Technical Report

Dept. for Pattern Recognition and Image Processing
Institute for Automation
Technical University of Vienna
Treitlstr. 3/1832
A-1040 Vienna AUSTRIA
Phone: +43 (222) 58801-8161
Fax: +43 (222) 569697
E-mail: api@prip.tuwien.ac.at

PRIP-TR-21 March 15, 1993

Neuromorphic Methods for Recognition of Compact Image Objects

Axel Pinz, Marek B. Zaremba [1], *Horst Bischof,*
Francois A. Gougeon [2], *Michel Locas* [1]

Abstract

The issue of the recognition of tree species from high resolution aerial images is addressed in this paper. An approach based on the use of neural networks is presented and discussed in more detail. The networks perform classification and recognition operations on compact image objects, obtained by applying different tree isolation procedures. The recognition capabilities of two classes of networks, multilayer feedforward networks and holographic networks, are compared and some results of the research carried out in Austria and Canada, using aerial photographs and multispectral scanner images, are given.

[1] Univ. du Québec à Hull, Dept. d'Informatique, Canada
[2] Petawawa National Forestry Institute, Ontario, Canada

1 Introduction

In this paper we present a general two-step approach for the recognition of compact objects in digital images. In the first step, the image is segmented into relatively small, compact regions, which are presumed to represent image objects. In the second step, these image objects are recognized using neuromorphic methods. Most of the paper deals with the neuromorphic part of the recognition process, both from a theoretical and a specific application point of view. The application involves the interpretation of forest data from aerial images. We discuss the segmentation of individual trees, and we show how the neuromorphic methods are applied to the recognition of individual tree species.

1.1 Forest Inventory

This work was originally motivated by a project on forest inventory in Austria using color infrared aerial photographs. The reported results of a two component recognition system – the Vision Expert System VES [21, 22, 23] and a neural network [26] – led to research cooperation between Austria and Canada.

In Canada, where 27 million hectares of forest are inventoried each year, the proposed use of multispectral digital aerial images instead of conventional aerial photographs will lead the way to computer-assisted on-screen image interpretation [16]. Georeferenced image enhancements, standardized interpretation keys, and links to geographical information systems will facilitate the production and update of forest inventories from powerful interpreter workstations. Thus, some short term research objectives are to develop the tools and the methodologies needed by this new approach. Eventually, with increasing computer assistance available, the winds will turn towards semi or complete automation of the process.

In this context, methods to isolate individual trees from each other and from the background vegetation, as well as methods to identify individual tree species are needed. These methods are substantially different from the pixel-based and forest stand-based approaches of the now 'traditional' satellite image-based digital remote sensing image analysis. With individual tree isolation (or delineation), the tree geographic positions and crown areas are immediately available. Tree heights are obtained from stereo image pairs and tree species from their multispectral reflectance. This leads to forest inventories on a tree by tree basis, either as a goal in itself, or as an intermediate but necessary step towards accurate forest stand inventories.

1.2 Neural Networks for Vision

A large amount of information about the world we live in is supplied by our visual system. Humans perform the task of vision effortlessly, without being aware of this complex process. The goal of computer vision (image understanding) is to build computers which are able to 'see', a goal that has not yet been reached [14]. Remarkable attributes of the human visual system are parallelism, robustness, and adaptivity. Similar attributes apply to neural networks (e.g. [19, 32]), a fact that makes them interesting for computer vision. The success of neural networks for pattern recognition tasks has been demonstrated in a variety of examples (e.g. [3, 36, 30, 29]). The increasing learning time when scaling the problem up in size [12] causes a severe limitation of neural networks. The use of all the pixels of an image as a neural network input is intractable, for example, in fully connected architecture, several million connections would be necessary per unit. One solution is to select a few features which describe the content of the image and use these features in the subsequent classification stage. Unfortunately, feature selection is a tedious and error prone process, and it is often impossible to find good features. Another solution is to use modular or hierarchical neural networks (e.g. [4, 8]), a further one, is to use only small portions of an image as input. In many cases, when the observed objects are *compact and small* compared to the image size, this is a reasonable approach.

A similar approach is also employed by the human visual system which does not process the whole image at full resolution. First, 'interesting' regions are identified, which are then analyzed in more detail. One of the main reasons is the inherent complexity of general purpose vision as has been shown by Tsotsos [38]. In this paper Tsotsos also points out, that most of the objects in the world are compact, so that a multistage approach is reasonable: Find interesting regions (i.e. the compact objects) in the first stage, and analyze them in more detail in subsequent stages. Since only a few objects in an image are of interest, compared to the original number of pixels, this approach can save a considerable amount of processing time. An advantage of focusing the attention on a particular region of the image is to avoid problems due to shifted inputs. In [7] we have proved, that hierarchical systems inherently suffer from shift variance problems. We have also shown, that this problem can be overcome by an appropriate attention mechanism. In [8] we have presented a general framework for a hierarchical and modular neural network architecture which is able to focus attention on particular locations in an image, and to use specific neural network modules to extract information at these locations.

The approach of this paper is a hybrid one: Compact image objects are found by conventional, non neural methods. At these locations, neural networks are used selectively.

2 Problem Specification

In both countries, Austria and Canada, a decision was made for a remote sensing based forest inventory. Due to the necessity of interpreting individual trees, low altitude high resolution sensors are required: color infrared aerial photographs in Austria, and MEIS aerial images in Canada. Forest inventory by using these sensors consists of the following steps:

1. *Identification of individual trees:* This is mainly a segmentation problem (see section 3). In most cases a monochrome image is sufficient for this task.

2. *Determination of the tree species:* This is a complex classification/recognition problem based on different clues: color and texture of the object in a multispectral image, 3D shape of the crown (stereo), altitude and topography of the terrain, as well as additional knowledge like dominating species in certain regions have to be considered. Preliminary investigations using statistical approaches to the classification of individual tree species, which typically implies the extraction of significant features or summarizing

the statistics from each tree, have had limited success. Classification results in the order of $62 - 74\%$ for five closely related coniferous species have been reported [10]. The advantages of neuromorphic methods in this respect are, that *it is not necessary to explicitly formulate the features* that have to be extracted from the image.

3. *Collection of additional attributes:* Besides many other attributes (e.g. age, crown diameter, slope and exposition of the terrain – in the case of the Austrian forest inventory more than 20 different attributes are determined), a very important one is the *tree vitality.* For coniferous trees the vitality is mainly a spectral characteristic and can be determined using multispectral classification. Since different species have different spectral signatures, the successful *prior* determination of the tree species is a necessity.

2.1 Austrian Inventory with Color Infrared Aerial Photographs

In the Austrian forest inventory, color infrared aerial photographs of a scale of approximately 1 : 15000 are used. The spectral characteristics of color infrared film are illustrated by Fig.1: Near infrared in the scene ($700 - 900nm$) is reproduced in red, red in green, green in blue, and blue in the scene is not reproduced at all. Due to the very high reflectivity of vegetation in the near infrared, as compared to visible wavelengths, this spectral band is a sensitive indicator for the assessment of vegetation. Figure 2 shows the green band (i.e. red in the scene) of such an image, digitized with $25\mu m$ pixelsize, corresponding with a resolution of approximately $40cm$ in the scene.

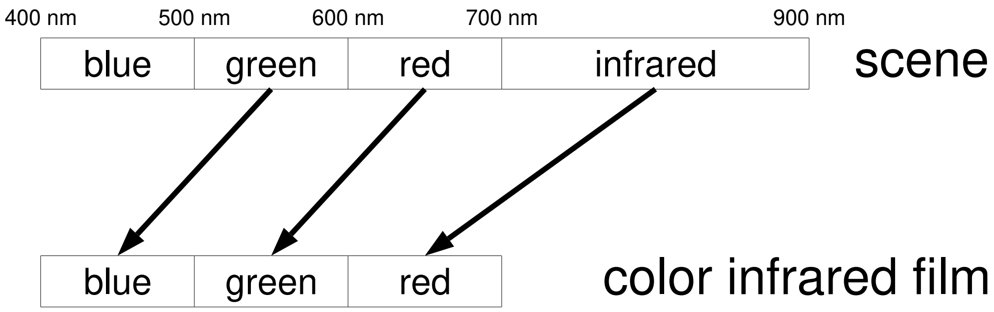


Figure 1: Spectral characteristics of color infrared film

The Austrian inventory is implemented in several phases. In the first one, which was completed in 1990, human interpreters have to interpret stereo pairs using analytical plotters. All three steps mentioned above (tree finding, species determination, attribute collection) are performed by the interpreters. In the second phase, which is close to completion now, a CCD camera is used for online extraction of the current image object, and several attributes (e.g. tree vitality) are computed automatically from the image. The human operator is still responsible for correct positioning (= tree finding) and species determination. Part of the preparatory research for phase three, where tree finding and species determination will be automated, is reported in this paper (see also [21, 26, 22, 23]).

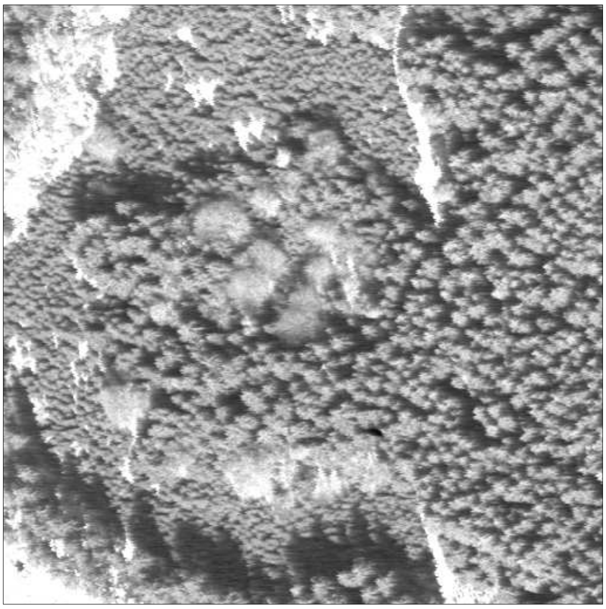


Figure 2: Green band of a color infrared aerial image

2.2 Aerial MEIS Images

The Multispectral Electro-optical Imaging Sensor (MEIS) is a Canadian remote sensing system capable of acquiring high quality multispectral digital images from an aircraft.

The present version, MEIS-2 [18], consists of eight linear CCD arrays, each with its own lens and changeable filter. The imagery and data from an inertial navigation system are collected on a fast speed very high density magnetic tape. The possible addition of mirrors in two of the eight optical paths permits the acquisition of fore-aft stereo pairs (typically via panchromatic filters). Sophisticated post-flight ground processing can output precise georeferenced images, ortho-images, and digital terrain models.

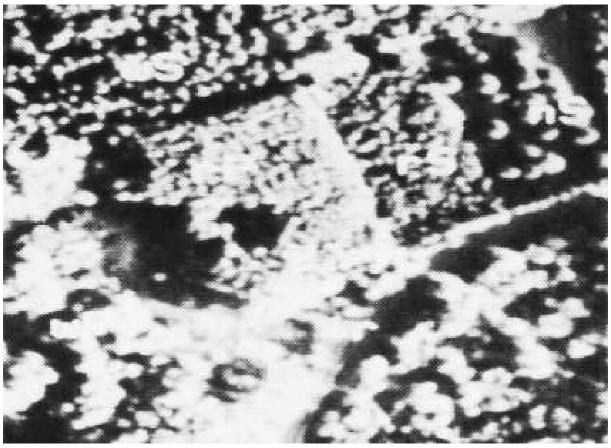


Figure 3: MEIS aerial image

In the work reported in this paper, the 433x512 images at the resolution of approximately $31cm$/pixel composed of four frequency bands (521, 673, 590 and $871nm$) were analyzed. An example of an image of an experimental plantation in Petawawa, Ontario is shown in Fig.3.

Five homogeneous regions can be clearly distinguished on the image. They correspond to white spruce, red spruce, Norway spruce, white pine, and red pine.

3 Segmentation Methods

In the context of our two stage hybrid recognition approach, the first step is the segmentation of small, compact image objects (i.e. focusing attention). Constraining ourselves to the application of tree finding, the scene object 'tree' corresponds to an image object 'bright blob'. We briefly introduce a region based and a contour based method to find bright blobs in an image, and we discuss a combination of these two methods. Once the image objects are found, they are 'enhanced' for the purpose of subsequent recognition.

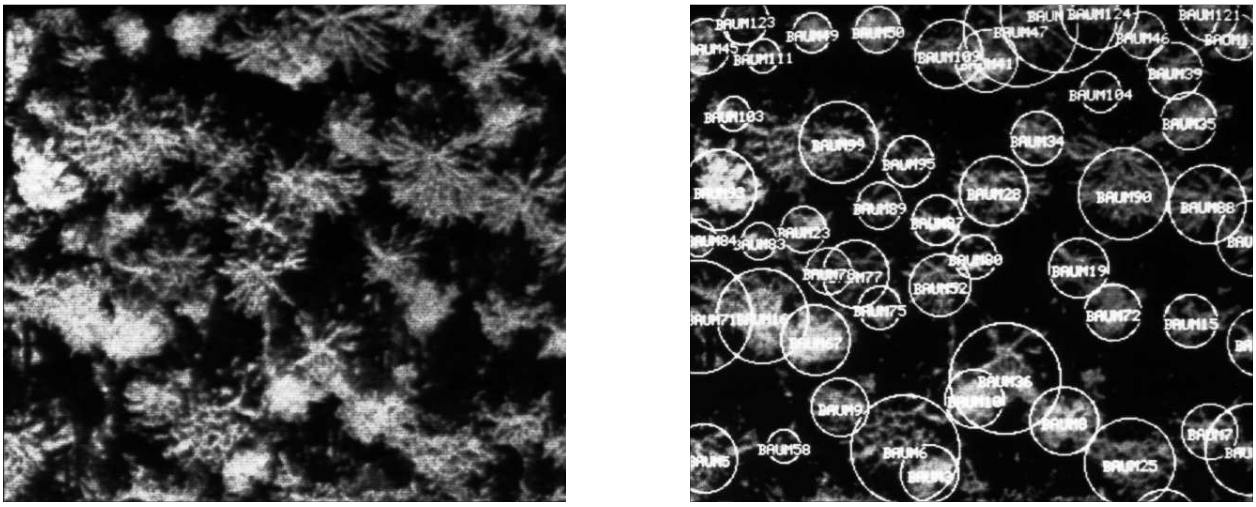
3.1 Region Based Tree Segmentation

In Austria, a region based segmentation approach to the tree finding task is implemented using a general purpose multilevel image understanding system (the first implementation was called Vision Expert System (VES), the second one Vision Station (VS)). We search for bright blobs in the image by applying a series of different lowpass filters followed by local maxima detection. Each of these maxima is a candidate for the center of a tree crown, so that a local window is examined at these locations at maximum spatial resolution. We use projections, called radial brightness distributions, to determine whether the shape of the object is tree-like or not. If this test holds, it delivers simultaneously an estimate for the crown radius, if not, the candidate is discarded. After this stage of processing, due to the series of different lowpass filters, several candidates remain for most of the trees (very close crown centers and similar radii), so that a fusion process has to be applied. Finally, a conflict resolution is carried out (e.g. where trees are growing too closely, contradictions such as trees in the middle of a road, etc.). Figure 4.a shows an original input image, Fig.4.b the result of the VES tree finder. In a dense forest image more than 90% of the trees are correctly identified. Details about the architecture of the system, its multilevel knowledge representation and control structure are reported in [21, 23].

3.2 Valley Following Method

This Canadian *contour based segmentation approach* is related to the behaviour of a photo interpreter performing tree delineation, who relies on three main features to separate trees from each other and from the background vegetation:

- a knowledge of color, texture, and structure differences in the crown of various species (mostly useful in mixed forest stands),



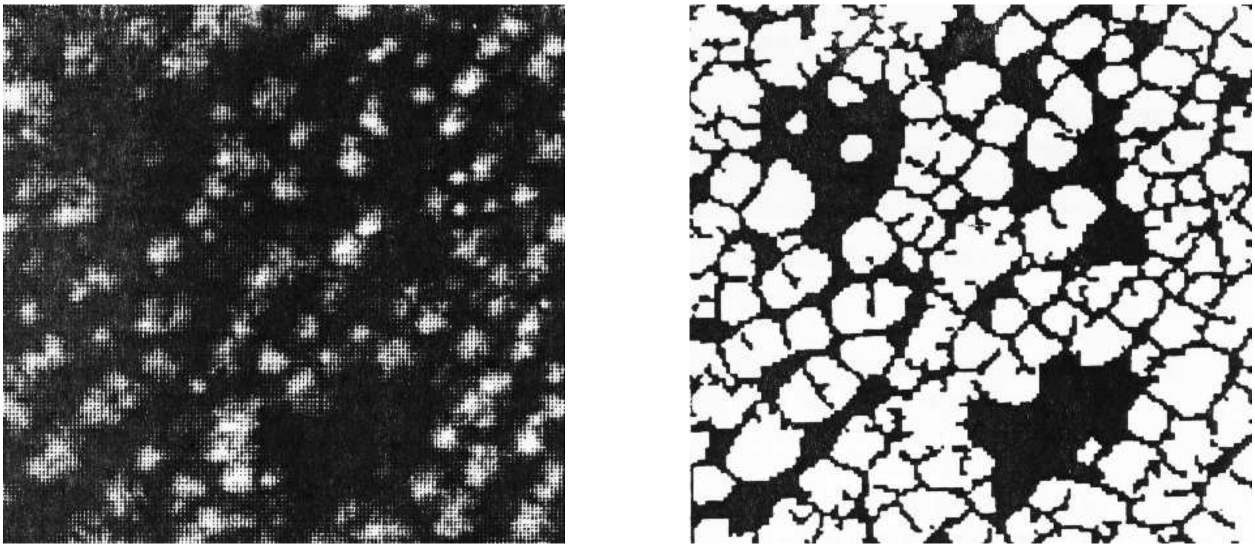
(a) Original input image

(b) VES result

Figure 4: The Vision Expert System VES tree finder

- a possible indication of crown boundaries by narrow (weak) darker lines of shaded material between adjoining trees (useful in dense stands of a single species), and
- an intuitive ability to follow contours when some parts of the crown outline are obvious.

The present version of the valley following method [11] partially emulates some of these human abilities and has potential for further improvement. Firstly, a simple parallelepiped classification process eliminates large open non-forested areas from further consideration. Secondly, local minima are found over the remaining image areas. They typically correspond to the darkest points in the shaded material that is often found between trees, almost labelling tree corners (if trees were polygons). By analogy, if the grey-level space of the image was to be interpreted as a digital elevation model, they would correspond to areas where water would accumulate as lakes or potholes. Thirdly, lines are grown from these local minima following 'valleys' in various directions. The growing process continues until another neighbouring local minimum is reached or progression is made impossible. As seen in Fig.5, the valley following process accomplishes most of the tree isolation work. Fourthly, rules are activated to finish incomplete tree crown contours. There are as many rules as there are different situations. These rules are organized in a hierarchical structure so that the safest rules, those requiring a minimum amount of work and leap of faith, are applied first. This fourth step is by far the most demanding and the most complex. However, as one realizes looking more closely at Fig.5, it is absolutely necessary since a majority of the tree crown contours are not closed at this point.



(a) Original input image (b) Valley following tree isolation result

Figure 5: White spruces from PNFIs Hudson plantation with $31cm$/pixel data from the near-infrared channel

3.3 Fusion of Different Tree Finders

As a concluding remark on the sections about finding a tree, we wish to state that no one single visual module will be sufficiently robust to solve any real application in vision. One solution to this problem is to integrate several visual modules [1]. We have developed a general concept of information fusion in image understanding [24, 25], where different levels of abstraction (e.g. image, image description, scene description) are identified as sources for the *process* of fusion.

As a consequence of the Austro-Canadian research cooperation there are two different tree finders available, so that a *combination* of these methods can be tried out. In a similar way as was reported in [2] for fusion at the image level, a fusion of tree finders at the image description and at the scene description level is expected to improve the results significantly. Especially the combination of a region based and a contour based segmentation approach seems very promising.

3.4 Object Preprocessing

For the purpose of recognition, the image objects have to be initially separated from the entire image. This is accomplished through the localization of the object, discussed above, and setting up a window around it.

In our application, there are several factors that necessitate some kind of further preprocessing of the image before using it as the network input. Forest regions are usually nonhomogeneous, with different tree species overlapping each other. The shape of the tree in the image is affected by the position of the sun, the intensity of solar radiation, and by the exposition and slope of the terrain. Those factors can significantly influence tree recognition.

A procedure, applied by the Austrian group for nonhomogeneous forest, consisted of finding, upon the determination of location of the tree, the radius of the tree crown, and blanking out the part of the window that extends beyond the circle. This has the effect of removing the pixels from neighbouring trees and the ground.

A preprocessing procedure that has a significant impact on the classification results is the expansion of the input data by the rotated images. Practically, the original raster window will be rotated by 90, 180 and 270 degrees. The main advantage of this procedure is improved robustness of the response generated by the network. However, increasing the input field also increases the training time.

Other modifications of the initial data fields are contrast enhancement and intensity difference. A nonlinear, sigmoid contrast enhancement operation offers better results than that of the original input image. Calculating the difference between the pixel intensities for different spectral frequencies and using the resulting values rather than the original ones as the network input compensates to certain extent for the variations in image brightness.

4 Neuromorphic Methods

Neural networks (Connectionist models, parallel distributed processing models) are information processing systems which consist of a large number of very simple yet highly interconnected processing elements called units. Information is processed in a parallel and distributed manner by neural networks. The above features give the networks a fault tolerant behavior which is, in general, referred to as graceful degradation [32]. This means, that such systems can operate successfully in noisy environments, and also in cases when some of their components are damaged. Neural networks can be seen as intermediate between statistical and structural pattern recognition methods, though they resemble the former more than the later. The ability to learn (i.e. changing the weights of connections between units) provides an interesting alternative to conventional statistical methods.

Using neural networks for recognition purposes, many different network architectures and representation schemes can be considered. In this section, we discuss two most prominent models in more detail: feedforward networks and associative memory networks.

4.1 Feedforward Networks

The feedforward neural network is one of the most commonly used network architectures. The units are arranged in a layered manner. The first layer, also called the *input layer*, receives information from the environment. The final layer, called *output layer*, represents the result of the computation of the network. The layers between input and output layer are called *hidden layers*. The units are connected in a feedforward manner only, i.e. the input units to the first hidden layer, ..., the last hidden layer to the output units. When all units of one layer are connected to all units of another, the architecture is called a *fully connected feedforward network*.

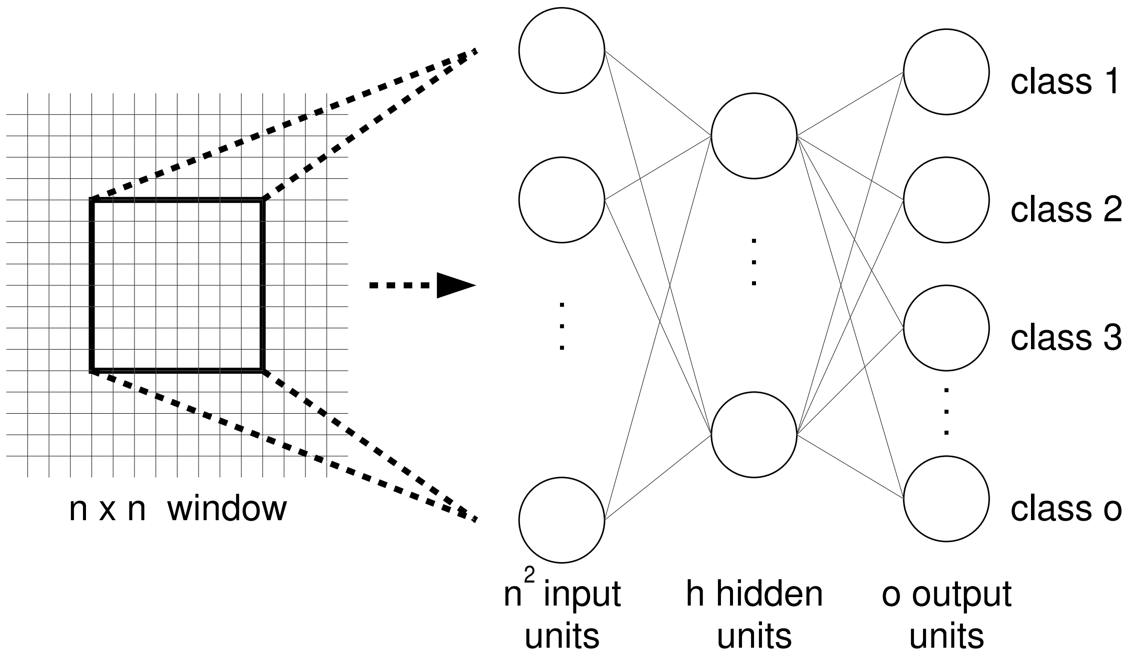


Figure 6: A feedforward neural network for image analysis

Figure 6 sketches a neural network architecure, where a $n \times n$ pixel window is used as input to a three layer feedforward neural network. The pixel matrix itself constitutes the input vector (n^2 input units), h hidden units and o output units are used to recognize o different classes of objects whenever they are present in the input window [5, 6, 26] (local representation for the output vector).

The function computed by a unit of a feedforward network is

$$\begin{aligned} net_i &:= \sum_j w_{ji} o_j + b_i \\ o_i &:= f(net_i) \end{aligned} \tag{1}$$

where w_{ji} is the weight of the connection from unit j to unit i. o_i, b_i and net_i are the output, bias, and effective input of unit i. f is the *activation function.* Typically

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

One of the most popular learning schemes for feedforward networks is the *back-propagation learning algorithm.* It was initially derived by Werbos [40] and later, independently by Parker [20], Le Cun [15], and Rumelhart et al. [33]. Back-propagation in its original version is designed to reduce an error between the actual and the desired output of a network in a gradient descent manner. The usual error measure, the summed squared error (SSE), is defined as

$$SSE = \frac{1}{2} \sum_p \sum_k (o_{pk} - t_{pk})^2 \tag{3}$$

where p ranges over all vectors in the training set and k denotes the output unit. Symbol o_{pk} indicates the output of unit k when the input vector is applied to the network, and t_{pk} is the corresponding target output. The change of the weights is achieved by a gradient descent search, thus

$$\Delta w_{ij} = -K \frac{\partial SSE}{\partial w_{ij}} \tag{4}$$

where Δw_{ij} is the weight change and K is the learning factor. A detailed derivation of the equations for the back-propagation algorithm can be found in [33].

Examples where back-propagation has been successfully used come from such diverse fields as text to phoneme conversion [34], protein structure prediction [29], backgammon [37], sonar target identification [9], tree species recognition, and many others.

There are several theoretical results which prove the strength of feedforward networks. Lippmann [17] showed, that a four layer network with threshold functions can represent any decision function. In [13], it was even shown, that a three layer network with sigmoid functions in the hidden units can approximate any Borel measurable function to any desired degree of accuracy (if enough hidden units are provided). Other results [31, 39] show, that a minimization of the mean (summed) squared error in multilayer feedforward networks approximates the a posteriori probabilities of the various classes. Recent results show that several statistical classifiers are special cases of neural networks [42, 43].

4.2 Associative Memory Networks

The learning time in feedforward networks using the backpropagation learning algorithm may turn out to be prohibitively long if large input data fields are involved in the recognition process. This has motivated the authors to search for other architectures that might offer the potential to effectively deal with large amounts of input information.

Particular attention was given to networks operating on the principles of holography [41]. The holographic models bear a formal similarity to correlation memory, a particular type of the associative memory. Stimulus-response associations are mapped directly to a correlation set. Large numbers of those associations may be superimposed on the same set, which allows the network to achieve high information density. The elements of the correlation set are complex numbers representing the input/output signals. The stimulus (**S**) and response (**R**) data fields may be represented as

$$\mathbf{S} = \{\lambda_1 e^{i\Theta_1}, \lambda_2 e^{i\Theta_2}, \ldots, \lambda_N e^{i\Theta_N}\} \tag{5}$$

and

$$\mathbf{R} = \{\gamma_1 e^{i\phi_1}, \gamma_2 e^{i\phi_2}, \ldots, \gamma_M e^{i\phi_M}\} \tag{6}$$

The analog value of an element of the signal is represented by the phase of a unit vector. The vector magnitude indicates a confidence level in that information. By establishing a confidence profile over stimulus and response and data fields, fuzzy logic operations are now made possible on those fields. Within HNeT Version 2.0 development system, used at l'Université du Québec à Hull, each data field may contain up to 1024 complex numbers.

As opposed to the error back-propagation learning, the individual associations in holographic nets are learned within one non-iterative transformation. The encoding process for multiple patterns may be represented in a canonical form by the following matrix formulation [35]:

$$\mathbf{X}+=\mathbf{S}^T\mathbf{R} \tag{7}$$

where $(+=)$ denotes complex valued multiply and accumulate operation. In the stimulus $[\mathbf{S}]$ and response $[\mathbf{R}]$ matrices, the information element index, related to the learned pattern, is on the horizontal, and the time element, related to consecutive learning cycles, is on the vertical axes. Thus:

$$\mathbf{S}=[s_{ij}]=[\lambda_{i,t_j}e^{i\Theta_{i,t_j}}] \tag{8}$$

$$\mathbf{R}=[r_{ij}]=[\gamma_{i,t_j}e^{i\phi_{i,t_j}}] \tag{9}$$

Assuming, for the sake of simplicity, only one neuron cell, the resulting correlation set is presented as a vector $[\mathbf{X}]$, where

$$\mathbf{X}=[x_j]=\sum_{t=1}^{l}\lambda_{i,t}\gamma_{i,t}e^{i(\phi_t-\Theta_{i,t})} \tag{10}$$

By applying the above process, the information content is preserved with the simultaneous enfolding of the time variable.

In its most elementary form, the process of learning a consecutive mapping has only a minimal influence on the previously learned mappings. The process can be modified by inclusion of, for instance, long term memory or higher order statistics.

5 Neuromorphic Results

5.1 Feedforward Networks

In the experiments reported here, we use a fairly standard type of three layer feedforward network. It is trained by the back-propagation learning algorithm to determine the species of trees in color infrared aerial photographs. Figure 7 shows a typical network with two 15×15 arrays of input units. Each of these units is used to represent one pixel of the original image. Since the blue and the green channels are strongly correlated, only the green and the red channel are used as input. Furthermore, 30 input units are used to feed a locally coded radius of the tree crown into the network, resulting in a total of 480 input units.

The pixel values of the original image are nonlinearly contrast stretched and transformed to the interval $[-0.5\ldots0.5]$. The training set is extended by rotating each tree by 90, 180, and 270 degrees. There are two benefits resulting from the use of this procedure: firstly, adverse effects caused by illumination and shadows are eliminated, and secondly, overfitting the data is avoided through the enlarged training set.

As initial experiments have shown, a network with 13 hidden units gives the best result in the case of a 15×15 input window and a discrimination between five different tree species. The output units use a local representation, i.e. each species (spruce, beech, fir, pine, larch) is represented by one dedicated output unit.

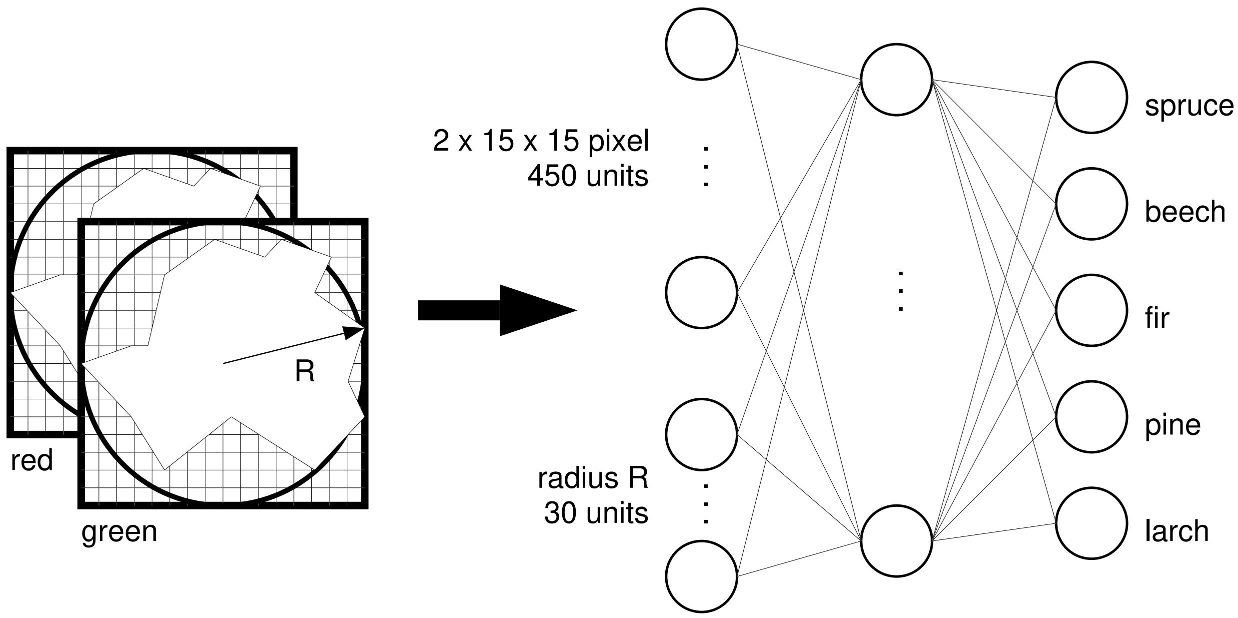


Figure 7: A tree species recognition network

The training set consists of 1024 trees, and the performance of the network is evaluated using an independent test set of 440 trees. To train the network, about 300 sweeps through the whole training set are necessary, resulting in a prediction accuracy of 86% for the training set and 85% for the test set for this basic network architecture.

We are able to further improve the above results by a method termed neural network surgery. The idea of this method is to combine several networks trained with different parameter settings. Hidden units which do not contribute to the classification are replaced by 'useful' hidden units from other networks. Details of this approach are described in [26, 27, 28]. Using neural network surgery, the prediction accuracy can be raised to 93% for the training set and 90% for the test set. Table 1 shows the confusion matrix for this network.

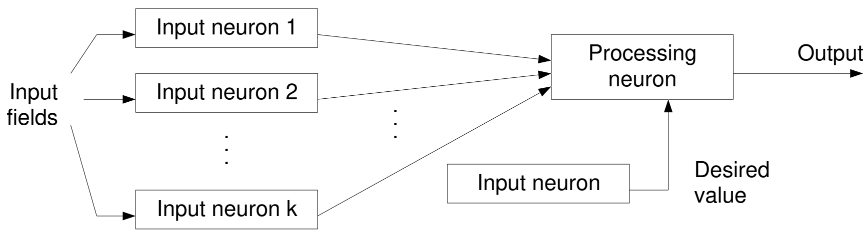
	classified as					correctly
species	spruce	beech	fir	pine	larch	classified
spruce	120	0	12	4	4	85.7%
beech	0	96	0	0	0	100.0%
fir	4	4	80	0	0	90.9%
pine	0	0	0	68	0	100.0%
larch	0	0	0	4	44	91.7%
						92.7%

Table 1: Confusion matrix for result of network surgery

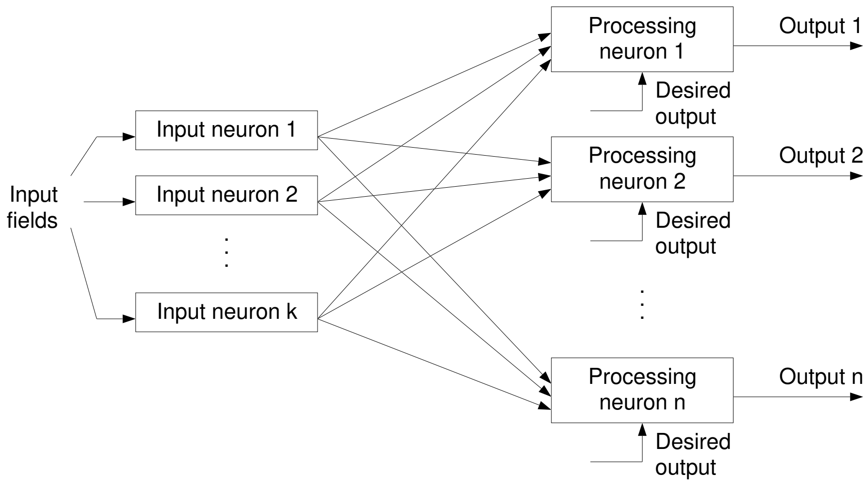
Summing up, the classification accuracy of the feedforward networks is very good but the training time is prohibitive, and the selection of appropriate parameters (e.g. number of hidden units) is a tedious process. Some of these defeciencies can be avoided by holographic networks.

5.2 Holographic Networks

Two types of architecture were investigated at the University of Quebec using HNeT simulator of AND America, Ltd.: one with a single neural cell and another, using multiple neural cell units (see Fig.8).



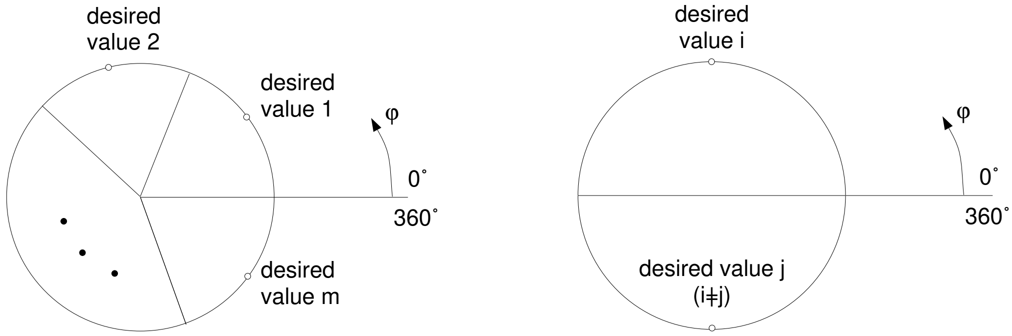
(a) Single neural cell



(b) Multiple neural cells

Figure 8: Holographic network architectures

The simpler architecture (Fig.8.a) consists of one neural cell performing encoding/decoding processes within the neural engine and four input neurons, one for each spectral frequency provided by the MEIS-II sensor. The architecture shown in Fig.8.b is characterised by one neural cell for every class of the output pattern, in our case – tree species. A different format for defining the desired output of the network during the learning phase is used for the two architectures. In the first, all the desired values are located in a single unit circle. This can be illustrated for the confidence level equal to 100%, by Fig.9.a. In the second architecture, each neural cell distinguishes between one particular class of objects and all other classes (Fig.9.b).



(a) Single neural cell architecture

(b) Multiple neural cells architecture

Figure 9: Definition of the desired output value

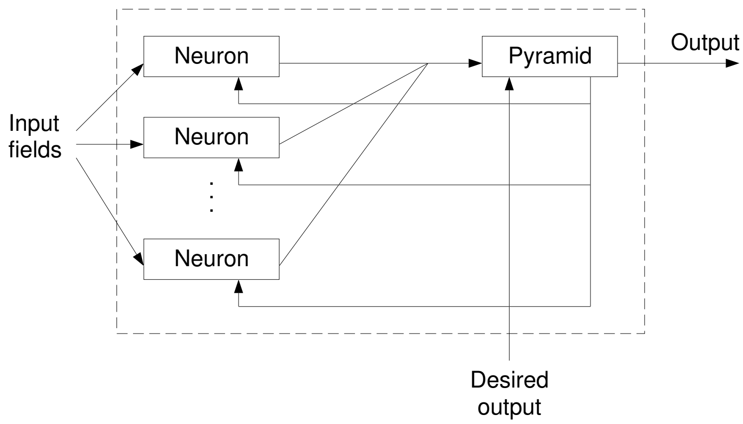
The results of the tests performed using the first type of network architecture showed a recognition rate of 50% - 60%. Consequently, more attention was given to the second type of architecture, where the networks with two compound cell structures, cortex and cerebellum, were analyzed (Fig.10).

The cortical cell is built upon multiple neurons executing the fundamental encode/decode function as a single pyramid cell fed by neuron cells. The cerebral cell executes the sigma-pi encoding/decoding process, and provides the option of expanding the stimulus field by using statistical terms. Some of the results of the application of the multiple neural cells architecture [44] are given in Table 2.

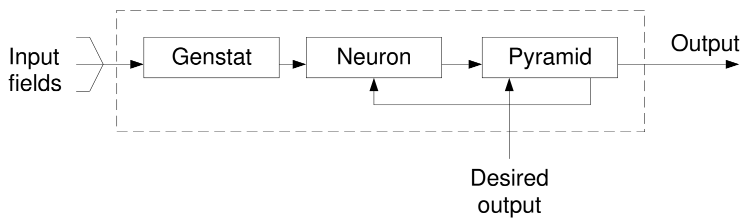
Neuron cell	Statistics		Correct Classification [%]		
	order	number of terms	red pine	red spruce	white spruce
Cortex			48	82	74
Cortex + sigmoid			76	96	88
Cerebellum	5	200	76	98	80
Cerebellum	4	200	72	98	82

Table 2: Classification rates for multiple neural cells architecture

The statistics, or higher order product terms, shown in Table 2 for cerebellum-type cells are the principal methods of achieving symmetry in an input data field, defined by a uniform probabilistic distribution of complex vectors oriented around the origin on a complex plane. Obtaining a highly symmetrical state over a wide class of input data distributions is an important condition for object discrimination. The two statistics parameters indicated in Table 2 signify the order of the statistics and the number of higher order product terms generated within the cell's output data field. These terms are generated as a function of the stimulus field size and the order of the statistics. They are particularly useful when the size of the input field is small.



(a) Cortex cell structure



(b) Cerebellum cell structure

Figure 10: Compound neural cells

As it can be seen from Table 2, the holographic networks do not provide a uniform and robust solution as the feedforward, error back-propagation networks. A high percentage recognition of certain objects, in this case red spruce, should be noted.

6 Conclusions

The contribution of this paper is twofold: Firstly, we have presented a general approach for recognition of compact image objects, based on a hybrid system. Interesting regions (i.e. regions which contain objects) are selected by knowledge based systems and then classified by neuromorphic methods. Secondly, we have presented two specific neuromorphic methods for the recognition as well as two specific methods for the location of trees in high spatial resolution aerial images.

Tree finding is accomplished by a region based approach which looks for bright blobs in the image and by a contour based approach which looks for the outline of a crown. The region based approach is successful in areas of dense forest ($> 90\%$ correctly located crowns), while the contour based approach works better in cases of sparse forest or at forest boundaries.

With respect to the two neuromorphic methods, we can say that feedforward networks trained with back-propagation have good prediction accuracy, whereas the training time is much lower for holographic networks (approximately 300 times).

Holographic networks, when directly applied to tree species recognition perform less well than the classical feedforward networks. However, due to other features, such as good approximation capabilities (shown in other experiments), memory-like operation, and fast learning time they may well be considered as a part of more complex, hybrid neural systems.

In both cases, tree finding as well as neuromorphic recognition, there are at least two different methods available for each task. A *combination* of these methods in the sense of information fusion in image understanding [24] seems very promising and will be one of our future research goals.

We are also interested in an 'all neural' system, where attention focusing is accomplished by modular and hierarchical neural networks [8, 4].

7 Acknowledgments

We are very grateful to Prof. Walter G. Kropatsch for his valueable suggestions on the homogeneity and the structure of this paper. This work was supported by the following grants:

- Österr. Waldzustandserhebung mittels Methoden der Fernerkundung, BMLF,
- Schwerpunkt Fernerkundung, S38/2, FWF,
- Image understanding and standardized image processing, P8785, FWF,
- International project initiation, A0052-PHY, FWF, and
- National Research Council Canada, grant BEF0131730.

References

[1] John Aloimonos and David Shulman. *Integration of Visual Modules. An Extension of the Marr Paradigm.* Academic Press, 1989.

[2] Renate Bartl and Axel Pinz. Information fusion in remote sensing: Land use classification. In *IAPR TC7 Workshop 92*, 1992. in press.

[3] Horst Bischof. Interpretation von Fernerkundungsdaten mit Hilfe von Backpropagation Netzwerken am Beispiel der Baumerkennung aus Farb-Infrarot Luftbildern. Master's thesis, TU-Vienna, 1989.

[4] Horst Bischof. Modular, Hierarchical, and Geometrical Neural Networks. Technical Report 9, Dept. for Pattern Recognition and Image Processing, TU-Vienna, December 1991.

[5] Horst Bischof and Axel Pinz. Verwendung von Neuralen Netzen zur Bestimmung der Baumart aus Digitalen Rasterbildern. In Axel Pinz, editor, *Wissensbasierte Mustererkennung*, volume 49 of *OCG-Schriftenreihe*, pages 149–161. Oldenbourg, 1989.

[6] Horst Bischof and Axel Pinz. Verwendung von Neuralen Netzwerken zur Klassifikation Natürlicher Objekte am Beispiel der Baumerkennung aus Farb-Infrarot-Luftbildern. In Georg Dorffner, editor, *Konnektionismus in Artificial Intelligence und Kognitionsforschung*, volume 252 of *IFB*, pages 112–119. Springer, 1990.

[7] Horst Bischof and Axel Pinz. The invariance problem for hierarchical neural networks. In *Neural and Stochastic Methods in Image and Signal Processing*, volume 1766, pages 118–129. SPIE, 1992.

[8] Horst Bischof and Axel Pinz. Neural networks in image pyramids. In *IJCNN'92*, volume IV, pages 374–379, 1992.

[9] P. Gorman and T. Sejnowski. Learned classifications of sonar targets using a massively parallel network. *IEEE Trans. Acoust. Speech Signal Processing*, 36:1135–1140, 1988.

[10] F.A. Gougeon. Individual tree identification from high resolution MEIS images. In *International Forum on Airborne Multispectral Scanning for Forestry and Mapping (with Emphasis on MEIS)*, Val-Morin, Quebec, 1992.

[11] F.A. Gougeon. A valley following approach to the automatic delineation of individual trees in high resolution MEIS images, 1993. submitted for publication.

[12] G. Hinton. Connectionist Learning Procedures. *Artificial Intelligence*, 40:185–234, 1989.

[13] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[14] Walter G. Kropatsch. Image Pyramids and Curves an Overview. Technical Report 2, Dept. for Pattern Recognition and Image Processing, TU-Vienna, March 1991.

[15] Y. Le Cun. Learning processes in an assymetric threshold network. In Bienenstock, Fogelman Souli, and Weisbruch, editors, *Disordered Systems and Biological Organization*. Springer, 1986.

[16] D.G. Leckie. Application of airborne multispectral scanning to forest inventory mapping. In *International Forum on Airborne Multispectral Scanning for Forestry and Mapping (with Emphasis on MEIS)*, Val-Morin, Quebec, 1992.

[17] R.L. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4:4–22, 1987.

[18] W.D. McColl, R.A. Neville, and S.M. Till. Multi-detector Electro-optical Imaging Scanner MEIS II. In *8th Can. Symp. on Remote Sensing and 4th Conf. of l'Association quénecoise de télédétection*, Montreal, Quebec, 1983.

[19] Yon Han Pao. *Adaptive Pattern Recognition and Neural Networks*. Addison-Wesley, first edition, 1989.

[20] D. B. Parker. Learning-logic. Technical report, Sloan school of management, MIT, Cambridge, MA, 1985.

[21] Axel Pinz. Final Results of the Vision Expert System VES: Finding Trees in Aerial Photographs. In Axel Pinz, editor, *Wissensbasierte Mustererkennung*, volume 49 of *OCG-Schriftenreihe*, pages 90–111. Oldenbourg, 1989.

[22] Axel Pinz. Wissensbasierte Informationsverarbeitung in der österreichischen Waldzustandsinventur. In *Handbuch der Umwelttechnik '91*, pages 228–232. Trend Commerz, 1990.

[23] Axel Pinz. A Computer Vision System for the Recognition of Trees in Aerial Photographs. In James Tilton, editor, *Multisource Data Integration in Remote Sensing*, volume 3099 of *NASA Conference Publication*, pages 111–124. NASA, 1991.

[24] Axel Pinz and Renate Bartl. Information fusion in image understanding. In *Proceedings of the 11.ICPR*, volume I, pages 366–370. IEEE Computer Society, 1992.

[25] Axel Pinz and Renate Bartl. Information fusion in image understanding: Landsat classification and ocular fundus images. In *SPIE Sensor Fusion V, Boston 92*, volume 1828, pages 276–287. SPIE, 1992.

[26] Axel Pinz and Horst Bischof. Constructing a Neural Network for the Interpretation of the Species of Trees in Aerial Photographs. In *Proc. of the 10.ICPR*, pages 755–757. IEEE Computer Society, 1990.

[27] Axel Pinz and Horst Bischof. Neural network 'surgery': Transplantation of hidden units. In Bernd Neumann, editor, *Proceedings ECAI'92*, pages 214–215. Wiley, 1992.

[28] Axel Pinz and Horst Bischof. Neural network 'surgery': Transplantation of hidden units. Technical Report PRIP-TR-15, PRIP, TU Wien, 1992.

[29] N. Qian and T.J. Sejnowski. Predicting the Secondary Structure of Globular Proteins using Neural Network Models. *Journal of Molecular Biology*, 202:865–884, 1988.

[30] M.W. Roth. Survey of Neural Network Technology for Automatic Target recognition. *IEEE Transaction on Neural Networks*, 1(1):28–43, 1990.

[31] D.W. Ruck, S.K. Rogers, M. Kabrisky, M.E. Oxely, and B.W. Suter. The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Trans. Neural Networks*, 1:296–298, 1990.

[32] David E. Rumelhart and James A. McClelland. *Parallel Distributed Processing*, volume 1. MIT Press, first edition, 1986.

[33] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[34] T.J. Sejnowski and C.R. Rosenberg. NETtalk: A parallel network that learns to read aloud. Technical Report JHU/EECS-86/01, John Hopkins University, Baltimore, MD, 1986.

[35] B. Soucek, editor. *Fuzzy, holographic, and parallel intelligence.* John Wiley & Sons, 1992.

[36] T.J. Sejnowski S.R. Lehky. Network model of shape-from-shading: Neural functions araises from both the receptive and projective field. *Nature*, 333:452–454, 1988.

[37] G. Tesauro and T.J. Sejnowski. A parellel network that learns to play backgammon. *Artificial Intelligence*, 39:357–390, 1989.

[38] J.K. Tsotsos. Analyzing vision at the complexity level. *BBS*, 13(3):423–469, 1990.

[39] E.A. Wan. Neural network classification: a Bayesian interpretation. *IEEE Trans. Neural Networks*, 1:303–305, 1990.

[40] P.J. Werbos. *Beyond regression: New tools for prediction and analysis in the behavioral sciences.* PhD thesis, Harvard University, Cambridge, MA, 1974.

[41] D. Willshaw. Holography, associative memory, and inductive generalization. In G.E. Hinton and J.A. Anderson, editors, *Parallel models of Associative Memory.* Lawrence Erlbaum, 1989.

[42] E. Yair and A. Gersho. Maximum *a posteriori* decision and evaluation of class probabilities by Boltzmann perceptron classifiers. *Proc. IEEE*, 78(10), 1990.

[43] H-C. Yau and M.T. Manary. Iterative improvement of Gaussian classifier. *Neural Network*, 3:437–443, 1990.

[44] M.B. Zaremba, M. Locas, M. Malak, and F.A. Gougeon. La reconnaissance des espèces de conifières sur une image MEIS haute définition par les rèseaux neuronaux. Technical Report RR 93/02-01, U.Q.A.H., Dèpt. d'Informatique, 1993.