91-12

OPTIMIZATION OF CONCEPT PROTOTYPES FOR THE RECOGNITION OF NOISY TEXTURE DATA

P.W. Pachowicz and J. Bala

Center for Artificial Intelligence
George Mason University
Fairfax, VA 22030

< submitted only to AAAI-91 >

Keywords: machine learning, concept optimization, recognition of noisy data, computer vision, texture recognition.

ABSTRACT

This paper presents an approach to the recognition of noisy vision concepts incorporating machine learning and concept optimization techniques. Noisy texture characteristics require the development of techniques that can improve the performance of a texture recognition system. We develop such techniques through the optimization of learned concept prototypes in order to remove noisy and less significant concept components. Three approaches to the optimization of concept prototypes are presented: (i) static truncation of concept components on a given optimization level, (ii) iterative optimization and matching through increasing optimization degrees with trend following decision making, and (iii) the application of optimized concept descriptions to select final learning examples. Evaluation criteria are specified according to texture recognition and image segmentation task and experiments are performed for noisy texture attributes. Acquired results are compared to the integrated learning system AQ16. The evaluation of developed and applied concept optimization techniques shows that some techniques successfully applied for simpler class distribution in the attribute space can no longer be effective for complex distributions. A newly introduced method that applies optimized concept prototypes to filtrate final training data gave the best recognition results for 12 classes of texture (where texture attributes were extracted incorporating Laws' energy masks and were averaged over a circular window of a radius of 7.5 pixels).

ACKNOWLEDGEMENT

Authors wish to thank George Tecuci and Eric Bloedorn for correction of an early draft of this paper. This research was done in the Artificial Intelligence Center of George Mason University. Research activities of the Center are sponsored in part by the Defense Advanced Research Projects Agency under grant, administrated by the Office of Naval Research, No. N00014-87-K-0874, in part by the Office of Naval Research under grant No. N00014-K0226, and in part by the Office of Naval Research under grant No. N00014-88-K0397.

In section 2, we discuss basic assumptions, requirements and the evaluation criteria given by our application domain; i.e., texture recognition. Section 3 presents approaches to the optimization of concept descriptions and cooperation schemas with other system modules. Experimental results are presented in section 5. Finally, we summarize our work in section 6.

2. PRELIMINARY ASSUMPTIONS AND EVALUATION CRITERIA

2.1 Application domain

We apply machine learning to increase operational flexibility and performance of autonomous intelligent systems. The most challenging aspects of the integration of machine learning and computer vision is the automatic acquisition of concepts and their evolution over time (Pachowicz, 1990, 1991). In this paper, we attack the problem of noise removal that is involved in the creation of an autonomous vision system. Integrating machine learning with texture recognition and segmentation, we incorporate concept representation based on the disjunctive normal form. Such a representation gives a system more flexibility in adaptation performed over time. A system can: (i) add a new class of concept, (ii) extend or shrink a list of attributes, and (iii) manipulate concept descriptions in order to merge/split concepts or to remove noisy/imperfect components. All these operations can be performed without storing training examples. We apply AQ programs (Michalski, 1973) to learn such concept descriptions despite of several disadvantages of these programs (Rendell, 1989).

The texture recognition and segmentation problem is viewed as the following three step schedule: (i) image processing to extract texture attributes (a single image element is then characterized by a vector of attributes), (ii) matching image elements with learned texture class descriptions in order to annotate them by most probable classification hypotheses, and (iii) local unification of classification hypotheses in order to segment an image into homogeneous areas corresponding to surfaces/objects.

Since we are not interested in low-level vision, we incorporate well known Laws' energy masks (Laws', 1980) to extract texture attributes with averaging circular window of a radius equal to 7.5 pixels. Extracted numeric attributes are quantized to integer values from 0 to 55. Extracted events are then presented to the learning system to acquire texture class descriptions or to the recognition system to annotate picture elements by their class memberships. Next, picture elements are unified to form object surfaces. A local spatial distribution of pixels is used when extracting attributes, recognizing their class membership, and unifying pixels into homogeneous surface/object areas. This bottom-up process smooths texture irregularities by incorporating characteristics of the neighborhood of a given picture element. Processed local information on one level is transferred onto the next abstraction level and represented as point-wise data.

2.2 Evaluation criteria

The quality of a segmented and annotated image depends on the quality of all three phases of the texture recognition and segmentation schema. Quality criteria for the evaluation of segmentation processes are precise (Zucker, et a., 1975, Davis, et al., 1981, Hsiao and Schawchuk, 1989, DuBuff, et al., 1990). An excellent system should: (i) preserve sharp and precise borders between different texture areas, (ii) smooth homogeneous texture surface areas, and (iii) preserve small objects against their removal from segmented image. These criteria, however, are too difficult for current computer vision systems. They are contradictory. For example, the criteria of smoothing texture surface areas requires the extension of a radius of local operators performed through all three phases of the recognition schema. On the other hand, if a radius is enlarged then the operators blur borders between texture areas which can remove small objects in an image.

What then are the evaluation criteria for the intermediate phase of our recognition system, i.e., the matching of events (attribute vectors) with texture class descriptions? We require that the best matching system should:

1) increase the classification confidence when matching a class description with data belonging to this class,
2) decrease the classification confidence when matching data with other class descriptions, and
3) perform on the similar confidence level for all classes when data is matched with their class descriptions (system stability criterion).

Let us consider that the matching system computes the distance measure between test data and concept descriptions in order to find the most probable classification decisions. The distance, however, can be the same for two (or more) concepts, especially when one considers that a measure is computed through the quantized attribute space or a distance tolerance is involved in the classification process (Reinke, 1984). Considering the above evaluation criteria, a confusion matrix should have the highest values for diagonal elements and the lowest through all other matrix elements. Moreover, the system stability criterion requires that each concept should have the same chance to be recognized. A highly negative effect is reached when one class can be recognized with the highest confidence (e.g., above 95%) and the other class with relatively low confidence (e.g., below 40%). Therefore, we evaluate the effectiveness of the optimization processes not only through the computation of the averaged recognition rate but also by monitoring the standard deviation and the recognition rate for the worst performing concept descriptions.

2.3 Requirements for a learning tool

The effectiveness of a performed matching processes depends mainly on the quality of acquired concept descriptions. This quality can be increased through the selection of most appropriate control parameters of the AQ programs. This selection has been done by the experimental observation of positive and negative effects of parameter changes.

Let us discuss the choice of one of these parameters, i.e., the degree of generality of acquired rule descriptions. A concept description that is learned by AQ programs is a disjunction of complexes covering positive examples only. A complex is a conjunction of selectors, and a selector represents the value range of a single attribute. Concept descriptions are created incorporating heuristic rules (Michalski,1983). These rules generalize a given positive example toward other positive examples through dropping a selector, extending the range of attribute value, etc. In this way, a description can be general or specific. A general description is more optimal than a specific description when the size of a description is considered. It does not mean that a general description is more effective.

General descriptions were used because (i) they are stored more efficiently, and (ii) the matching process can be performed faster than for specific descriptions. We have found that general descriptions can perform better when generalized over noisy examples. Unfortunately, such general descriptions increase also the number of miss-classifications when descriptions of counter classes are matched. The confusion matrix has higher values through the main diagonal but values out of the main diagonal increase quickly (even twice as fast). Such an effect is harmful to texture recognition and segmentation, where classification hypotheses are processed later in order to derive homogeneous surface areas. Therefore, we decided the learning of specific concept descriptions rather than general descriptions would be better.

3. OPTIMIZING CONCEPT DESCRIPTIONS

A parametrized learning tool can be flexible in the acquisition of concept descriptions. Such tuning, however, is limited and no further increase in system performance can be reached through the learning phase only. Therefore, we manipulate acquired prototypes in order to gain further tuning of improvement. Optimization of concept descriptions can be executed in order to fulfill one or more of the following tasks:

- to improve the overall performance of concept prototypes through the elimination of their noisy and less significant components,
- to improve system stability through the increase in recognition effectiveness of the worse performing prototypes,
- to compress concept descriptions for their effective storage and faster performance.

The choice of one task or a combination of above tasks depends on the application domain and technical support. In experiments presented in this paper, we focus primarily on the improvement of overall performance of concept prototypes and on the improvement of system stability rather than on the compression of concept descriptions (a secondary effect of concept optimization).

3.1 Direct optimization of concept descriptions

We consider concept optimization methods that were developed based on the theory of Two-Tiered Representation (TT) of imprecise concepts (Michalski, 1987). The theory assumes that an acquired concept description can be transformed to its TT representation through a separation of the most significant concept properties (Base Concept Representation) from exceptions to these properties (Inferential Concept Interpretation). The cooperation schemas between learning, optimization and recognition modules are presented in Figure 1 for each discussed method.

Considering a separation process, we would like to separate the most significant concept components from those that are noisy or redundant. Since the concept descriptions learned by the family of AQ programs are composed of ordered complexes (from most to less significant complexes), we can truncate such descriptions by removing some less significant complexes. The truncation degree can then be controlled by a parameter corresponding to the number of removed complexes or to the percentage number of events covered by removed complexes. Such optimized concept descriptions are more specific. They can improve the performance of the recognition system because test data will no longer match removed noisy concept components.

The choice of the truncation degree is very difficult without evaluation of noise in the training data. Such evaluation can sometimes be impossible to perform. A separate set of tuning data can be useful to set up the optimization degree. If the concept descriptions are over-optimized, they can give worse results than non-optimized descriptions. Therefore, we developed a method that follows recognition results over several relatively low optimization values without looking for a specific value of the optimization degree.

The method (Bala and Pachowicz, 1990) is based on the assumption that if optimized descriptions can perform better, then the system can track their performance over low optimization degrees. The system increases optimization degree and matches optimized concept prototypes with test data. Acquired recognition curves (versus increased optimization degree) are followed by a system in order to make the final decision. If the recognition rate is high (e.g., above 70%) for one class and all other class descriptions are matched with test data on a relatively low confidence level (e.g., below 20%) then the system does not follow recognition curves. However, if the difference between the highest and the second classification hypothesis is relatively low (e.g., 44% for class A and 39% for class B) then the system follows recognition curves over gradually increased optimization degrees. The final classification decision is made yielding a class for which the classification confidence increases with an increase in optimization degree. If there are multiple classes for which classification confidence increases, then the system yields the one with the highest confidence values.

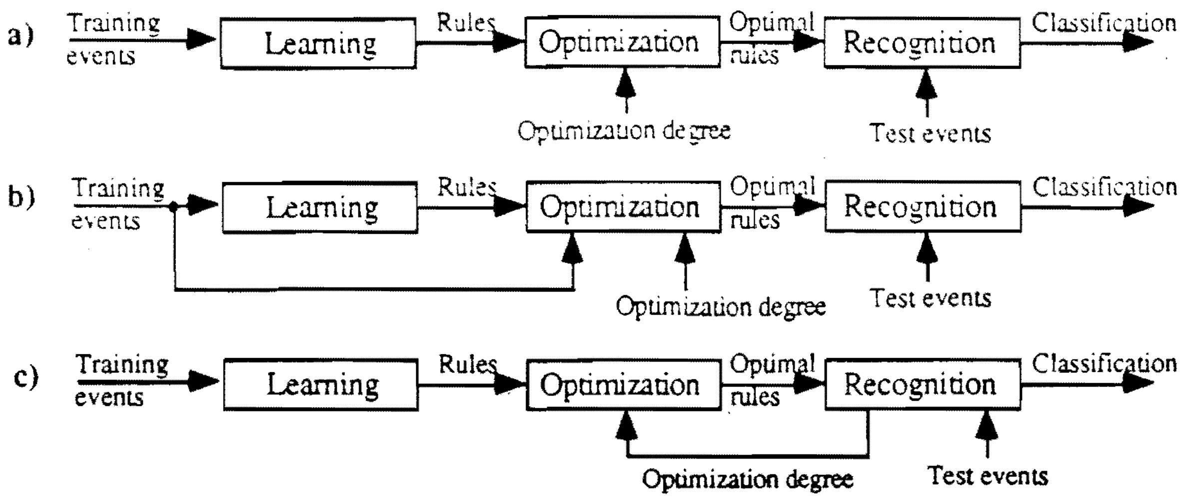

Fig. 1 Cooperation schemas for direct optimization of concept descriptions incorporating: a) simple truncation of concept components, b) generalization of concept descriptions over negative examples (the AQ16 system), and c) iterative optimization and matching of concept descriptions.

Simple truncation of less significant complexes specializes concept prototypes. Another method applied to improve the performance of concept prototypes incorporates generalization operators and was implemented within the AQ16 integrated system (Zhang, 1990). Such generalization is performed through dropping a selector or extending the range of selector attribute values. In this way, a complex covers not only positive training examples but it can cover some of negative examples. The degree of such allowed coverage is controlled by optimization parameters.

3.2 Indirect optimization of concept descriptions

Optimization methods introduced in the above section work well with data chosen for those methods. Unfortunately, they are not universal. We found that some can work well with simple data but fail when applied to complex data. Advancing the optimization of concept descriptions, we focus on more complex data. This advancing is based on the following two observations: (i) the family of AQ programs learns concept descriptions by drawing a cover over positive examples only, and (ii) the optimization process exactly eliminates the influence (on concept description) of the training events that are covered by removed concept components. Optimized concept descriptions can still be non-optimal, for example, because some of their components can be again partitioned by the existence of negative examples during the learning phase. If some of these examples are noisy and can be filtered from the initial set of training examples then newly learned concept descriptions (learned from a filtered set of examples) should perform better during the recognition phase.

Thus, acquired primary concept descriptions can be optimized and applied to filter the final set of training data. Training examples are passed if they are covered by optimized descriptions. If they are not covered then they are removed from the final set of training examples. The learning process is then repeated for selected training events in order to acquire the final concept descriptions. Considering the learning process employed by the AQ14 program (i.e., the above explained fact that clusters of positive examples are connected and described by one concept component), we decided to change the control parameters of the AQ14 program. This change is applied to force the formation and description of single (non-connected) clusters rather than connected clusters that weaken the removal of noise (isolated events) during the optimization process. The optimization is executed through truncation of less significant concept components. The optimization degree points out the percentage number of training events covered by removed complexes.

Since the choice of the most appropriate optimization degree is difficult, we monitor the performance of the introduced method for a wide range of optimization values. The cooperation schema between learning, optimization, filtration and recognition modules is presented in Figure 2. The closed-loop of optimization process (dashed line) implements a method of the iterative optimization and recognition of noisy concepts (Bala and Pachowicz, 1990).

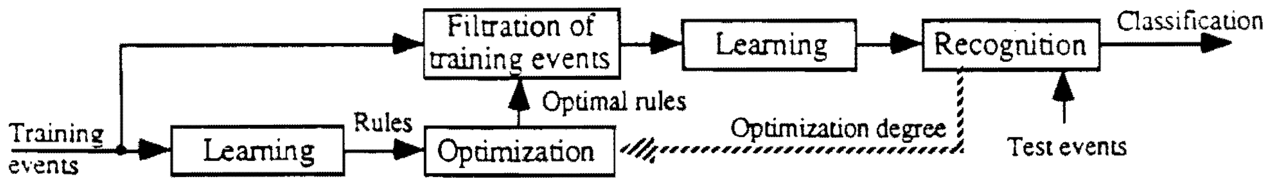

Fig. 2 Cooperation schemas for indirect optimization of concept descriptions

4. EXPERIMENTAL RESULTS

4.1 Application domain

In this paper, texture recognition is the application domain for the machine learning methodology. The task is to assign to test data a class membership of texture based on computed characteristics of texture samples. Small sections of twelve classes of texture used in our experiments are presented in Figure 3. Texture characteristics involve the extraction of texture attributes that in most studies are typical numeric attributes. The learning process is applied to describe clusters of training examples in the attribute space. Considering the nature of texture, extracted texture attributes contain noise.

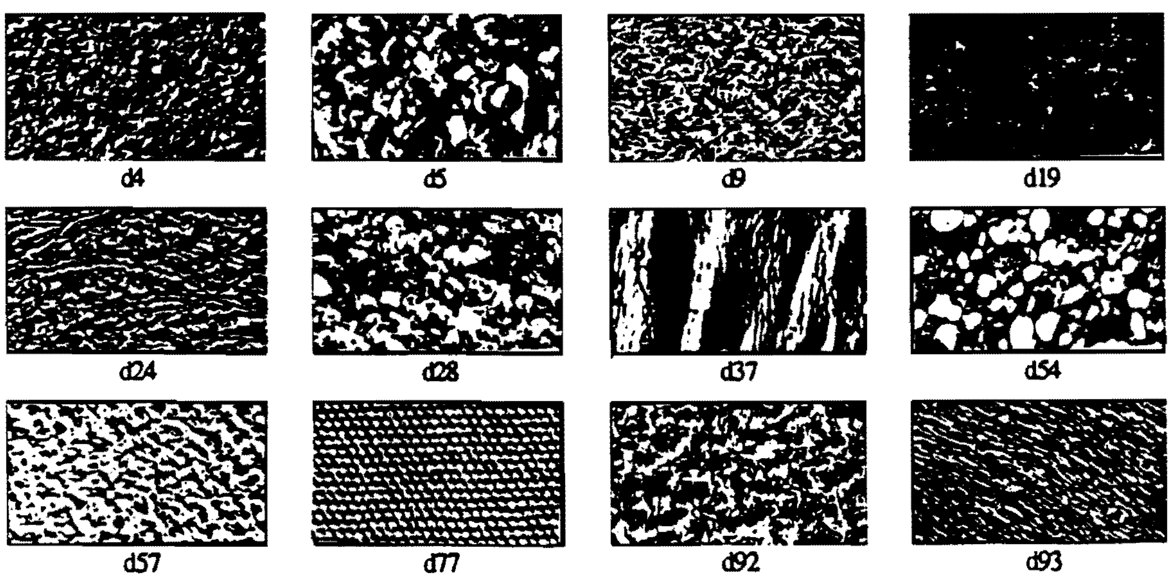

Fig. 3 Samples of twelve classes of texture (taken from Brodatz album, 1966)

Extraction of texture attributes was performed incorporating Laws' energy masks (Laws', 1980). Eight masks were convoluted with texture images and their responses were averaged over a local window of a radius equal to 7.5 pixels. This averaging smooths mask responses over a local area. If the window is large enough then extracted attributes form better clusters. If it is small, then the attribute space contains clusters with neither a clear center nor boundaries. From the image segmentation point of view, the extension of the local averaging window is limited because it blurs borders between different texture areas and eliminates smaller objects from the image. So, learning

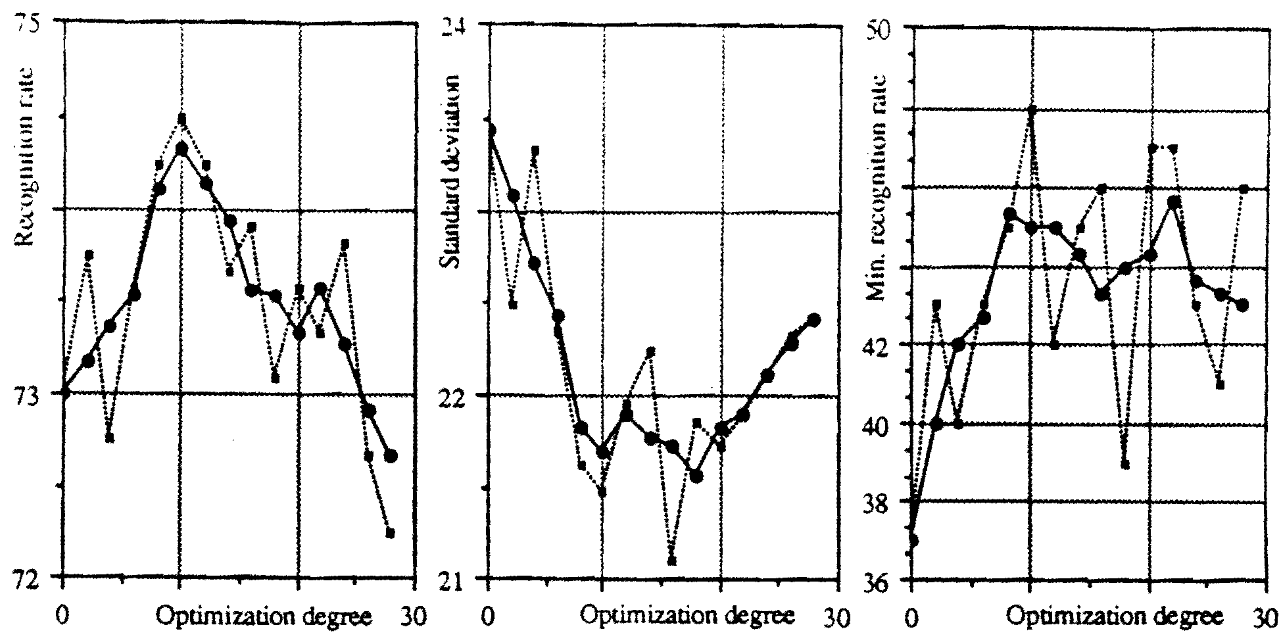

Fig.5 Recognition results for the indirect optimization of concept descriptions

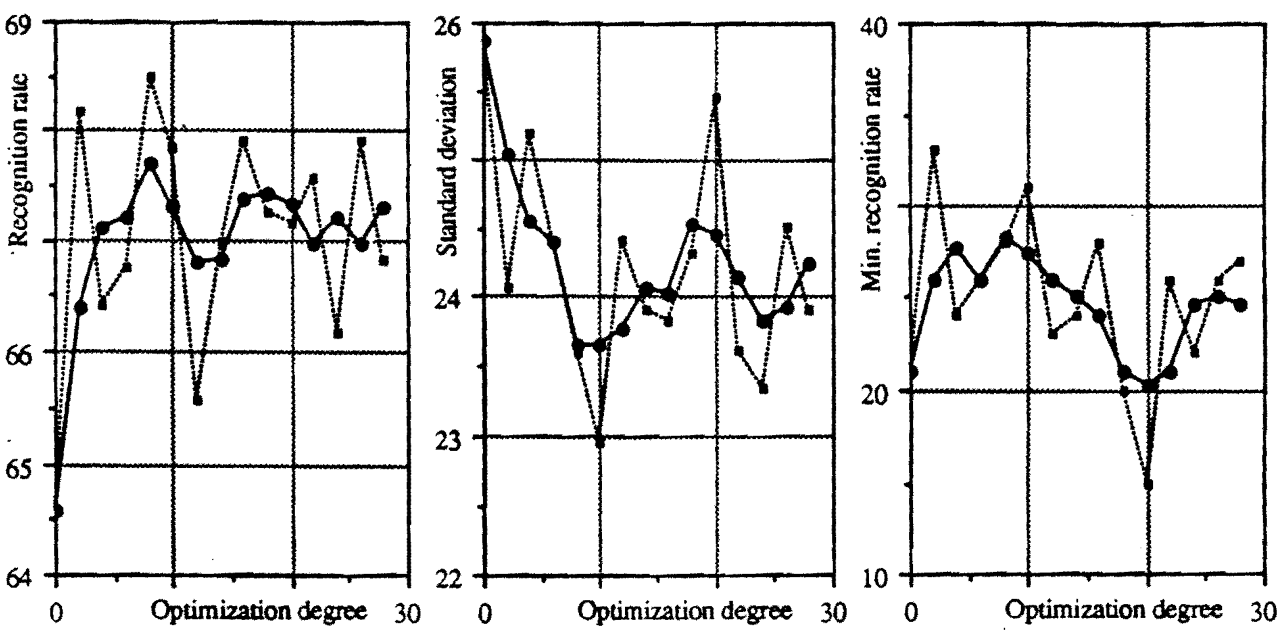

Fig.6 Recognition results for texture images affected by added Gaussian noise (SNR=100) and for applied indirect optimization of concept descriptions

The average recognition rate was increased fast from the 73% level up to 74.5% before a slower decrease to slightly below the 73% level. At the same time, the standard deviation decreased from above 23 to below 22. Minimum recognition rate increased significantly from 37% to above 45%. For higher optimization degrees the minimum recognition rate decreased slowly but still to a level well above the initial 37%. We find this effect very positive.

Pachowicz, P.W., "Advancing Concept Learning Methodology for Intelligent Autonomous Systems", Report of the Center for Artificial Intelligence, George Mason University, 1991 (to appear).

Quinland, J.R., "Induction of Decision Trees", Machine Learning, Vol.1, pp.81-106, 1986.

Reinke, R.E., "Knowledge Acquisition and Refinement Tools for the Advice Meta-Expert System", ISG 84-4, Department of Computer Science, University of Illinois, Urbana, 1984.

Rendell, L, Cho, H. and R. Seshu, "Improving the Design of Similarity-Based Rule-Learning Systems", International Journal of Expert Systems, Vol.2, No.1, pp.97-133, 1989.

Zucker, S.W., Rosenfeld, A. and Davis, L.S., "Picture Segmentation by Texture Discrimination", IEEE Trans. Comput., Vol. C-24, pp.1228-1233, 1975.

Zhang, J. and R.S. Michalski, "Rule Optimization via SG-TRUNC Method", Proc. of the 4th European Working Session on Learning, pp.251-262, Montpellier, Dec. 1989.

Zhang, J., "Learning Flexible Concepts from Examples: employing the ideas of Two-Tiered Concept Representation", Ph.D. Thesis, Computer Science Department, University of Illinois, Urbana, 1990.