We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Confocal Laser Scanning Microscopy in Dermatology: Manual and Automated Diagnosis of Skin Tumours

Wiltgen Marco

Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz
Austria

1. Introduction

The number of melanocytic tumours has increased in the last decades, whereby the frequency of melanoma doubles every 20 years. At present there is a risk of 1:100 to falling sick with a melanoma. According to a WHO report, in the year 2006, about 48,000 melanoma related deaths occur worldwide per year (Lucas et al., 2006). The malignant melanoma is one of the less common types of skin cancer but causes the majority (75%) of skin cancer related deaths. Most melanomas are brown to black looking lesions (Friedman et al., 1985). Warning signs that might indicate a malignant melanoma include change in size, shape and color. Early signs of melanoma are changes to the shape or color of existing moles (ABCD rule). Skin cancer has many potential causes, including: overexposure to UV-radiation (extreme sun exposure during sun-bathing), which may cause skin cancer either via the direct DNA damage or via the indirect DNA damage mechanism; smoking tobacco can double the risk of skin cancer; chronic non-healing wounds; genetic predisposition and the human papilloma virus (HPV), which is often associated with squamous cell carcinoma of the genitals, anus, mouth and fingers (Oliveria et al., 2006) The detection of malignant changes of skin tissue in the early stages will augment the success of the therapy (Marcovic et al., 2009). Metastasis of the melanoma in a progressive stage may spread out to the lymph nodes or even more distant places like: lungs; brain; bone and liver. Such metastatic melanoma may cause general symptoms like: fatigue; vomiting and loss of appetite. The greatest chance of cure is in the early surgical resection of thin melanomas. Therefore, a periodical screening of risk persons is necessary. In the fight against skin cancer, researchers have high hopes in improved and fast provisional screening methods for the clinical routine.

Confocal laser scanning microscopy (CLSM) is a novel imaging device enabling the non-invasive examination of skin lesions in real-time (Rajadhyaksha et al., 1999). Therefore, CLSM is very suitable for routine screening and early recognition of skin tumours. The CLSM technique allows the viewing of micro-anatomic structures and individual cells. In contrast to the conventional examination, where suspicious skin tumours have to be excised, embedded in paraffin and stained, this method is much more agreeable for the patient and faster. However, training and experience is necessary for a successful and accurate diagnosis in this new and powerful imaging technique. To diminish the need for training and to improve diagnostic accuracy, computer aided diagnostic systems are required by the derma pathologists.

Computer aided diagnosis means that the harmless (nevi) and malignant cases are discriminated by automatic analysis on a computer, providing optimised preventive medical checkups and accurate and reliable detection of skin tumours. Automated diagnostic systems need no input by the clinician but rather report a likely diagnosis based on computer algorithms. A main task in automated image analysis is the selection of appropriate features for a "computer friendly" description of the tissue. The choice and development of the features is driven by the diagnostic guidelines of the derma pathologists. In the diagnosis of CLSM views of skin lesions, architectural structures at different scales play a crucial role. The images of benign common nevi show pronounced architectural structures, such as: arrangements of nevi cells around basal structures and tumour cell nests. The images of malign melanoma show melanoma cells and connective tissue with few or no architectural structures. Features based on the wavelet transform have been shown to be particularly suitable for the automatic analysis of CLSM images because they enable an exploration of images at different scales (Wiltgen et al., 2008). A further task in automated analysis is the choice of the machine learning algorithm for classification, which enables it, after a previous training, to predict the class of a lesion (nevi or malignant melanoma). For medical diagnosis, the algorithm should duplicate the automated diagnostic process by making it understandable for the human interpreter. By the CART (Classification and Regression Tree) algorithm, the inferring rules are automatically generated during the training of the algorithm. The generated rules have a syntax that is understandable for the human interpreter and they can be discussed and explicitly used as diagnostic rules. The classification results are relocated to the images by use of the inferring rules as diagnostic aid. The regions enabling a high discrimination power are highlighted in the images, showing tissue with features in good accordance with typical diagnostic CLSM features.

In this paper, we introduce the basic principles of automated diagnosis of CLSM images of skin lesions. Special attention is given to the wavelet transform for the description of the tissues in a way that conforms with the guidelines of the human interpreter. Further, the machine learning algorithm for the class prediction and its diagnostic rules is discussed in some detail. The application and performance of the discussed methods is demonstrated by a selected study.

The procedure for image analysis presented in this paper, was developed with the "Interactive Data Language" software tool IDL, which is a computing environment for image analysis, data visualization, and software application development (IDL 7.1, ITT Visual Information Solutions (ITT VIS), formerly known as Research Systems Inc. (RSI), http://www.ittvis.com/). IDL belongs to the Fourth Generation Languages (4GL) and includes image processing procedures and tools for rapid prototyping and rapid application development. All software runs on a PC under Windows and supports the development of applications with graphical user interface (GUI). The CART (Classification and Regression Trees) analysis was done with the software from Salford Systems, San Diego, USA.

2. Confocal laser scanning microscopy

The principle of confocal microscopy was developed by Marvin Minsky in 1957. After the development of lasers, confocal laser scanning microscopy became a standard technique toward the end of the 1980s (Patel et al., 2007). Confocal laser scanning microscopy (CLSM) is a technique for obtaining high-resolution optical images with depth selectivity (Paoli et al., 2009). This technique enables the acquisition of in-focus images from selected depths

(Pellacani et al., 2008). Therefore, for non-opaque specimens, such as biological tissue, interior structures can be imaged. The images are acquired by a point-by-point scanning and reconstructed using a computer. For interior imaging, the quality of the image is greatly enhanced over conventional microscopy because the image information from multiple depths in the specimen is not superimposed.
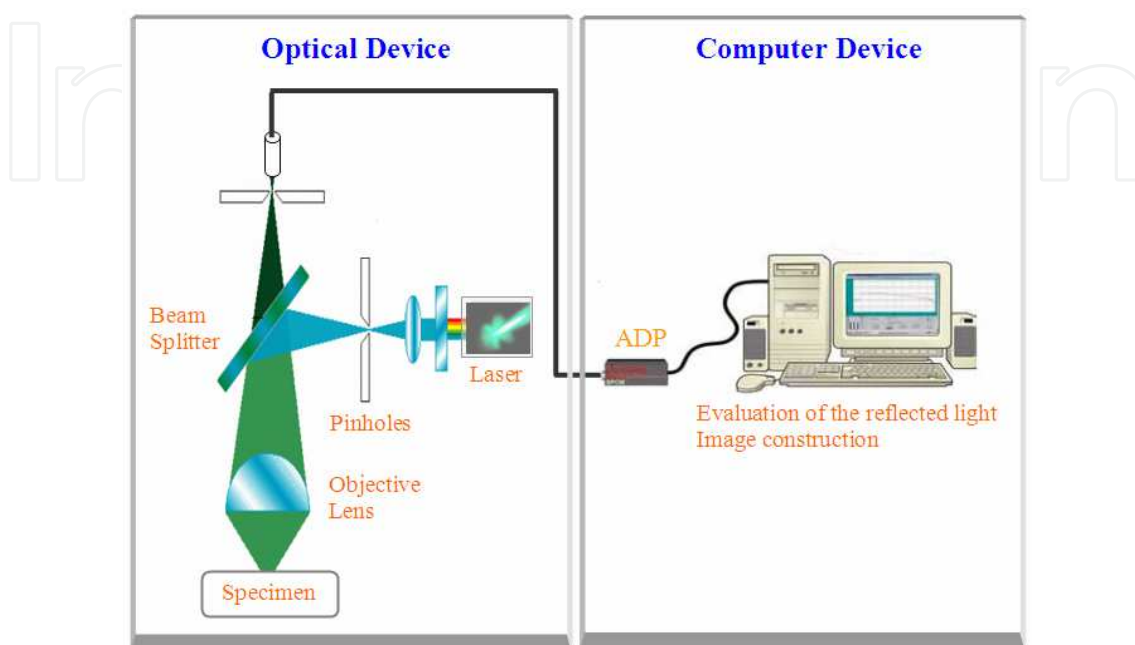

Fig. 1. Principle of the confocal laser scanning microscope

In a confocal laser scanning microscope, a laser beam passes trough a pinhole and is directed by a beam splitter (dichroic mirror) to the objective lens where it is focused into a small focal volume at a layer within the biological specimen (Fig. 1). The scattered and reflected laser light from the illuminated spot is then re-collected by the objective lens. The beam splitter separates it from the incident light and deflects it to the detector. After passing through a further pinhole, the light intensity is detected by the photon detection device (usually a photomultiplier tube or avalanche photodiode), transforming the intensity of the reflected light signal into an electrical one that is recorded by a computer. The depth of the layer (its vertical position) inside the specimen is controlled by the pinhole at the laser source. Each layer shows horizontal sections of the lesions. The electric signal, which is obtained from the intensity of the reflected light out of the illuminated volume element at a layer within the specimen, represents one pixel in the resulting image (Fig. 1). As the laser scans over the plane of interest, a whole image is obtained pixel-by-pixel and line-by-line. The brightness of a resulting image pixel corresponds to the relative intensity of the reflected light. The contrast in the images results from variations in the refractive index of microstructures within the biological specimen.

In this paper, the confocal laser scanning microscopy is performed with a near-infrared reflectance convocal microscope (Vivascope 1000, Lucid Inc., Rochester, NY, USA, http://www.lucid-tech.com/). The microscope uses a diode laser at 830 nm wavelength and a power of <35mW at tissue level. A x30 water-immersion objective lens with a numerical aperture of 0.9 is used with water (refractive index 1.33) as an immersion medium. The spatial resolution is 0.5-1.0 μm in the lateral and 3-5 μm in the axial

dimension. The images contain a field-of-view of 0.5x0.5 mm on the skin tissue providing insights into cellular structures. Up to 16 layers per lesion can be scanned. Usually an examination depth of 350 μ m can be reached (then the reflected light is absorbed by the preceding layers). All images, stored in BMP file format, are monochrome images with a spatial resolution of 640x480 pixels and a gray level resolution of 8 bits per pixel. (For the image analysis, the image matrix is rescaled to 512x512 pixels). The images are taken from the centre of the tumours representing the dermo-epidermal junction.

3. Automated image analysis

Prior to the automated image analysis, the CLSM views of the lesions are dissected into square elements. The size of the square elements ranges from 128x128 pixels to 512x512 pixels (whole image). This enables the analysis of different parts of the images which are of interest for the evaluation of the diagnostic relevant regions in the CLSM views. For an automated, or computer aided, diagnosis of CLSM views of skin lesions, the following steps are necessary (Wiltgen et al., 2003). First, the images must be appropriately analysed to enable a formulation of suitable features (Fig. 2).
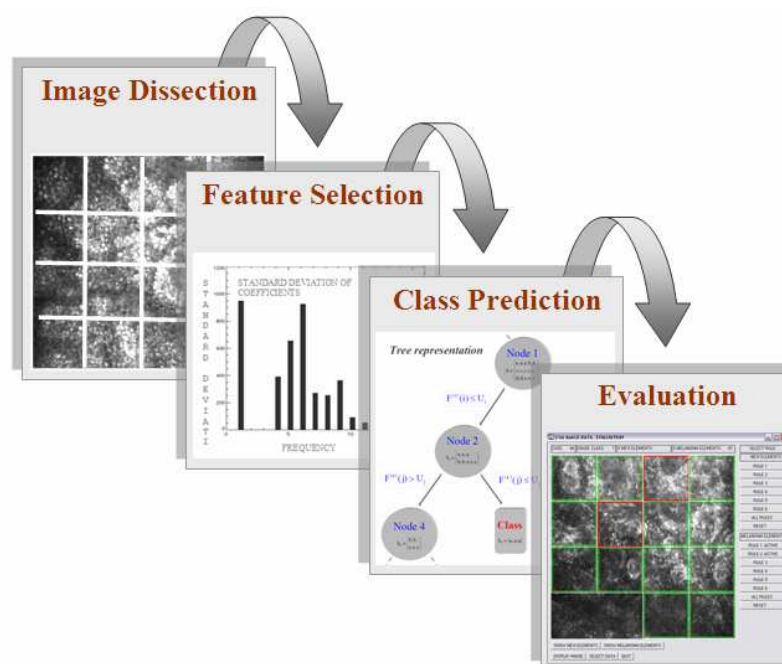

Fig. 2. Steps in the image analysis system

The formulation of appropriate features follows the derma pathologists´ diagnostic guidelines for CLSM views. Second, the images must be categorized according to their features, enabling a class prediction for every CLSM view of skin lesions. For medical purposes, it is important that the knowledge acquired during the training phase of the machine learning algorithm is represented in an understandable and readable form. Knowledge representation by rules, with a syntax that permits their use as diagnostic rules, is suitable. Third, the class prediction performance is evaluated. To this purpose, the single square elements are superimposed (relocated) onto the corresponding images by use of the diagnostic rules generated by the machine learning algorithm. The categorized square elements are highlighted in the images, enabling an identification of diagnostically highly

relevant regions and an evaluation of typical diagnostic CLSM features by comparing them with the diagnostic decisions by the human observer.

Fig. 3. Preventive medical skin checkups are of special importance in dermatology

4. Diagnosis of CLSM images

CLSM is a non-invasively method which takes 3-5 minutes per lesion whereby the microscope optical unit is positioned on the patient's skin (Fig. 3). An important step for a successful automated image analysis of histological tissue is the choice of the appropriate texture features. The selection of the features can be made according to the diagnostic guidelines used by the derma pathologist. For the diagnosis of the corresponding CLSM views, architectural structures such as: micro-anatomic structures; cell nests etc. play an important role in the diagnosis (Fig. 4).

Melanocytic cytomorphology and architecture and keratinocyte cell borders are taken into account for diagnostic decisions by the derma pathologist (Pellacani et al., 2008). Due to the high refraction of melanin, basal keratinocytes appear very intensive. The images of benign common nevi show, beside the nevi cells, pronounced architectural structures, whereas images of malign melanoma show melanoma cells and connective tissue with little or no architectural structures (Fig. 4). Therefore the information at different scales (from coarse structures to details) plays a crucial role in the diagnosis of CLSM images of skin lesions (Scope et al., 2007). This can be compared with the opinion of a human observer at different distances from an image. When the observer is close enough to the image he can study details. The greater the distance between the observer and the image, the better he can study the image as whole without being overwhelmed by the details (the details are smoothed out). In other words, he studies the images at different scales. This procedure is mathematically reflected by the wavelet transform. Therefore, features based on the properties of the wavelet transform enable an exploration of architectural structures, of different sizes, at different spatial scales. This makes the wavelet transform suitable for the

automatic analysis of CLSM views of skin lesions (Wiltgen et al., 2008). The diagnostic guide lines are reflected by different groups and properties of wavelet coefficients.

5. Wavelet analysis and wavelet transform

A wavelet is a fast-decaying wave-like oscillation that starts with an amplitude of a given (non-zero) amount and decreases rapidly to zero. This means that a wavelet is zero-valued outside of a defined interval. It can typically be visualized as a "brief oscillation".
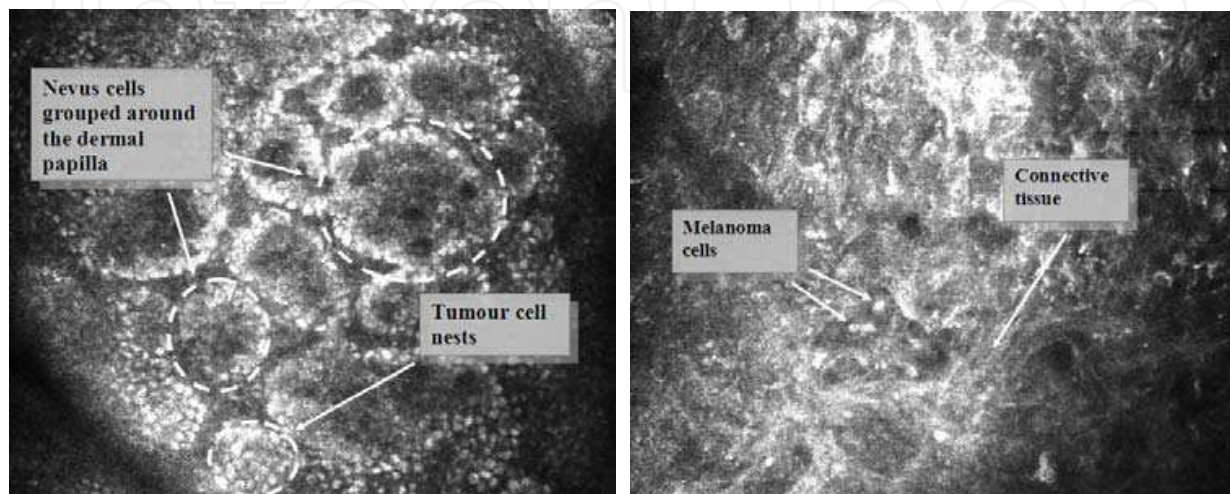

Fig. 4. CLSM images of common benign nevi (left) and malignant melanoma (right)

The wavelet transform is the representation of a given function by wavelets (Press et al., 1992). The basis wavelets are scaled and translated copies ("daughter wavelets") of the finite-length waveform ("mother wavelet"). The goal of the transformation is to represent data in another way, one which is more suitable for the analysis and interpretation (Prasad et al., 1997). Similar to the Fourier transformation, the wavelet transformation checks a given function for conformities with the basis functions and calculates the transformation from it. But in contrast to the Fourier transformation, where the basis functions (sine and cosines) are non-local in the spatial domain (and stretch out to infinity) and global properties of the analysed function are localized in prominent peaks of the frequency spectrum, the wavelet basis functions are localized in both: the spatial and the frequency domain. Wavelet transforms have advantages over the traditional Fourier transforms for accurate decomposition and reconstructing of finite, non-periodic functions. This is a crucial advantage for the analysis of the local properties of a function or a texture (Chui, 1992). By use of the wavelet basis functions with increasing spatial extension, the function or texture can be analysed at different scales showing detail (local) to coarse (global) properties (Burrus et al., 1988). This feature gives the wavelet transform a multi resolution property, enabling the study of functions and textures at varying resolutions (Daubechies, 1992).

The wavelet transforms can be divided into three classes: continuous transforms (CWT), discrete transforms (DWT) and multi-resolution-analysis (MRA) based transforms. The continuous wavelet transforms operate over every possible scale and translation whereas the discrete wavelet transforms use a discrete subset of scale and translation values. Whereas the CWT´s are mainly used in the mathematically analysis, the DWT´s have many practical applications, such as image processing. The MRA represents a design method of most of the relevant discrete wavelet transforms.

5.1 The continuous wavelet transforms

The wavelet transform (integral transform) is a linear and often invertible operation that defines a relationship between the spatial and the frequency domain. If the wavelet transform is invertible then, beside the analysis of a function, the synthesis of the function is also possible. The relationship between the spatial and the frequency domain is done by the wavelet function ("mother wavelet") Φ. The wavelet function provides the source function for the generation of the wavelets ("daughter wavelets"), which are the translated and scaled versions of the wavelet function:

$$\Phi_{s,l}(x) = \frac{1}{\sqrt{s}} \Phi\left(\frac{x-l}{s}\right) \tag{1}$$

The scale index s defines the width of the wavelet and the location index l its position. In this way, wavelets are able to localize behaviour in space (via translation) and at a characteristic scale (via dilation/contraction). The "wavelet family" is then given by:

$$\left\{ \Phi_{s,l}(x) = \frac{1}{\sqrt{s}} \Phi\left(\frac{x-l}{s}\right) \middle| (s,l) \in \mathbb{R}_+ \times \mathbb{R} \right\} \tag{2}$$

There exists a variety of different types of wavelet functions; such the Hermitian wavelet, the Shanon wavelet, the Morlet wavelet, the Mexican-hat wavelet, the Beta wavelets etc. The different wavelet types have different properties. For illustration, the Mexican-hat wavelet is presented (Fig. 5). By the choice of different values for s and l, the wavelet basis functions have different widths and are located at different spatial positions (Fig. 5). In continuous wavelet transforms, a given function is projected on a continuous family of frequency bands. The integral wavelet transform of a function $f(x) \in L^2(\mathbb{R})$, at a scale $s \in \mathbb{R}_+$ and translational value $l \in \mathbb{R}$, is expressed as convolution of the function with the wavelet:

$$\left(W_\Phi f\right)(s,l) = \int_{-\infty}^{+\infty} f(x) \Phi_{s,l}^*(x)\, dx \tag{3}$$

Whereby, $\Phi_{s,l}^*$ is the complex conjugate of the wavelet function. $\left(W_\Phi f\right)(s,l)$ is the continuous wavelet transform of $f(x)$ at position l and scale s. It is the coefficient that measures the strength of the features at scale s and position l. In order to satisfy analytical requirements for the continuous wavelet transform and for the discrete wavelet transform, the wavelet functions are chosen from a subspace of the space: $\Phi(x) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. This is the space of measurable functions that are absolutely integrable ($L^1(\mathbb{R})$) and square-integrable ($L^2(\mathbb{R})$):

$$\int_{-\infty}^{+\infty} |\Phi(x)|\, dx < \infty \;\; ; \;\; \int_{-\infty}^{+\infty} |\Phi(x)|^2\, dx < \infty \tag{4}$$

The functions in this space guaranty the admissibility requirement and the square norm. In principle, every finite-length function or fast-decaying oscillating waveform can be used as a wavelet function. But if the wavelet transform is required to be invertible, not every function can be used. For the wavelet transform to be invertible, the wavelet must satisfy the admissibility criterion:

$$0 < 2\pi \int_{-\infty}^{+\infty} \frac{\left|\hat{\Phi}(\omega)\right|^2}{|\omega|} d\omega < +\infty \tag{5}$$

To determine the admissibility criterion for a given wavelet $\Phi(x)$, first its Fourier transform $\hat{\Phi}(\omega)$ must be calculated. If a wavelet function fulfils the admissibility criterion, it is called an admissible wavelet. Because the integrand in the admissibility criterion has a null in the denominator for $\omega = 0$, the Fourier transform $\hat{\Phi}(0)$ must be zero. Then the admissibility criterion can be simplified to:

$$\hat{\Phi}(0) = \int_{-\infty}^{+\infty} \Phi(x)dx = 0 \tag{6}$$

Thereby the admissibility criterion can be reconsidered with the original wavelet $\Phi(x)$. By the integral wavelet transform, the wavelet functions, with given translation and scaling values, are compared successively with different sections of the function (Fig. 6). The wavelet transform coefficients are determined by scanning the function with the wavelets at different scales. The coefficients describe the conformity of the function at every position along the function with the analyzing wavelets.
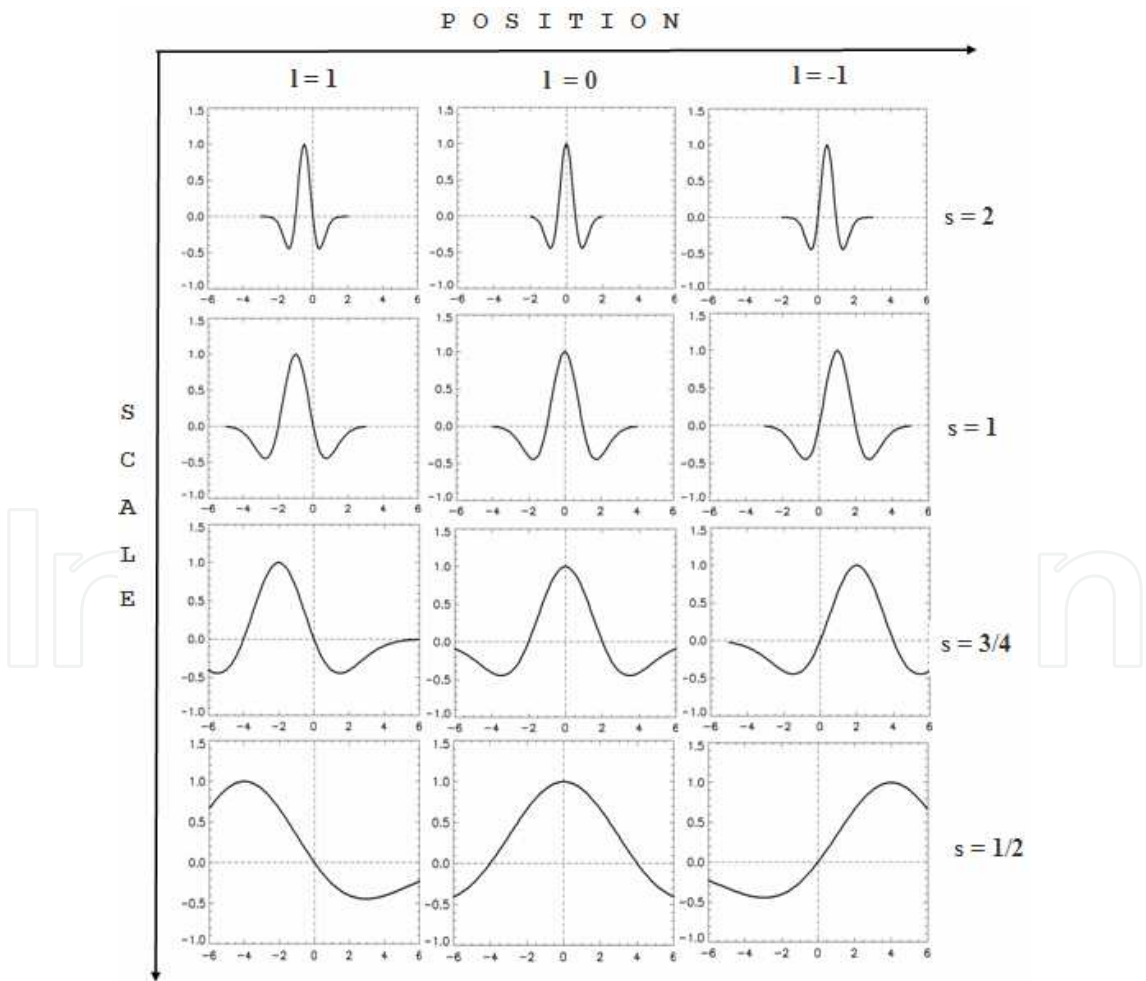

Fig. 5. Starting from the "mother function", the Mexican-hat basis functions are generated by selecting values for l and s

By analyzing a function with differently scaled wavelets, the degree of conformity is determined. A small wavelet (lower scale) at a certain position is well localized and delivers detailed information at this part of the function. An expanded wavelet (larger scale) has a lower spatial resolution and delivers coarse (details are smoothed out) information of the function. By the translation of the wavelets, the function is then analyzed at every position (Fig. 6). The wavelet transform coefficients reflect the degree of correspondence between the wavelet (at a certain position and scale) and the function. The product of a wavelet with a function section gives a new function and the enclosed area (by integrating out) is the corresponding wavelet coefficient. By use of wavelets with different widths, the function is analyzed at different scales (spatial resolution). With the wavelet transform coefficients together with the corresponding wavelets, the original function $f(x)$ can be recovered by the inverse continuous wavelet transform:

$$f(x) = \int\limits_{0}^{+\infty} \int\limits_{-\infty}^{+\infty} \frac{1}{s^2} (W_\Phi f)(s,l) \phi_{s,l}(x) \, dl \, ds \tag{7}$$

$\phi_{s,l}(x)$ is the dual function of $\Phi_{s,l}(x)$ satisfying the condition:

$$\int\limits_{0}^{+\infty} \int\limits_{-\infty}^{+\infty} \frac{1}{|s^3|} \Phi_{s,l}(x') \phi_{s,l}(x) \, dl \, ds = \delta(x - x') \tag{8}$$

The function is synthesized by the superposition of the analyzing wavelets, weighted by their coefficients.

5.2 The discrete wavelet transforms

The discrete wavelet transform is given by extracting a discrete subset of the set of wavelet functions in the continuous wavelet transform. In other words: only wavelet functions with discrete scaling and translation parameters are used. That means the wavelet functions are still continuous, but they exist only at discrete positions with discrete scales (Daubechies, 1988). The basis wavelet functions are translated and dilated by whole numbers $m \in \mathbb{Z}_+$ and $n \in \mathbb{Z}$.

$$\Phi_{m,n}(x) = 2^{-\frac{m}{2}} \Phi(2^{-m} x - n) \tag{9}$$

In spit of the reduction to a discrete subset, it can be shown that the information remains completely preserved. For many applications of the discrete wavelet transform, orthonormal basis wavelets are used.

A function $\Phi \in L^2(\mathbb{R})$ is called an orthonormal wavelet if it can be used to define a Hilbert basis (an orthonormal and complete basis) for the Hilbert space $L^2(\mathbb{R})$ of square integrable functions. The Hilbert basis is constructed as the family of basis functions:

$$\left\{ \Phi_{n,m}(x) = 2^{-\frac{m}{2}} \Phi(2^{-m} x - n) \, \middle| \, (m,n) \in \mathbb{Z}_+ \times \mathbb{Z} \right\} \tag{10}$$
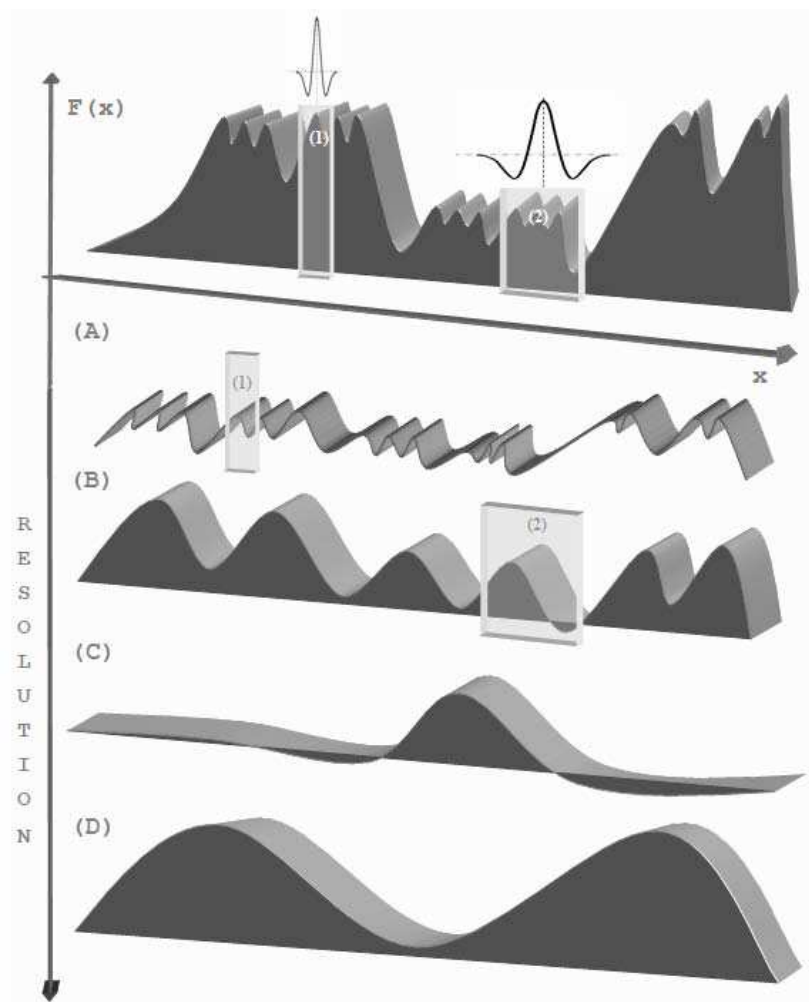
Fig. 6. In wavelet analysis, the wavelet basis functions are compared successively with different sections of the function

by means of dyadic translations and dilations of the wavelet "mother function" Φ. Or in other words: the mother wavelet is the generating function of an affine function system which constitute a Hilbert basis in $L^2(\mathbb{R})$. The family of wavelet basis functions is an orthonormal system if it is orthonormal under the inner product:

$$\left\langle \Phi_{m,n}, \Phi_{m',n'} \right\rangle = \delta_{mm'} \delta_{nn'} \tag{11}$$

Thereby, δ_{nm} is the Kronecker delta and $\left\langle \Phi, \Phi' \right\rangle$ is the standard inner product on $L^2(\mathbb{R})$:

$$\left\langle \Phi_{m,n}, \Phi_{m',n'} \right\rangle = \int\limits_{-\infty}^{+\infty} \overline{\Phi}_{m,n}(x) \cdot \Phi_{m',n'}(x) \, dx \tag{12}$$

The requirement of completeness is that every given function $f \in L^2(\mathbb{R})$ can be decomposed into a linear combination of the basis wavelet functions weighted by wavelet coefficients:

$$f(x) = \sum_{n,m=-\infty}^{+\infty} c_{n,m} \Phi_{n,m}(x) \tag{13}$$

The convergence of the series is understood to be convergence in norm. Such a representation of a function $f(x)$ is known as a wavelet series and the basis functions are the building stones of the function (an orthonormal wavelet is self-dual). Then, $L^2(\mathbb{R})$ is spanned by the discrete set of basis wavelets $\{\Phi_{m,n}|(m,n)\in\mathbb{Z}_+\times\mathbb{Z}\}$, and the discrete wavelet coefficients $\{c_{m,n}=\langle f,\Phi_{m,n}\rangle|(m,n)\in\mathbb{R}_+\times\mathbb{R}\}$ of a function $f\in L^2(\mathbb{R})$ are determined by the discrete wavelet transform:

$$\langle f,\Phi_{m,n}\rangle = \int_{-\infty}^{+\infty} f(x)\bar{\Phi}_{m,n}(x)\,dx \tag{14}$$

Any function $f\in L^2(\mathbb{R})$ is therefore characterized by the set of its discrete wavelet coefficients $\{\langle f,\Phi_{m,n}\rangle|(m,n)\in\mathbb{Z}_+\times\mathbb{Z}\}$ and it is possible to recover any function from its discrete wavelet coefficients in a numerically stable procedure. (Numerically stable means that small perturbations in the wavelet coefficients correspond to small perturbations of the function).
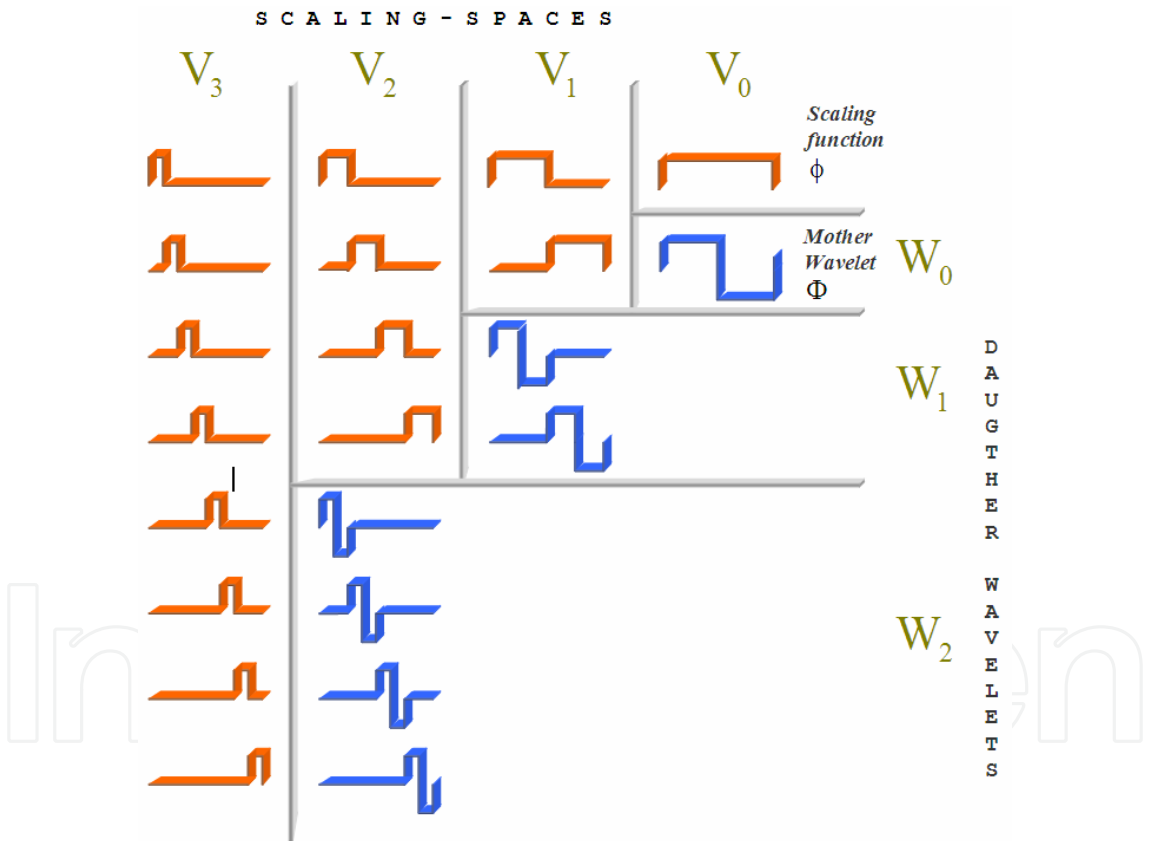

Fig. 7. The subspaces V_j represent the space of the scaling function and the W_j are spanned up by the wavelets

For any admissible wavelet, spanning up a basis $\Phi_{m,n}$ for $f\in L^2(\mathbb{R})$, the following condition holds:

$$\sum_{m,n}|< f,\Phi_{m,n}>|^2 \le C\|f\|_2^2 \tag{15}$$

Thereby, C is a given constant value. Then, the discrete wavelet coefficients belong to:

$$l^2(\mathbb{Z}_+ \times \mathbb{Z}) = \left\{ c_{m,n} \mid \sum_{m,n} |c_{m,n}|^2 < \infty \right\} \tag{16}$$

Therefore the discrete wavelet transform $DWT : f \to \left\{ \langle f, \Phi_{m,n} \rangle \right\}$ is a mapping from $L^2(\mathbb{R})$ (the space of square integrable functions) into $l^2(\mathbb{Z}_+ \times \mathbb{Z})$ (the space of square-summable, infinite sequences of numbers). For the wavelet coefficients, the following conditions are valid:

$$\forall_{m,n \in \mathbb{Z}_+ \times \mathbb{Z}} : \langle f, \Phi_{m,n} \rangle = \langle g, \Phi_{m,n} \rangle \Leftrightarrow f = g ; \quad \forall_{m,n \in \mathbb{Z}_+ \times \mathbb{Z}} : \langle f, \Phi_{m,n} \rangle = 0 \Leftrightarrow f = 0 \tag{17}$$

Then, numerical stability is defined by: given two functions $f \in L^2(\mathbb{R})$ and $g \in L^2(\mathbb{R})$, the characterization of the functions by their wavelet coefficients in $l^2(\mathbb{Z}^2)$ is stable if: whenever two sequences of wavelet coefficients in $l^2(\mathbb{Z}^2)$ are close, the corresponding functions are close in $L^2(\mathbb{R})$. As in the case of continuous wavelets, there also exists a variety of discrete wavelets; the Coiflet wavelets, the Legendre wavelets, the Haar wavelets, etc. For illustration, the Haar wavelet is presented (Fig. 7). It is now recognized as the first known wavelet and proposed 1909 by A. Haar (Alfréd Haar was a Hungarian mathematician). The Haar wavelet is also the simplest possible wavelet. The Haar wavelets give an example of a countable orthonormal system for the space of square-integrable functions on the real line. A large subclass of wavelets arises from special structures imposed on $L^2(\mathbb{R})$, known as multi-resolution-analysis (MRA). MRA is a framework for understanding and constructing wavelet bases, which generates discrete wavelet families of dilations and translations that are orthonormal bases in $L^2(\mathbb{R})$.

5.3 Multi-resolution analysis

Multi-resolution analysis (MRA) is a design method to generate a lot of practically relevant discrete wavelet transforms (DWT) and is the mathematical basis of the fast wavelet transform (FWT). The MRA satisfies certain self-similarity relations in spatial and the scale domain as well as completeness and regularity relations (Daubechies & Lagarias, 1991). The MRA describes the approximation properties of the discrete wavelet transform. The generated wavelet transformation is iterative and the analyzed function is split in successive "smoother" versions, which contain by progressive iterations successively poorer information. Of special importance in the multi-resolution analysis are the so called two scale equations for a scaling function and a wavelet function. The MRA is defined as follows: A MRA of the space $L^2(\mathbb{R})$ consists of a sequence of nested closed subspaces: $\left\{ V_n \mid n \in \mathbb{Z} \right\}$ with the following properties:

a. Nesting property: $\{0\} ... \subset V_0 \subset ... V_n \subset V_{n+1} \subset ... \subset L^2(\mathbb{R})$

b. Invariance under translation: Self-similarity in the spatial domain requires that each subspace V_k is invariant under shifts by integer multiples of 2^{-k}:

$$f(x) \in V_k \Leftrightarrow f(x + m2^{-k}) \in V_k ; \; (m \in \mathbb{Z})$$

c. Scaling property: Self-similarity in scale requires that all nested subspaces $V_k \subset V_l$ with $k < l$ are scaled versions of each other, with the scaling factor 2^{-k}:

$$f(x) \in V_k \Leftrightarrow f(2^{l-k} x) \in V_l$$

d. Existence of a scaling function: This requires the existence of an orthogonal basis $\{\phi_{0n} | n \in \mathbb{Z}\}$ for the subspace $V_0 \subset L^2(\mathbb{R})$, where:

$$\phi_{jk}(x) = 2^{-j/2} \phi(2^{-j} x - k)$$

e. Completeness requires that the nested subspaces V_m fill the whole space, i.e., their union should be dense in $L^2(\mathbb{R})$:

$$\bigcup_{m \in \mathbb{Z}} : V_m = L^2(\mathbb{R})$$

f. The lack of redundancy requires that the intersection of the nested subspaces V_m should only contain the zero element:

$$\bigcap_{m \in \mathbb{Z}} V_m = \{0\}$$

The generating functions ϕ are known as scaling functions or "father wavelet". The structure of the MRA allows several conclusions for the construction of wavelet bases for practical applications. The orthonormal basis of each subspace, their scaling properties and resolutions follow from the MRA definition. The scaling property (c) implies that each subspace V_j is a scaled version of the central subspace V_0 (Fig. 7). Together with the invariance under translation (b) and the existence of a scaling function (d) it implies that $\{\phi_{jn} | n \in \mathbb{Z}\}$ is an orthonormal Hilbert basis of the subspace V_j:

$$\langle \phi_{j,k}, \phi_{j,k'} \rangle = \delta_{kk'} \qquad (18)$$

Then, the sequence of scaling subspaces can be defined by spaces that are spanned off by a proprietary orthonormal basis system:

$$V_i = span\{\phi_{i,k} | k \in \mathbb{Z}\} \qquad (19)$$

Further, the scaling property (c) implies that the resolution of the l-the subspace V_l is higher than the resolution of the k-the subspace V_k ($\forall_{k<l} : V_k \subset V_l$). In the case that there exists only one scaling function ϕ in the MRA which generates a Hilbert basis in V_0, the scaling function satisfies the two scale equation (or refinement equation):

$$\phi(x) = \sum_{k=-N}^{N} a_k \phi(2x - k) \qquad (20)$$

Because of the nesting property $V_0 \subset V_1$, there exists a finite sequence of coefficients $a_k = 2\langle \phi(x), \phi(2x - k) \rangle$, for $|k| \leq N$ and $a_k = 0$ for $|k| > N$ |. This sequence of coefficients (real numbers) is called scaling sequence or filter (scaling) mask and is given by:

$$\{..., 0, a_{-N}, ..., a_0, ..., a_N, 0, ...\} \qquad (21)$$

To fulfil filter properties, several conditions must be imposed on the coefficients of the scaling sequence. To demonstrate this, first the Fourier transform $\hat{\phi}$ of the scaling function ϕ is calculated:

$$
\begin{aligned}
\hat{\phi}(\omega) &= \frac{1}{2}\sum_k a_k e^{-i\frac{1}{2}\omega k}\int \phi(2x-k)e^{-i\omega x}dx \\
&= \frac{1}{2}\sum_k a_k e^{-i\frac{1}{2}\omega k}\hat{\phi}(\tfrac{\omega}{2})
\end{aligned}
\tag{22}
$$

With: $\hat{a}(\omega) = \frac{1}{2}\sum_k a_k e^{-i\omega k}$ and $\hat{\phi}(\omega) = \int \phi(2x-k)e^{-i\omega x}dx$, the resulting equation can be formulated as follows:

$$
\hat{\phi}(\omega) = \hat{a}(\tfrac{\omega}{2})\hat{\phi}(\tfrac{\omega}{2})
\tag{23}
$$

The requirements for a low-pass filter are that the Fourier series must have the value 1 at the zero point $\omega = 0$ ($\hat{a}(0) = 1$) and the value 0 at $\omega = \pi$ ($\hat{a}(\pi) = 0$). From these requirements follows immediately the following restrictions for the coefficients of the scaling sequence:

$$
\sum_{k=-N}^{N} a_k = 2 \;\; ; \;\; \sum_{k=-N}^{N} (-1)^k a_k = 0
\tag{24}
$$

One task of the wavelet design is to impose several conditions on the coefficients a_k in order to obtain the desired properties of the scaling function ϕ. For example: if ϕ is required to be orthogonal to all dilations of itself, the coefficients of the scaling sequence must fulfil the following conditions:

$$
\forall_{m \in \mathbb{Z}\setminus\{0\}} : \sum_{k=-N}^{N} a_k a_{k+2m} = 0 \;\; ; \;\; \sum_{k=-N}^{N} a_k^2 = 2
\tag{25}
$$

This can easily be seen, since: $\langle \phi, \phi \rangle = 1$ and $\langle \phi_{j,k}, \phi_{j,k'} \rangle = \delta_{kk'}$. The requirement that the scaling sequence is orthogonal to any shifts of it by an even number of coefficients is the necessary condition for the orthogonality of the wavelets. In terms of the Fourier transform the orthogonality is given by the relation:

$$
|\hat{a}(\omega)|^2 + |\hat{a}(\omega + \pi)|^2 = 1
\tag{26}
$$

The trigonometric polynomial $\hat{a}(\omega)$ (filter function or transfer function) plays an important role in wavelet theory.
The "mother wavelet" is defined by a similar two scale equation as follows:

$$
\Phi(x) = \sum_{k=-N}^{N} b_k \phi(2x-k)
\tag{27}
$$

Where the sequence of coefficients, called a wavelet sequence or wavelet mask, is given by:

$$\{...,0,b_{-N},...,b_0,...,b_N,0,...\} \text{ with: } b_k = (-1)^k a_{K-k} \tag{28}$$

Whereby, $K \in \mathbb{Z}$ is an arbitrary odd number. The corresponding wavelet subspaces W_i are spanned up by:

$$W_j = span\{\Phi_{j,k}(x) = 2^{-j/2}\Phi(2^{-j}x - k) | k \in \mathbb{Z}\} \tag{29}$$

The space $W_0 \subset V_1$ is defined as the linear hull of the mother wavelets integer shifts. W_0 is the orthogonal complement to V_0 inside V_1 (that means: V_1 is the orthogonal sum of W_0 and V_0). By successive application of the orthogonal sum, the orthogonal dissection of the scaling spaces is obtained:

$$V_2 = V_1 \oplus W_1 = V_0 \oplus W_0 \oplus W_1 \tag{30}$$

(Explicitly, this formal notation is given by: $\forall_{u \in V_{i+1}} : u = v + w$ with: $v \in V_i$ and $w \in W_i$). This is illustrated for the Haar wavelet in Figure 7. By self-similarity, there exist scaled versions W_k of W_0 and by completeness one has:

$$L^2(\mathbb{R}) = closure \, of \bigoplus_{k \in \mathbb{Z}} W_k \tag{31}$$

This is the basic analytical requirement for an MRA. Starting from the two scale equation, respectively the equation for the "mother wavelet", the multi-resolution decomposition of $f \in L^2(\mathbb{R})$ follows by calculating the wavelet coefficients.

5.4 Multi-resolution wavelet decomposition of functions

The next task is to compute the wavelet coefficients $\langle f, \Phi_{lk} \rangle$ of the discrete wavelet transform. Starting point is the two scale equation for the "mother wavelet" (formula 27). Then the "daughter wavelets" $\Phi_{j,k}$ are generated from the "mother wavelet" Φ and inserted into the equation (by use of $x = 2^{-j}x - k$ in $\phi(2x - k)$) :

$$\begin{aligned} \Phi_{j,k} &= 2^{-j/2}\Phi(2^{-j}x - k) \\ &= 2^{-j/2}\sum_n b_n 2^{1/2}\phi(2^{-j+1}x - 2k - n) \end{aligned} \tag{32}$$

The expression can be rewritten as:

$$\Phi_{j,k}(x) = \sum_n b_{n-2k}\phi_{j-1,n} \tag{33}$$

By use of the inner product, the wavelet coefficients are then calculated immediately by:

$$\langle f, \Phi_{j,k} \rangle = \sum_n b_{n-2k}\langle f, \phi_{j-1,n} \rangle \tag{34}$$

In other words, if the sequence of coefficients $\left\{\left\langle f,\phi_{j-1,n}\right\rangle \middle| n\in\mathbb{Z}\right\}$ is known, then the sequence $\left\{\left\langle f,\Phi_{j,k}\right\rangle \middle| k\in\mathbb{Z}\right\}$ can be obtained by convolution of $\left\{\left\langle f,\phi_{j-1,n}\right\rangle \middle| n\in\mathbb{Z}\right\}$ with the sequence $\left\{b_{-n}\middle| n\in\mathbb{Z}\right\}$ where only the even indexed terms of the resulting sequence are retained. The same procedure can be done for the scaling function by use of its two scale equation (formula 20). As in the case of the "mother wavelet", results for the scaling function ("father wavelet"):

$$\phi_{j,k}(x)=\sum_{n}a_{n-2k}\,\phi_{j-1,n}(x) \tag{35}$$

And the wavelet coefficients are given by the convolution:

$$\left\langle f,\phi_{j,k}\right\rangle=\sum_{n}a_{n-2k}\left\langle f,\phi_{j-1,n}\right\rangle \tag{36}$$

The connection between both procedures is illustrated by the following:

$$
\begin{array}{cccc}
\left\{\left\langle f,\phi_{0,n}\right\rangle\middle| n\in\mathbb{Z}\right\} \to & \left\{\left\langle f,\phi_{1,n}\right\rangle\middle| n\in\mathbb{Z}\right\} \to & \left\{\left\langle f,\phi_{2,n}\right\rangle\middle| n\in\mathbb{Z}\right\} \to & \dots \\
\searrow & \searrow & \searrow & \\
\left\{\left\langle f,\Phi_{1,n}\right\rangle\middle| n\in\mathbb{Z}\right\} & \left\{\left\langle f,\Phi_{2,n}\right\rangle\middle| n\in\mathbb{Z}\right\} & \left\{\left\langle f,\Phi_{3,n}\right\rangle\middle| n\in\mathbb{Z}\right\} & \dots
\end{array} \tag{37}
$$

Thus, successively coarser approximations of the function $f\in L^2(\mathbb{R})$ are computed along with the difference in information between successive levels of approximation. The wavelet decomposition can be considered as an orthonormal basis transformation:

$$\left\{\phi_{j,n}\middle| n\in\mathbb{Z}\right\}\to\left(\left\{\phi_{j+1,n}\middle| n\in\mathbb{Z}\right\},\left\{\Phi_{j+1,n}\middle| n\in\mathbb{Z}\right\}\right) \tag{38}$$

on the coefficients of the projections. f^j is the projection of f onto the subspace V_j ($f^j=\sum_{k}\left\langle f,\phi_{j,k}\right\rangle\phi_{j,k}$), g^j is the projection of f onto the subspace W_j ($g^j=\sum_{k}\left\langle f,\Phi_{j,k}\right\rangle\Phi_{j,k}$). Then it results for the projections:

$$f^{j-1}=f^j+g^j \tag{39}$$

The wavelet decomposition of a function f means that f is successively projected onto the subspaces V_j and W_j. It is a fine-to-coarse decomposition. By use of $s^j=\left\{s_n^j=\left\langle f,\phi_{j,n}\right\rangle\right\}$ and $d^j=\left\{d_n^j=\left\langle f,\Phi_{j,n}\right\rangle\right\}$, the equations for the wavelet coefficients can be rewritten as:

$$s_k^j=\sum_{n}(a_{n-2k})s_n^{j-1}\;;\;d_k^j=\sum_{n}(b_{n-2k})s_n^{j-1} \tag{40}$$

The d^j are the coefficients of the projection g^j onto the subspace W_j. d^j is the sequence of wavelet coefficients representing the difference in information between two consecutive

levels of resolution of the function f. The coefficients d^j represent detail (fine) information about the function f. s^j is the sequence of wavelet coefficients of the projection of f onto V_j. The coefficients s^j represent smoothed (coarse) information of the function. From the consecutive decompositions of the function f results, after a finite number of iterations, a finite sequence of coefficients:

$$s^0 \rightarrow \left(d^1, d^2, d^3, ..., d^j, ..., d^N, s^N \right) \tag{41}$$

s^0 is the coefficient sequence of the function f^0 (projection in V_0). The procedures discussed above define wavelets abstractly by their properties. They enable a systematical construction of wavelet bases with certain desired properties. The abstract formulation can be illustrated by a wavelet decomposition of a function with filter cascades. This enables also the implementation of the fast wavelet transform. The fast wavelet transform is an algorithm, which implements the discrete wavelet transform by use of the multi scale analysis. By this procedure, the calculation of the inner product of the function with every basis wavelet is replaced by a successive dissection of the frequency bands. This is realized by a sequence of filters. The next step is to determine the scaling sequence $\{a_k\}$ and the wavelet sequence $\{b_k\}$.

5.5 Construction of wavelets

The Z-transform enables it to express the decomposition procedure in a form that is more suitable for the discussion in terms of sub-band filtering. The Z-transform converts a discrete function (sequence of numbers), into a special complex representation. Given a discrete sequence $\{a_n | n \in \mathbb{Z}\}$, the Z-transform is defined by:

$$a(Z) = \sum_n a_n Z^n \tag{42}$$

Z is in general a complex number ($Z = Ae^{i\varphi}$, where A is the magnitude of Z, and φ is the complex argument). In terms of the Z-transform, the transfer function $\hat{a}(\omega)$ is expressed as:

$$\hat{a}(\omega) = a(e^{-i\omega}) \tag{43}$$

Then the conditions for the coefficients of the scaling sequence to fulfil a low-pass filtering are (formula 24):

$$a(1) = 2 \quad \text{and} \quad a(-1) = 0 \tag{44}$$

We demonstrate the construction of orthogonal wavelets on hand of the Daubechies wavelets (named after Ingrid Daubechies, a Belgian physicist and mathematician).
In other words, the scaling sequence $\{a_k\}$ and in consequence the wavelet sequence $\{b_k\}$ are determined (Fig. 8). The Daubechies wavelet transform can be easy implemented for practical purposes by use of the fast wavelet transform. Ingrid Daubechies assumes that $\hat{a}(\omega)$ has an A-fold zero at the values $\omega = \pm\pi$. Therefore she formulated the following expression for $\hat{a}(\omega)$:
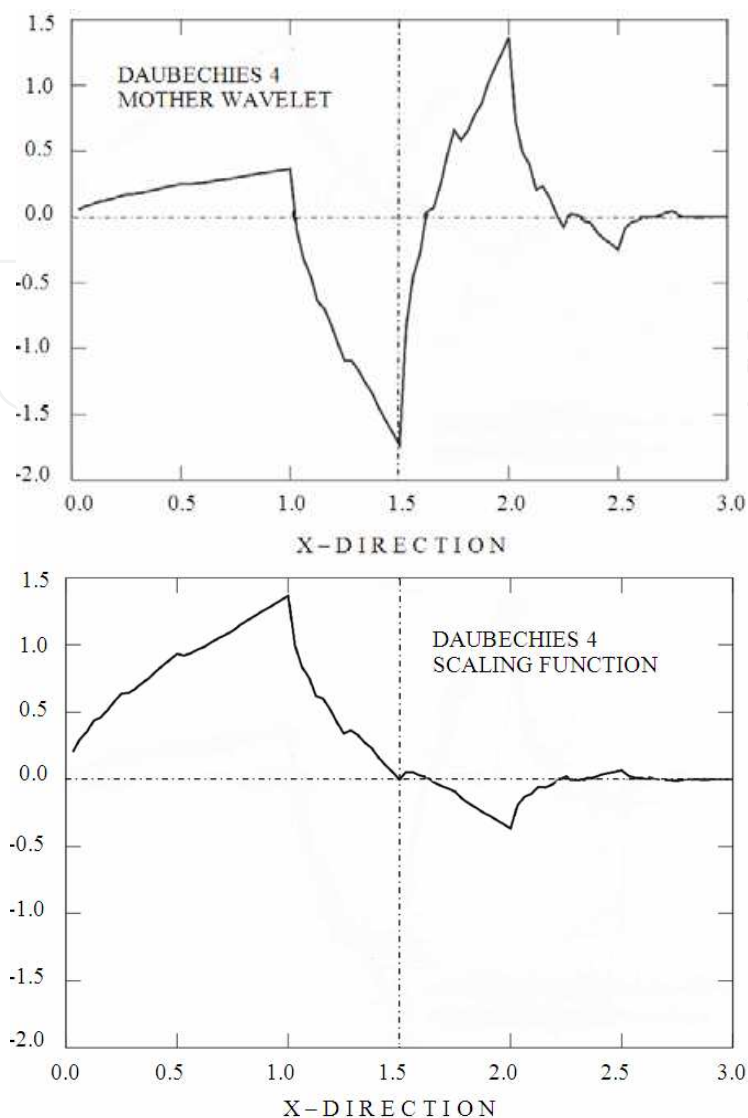
Fig. 8. Daubechies wavelets can not be represented analytically; they are generated by an iteration procedure

$$\hat{a}(\omega) = \left(\frac{1 + e^{-i\omega}}{2} \right)^{A} \sum_{n=0}^{A-1} q_n e^{-in\omega} \tag{45}$$

This is called the factorization of the scaling sequence. As consequence, there is an A-fold zero of $\hat{a}(\omega + \pi)$ and $\hat{a}(-\omega - \pi)$ at $\omega = 0$. The general representation for a scaling sequence of an orthogonal discrete wavelet transform with approximation order A, is given by the following factorization of the scaling sequence:

$$a(Z) = 2^{-A} \left(1 + Z \right)^{A} p(Z) \tag{46}$$

$p(Z)$ is a polynomial in Z with degree A-1 and $p(1) = 1$.

$$p(Z) = \sum_{n=0}^{A-1} q_n Z^n \tag{47}$$

The equation for $\hat{a}(\omega)$ results by setting $Z = e^{-i\omega}$. The Daubechies wavelets are characterized by a maximal number of vanishing moments (A) for some given support. The maximal potency A, enabling $(1+Z)^A$ to be a factor of $a(Z)$ is called the polynomial approximation degree. This degree reflects the ability of the scaling function ϕ to represent polynomial until the degree A-1 as a linear combination of translations (by whole numbers) of the scaling function. Then, by use of the Z-transform, the orthogonality condition (formula 25) can be written as:

$$a(Z)a(Z^{-1}) + a(-Z)a(-Z^{-1}) = 4 \tag{48}$$

The index number of a Daubechies DN wavelet refers to the number N of coefficients. Each wavelet has a number of vanishing moments (zero moments) equal to half the number of coefficients. (N=2A). In the case of Daubechies 4 (A=2), the polynomial p(Z) is linear and the following function can be used:

$$a(Z) = \frac{1}{4}(1+Z)^2\left((1+Z) + c(1-Z)\right) \tag{49}$$

By use of the factorization of the scaling sequence, the orthogonality condition can be formulated as Laurent polynomial:

$$(1-u)^A p(Z)p(Z^{-1}) + u^A\left(p(-Z)p(-Z^{-1})\right) = 1 \tag{50}$$

Whereby u is defined as: $u := \frac{1}{4}\left(2 - Z - Z^{-1}\right)$. By use of: $P_A(u) = p(Z)p(Z^{-1})$, the equation can be rewritten as:

$$(1-u)^A P_A(u) + u^A P_A(1-u) = 1 \tag{51}$$

A polynomial solution of this equation is the following binominal expansion:

$$P_A(u) = \sum_{k=0}^{A-1}\binom{A+k-1}{k}u^k \tag{52}$$

This polynomial plays an important role in the construction of the Daubechies wavelets. There exists a close connection between the zeros of $P_A(u)$ and the N scaling coefficients (filter coefficients) a_k of the Daubechies wavelet DN. When $P_A(u)$ is known, the product $p(Z)p(Z^{-1})$ can be calculated and the coefficients q_n be determined. In the case of N=4 (A=2, Daubechies D4), the polynomial till degree A-1 is given by $P_2(u) = 1 + 2u$ and the following equation is valid (with $P_2(u) = p(Z)p(Z^{-1})$):

$$1 + 2u = (q_0 + q_1)^2 - 4u q_0 q_1 \tag{53}$$

This equation can be solved for q_0 and q_1. By use of $\hat{a}(0) = 1$ results: $q_0 + q_1 = 1$ and furthermore, by use of $\hat{a}(\omega)$ and the factorization of the scaling sequence (till A=2) results: $q_0 - q_1 = \sqrt{3}$. Then q_0 and q_1 are given by adding and subtracting the two relations:

$$q_0 = \frac{1+\sqrt{3}}{2} \;;\; q_1 = \frac{1-\sqrt{3}}{2} \tag{54}$$

The scaling coefficients are then calculated by equating $\hat{a}(\omega)$ and the factorization of the scaling sequence:

$$(a_0 + a_1 Z^{-1} + a_2 Z^{-2} + a_3 Z^{-3}) = \left(\frac{1+Z^{-1}}{2}\right)^2 (q_0 + q_1 Z^{-1}) \tag{55}$$

This gives the coefficients of the scaling sequence for Daubechies D4:

$$a_0 = \frac{1+\sqrt{3}}{4}, a_1 = \frac{3+\sqrt{3}}{4}, a_2 = \frac{3-\sqrt{3}}{4}, a_3 = \frac{1-\sqrt{3}}{4} \tag{56}$$

The wavelet coefficients b_k follow from the scaling coefficients a_k by use of $b_k = (-1)^k a_{K-k}$ (formula 28). In the case of Daubechies 4, the index K has the value 3. In the next chapter, the Daubechies 4 scaling and wavelet coefficients are used as filter coefficients to realize the wavelet transform as filter operations.

5.6 The relation between the discrete wavelet transform and digital filters

By use of discrete (digitized) data, the function is represented by a data vector with N components:

$$f = \left(f(1), f(2), \ldots, f(N-1), f(N)\right)^{\dagger} \tag{57}$$

N is the dimension of the data vector or the number of the discrete data points. There exists a relation between the wavelet theory and the digital filter theory (Strang & Nguyen, 1996). Like the wavelet transform, digital filters extract details (high-pass filter) or smooth out details (low-pass filter). This correspondence between wavelets and digital filters enables the realization of a wavelet transform operation without an explicit formulation of basis wavelet functions. Just the scaling coefficients $\{a_n\}$ and the wavelet coefficients $\{b_n\}$ are necessary and in addition the transformation of the coefficients $s_n^j = \langle f, \phi_{j,n} \rangle$ and $d_n^j = \langle f, \Phi_{j,n} \rangle$, resulting from the two scale equations (formula 20 and formula 27). The transformation can be realized by definition and use of adequate filter operations. High-pass filters, which detect details (variations) in the data vector, are realized by subtraction operations. Along the data vector, the differences between the values of neighbour data components are considered: if there are great differences, the subtraction operation equals a finite number, if the values are nearby equal, the subtraction operation equals zero. Low-pass filters are realized by taking the mean value of neighbour data components: variations of the values are smoothed out. This corresponds with the dissection of a function f in a high-pass part in the subspace W_i and a low-pass part in the subspace V_i (beginning with the subspace V_{i+1}). The task of the scaling function is the generation of mean values of the data components (smoothing). It defines the global level of operation. The wavelet function detects the details of the data components.

The discrete wavelet transform can therefore be implemented and performed as convolution of the data vector with a filter bank. To realize a discrete wavelet transform by a filter operation, first a transformation matrix with appropriate filter properties is defined. In terms of the matrix operation, the scaling and wavelet functions apply to matrix rows (row basis). The transformation matrix consists of a set of filter coefficients that define special filter characteristics along the matrix rows. The matrix structure is determined by these filter coefficients, which are ordered using two dominant patterns: one works as a smoothing filter, the other works as a sharpening filter to bring out the data detail information. The associated wavelets are represented by the filter coefficients. Alternating, the even rows act as a high-pass filter and the odd rows act as a low-pass filter. The filter coefficients are associated with the corresponding "mother wavelet" and the scaling function, which determine the number of filter coefficients. The filter matrix for Daubechies 4 is defined as:

$$WT = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & & & & \\ b_0 & b_1 & b_2 & b_3 & & & & \\ & & a_0 & a_1 & a_2 & a_3 & & \\ & & b_0 & b_1 & b_2 & b_3 & & \\ & & & & & & \ddots & \\ & & & & a_0 & a_1 & a_2 & a_3 \\ & & & & b_0 & b_1 & b_2 & b_3 \\ a_2 & a_3 & & & & & a_0 & a_1 \\ b_2 & b_3 & & & & & b_0 & b_1 \end{pmatrix} \qquad (58)$$

The filter coefficients $[a_0, a_1, a_2, a_3]$ define the scaling function (low-pass filter) and the filter coefficients $[b_0, b_1, b_2, b_3]$ define the wavelet function (high-pass filter). Therefore, a data vector is decomposed into parts containing "smooth" information and parts with "detail" information. The filter coefficients are related according formula (28): $b_k = (-1)^k a_{3-k}$ and the transformation is orthogonal if the filter coefficients satisfy formula (25): $\sum_k a_k^2 = 1$ and $\sum_k a_k a_{k+2} = 0$. The transformation is then an image decomposition through the filter bank, in which the low-passed (smooth) and the high-passed (detail) parts serve as input for the next level (Fig. 9). This is called a pyramidal algorithm and the multi-resolution analysis is the theoretical foundation of the filter bank. This is an iterative procedure where the resulting output data vectors are used successively as input vectors (i) for the matrix operation in the next level (i+1):

$$f^{i+1} = WT \cdot f^i \qquad (59)$$

The dimension of the transformation matrix is defined by the length of the input data vector. It should be noticed that by the first iteration, the data vector, with the analyzing data, is used as the input vector and the output vector contains the smooth and detail parts of the data components.

5.7 The 1-dimensional and 2-dimensional discrete wavelet transform

The matrix is applied to the data and produces alternating "smooth" (even rows) and "detail" (odd rows) components:

$$c^i(n) = \begin{cases} d^i(n); \ High-pass\ filtering\ with: [b_0, b_1, b_2, b_3] \\ s^i(n); \ Low-pass\ filtering\ with: [a_0, a_1, a_2, a_3] \end{cases} \qquad (60)$$

The resulting output data vector is subsequently arranged into a part with "detailed" components and a part with "smooth" components (Fig. 9). Then the matrix is applied on the "smooth" part (halved vector, down sampling) of the output vector, then again on the "smooth-smooth" part of the resulting new output vector and so on until a trivial number of "smooth-…-smooth" components remain.

$$\begin{pmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \\ f(5) \\ f(6) \\ f(7) \\ f(8) \\ f(9) \\ f(10) \\ f(11) \\ f(12) \end{pmatrix} \xrightarrow[WT]{\substack{LEVEL\ 1 \\ Multipl.}} \begin{pmatrix} s^1(1) \\ d^1(1) \\ s^1(2) \\ d^1(2) \\ s^1(3) \\ d^1(3) \\ s^1(4) \\ d^1(4) \\ s^1(5) \\ d^1(5) \\ s^1(6) \\ d^1(6) \end{pmatrix} \xrightarrow[Coeff.]{Sort.} \begin{pmatrix} s^1(1) \\ s^1(2) \\ s^1(3) \\ s^1(4) \\ s^1(5) \\ s^1(6) \\ \hline d^1(1) \\ d^1(2) \\ d^1(3) \\ d^1(4) \\ d^1(5) \\ d^1(6) \end{pmatrix} \xrightarrow[WT]{\substack{LEVEL\ 2 \\ Multipl.}} \begin{pmatrix} s^2(1) \\ d^2(1) \\ s^2(2) \\ d^2(2) \\ s^2(3) \\ d^2(3) \\ \hline d^1(1) \\ d^1(2) \\ d^1(3) \\ d^1(4) \\ d^1(5) \\ d^1(6) \end{pmatrix} \xrightarrow[Coeff.]{Sort.} \begin{pmatrix} s^2(1) \\ s^2(2) \\ s^2(3) \\ \hline d^2(1) \\ d^2(2) \\ d^2(3) \\ \hline d^1(1) \\ d^1(2) \\ d^1(3) \\ d^1(4) \\ d^1(5) \\ d^1(6) \end{pmatrix} \xrightarrow[WT]{\substack{LEVEL\ 3 \\ Multipl.}} \bullet\bullet\bullet \quad (61)$$

The "detailed" parts obtained always remain conserved at the next operation levels. During the iterative procedure, the "smooth" parts of the resulting output vectors are successively filtered and their components rearranged in smooth ($s^i(n)$) and detail ($d^i(n)$) parts in each operation level. The output of the wavelet transform consists of the remaining "smooth-…-smooth" components and all the accumulated "detailed" components. The application of the filter bank to the smooth part, successively decimated by factor 2, corresponds to a successive stretching of the scaling function. In other words: the data vector is analyzed at successively larger scales (from high resolution to low resolution). At every operation level, the input data vectors are reduced to the half resolution (sub sampling). By the halving of the resolution at every level, the respective output vector always has the same number of components (wavelet coefficients) as the respective input vector. There is no loss of information during the procedure.

The 2-dimensional wavelet transforms works similarly to the 1-dimensional case. In the first step (or level), the rows of the input image are filtered by a high-pass filter and independently by a low-pass filter (Fig. 10). From both operations two images (sub-bands) result: one shows details and the other is smoothed out. Every second column is then
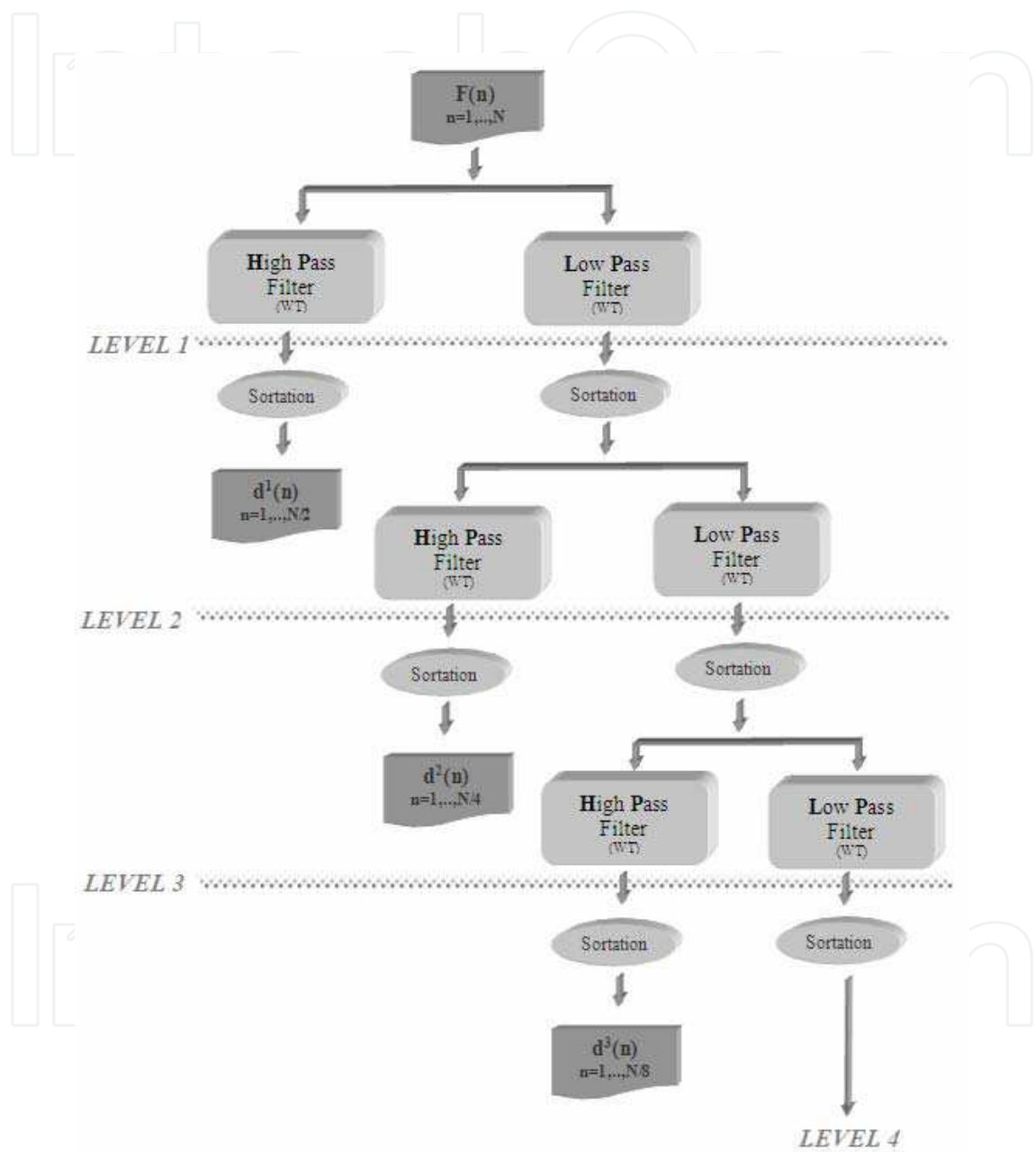
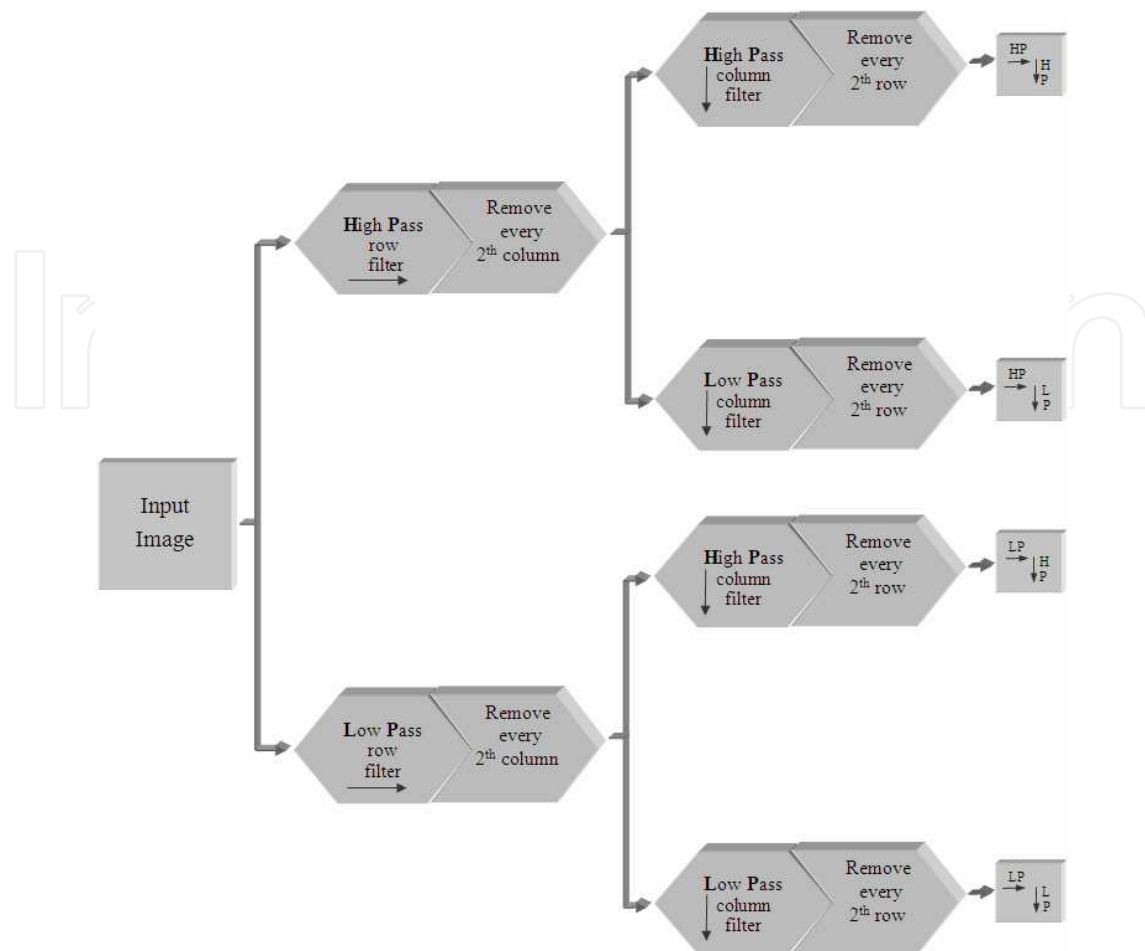Fig. 9. The 1-dimensional wavelet transformation as filter operation

Fig. 10. The 2-dimensional wavelet transformation as filter operation

removed in both sub-bands, which corresponds to a sub sampling (half resolution). Subsequently the columns of both sub-bands are high-pass filtered and low-pass filtered. Again two sub-bands result from each of the two input sub-bands. A total of four sub-bands, which differ by the kind of filtering, are obtained in the first step. At the end of the first step, every second row is removed from each of the four sub-bands. Only the double low-passed sub-band is used in the next step, the other three sub-bands remain conserved for the next operation steps. During the next steps, the same procedure is used again and again. At each step the double high-passed sub-band, which results from the previous step, is used as an input sub-band. At every step, the resulting sub-bands are reduced to the half resolution. The sub-bands with higher spatial resolution contain the detail information (high frequencies) whereas the sub-bands with the low resolution represent the large scale coarse information (low frequencies). After the dissection of the quadratic sub-bands, they are usually arranged in a quadratic configuration where the sub-bands of the first level fill 3/4 of the square; the sub-bands of the second level fill 3/16 of the square; etc. (Fig. 11).

As in the case of the 2-dimensional function, for the 2-dimensional image array, the wavelet transform is computed along each dimension, e.g.: the array is first filtered on its first index (row), then on its second (column) and so on. The multi resolution analysis takes scale information into consideration and successively decomposes the original image into approximations (smooth parts) and details. That means, by the wavelet transformation, the two dimensional image array is split up into several frequency bands (containing various

numbers of wavelet coefficients), which represent information at different scales. At each scale the original image is approximated with more or fewer details.
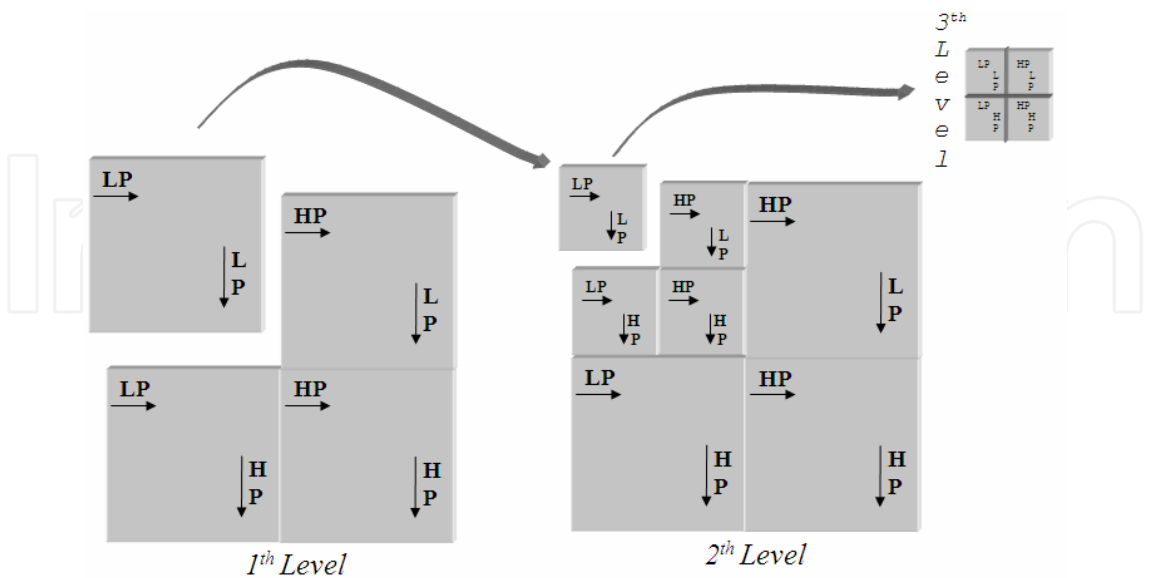

Fig. 11. Dissection of the sub-bands in the 2-dimensional wavelet transformation
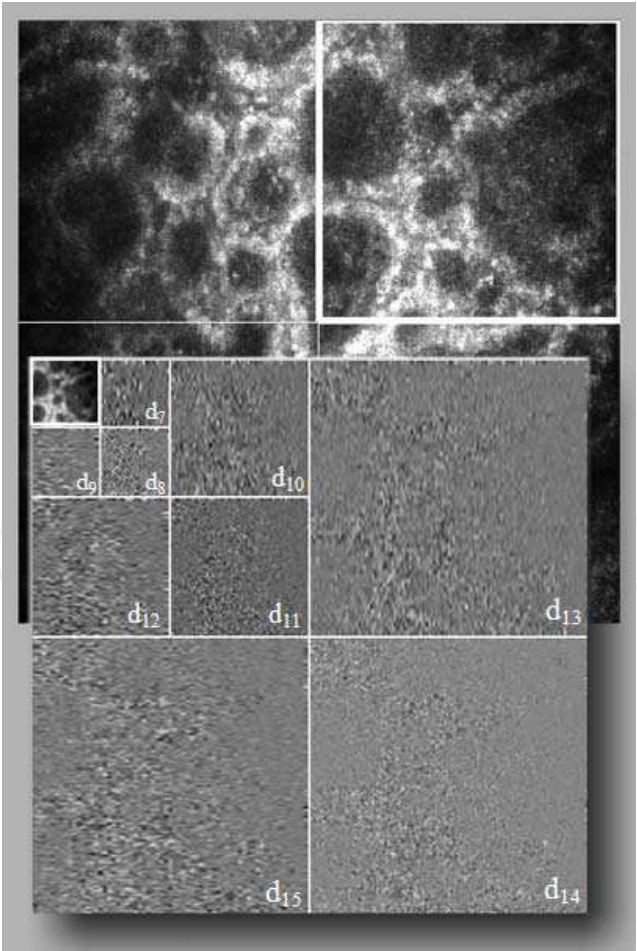

Fig. 12. Wavelet frequency band structure in the 2-dimensional wavelet transformation

The output of the last low-pass filtering is the mean gray level of the image. The frequency bands, representing information at large scale, are labelled with low indices and the frequency bands representing successively decreasing scales are labelled with higher indices (Fig. 12). Then the architectural structure information is accumulated along the energy bands (from course to fine), enabling the analysis of a given texture by its frequency components:

$$d_i = (c_i(k,l)) \tag{62}$$

where the subset of coefficients ($(c_i(k,l))$) are the coefficients contained in the i-th frequency band. At increasing frequency bands the detail structures are smoothed out and large architectural structures of the texture are analysed. At decreasing frequency bands, successively finer structures and details are registered. In wavelet texture analysis, the features are mostly derived from statistical properties of the wavelet coefficients. For the texture analysis, we choose the Daubechies 4 wavelets transform, because this has good localization properties in space and is computationally efficient (fast wavelet transform).

5.8 Wavelet texture analysis and texture features

After the wavelet transformation of the CLSM images, it is necessary to define texture features which reflect and describe the tissue properties. These features must be defined so that they allow a unique distinction between common benign nevi and malign melanoma tissues in the square elements. The texture features are based on the variations of the wavelet coefficients within the frequency bands and the distribution of the energy of the frequency bands in the power spectrum. As features, the standard deviations of the coefficients inside the frequency bands and the energy and entropy of the different frequency bands are used. The standard deviations of the coefficients in the frequency bands are calculated by:

$$F_{STD}(i) = \sqrt{\frac{1}{n_i} \sum_k \sum_l (m_i - c_i(k,l))^2} \tag{63}$$

where n_i is the number of coefficients in the i-th frequency band and $m_i = \frac{1}{n_i} \sum_k \sum_l c_i(k,l)$ is the mean value within the band. The energy of the i-th frequency band is determined by:

$$F_E(i) = \frac{1}{n_i} \sum_k \sum_l (c_i(k,l))^2 \tag{64}$$

The entropy of the i-th frequency band is calculated as:

$$F_{ENT}(i) = (-1) \sum_k \sum_l w_i(k,l) \log_2 w_i(k,l) \tag{65}$$

with the normalized values: $w_i(k,l) = \dfrac{c_i(k,l)^2}{\sum_k \sum_l (c_i(k,l))^2}$. The features in the different frequency bands reflect architectural structures and cell structures at different scales (Fig. 12). In total,

39 different features are calculated, representing large scale and low scale information. The features are calculated for 16 frequency bands (labelled from 0 to 15). The mean value is calculated from the 4 first frequency bands, therefore 13 values result for each of the 3 features (Fig. 12). The highest frequency bands contain only information about very fine grey level variations, such as noise, and are therefore not considered for the image analysis. The single square elements are represented by a feature vector:

$$x_n = \left(F_{STD}^{(n)}(0), F_{STD}^{(n)}(1), ..., F_{ENT}^{(n)}(15) \right) \tag{66}$$

The index n refers to the n-th square element. The next step is the splitting of the square elements, on hand of the features, in square elements with common benign nevi tissue and square elements with malign melanoma tissue.

6. Classification and class prediction

By the classification procedure, the inhomogeneous set of square elements (referring to the feature values) is split into homogeneous sub-sets, which are assigned to one of the two classes (common benign nevi or malignant melanoma). A homogeneous subset means that it contains only square elements with similar feature values. In the following, we will name the square elements as instances. (The input to a machine learning scheme is, in its general form, a set of instances. The instances are the things to be classified). In the context of machine learning the features are generally called attributes. The thing to be learned, in our case the discrimination of lesion tissues, is called the concept. We will use this terminology in the following. Classification is done by machine learning algorithms (Witten & Frank, 2005). After a training phase, these algorithms enable the automated prediction of the class of a new instance. Machine learning can be defined operationally as: machine learning algorithms learn when they change their behaviour in a way that makes them perform better in the future. In other words, the algorithm learns on hand of a training set how to assign the instances to given classes. Then, in future, the algorithm can apply the gained knowledge to predict the class of unknown instances. The task of machine learning algorithms is in general the search of patterns in big amounts of data, known as data mining (Van Rijsbergen, 1979). The set of square elements is used as instances:

$$I = \left\{ x_1, ..., x_N \right\} \tag{67}$$

To simplify the expressions, the features (attributes) in the feature vector are rewritten as:

$$x_n = (a_0^{(n)}, a_1^{(n)}, ..., a_{L-1}^{(n)}) \tag{68}$$

Here $a_i^{(l)}$ is the i-th attribute of the l-th instance (L is the number of attributes). As an example, the attribute $a_0^{(n)}$ refers to $F_{STD}^{(n)}(i)$. The set of square elements is split into more or less homogeneous subsets, which are assigned to classes. The partition of the set of instances into g classes is then given by:

$$P = \left\{ C_1, ..., C_g \right\} \tag{69}$$

For every class C_l the following conditions should be satisfied:

$$\text{i) } \bigcup_{k=1}^{g} C_k = I \quad \text{ii) } \forall_{i \neq j} : C_i \cap C_j = \varnothing \tag{70}$$

The first condition (i) means that every instance is assigned to a specific class. All the instances are classified, and then the union of all classes equals the set of instances. Or in other words, all the instances are assigned to a class, no one is left over. The second condition (ii) means that no two different classes contain the same instances (the classes are disjunct). Or in other words, no instance belongs to two or more classes. If the attributes have numerical values (as in our case) then for an optimal partition, the Euclidian distance of every feature vector of a class k to the class mean value $\overline{x}_k$ is less or equal to the mean values of every other class (j=1,…,g):

$$\left\| x_n - \overline{x}_k \right\|^2 \leq \left\| x_n - \overline{x}_j \right\|^2 \tag{71}$$

Whereby the mean value of a class k containing K instances is $\overline{x} = \dfrac{1}{K} \sqrt{\sum_{i=1}^{K} (x_i)}$ and

$\left\| x_i - x_j \right\| = \sqrt{\sum_{l=1}^{L} \left(a_l^{(i)} - a_l^{(j)} \right)^2}$ is the Euclidian distance. The first step for a successful use of

machine learning algorithms is the representation of the knowledge. In computers, knowledge can be represented in different ways, for example: by numeric values, trees, rules or instance based (Brachman & Levesque, 1985). For medical diagnosis purposes it is important to duplicate the automated diagnostic process. Therefore, the rules that the algorithm uses to predict the class of an instance should be understandable for the human interpreter. Here we chose the CART (Classification and Regression Trees) algorithm for classification and class prediction, which uses a tree representation but where the inferring rules are automatically generated out of the tree.

6.1 Classification and Regression Trees (CART)

The CART algorithm was published for the first time in the year 1984 by Leo Breiman et al. CART is an algorithm which is used for optimal decision finding (Breiman et al., 1993). It is based on the generation of so called binary decision trees. The choice of the attributes is guided by the optimization of the information measure at each decision step. An important feature of the CART algorithm is that for every decision in a given node of the generated tree, the node is split only in two sub nodes (every node has maximally two child nodes). This is called a binary decision and the generated tree is called a binary tree. The central task of the algorithm is to determine the optimal binary decision for separation of the attributes at each step. Therefore, the special merit of the CART algorithm is the ability for the optimal separation of the data relating to classification purposes. Furthermore, the CART algorithm has the ability to capture the decision structure explicitly. This means that the decision rules generated from the tree are intelligible in that way that they can be understood discussed and explicitly used as diagnostic rules. A binary tree has a root node (starting point), several leaves (end points) and inner nodes which are branched into two succeeding nodes (Fig. 13). For a given node, the preceding node is the parent node and its successors are the child nodes. The root node of a tree is the node with no parents. There is, at most, one root node
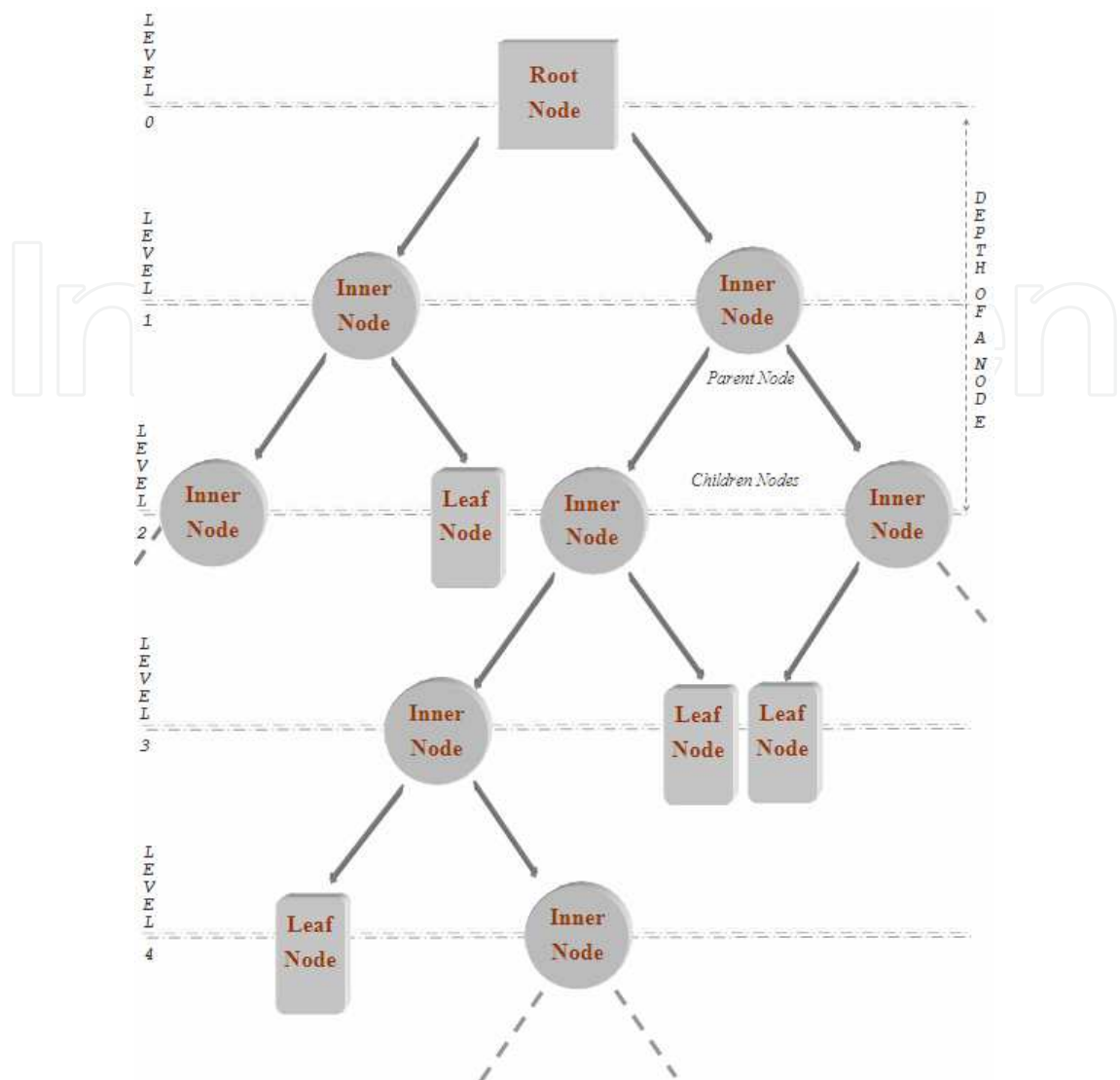
Fig. 13. In computers, knowledge can be represented as binary trees, consisting of a root node, inner nodes and leaves

in a rooted tree. A leaf node has no children. The depth of a node in the tree is the length of the path from the root node to the considered node. The set of all nodes at a given depth is called a level of the tree. The height of a tree is the length of the path from the root to the deepest leaf node in the tree. To classify an unknown instance, it is routed down the tree according to the values of the different features (Steinberg & Colla, 1995).

6.2 Generation of the decision tree

Divide-and-conquer (lat. divide et impera) algorithms are used to produce decision trees in a recursive way. A divide and conquer algorithm works by recursively breaking down a set of different objects (distinguishable by their specific feature values) into two (or more) sub-sets of similar kinds of objects, until the sub-sets contain only objects of the same kind. A node in the decision tree involves the testing of a particular feature. The root node, as the first node in a decision tree, contains all instances (square elements). A terminal (leaf) node contains instances that belong to the same class. The divide-and-conquer algorithm for constructing a decision tree consists in principal of three parts:

1.   The determination of the optimal splitting at every node;
2.   The decision whether the node is a leaf node or a inner node;
3.   The assignment of a leaf node to a specific class.

First a feature (attribute) is selected at the root node and a branch is made for each possible value. Through this operation the set of the instances is split up into subsets. Then this process is repeated recursively for every branch, at an inner node, with the instances in the corresponding subset (Fig. 14). If the nodes cannot be divided anymore, the process is stopped and all the instances at a terminal node belong to the same class. The features that produce the purest daughter nodes are used for an optimal splitting. In the case of numerical features, the splitting is usually done by use of numeric thresholds that divide the range of the feature values. Trees with averaged numerical values are called regression trees. Trees with tests involving more than one feature at a time are called multivariate decision trees (in contrast to univariate trees, where only one feature is used).

Determining the optimal splitting: Each step in the tree construction consists of a binary decision leading to two daughter nodes. At every inner node, the actual instance set (parent set: t) is split into two children sets (t_R and t_L).

$$t \xrightarrow{\quad s \quad} \begin{cases} t_R = \{x \,|\, F_{STD}^{(n)}(i) \le U_i\} \\ t_L = \{x \,|\, F_{STD}^{(n)}(i) > U_i\} \end{cases} \tag{72}$$

The set of splits $S = \{s_i\}$ is generated by a set of tests $Q = \{q_i\}$. The tests $q_i \in Q$ consist of binary decisions, where each split $s_i \in S$ depends on the value of a single feature ($F_{STD}^{(n)}(l)$). The splitting operation is done by use of numerical threshold values U_i, which are selected for every attribute. The decisions are made by comparing the feature values with the threshold values separating the range of the features into two parts. The decision thresholds result from the optimization of the information measure. The fundamental principle for the construction of the decision tree is to select the split of a set in such a way that the instances of the children sets are purer than the instances of the parent set. The information measure $I(t)$ is commonly used for calculating the purity or impurity of the nodes. According to Shannon, the information measure of a set of objects $S = \{z_1, z_2, ..., z_n\}$ is defined by:

$$I = -\sum_{z \in S} [p_z \log_2(p_z)] \tag{73}$$

Whereby the probability of occurrence of an element in the set is $p_z = \dfrac{N(z)}{N}$ and $N(z)$ is the number of the elements z in the set with N elements. The sum of the probabilities is one ($\sum_z p_z = 1$). The information measures the amount of uncertainty associated with each element (object) in a given discrete set, depending on the probability of occurrence in the set. For a pure node (only elements of one kind), the information measure has a minimum (=0) value. For an impure node with equal distribution of the elements, the information measure has a maximum value.
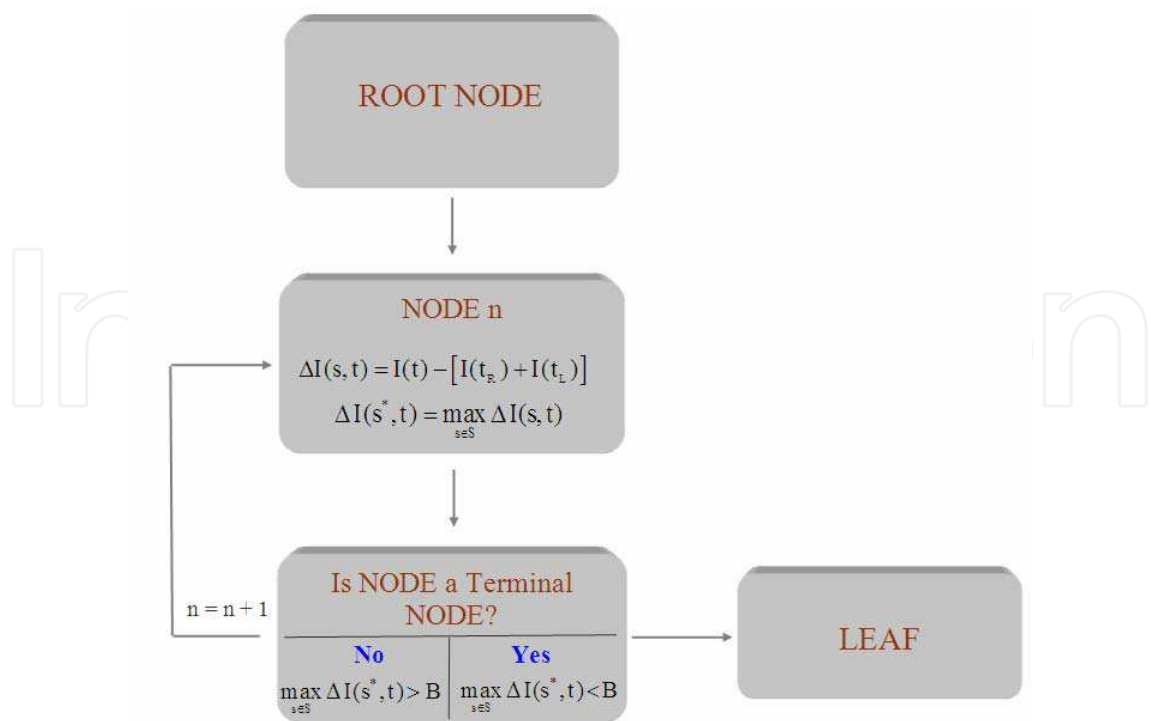
Fig. 14. The decision tree is generated recursively

For sets with different distributions of the elements, the value of the information measure lies between these two extremes. If the information measure is calculated by use of the binary logarithm, which is closely connected to the binary numeral system, the units of the corresponding values are expressed in bits. The information measure of an attribute is considered to be high when it enables a set splitting in such a way that a classification can be done with a high score. The higher the information measure of an attribute is, in relation to the classification values, the higher in the tree the attribute is selected. The difference between the impurities of the parent and the children sets is given by:

$$\Delta I(s,t) = I(t) - \left[I(t_R) + I(t_L) \right] \tag{74}$$

Then the parent set splits into children sets, by use of a splitting rule s^* generated by a test $q^* \in Q$, in such a way that the difference between the impurities of the parents and the children sets has the maximum value:

$$\Delta I(s^*,t) = \max_{s \in S} \Delta I(s,t) \tag{75}$$

This means that the set of parent square elements is split into children sets so that the feature values for the square elements in each of the children sets are as similar as possible. At every node, a feature is interrogated and a decision is made about the selection of the successor nodes . This procedure is repeated until a leaf is reached.
Decision whether the node is a leaf node. The decision whether the node is a terminal node or an inner node depends on the purity of a children node and is given by the stop rule:

$$\max_{s \in S} \Delta I(s^*,t) < B \tag{76}$$

The threshold B is a measure of how homogenous the terminal nodes must be. Ideally, the information measure for a set of equal elements is zero.

<u>Assignment of a leaf node to a specific class.</u> When a leaf node is reached, the instance is classified according to the class assigned to the leaf. Because the texture features are represented by numeric values, the classification task is to predict a numeric quantity by which the instance is assigned to the averaged numeric value of a class.

6.3 Inferring rules

The divide-and-conquer algorithm for generating decision trees enables a good separation of the instances. But the knowledge representation by binary trees is not as suitable for the interpretation of the classification process. In contrast, the knowledge representation by rules is very suitable in medicine because rules are easily understandable for the human interpreter. By the CART algorithm, the knowledge is extracted from the dataset in the form of inferring rules, which can easily be implemented as diagnostic rules. The inferring rules result from the splitting rules and are automatically derived directly off the decision tree, where one rule is generated for each terminal node. The rules have a syntax consisting of an antecedent (precondition) and a consequent (conclusion) part:

$$\forall_{x \in I} : P(x) \rightarrow C(x) \tag{77}$$

For all instances, which fulfil the precondition $P(x)$, follow the conclusion $C(x)$. The rule, represented by the formula, is called "Modus Ponens" meaning that: from $P(x)$ infer $C(x)$ or in other words "if $P(x)$ then $C(x)$". This rule is the formal description of the diagnostic process. The precondition may include several parts (single sub-conditions, for example clinical symptoms) $P(x) = p_1(x) \wedge p_2(x) \wedge ... \wedge p_N(x)$. If all the single sub-conditions ($p_n(x)$) are true then the precondition $P(x)$ is valid. The antecedent of a rule includes a condition for every node on the path from the root node to a specific terminal node. The consequence of the rule is the class assigned to the terminal node. For the implementation on a computer they are generally expressed as "IF-THEN" rules (IF {precondition} THEN {conclusion}). These rules represent knowledge in a form that is easily understandable for the human observer. It enables him to understand why a specific square element is assigned to a benign common nevi or malignant melanoma. Such inferring rules are used as diagnostic rules. The rules can be compared with the diagnostic guidelines of the derma pathologist. Furthermore they enable the computer scientist to validate the gained knowledge and to combine it eventually with previously known facts. An example of such a (simplified) inferring rule is:

$$\text{IF} \begin{cases} \text{Standard deviation in frequency band (5)} > a \\ .and. \\ \text{Energy of highest frequency band } (0,1,2,3) \geq b \end{cases} \text{THEN} \{\text{tissue} := \text{nevus tissue}\} \tag{78}$$

According to the guidelines for the diagnosis, the rule can be translated in: if the tissue contains large and medium structures, for example: nevi cells grouped around basal structures, then it is nevus tissue. Generally, such rules are far more complex than necessary and therefore they are usually pruned to remove redundant tests. Pruning is a technique in machine learning that reduces the size of decision trees by removing sections of the tree that provide little power to classify instances. The goal of pruning is to reduce complexity of the

final classifier and to improve the classification accuracy by the reduction of over fitting and removal of sections of a classifier that may be based on erroneous data. Pruning should reduce the size of a learning tree without reducing predictive accuracy as measured by a test set or using cross-validation.

7. Study

The discrimination power of texture features based on the wavelet transform and the performance of the CART algorithm are demonstrated in the following study (square size 512x512). Overall, 857 images of benign common nevi (408 images) and malignant melanoma (449 images) are used as study set. To get more insights into the classification performance, a percentage split was performed by using 66% of the dataset for training and the remaining instances (34%) as the test set. The classification results of 572 cases (276 benign common nevi, 296 malignant melanomas) in the training set and 285 cases (132 benign common nevi, 153 malignant melanomas) in the test set are shown in Table 1.

CART	Training Set			Test Set		
	% Correct	NZ	MM	% Correct	NZ	MM
NZ	96.6	267	9	78.0	103	29
MM	98.0	6	290	84.1	24	129

Table 1. Classification result for malignant melanomas (MM) and benign common nevi (NZ) in the training and test set

8. Discussion

Delayed recognition of skin malignancies puts the patient at risk of destructive growth and death from disease once the tumour has progressed to competence for metastasis. Therefore, preventive and periodical skin checkups are of special importance. Technological advancements in imaging systems have led to the development of convocal laser scanning microscopy (Ericson et al., 2008). This technique enables the examination of skin lesions in vivo and significantly higher prediction success than reported for dermoscopic examination can be achieved for the diagnosis of melanoma (Rajadhyaksha, 2009). However, due to the fact that the CLSM method is relatively new, there is still a lack of experiences with the diagnostic features and an intensive training is necessary for the clinician.

The study demonstrates the applicability of the automated diagnosis system for the discrimination of CLSM views of skin lesions (Table 1). The image analysis, based on the wavelet transform, together with tree based machine learning algorithms provide a powerful tool for automated diagnosis of CLSM images of skin lesions. For the diagnosis of the CLSM views, architectural structures such as: micro-anatomic structures; cell nests etc., are used as guidelines by the derma pathologist. Therefore features based on the spectral properties of the wavelet transform, enabling an exploration of architectural structures at different spatial scales are suitable for the automatic analysis. The images of benign common nevi show pronounced architectural structures whereas images of malign melanoma show melanoma cells and connective tissue with few or no architectural structures. These guide lines are reflected by the wavelet coefficients inside the different frequency bands. The standard deviations of the wavelet coefficients in the lower and medium frequency bands show higher values for the benign common nevi than for malignant melanoma tissue,

indicating more pronounced structures at different orders of magnitude (Fig. 15). The energy of the lowest frequency band (large scale architectural structures) is higher for the CLSM views of benign common nevi than for malignant melanoma. The CART algorithm has the ability to capture the decision structure explicitly. This means that the, from the tree, generated decision rules are intelligible in that way that they can be understood, discussed and explicitly used as diagnostic rules.
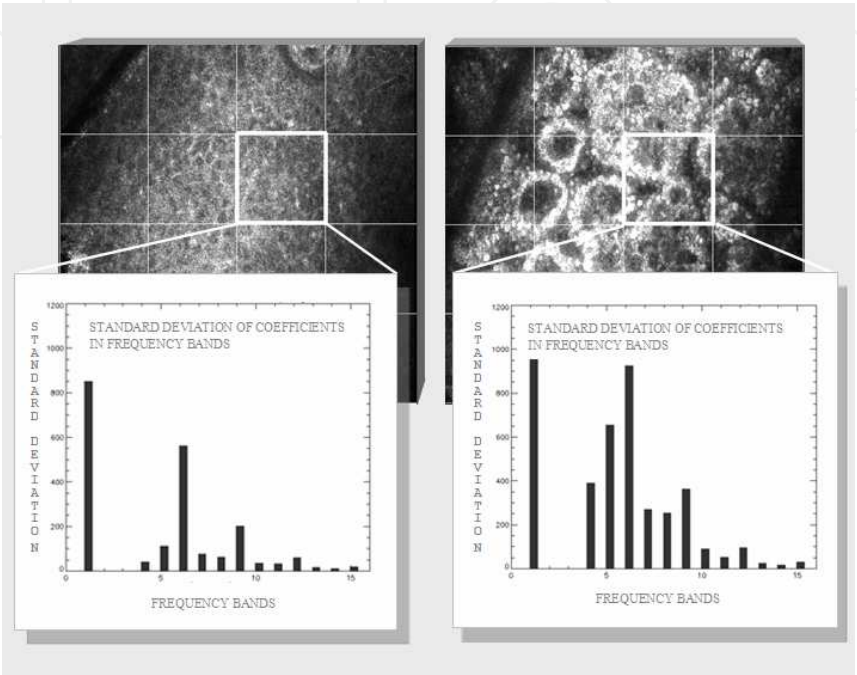

Fig. 15. The standard deviations of coefficients in the wavelet frequency bands for melanoma (left) and nevi (right)
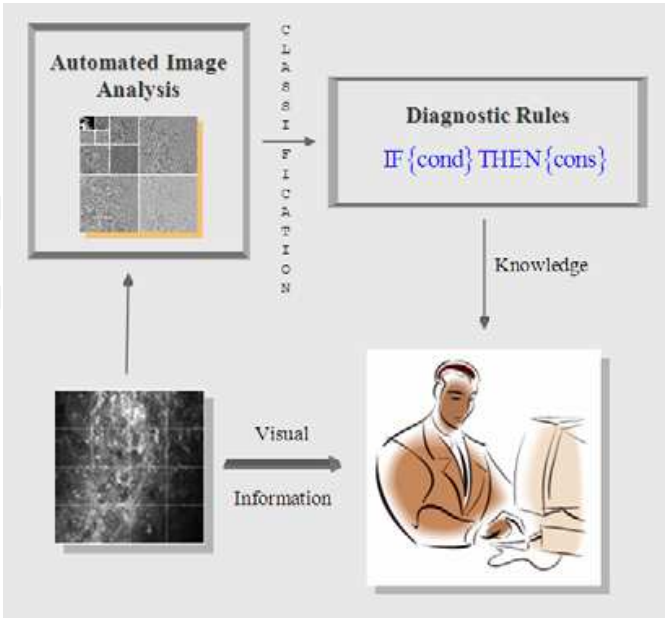
Fig. 16. From the CLSM image the features, based on wavelet transform, are extracted and automatically analyzed

Computer aided diagnosis, providing automated decisions, can be used as an expert second opinion or help and assist the non-experienced physician in the diagnostic procedure (Fig. 16). Automated diagnosis is the principal performance of medical expert systems. Although the handcrafted diagnostic rules in the inference machine of expert systems perform well in medical applications, machine learning has the advantage that the rules are generated automatically for systems where the producing of manual rules is too labour intensive and there is a lack of human expert knowledge and experience. The generating of manual rules requires expert knowledge whereas rules generated by machine learning algorithms represent knowledge that can be used to analyse and refine the diagnostic process.

To verify the performance of the method and to interpret the diagnostic process, the automatically generated inferring rules are implemented in specially developed viewer software (Fig. 17). By this viewer the classified square elements are indicated in the corresponding CLSM image in order to judge the performance of the analysis. For these purposes square elements of size 128x128 are used, because they enable a good localization of the different regions into the images and the texture features have still enough discrimination power (sensitivity and specificity of 88.12% for malignant melanoma and 84.75% for benign common nevi). The procedure is illustrated in the case of malignant melanoma tissue (Fig. 17). Square elements resulting from terminal nodes, with 100% of discrimination power, are taken as highly significant and are drawn with red margins at the graphical user interface of the viewer and labelled with the number 1. Square elements with discrimination power of 80-99% are drawn with green margins and labelled with the number 2. The relocated elements mainly show polymorphic tumour cells with structural
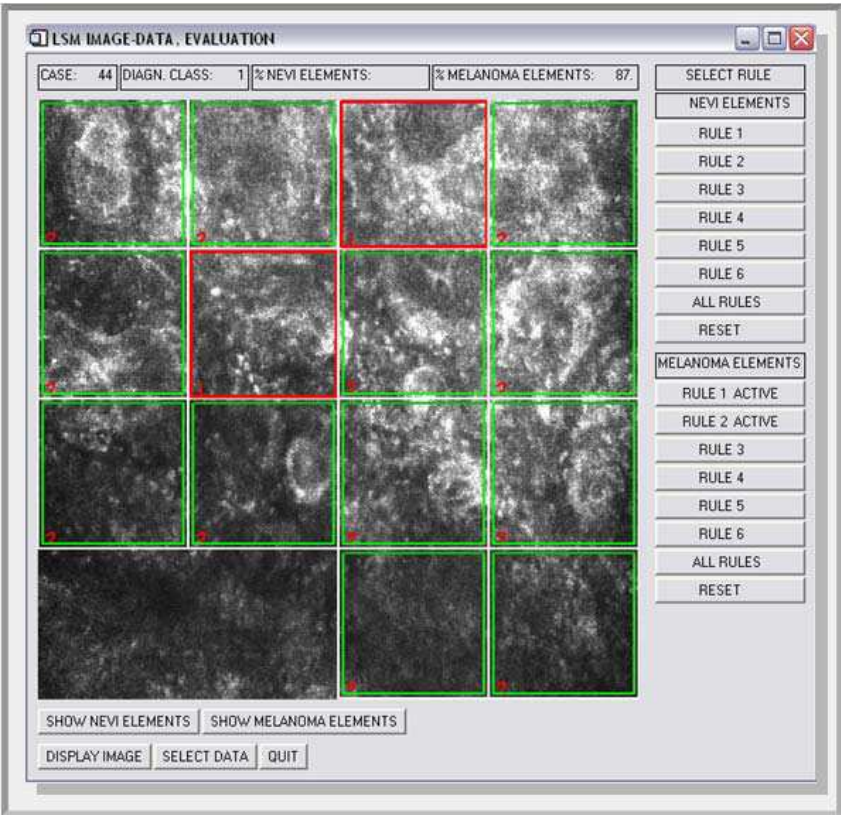

Fig. 17. Square elements containing diagnostic significant regions in a CLSM image of a malignant melanoma

disarray and are in good accordance with previously published diagnostic CLSM features. Due to the fact that the method of confocal laser scanning microscopy is relatively new, the diagnostic features for the clinician have not yet been completely investigated. Therefore the automated image analysis is a fundamental and important step toward the general assessment of the CLSM images by the clinician. The viewer software is further used for the interaction of the clinician with the automated diagnostic system. Then he can use it as computer aided tool for a second opinion or a non-experienced user can consult the system, learn from the system and ameliorate his diagnostic skills. The results of the inferring mechanism (the classification) are visualized by the graphical user interface enabling an interpretation and evaluation of the diagnostic process.

In this review, we demonstrated how diagnostic guidelines and experiences are represented by mathematical structures and implemented on a computer. To enable the human interpreter an interpretation of the results of the automated diagnostic, the output is visualized in an appropriate manner. Since the highlighted regions in the CLSM images show tissue structures which are in good accordance with the already known diagnostic guidelines, no explicit formulation of the diagnostic rules is necessary. Due to the fact that the tissue features are based on the wavelet transform, reflecting the guidelines of the derma pathologists, the highlighted regions are self explaining. Although it should be noted that, beside the visual interpretation, it seems that features not accessible to the human eye contribute to the automated diagnostic process.

9. Conclusion

In conclusion, image analysis, based on the wavelet transform, together with a tree based machine learning algorithm provide a powerful tool for automated diagnosis of CLSM images of skin lesions. Already known, but subjective CLSM criteria are objectively reproduced. The system enables the identification of highly significant parts in CLSM views of malignant melanoma. In a clinical application, the system can be used as a screening tool to improve preventive medical checkups and the early recognition of skin tumours. The automated decisions provided can be used as an expert second opinion and as a training system for inexperienced or student derma pathologists. In another clinical onset, the system may automatically pre-select the cases in such a way that the critical cases are first interpreted by the clinician.

10. References

Brachman, R.J.; Levesque, H.J. (1985). *Readings in knowledge representation*. Morgan Kaufmann, San Francisco

Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.F. (1993). *Classification and Regression Trees*. Chapman & Hall, New York, London

Burrus, C.S.; Gopinath, R.A.; Guo, H. (1988). *Introduction to Wavelets and Wavelet Transforms:A Primer*. Prentice-Hall

Chui, C. K. (1992). *An Introduction to Wavelets*, Academic Press, San Diego

Daubechies I. (1988). Orthonormal bases of compactly supported wavelets, *Comm. Pure and Appl. Math.* 41

Daubechies, I.; Lagarias, J. (1991). Two-scale equations: Existence and global regularity of solutions. *SIAM J. Math. Anal*, 22

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM

Ericson, M.B.; Simonsson, C.; Guldbrand, S.; Ljungblad, C.; J. Paoli, J.; Smedh, M. (2008) Two-photon laser- scanning fluorescence microscopy applied for studies of human skin. *J Biophotonics*, 1, 4, 320-330

Friedman, R.; Rigel, D.; Kopf, A. (1985). Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA Cancer J Clin,* 35, 3, 130–151

Lucas, R. (2006). Global Burden of Disease of Solar Ultraviolet Radiation. *Environmental Burden of Disease Series*, July 25, 13. News release, World Health Organization

Markovic, S.N.; Erickson, L.A.; Flotte, T.J.; Kottschade, L.A. (2009). Metastatic malignant melanoma. *G Ital Dermatol Venereol.* 144(1) 1-26.

Oliveria, S.; Saraiya, M.; Geller, A.; Heneghan, M.; Jorgensen, C. (2006). Sun exposure and risk of melanoma. *Arch Dis Child,* 91, 2, 131–138

Paoli , J.; Smedh, M.; Ericson, M.B. (2009). Multiphoton laser scanning microscopy--a novel diagnostic method for superficial skin cancers. *Semin Cutan Med Surg*, 28, 3, 190-195

Patel, D.V.; McGhee, C.N. (2007). Contemporary in vivo confocal microscopy of the living human cornea using white light and laser scanning techniques: a major review. *Clin. Experiment. Ophthalmol,* 35, 1, 71–88

Pellacani, G.; Longo, C.; Malvehy, J.; Puig, S.; Carrera, C.; Segura, S.; Bassoli, S.; Seidenari, S. (2008). In vivo confocal microscopic and histopathologic correlations of dermoscopic features in 202 melanocytic lesions. *Arch Dermatol*. 144, 12, 1597-608

Pellacani, G.; Vinceti, M.; Bassoli, S.; Braun, R.; Gonzalez, S.; Guitera, P.; Longo, C.; Marghoob, A.A.; Menzies, S.W.; Puig, S.; Scope, A.; Seidenari, S.; Malvehy, J. (2009). Reflectance confocal microscopy and features of melanocytic lesions: an internet-based study of the reproducibility of terminology. *Arch Dermatol,* 145, 10, 1137-1143

Prasad, L.; Iyengar, S.S. (1997).*Wavelet analysis with applications to image processing.* CRC, Press Boca Raton, Boston, London, New York, Washington D.C.

Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* 2nd ed. Cambridge University Press, 591-606

Rajadhyaksha, M.; Conzales, S.; Zavislan, J.M.; Anderson, R.R.; Webb, R.R. (1999). In vivo confocal scanning laser microscopy of human skin: advances in instrumentation and comparision with histology. *J Invest Dermatol*, 113, 293-303

Rajadhyaksha, M. (2009). Confocal microscopy of skin cancers: Translational advances toward clinical utility. *Conf Proc IEEE Eng Med Biol Soc.* 1, 3231-3233

Scope, A.; Venute-Andrade, C.; Agero A.L. (2007). In vivo reflectance confocal microscopy imaging of melanocytic skin lesions: consensus terminology and illustrative images. *J Am Acad Dermatol,* 57, 644-658

Steinberg, D.; Colla, P. (1995). *CART: Tree-structured non-parametric data analysis.* Salford Systems, San Diego

Strang, G.; Nguyen, T. (1996). *Wavelets and Filterbanks.* Wellesley-Cambridge Press

Van Rijsbergen, C.A. (1979). *Information retrieval.* Butterworths, London

Wiltgen, M.; Gerger, A.; Smolle, J. (2003). Tissue counter analysis of benign common nevi and malignant melanoma. *International Journal of Medical Informatics,* 69, 17-28

Wiltgen, M.; Gerger, A.; Wagner, C.; Bergthaler, P.; Smolle, J. (2003). Discrimination of benign common nevi and malignant melanoma lesions by use of features based on spectral properties of the wavelet transform. *Anal Quant Cytol Histol,* 25, (5), 243-253

Wiltgen, M.; Gerger, A.; Wagner, C.; Smolle, J. (2008). Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumours. *Methods Inf Med,* 47, 15-25

Witten, I.H.; Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Technique.* Elsevier, Morgan Kaufmann Publishers

Laser Scanning, Theory and Applications

Edited by Prof. Chau-Chang Wang

Ever since the invention of laser by Schawlow and Townes in 1958, various innovative ideas of laser-based applications emerge very year. At the same time, scientists and engineers keep on improving laser's power density, size, and cost which patch up the gap between theories and implementations. More importantly, our everyday life is changed and influenced by lasers even though we may not be fully aware of its existence. For example, it is there in cross-continent phone calls, price tag scanning in supermarkets, pointers in the classrooms, printers in the offices, accurate metal cutting in machine shops, etc. In this volume, we focus the recent developments related to laser scanning, a very powerful technique used in features detection and measurement. We invited researchers who do fundamental works in laser scanning theories or apply the principles of laser scanning to tackle problems encountered in medicine, geodesic survey, biology and archaeology. Twenty-eight chapters contributed by authors around the world to constitute this comprehensive book.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Wiltgen Marco (2011). Confocal Laser Scanning Microscopy in Dermatology: Manual and Automated Diagnosis of Skin Tumours, Laser Scanning, Theory and Applications, Prof. Chau-Chang Wang (Ed.), ISBN: 978-953-307-205-0, InTech, Available from: http://www.intechopen.com/books/laser-scanning-theory-and-applications/confocal-laser-scanning-microscopy-in-dermatology-manual-and-automated-diagnosis-of-skin-tumours

INTECH

open science | open minds