Knowledge Acquisition by Symbolic Decision Tree Induction for Interpretation of Digital Images in Radiology

Petra Perner*, Tatjana B. Belikova, Nadeszda I. Yashunskaya*****

*** Institute of Computer Vision and applied Computer Sciences, PSF 1519, 04257 Leipzig, Germany, perner@imise.uni-leipzig.de**
**** Russian Acad. of Sciences, Inst. of Information Transmission Problems, Ermolovoy 19, Moscow GSP 101447, tatjana@lum.fian.msk.su**
***** Moscow Medical Academy, Bolshaza Pirogovskaja 2/6, Moscow GSP 101447**

Abstract

Knowledge Acquisition is an important task when developing image interpretation systems. Whereas in the past this task has been done by interviewing an expert, the current trend is to collect large data bases of images associated with expert description (known as picture archiving systems). This makes it possible to use inductive machine learning techniques for knowledge acquisition of image interpretation systems. We use decision tree induction in order to learn the symbolic knowledge for image interpretation. We applied the method to interpretation of x-ray images for lung cancer diagnosis. In the paper, we present our methodology for applying inductive machine learning. We discuss our results and compare it to other knowledge acquisition methods.

1 Introduction

Knowledge acquisition is the first step when developing an image interpretation system. The kind of method used for knowledge acquisition depends on the inference method the image interpretation system is based on.
The knowledge acquisition process for rule-based system is usually manually done by interviewing a human expert [Per94] or by employing interactive knowledge acquisition tools like for e.g. repertroy grid [BSB89].
In model-based systems, the knowledge about the objects is represented based on semantic nets that structure the knowledge into concepts and their relations. The language of the semantic net determines the way how new knowledge is elicitated. Kehoe et al. [KeP91] describe a model based system for defect classification of welding seams. The knowledge base is manually mantained by specializing or generalizing the defect classes, their attributes, and attribute values. Schröder et al. [SNS88] described a system where knowledge acquisition is done automatically based

on the language of the semantic net. Although semantic nets seem to be the most convinient way of representing and eliciting knowledge, this method requires a deep understanding of the domain, which is not given a-priori for all applications.
When generalized knowledge is lacking, then case based reasoning [KSS85] seems to be a proper method. The system is based on a case base consisting of a set of cases. An interpre-tation is done by determining the closest case or cases in the case base to the actual case and by displaying the value of the closeness measure and the interpretation associated with the similar case of the case base. How to interpret the closeness measure is left to the user. The limited explanation capabilities are the main drawback of case based reasoning systems. A system which uses case based reasoning for image interpretation is described in [PeP95].
The aim of our work is to develop a knowledge acquisition method for such applications where no generalized knowledge about the domain is available but a large data base of images associated with expert description and interpretation. If we think of the recent trend to picture archiving systems in medicine and other domains, such a task becomes quite important. The relevant attributes for interpretation and the decision model should be learnt by applying symbolic decision tree induction methods to the data base.
In the paper, we present our methodology for applying inductive machine learning methods for image interpretation. In Section 2, we describe the decision tree induction algorithm used for the investigation. The chosen domain vocabular and the experiment set up is given in Section 3. We discuss our results in Section 4 and compare the method and the results with other knowledge acquisition methods in Section 5.

2 Knowledge Acquisition by Decision Tree Induction

Decision trees partition decision space recursively into subregions based on the sample set. By doing so they recursively break down the complexity of the decision space. The represen-tation form which comes out is a format which naturally covers the cognitive strategy for human decision making process.
A decision tree consists of nodes and branches. Each node represents a single test or decision. In the case of a binary tree, the decision is either true or false. Geometrically, the test describes a partition orthogonal to one of the coordinates of the decision space. The starting node is usually referred to as the root node. Depending on whether the result of a test is true or false, the tree will branch right or left to another node. Finally, a terminal node is reached (sometimes referred to as a leaf), and a decision is made on the class assignment. All the paths in a tree are mutually exclusive. For any new case, always one and only one path in the tree has to be satisfied. Also nonbinary decision trees are widely used. In these trees, more than two branches may leave a node, but again only one branch may enter a node. In this type of tree, a test performed at a node results in a partition of two or more disjoint sets that cover every possibility, i.e., any new case must fall into one of the disjoint subsets. For any tree, all paths lead to a terminal node corresponding to a decision rule that is a conjunction (AND) of various tests. If there are multiple paths for a given class, then the paths represent disjunctions (ORs) [WEK91].

The most used criterion for automatic splitting of the sample set [Quin86][VGO94] that is simple to calculate and performs well, is the Shannon entropy:

$$I = - \sum_i p_i \log_2 p_i \qquad (1).$$

Note that I has a maximum value when all the p_i's are equal; that is, the classes are uniformly distributed throughout the set. This means that there is still a lot of information in this set. I is minimized to zero if one of the p_i's is unity and the others are zero: In this case all examples belong to one class, and there is no more information in the set.
Now if we subdivide the subset according to the values of an attribute, we shall have a number of subsets. For each of these subsets we can compute the information value. Let the information value of subset n be i_n, then the new information value is given by

$$I_i = \sum q_n \; i_n \qquad (2),$$

where q_n is the portion of examples having attribute values n. I_i will be smaller than I, and the difference $(I - I_i)$ is a measure of how well the attribute has discriminated between different classes. That attribute that maximizes this difference will be selected.
Since only relevant attributes are chosen as decision rules, decision tree induction can also be considered as a method for attribute selection. However, the entropy in Eq. 1 requires uncorrelated attributes. Two linear correlated attributes would bring nearly the same result but only the first appearing attribute, which might not be the truly relevant attribute, is chosen for the next node. The second attribute, which has not been chosen for the node, is not sorted out, it is still left in the sample set and gets still processed during the tree building process.
The recusive partitioning method of constructing decision trees will continue to sub-divide the set of training cases until each subset in the partition contains cases of single classrs, or until no test offers any improvement. For this tree based on the sample cases the error rate is:

$$E = S_m / N, \qquad (3)$$

where S_m is the number of samples that were misclassified and N is the whole number of samples.
The result is often a very complex tree that "overfits the data" by inferring more structure than is justified by the training cases. Therefore, pruning techniques are used which simplify the tree by discarding one or more subtrees and replacing them with leaves. We use a reduced-error pruning technique [Quin87] which accesses the error rates of the tree and its components directly on the set of cases. The predicted error rate is

$$E_{pred} = \sum_i N_i \, U_{CF} (E_i, N_i), \qquad (4)$$

where N_i is the number of sample cases covered by the leave, E_i is the number of misclassified sample cases covered by the leave, $U_{CF}(E_i, N_i)$ is the upper limit on the probability for a chosen confidence value CF for the binominal distribution and i is the number of leaves. This implies that a tree with fewer leaves will have a lower error rate than a tree with more leaves.
We also calculate the class identification

$$E_{id} = S_{CLm} / N_{CL}, \tag{5}$$

where S_{CLm} is the number of misclassified samples of particular classes and N_{CL} is the number of all samples of particular classes.
Sensitivity for Class 1 and Specifity for Class 2 were calculated as well:

$$E_{sens/spec} = S_{CL} / N_{CL} \tag{6}$$

where S_{CL} is the number of correctly classified samples and N_{CL} is the number of all samples of Class_1 and Specifity for Class_ 2 respectively.
Decision trees can be built up top-down [BaM95] or bottom-up [VGO94]. Our decision tree is built top-down. Whereas the intension of the work of most others is to develop decision tree construction methods that outperform other classifiers in correct classification and excecution time, the intension of our work is more on how decision tree induction can be used for knowledge acquisition.

3 Experiment Set up

The algorithm described in Section 2 was realized by a tool for inductive machine learning, called SALOMON [TrP95]. The devoloped tool satisfies experts needs by many functions for carrying out induction experiments for knowledge acquisition.
For the database tomograms of 250 patients with verified diagnoses were used (80 cases with benign disease and 138 cases with cancer of lung). Patients with small pulmonary nodules (up to 5 cm size) were selected for this test. Conventional (linear) coronal plane tomograms with 1 mm thickness of section were used for specific diagnosis.
Original linear tomograms were digitized with step of 100 micron (5.0 line pairs per millimeter) to get 1024 x 1024x 8 bits matrices with 256 levels of gray, see Fig. 1. The use of linear tomograms and such a digitization enabled an acquisition of high spatial resolution of anatomical details that were necessary for the specific diagnosis of lung nodules.
To improve results of specific diagnosis of small solitary pulmonary nodules we used optimal digital filtering [BYK94] and analysis of post-processed images. The processing emphasized diagnostically important details of the nodule and thus helped to improve the reliability of image analysis: the physician was more certain in feature reading and interpretation. The radiologist worked as an expert on this system.

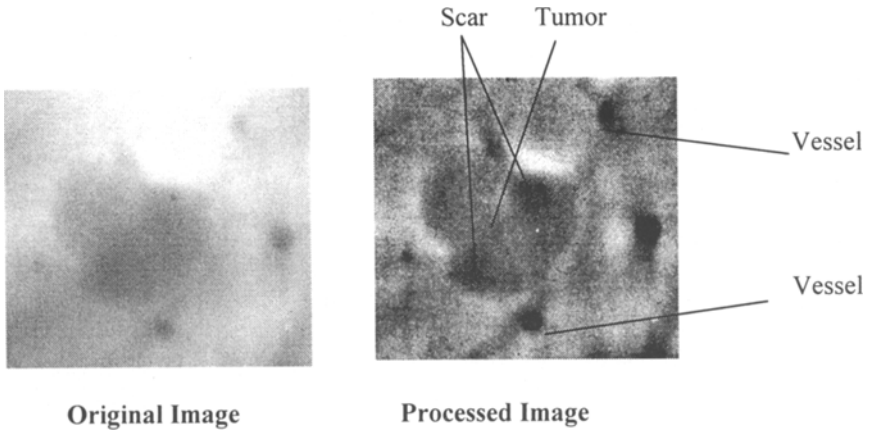

Fig. 1 Original and Processed Image with Description of Image Details

First, together with the expert an attribute list was set up , which covered all possible attributes used for diagnosis by the expert as well as the corresponding attribute values, see Table 1.
Then, the expert collected the database and communicated with a computer answering to its requests. He determined whether the whole tomogram or its part had to be processed and outlined the area of interest with overlay lines and he also outlined the nodule margins. The parameters of optimal filter were then calculated automatically . A radiologist watched the processed image (see Fig. 1), displayed on-line on a TV monitor, evaluated its specific features (character of boundary, shape of the nodule, specific objects, details and structures inside and outside the nodule, etc.), interpreted these features according to the list of attributes and inputted the codes of appropriate attribute values into computer with Excel program. Hard copies of the previously processed images from the archive have been used in this work as well.
The collected data set was given as a dBase-file to the inductive machine learning tool.

4 Results

The induced tree is shown in Fig. 2 (The tool SALOMON actually shows the tree as a directed graph on the monitor).
The unpruned tree consits of 40 leaves. The pruned tree consits of 11 leaves, see Fig. 3. The expert liked the unpruned tree much more since nearly all attributes he is using for decision making appeared in the tree. The expert told us that the attribute ***Structure*** is very important, also the ***attribute Scar-like changes inside the nodule***.

No.	Attr.type	Attr. name	Shortname	No.	Attr. value
1	boolean	Class	Class	1 2	malignant benign
2	categorical	Structure inside the nodule	StrInsNod	1 2 3 4 5 6 7	Inhomogeneous with disorderly structures Inhomogeneous with orderly structure: regularly decreasing film density along the periphery of the nodule Areas with calcifications Enough homogeneous structures Inhomogeneous with calcifications Inhomogeneous with orderly structures and calcifications Inhomogeneous with cavities
3	categorical	Scar-like changes inside the nodule	ScrLikeChan	1 2 3	Irregularly shaped fragmentary dense shadow Regular dense shadow along pheriphery None
4	categorical	Shape	Shape	1 2 3 4 5	Nonround Round Oval Lobular Angular
5	categorical	Margin	Margin	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	Nonsharp Sharp Nonsmooth Smooth Lobular Angular Spicular Nonsharp-sharp: in some regions it is nonsharp, in others it is sharp Nonsharp and Nonsmooth Nonsharp and Angular Nonsharp and Spicular Sharp and Smooth Sharp and Lobular Sharp and Angular Nonsharp-sharp and Angular
6	categorical	Convergence of vessels	ConvofVes	1 2	Vessels constantly exist converging to the nodule Vessels are forced away the nodule
7	categorical	Outgoing Shadows in surrounding tissues	OutgoShadinTis	1 2 3 4	Chiefly vascular Outgoing sharp tapelines (septa) None Invasion into surrounded tissues
8	continous	Size of Nodules	Size		Values in cm (e.g. 1,2 := 1,2 cm)
9	categorical	Character of the Lung pleura	CharPleu	1 2 3 4	Thickening Withdrawing None Thickening with Withdrawing

Table 1 Attribute List and Classes

However the expert wonders why other features such as ***Shape*** and some others didn't work for classification. The expert told us that he usually analyzes a nodule starting from its ***Structure***, then tests ***Scar-like changes inside the nodule***, then ***Shape*** and ***Margin***, ***then Convergence of Vessels*** and ***Outgoing Shadow in Surrounding tissues***. On his opinion, in many cases they are important for the final decision as well.

```
Scar-likeChan = 1
I       ConvofVes = 1   Class_1
I       ConvofVes = 2   Class_2
I       ConvofVes = 3
I               Margin = 1  Class_1
I               Margin = 2  Class_2
I               Margin = 3  Class_1
I               ...
I               Margin = 15 Class_1
Scar-likeChan = 2
I       Structure  =1  Class_1
I       Structure  =3  Class_2
I       ...
I       Structure = 6  Class_2
I       Structure = 7  Class_1
I       Structure = 2
I       I       OutgoShadinTis = 1  Class_2
I       I       ...
I       I       OutgoShadinTis = 3  Class_2
I       I       OutgoShadinTis = 4  Class_1
        Scar-likeChan = 3
I       Structure = 1  Class_1
I       Structure = 2  Class_1
I       Structure =3   Class_2
I       ...
I       Structure = 6   Class_2
I       Structure = 7   Class_1
```

Fig. 2 Decision Tree

```
Scar-likeChan = 1  Class_1
Scar-likeChan = 2  Class_2
Scar-likeChan = 3
I       Structure = 1  Class_1
I       Structure =2  Class_1
I       Structure = 3   Class_2
I       ...
I       Structure = 6  Class_2
I       Structure = 7  Class_1
```

Fig. 3 Pruned Tree

Although decision trees represent the decision in a human understandable format, the decision tree might not represent the strategy used by an expert since always the attribute appearing first in data base and satisfying the splitting criteria, is chosen.

Therefore, we looked for the error rate as main criterion, see Tab. 2 and Tab. 3.

We did not come close to the expert's performance. One reason might be the choice of attribute values. For some categorical attributes, there are too many categorical values. That causes that during the tree building process the training set gets split up into too many subsets with few data samples. As a result the tree building process will stop very soon since no discrimination power is left in the remaining data samples.

	Before Prunning		*After Prunning*		
	Size	**Error**	**Size**	**Error**	**Estimated Error**
(1)	11	1,0%	11	1,0%	11,2%
(2)	11	11,4%	11	14,4%	11,2%

Tab. 2 Result (1) and Evaluation of Decision Tree on Test Data (2)

Since the attributes are nominal we cannot find an ordering on attribute values. Therefore, we cannot summarize attributes values to a more general attribute value. For e.g., think of an attribute ***intensity*** with attribute values „*black, dark grey, grey,*

light grey, white". The attribute values *dark grey, grey and light grey* we can be gereralized to *grey*. First, we can use the gerneralized attribute value for tree

Accuracy		*Sensitivity/Specitity*			
		Class_1		Class_2	
Human	DT	Human	DT	Human	DT
94,7%	89,6%	97,2%	88,8%	91%	88%

Tab. 3 Comparison between Human Expert and Decision Tree Classification

building process and if we notice in the induced tree that further destinction between the attribute values is necessary then we can carry out another induction experiment based on the specialized attribute values starting with the data set corresponding to the leaf of the tree with the generalized attribute value. This approach is proposed by Shapiro [Sha85]. In one classification problem he studied, this method reduced a totally opaque, large decision tree to a hierarchy of nine small decision trees, each of which 'made sense' to an expert. The way we chose was the construction of new attributes. For the attributes with many attribute values we tried to find attributes representing one or two of the attribute values in a boolean fashion or with lower attribute values, see Table 4.

No. of Attr.	Attr. type	Attr. Name	Short Attr. name	No. of Values	Attr. value
...	...	...	...	...	...
8	categorical	Sharpness of Margins	SharpMar	1 2 3	nonsharp mixedsharp smooth
9	categorical	Smoothness of margin	SmoothMar	1 2 3	nonsmooth mixedsmooth smooth
10	boolean	Lobularity of Margin	LobMar	0 1	nonlobular lobular
11	boolean	Angularity of Margin	AngMar	0 1	nonangular angular
12	boolean	Spicularity of Margin	SpicMar	0 1	nonspicular spicular
...	...	...	...	...	...
14	boolean	Vascular Outgoing Shadow	Vascshad	0 1	none chiefly vascular shadows
15	boolean	Outgoing Sharp thin tape lines	OutgoSha	0 1	none Outgoing shapr thin tape lines
...	...	...	...	...	...
18	boolean	Thickening of lung pleura	ThLungPl	0 1	none thickening
19	boolean	Withdrawing of lung pleura	WithLupl	0 1	none withdrawing
...	...	...	...	...	...

Tab. 4 Second Attribute List and Classes

In order to make sure that we did not develop many redundant and highly correlated attributes we checked the reliablity of features by calculating a proximity matrix based on Kruskal's tau [Agr90] from the new data base and gouping the set of features into functional groups based on an average link hierarchical clustering method [JaC88] . Only between the attributes ***Charlung*** and ***Withlupl*** we had high correlation. For all other attributes we were satisfied with the result. The resulting decision tree (see Fig. 4) performs better then the first decision tree. However, the decision tree is harder to interpret from a human point of view. From the expert's point of view there are too less attributes.

```
Angmar = 0
I        Spicmar = 1  Class_1
I        Spicmar = 0
I        I        Invsourtis = 1  Class_1
I        I        Invsourtis = =
I        I        I        Scarinsnod = 0
Class_2
I        I        I        Scarinsnod = 1
Class_2
I        I        I        Scarinsnod = 2
Class_1
Angmar = 1
I        Siofnod > = 0.5  Class_1
I        Siofnod <= 0.5
I        I        Regstrins = 0  Class_1
I        I        Regstrins = 1  Class_1
I        I        Regstrins = 2  Class_2
I        I        Regstrins = 3  Class_1
```

Fig. 4 Decision Tree

Therefore, it was interesting to see how the tree performed on test data set.

The error rate on test data set was better than the error rate of the first tree, see Table 5. Also compared to a high level expert's performance the tree performed better, see Table 6. In another test, we used test data set obtained by image reading from a middle level expert on difficult cases of images. This expert had no special courses of processed image reading and interpretation. It was our special intension to obtain as much uncertainty and noise in the data as possible. As the expert did not know how to read a new roentgenological picture that appeared after digital image processing, in many cases he could not choose which attribute value was suitable for this case (marked as missing attribute value in data base).

	After Prunning		*Before Prunning*		
	Size	**Error**	**Size**	**Error**	**Estimated Error**
(1)	11	1,3%	11	1,3%	8,0%
(2)	11	4,2%	11	4,2%	8,0%

Tab. 5 Results (1) and Evaluation of Decision Tree on Test Data (2)

Accuracy		*Sensitivity/Specifity*			
		Class_1		**Class_2**	
Human	**DT**	**Human**	**DT**	**Human**	**DT**
94,5%	**95,8%**	**96,2%**	**93,75%**	**90%**	**100%**
55,2%	**73%**	**61,1%**	**75%**	**50%**	**72%**

Tab. 6 Comparison between Human Expert and Decision Tree Classification
(1) high-level Expert (2) middle-level Expert

These readings were given to the decision tree for classification. The resulting error rate showed that classifier based on decision tree gives reliable error rate even by bad image readings, see Table 6.

5 Comparision to other Knowledge Acquisition Methods

As mentioned in the introduction of this paper, other methods like for e.g. model-based reasoning are not applicable to this problem since they require to have a model of the domain. Case based reasoning can be used as well, but has the drawback that the knowledge used for classification cannot be made explicit.
We compared the performance of the induced tree classifier to a rule based system built with the help of an interactive knowledge acquisition tool [Pere92]. The tool employs questionnaire strategies for determining symptoms and classes by interviewing an expert and data analysis methods like clustering methods and multi-layer threshold networks. The knowledge base of the system consists of a set of decision rules as a multi-layer network of threshold elements and a voting rule scheme [BStY95]. Expert knowledge is used to control the decision-making process. The knowledge base has some redundancy in knowledge for improving the reliability of the decision-making model. In contrast to that, the rules obtained by decision tree induction are built up according on the minimum description lenght principle, i.e. we ask for the minimal number of bits needed to be coded for inferring a particular class.
The test based on image readings from a high-level expert shows that the decision tree method performs much better than expert system, see Table7. Only in the case of the middle-level expert where the data contain many missing and incorrect values the expert system performs better. An investigation shows that misclassification of samples is mostly based on incorrectly chosen attributes not on missing attributes, since the classifier is based only on a few attributes that are the most important ones and that always appear in the image. However, the decision tree classifier has no special strategy to deal with such kind of knowledge. In general, we believe that rather the inclusion of special search strategies [LiE89] over decision tree in the inference process than the adding of redundant rules to the rule set can improve the performance.
Unlike in rule based systems where we can add a rule to the rule base without any problem, in decision tree induction we cannot readily update the tree without having to rebuild the entire tree. However, there is some research going on for incremental decision tree induction [Utg89], but that is not the subject of this paper.

Accuracy			*Sensitivity/Specifity*			
			Class_1		Class_2	
Expert System ES		**Decision Tree DT**	**ES**	**DT**	**ES**	**DT**
(1)	96%	98,7%	98,4%	98,9%	91,6%	98%
(2)	84%	73%	90%	75%	77%	72%

Tab. 7 Comparison between Expert System and Decision Tree Classification
(1) Data Readings from high-level Expert (2) middle-level Expert

6 Conclusion and Further Work

In the paper, we presented our methodology for knowledge acquisition for image inter-pretation. We assumed to have a large enough data base with images and associated with expert descriptions. From this data base we could learn the important attributes needed for inter-pretation and the way how they were used for decision making by applying inductive machine learning methods. We showed how the domain vocabular should be set up in order to get good results and which techniques could be used in order to check reliability of the choosen features.
Explanation capability of the induced tree was reasonable. The attributes included in the tree represented the expert knowledge.
We compared our methodology to standard rule based system. The error rate for classification based on decision tree was better than the error rate obtained by using standard knowledge based system. Only for handling very uncertain datas the performance of the decision tree got worsen than the expert system. However, we believe that special search strategies for the decision tree method during the decision making process would improve our results. But this is left for further research.

References

[Agr90] A. Agresti, „Categorical Data Analysis," John Wiley Inc. 1990

[BaM94] H.S. Baird and C.L. Mallows, „ Bounded-Error Preclassification Trees," In: Shape, Structure and Pattern Recognition, Eds. D. Dori and A. Bruckstein, World Scientific Publishing Co.,1995, pp. 100-110.

[BSB89] J.H.Boose, D.B. Shema, and J.M. Bradshaw, „Recent progress in Aquinas: a knowledge acquisition workbench," Knowledge Acquisition 1 (1989): 185-214.

[BStY95] T.P. Belikova, I.I. Stenina, and N.I. Yashunskay, „Image Processing and syndrom features analysis for enhancement of expert diagnostic abilities," Pattern Recognition and Image Analysis, 1995, vol. 5, no. 3, pp. 402-409.

[BYK94] T.P. Belikova, N.I. Yashunskaya, and E.A. Kogan, „Computer Analysis for differential diagnosis of small pulmonary nodules," In Proc. of Intern. Congress for lung cancer, Moduzzi (Eds.), Athens Greece, 22-26 June 1994, pp. 93-98.

[JaC88] A.K. Jain and R. Dubes, „Algorithms for Clustering Data," Prentice Hall, Englewood Cliffs, NJ 1988

[KeP91] A. Kehoe and G.A. Parker, „An IKB defect classification system for automated industrial radiographic inspection," IEEE Expert Systems (1991) 8, pp. 149-157.

[KSS85] J.L. Kolodner, R.L. Simpson, and K. Sycara, „ A Process Model of Case-Based Reasoning in Problem Solving," Proc. 9th Intl. Joint conf. Artificial Intelligence, Los Angeles, CA, 1985.

[LiE89] Cl.-E. Liedtke and M. Ender, Wissensbasierte Bildverarbeitung, Springer-Verlag 1989

[PeP95] P. Perner and W. Pätzold, „ An Incremental Learning System for Image Interpretation," In : Shape, Structure and Pattern Recognition, Eds. D. Dori and A. Bruckstein, World Scientific Publishing Co.,1995, pp. 311-323.

[Pere92] V.S. Pereverzev-Orlov, „ A Partner System and the Concept of Recognition Learning," Pattern Recognition and Image Analysis, vol. 2, No. , 1992 pp. 420-437.

[Per94] P. Perner, „A knowledge-based image inspection system for automatic defect recognition, classification, and process diagnosis," Int. Journal on Machine Vision and Applications, 7 (1994): 135-147.

[Quin86] J.R. Quinlain, „Induction of Decision Trees," Machine Learning 1 (1986): 81-106.

[Quin87] J.R. Quinlain, „Simplifying decision tree," Intern. Journal on Man-Machine Studies, 27 (1987); 221-234.

[Sha85] A. Shapiro, „Structured Induction,", John Wiley Inc. 1985.

[SNS88] S. Schröder, H. Niemann, G. Sagerer, „ Knowledge acquisition for a knowledge based image analysis system," In: Proc. of the European Knowledge-Acquisition Workshop (EKAW88), Bosse, J. and Gaines, B. (ed.), GMD-Studien Nr. 143, Sankt Augustin, 1988.

[TrP95] S. Trautzsch and P. Perner, „User Handbook for Machine Learning Tool SALOMON," Institute of Computer Vision and applied Computer Sciences Leipzig 1995.

[Utg89] P.E. Utgoff, „Incremental Induction of Decision Trees," Machine Learning (4) 1989, pp. 161-186.

[VGO94] P. Vanroose, Luc Van Gool, and Andre Oosterlinck, „ BUCA: A new pattern classification algorithm," In Proc. 12th Prague Conference on Information Theoretical Decision Functions and Random Processes, Aug. 29 - Sept. 1, 1994.

[WEK91] S. M. Weiss and C.A. Kulikowski, „ Computer Systems that learn," Morgan Kaufmann, 1991.