gland into two classes: normal tissue and lymphocytic thyroiditis (LT). For this purpose, images were represented by features that should extract important textural character of image. So far, it was reported that Haralick features are able to distinguish between normal and LT class. However, it was done for small dataset containing merely several subjects. This thesis focuses to use Haralick features and histograms on dataset consisting of 1405 images from 71 subjects and to use classification that suppose no assumption on data distribution: K-nearest-neighbour rule.

Classification was done using leave-one-out and resubstitution methods, which give estimates of upper and lower bounds on the Bayes error. Dependence of these errors over K was analyzed and data structure discussed. Results for Haralick features were not encouraging: leave-one-out error was 45%. Classification of histograms achieved the following results: leave-one-out error approached 12%. Experiments were performed on dataset of 71 subjects, i.e. 33 normal and 38 subjects with lymphocytic thyroiditis. These results were confirmed by parametric methods, Fisher linear discriminant and multicorrelation coefficient.

We conclude that Haralick features (texture entropy, texture correlation and texture probability of run length of 2) derived from co-occurrence matrix for distance vector $d = [11, 0]$ are not efficient enough for our purpose. They provide feature space in which two clusters of normal and LT tissue are non-separable. Our results suggest that information needed to distinguish normal from LT tissue can be obtained using first-order statistics: 1-dimensional histograms consisted of 32 bins. Automatic classification based on histograms can improve the diagnosis reliability in distinguishing between normal tissue and Hashimoto's thyroiditis. Better results might be obtained by using higher-level statistics. This is a subject of ongoing work. Preliminary results

with histogram-based probability density estimation show that the leave-one-out error drops to 9.6% for fourth-order statistics optimal in the sense of (3.1).

Bibliography

[1] M. Švec and R. Šára. Analýza textury sonografických obrazů difúzních procesů parenchymu štítné žlázy. Research Report CTU–CMP–1999–12, Center for Machine Percepcion, FEE CTU in Prague, Dec 1999.

[2] 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. In Svoboda T., editor, *Proceedings of the Czech Pattern Recognition Workshop 2000*, pages 45–47. Czech Pattern Recognition Society Praha, Feb 2000.

[3] 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.

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

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

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

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

[8] 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.

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

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

[11] 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.

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

[13] 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.

[14] A. Kakemura, T. Higashi, and K. Irie. Texture characteristic variables based on virtual volume. *Systems and Computers in Japan*, 29(6):38–48, June 1998.

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

[16] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on PAMI*, 20(3):226–239, March 1998.

[17] E. M. Kleinberg. On the algorithmic implementation of stochastic discrimination. *IEEE Transactions on PAMI*, 22(5):473–490, May 2000.

[18] 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. Springer-Verlag, 2000.

[19] 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, SPIE.

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

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

[22] R. Uppaluri, T. Mitsa, M. Sonka, E. A. Hoffman, and G. McLennan. Quantification of pulmonary emphysema from lung computed tomography images. *American Journal of Respiratory and Critical Care Medicine*, 156:248–254, 1997.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

[31] R. Šára, D. Smutek, P. Sucharda, and Š. Svačina. Systematic construction of texture features for Hashimoto's lymphocytic thyroiditis recognition from sonographic images. In S. Quaglini, P. Barahona, and S. Andreassen, editors, *Artificial Intelligence in Medicine*, LNCS, Berlin-Heidelberg, Germany, 2001. Springer. Accepted.

[32] E. A. Toufik. Automatic classification of the thyroid gland diseases by a histogram. Master's thesis, Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic, Jan 2001.

[33] R. Šára. Sonograph images: Texture analysis [online]. c1998, last revision 9th of November 2000 [cit. 2001-4-25]. `http://cmp.felk.cvut.cz/~sara/Sono/sono.html`.

[34] 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 Prokosch H.-U., editors, *Telematics in Health Care – Medical Infobahn for Europe, Proceedings of the MIE2000/GMDS2000 Congress*, volume CD-ROM, Berlin, Germany, August/September 2000. Quintessenz Verlag.

[35] D. Smutek, T. Tjahjadi, R. Šára, M. Švec, P. Sucharda, and Š. Svačina. Kvalitativní ukazatelé ultrazvukového vyšetření štítné žlázy. *Diabetologie, Metabolismus, Endokrinologie, Výživa*, 3(Supplementum 2):16, December 2000. Proceedings XXIII. Endokrinologické dni, Košice, 5-7 October, Slovak Republic.

[36] 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. Brno University of Technology VUTIUM Press.

[37] D. Smutek, R. Šára, P. Sucharda, and Š. Svačina. Quantitative tissue characterization in sonograms of thyroid gland. In *Proceedings Conf. MEDINFO 2001*, September 2001.

[38] Z. Kotek, I. Brůha, V. Chalupa, and J. Jelínek. *Adaptivní a učící se systémy.* SNTL, 1980.

[39] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 2nd edition, 1990.

[40] C. M. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, 1995.

[41] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* A Willey-interscience publication, 1973.

[42] J. Anděl. *Statistické metody.* Matfyzpress, 2nd edition, 1998.

[43] Z. Kotek, P. Vysoký, and Z. Zdráhal. *Kybernetika.* SNTL, 1990.

[44] Knoll Pharmaceutical Company. Gland central [online]. [cit. 2001-4-12]. `http://www.glandcentral.com/home/`.

[45] B. B. Tempkin. *Ultrasound Scanning: Principles and Protocols.* W B Saunders Co., 2nd edition, 1999.