Pattern Recognition Letters 14 (1993) 241-255 March 1993
North-Holland

11th International Conference on Pattern Recognition August 31 - September 3, 1992

On pattern, categories, and alternate realities

Laveen N. Kanal*

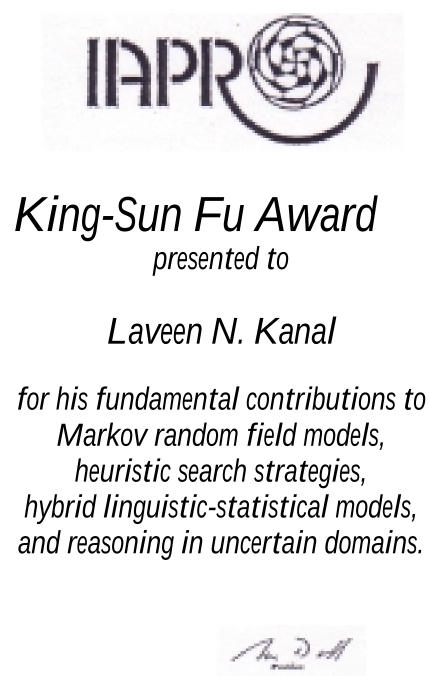
Institute for Advanced Computer Studies and Department of Computer Science, University of Maryland, College Park, MD 20742, USA

This is the text of the talk given by the author on August 31, 1992 at the 11th International Conference on Pattern Recognition when the author received the King-Sun Fu award of the International Association for Pattern Recognition (IAPR).

Preamble

I thank all the individuals involved in the nomination and selection process of IAPR for this honor of the 1992 King-Sun Fu award. In addition, I thank my many collaborators from all over the world over the last three decades, as this award also recognizes their contributions.

If memory serves me correctly it was in the early summer of 1961 a little over 31 years ago that I first met King-Sun Fu. King-Sun had joined Purdue as an Assistant Professor in 1960 the year I received my Ph.D. and joined General Dynamics/ Electronics (GD/E). Thanks to the excitement generated by Frank Rosenblatt's Perceptrons, my dissertation titled Stochastic Models for Learning, and Alfred Wolf a former classmate, I had been hired by General Dynamics with the fancy title of Manager, Machine Intelligence Laboratory. It sounded important. In the summer of 1961, while King-Sun was working at IBM, he visited me. I don't remember what we talked about; what I do remember is that we spent a fair amount of time talking in a Chinese restaurant. In the years that followed we collaborated in the area of professional activities in pattern recognition, working together on various IEEE committees. An especially memorable event was the (first) workshop on pattern recognition organized by the IEEE Computer



* Preparation supported in part by NSF grant IR18802419 and in part by LNK Corporation, Inc.

Elsevier Science Publishers Q_V.

Society's pattern recognition committee in Puerto Rico in October, 1966. King-Sun, Azriel Rosenfeld, Herb Freeman, Dick Duda, Tom Cover, Bernie Widrow, Leonard Uhr, Russell Kirsch, George Nagy, C.K. Chow, Marvin Minsky, Seymour Papert, Bela Julesz and many others who have become well known in the field were at that workshop (1968). It was through the IEEE activities I came to appreciate that King-Sun was not only very good at getting things done but was a generous spirit who liked to include a broad spectrum of people and research approaches in the activities that he helped organize.

About a decade after I first met him, a little more than twenty years ago now, King-Sun, Herb Freeman, Azriel Rosenfeld, I, and a few others met at a place in Virginia-I think it was the Airlie House, and talked about organizing the first International Joint Conference on Pattern Recognition (IJCPR). As I recall, Herb Freeman nominated me to head the organizing group and King-Sun seconded the nomination. At that time I made a most important contribution to the subsequent success of IJCPR and its successor organization, IAPR. I declined the nomination and nominated King-Sun. In the years that followed, his energy, dedication, sponsorship, and support made IAPR, the IEEE Transactions on Pattern Analysis and Machine Intelligence, and many other pattern recognition activities flourish. A word to describe such a champion of the field is patron.

Pattern

It is interesting to note that the English word pattern derives from the old French word patron and one of the first usages of pattern was to denote a person or a thing so ideal as to be worthy of imitation or copying, i.e., a model. We find this usage quite frequently in 19th century English, as in Robert Browning's poem "The Lost Leader", when he says:

> "We ... Made him our pattern to live and to die!"

Earlier, it appears in Shakespeare's *King Lear*. Of course Shakespeare, being the master that he was, could be expected to use the word in several interesting ways. In some books on pattern recognition I have read statements to the effect that the use of pattern as an example or sample representing a class, type, or concept, is a modern usage. But in *Henry VIth* Shakespeare wrote:

> "For what is wedlock forced, but a hell
> An age of discord and continual strife
> Whereas the contrary bringeth bliss
> And is a pattern of celestial peace."

In *Othello,* the example is more complex. Under the light of a flaming torch, gazing at the sleeping Desdemona whom he is contemplating killing, Othello says:

> "Put out the light, and then put out the light:
> If I quench thee, thou flaming minister
> I can thy former light restore,
> Should I repent me; but once put out thy light,
> Thou cunning'st pattern of excelling nature,
> I know not where is that Promethean heat
> That can thy light relume."

Among 19th century English quotations on pattern I encountered the following which triggered thoughts leading to a connection with some current topics in pattern recognition.

> "Take a mere beggar-woman, lazy, ragged, filthy, and not over scrupulous of truth-but if she be chaste, and sober and cheerful, and goes to her religious duties-she will in the eyes of the church, have a prospect of heaven, quite closed and refused to the state's pattern-man, the just, the upright, the generous, the honorable, the conscientious, if he be all this, not from supernatural power-but from mere natural virtue."
>
> John Henry, Cardinal Newman, Lecture VIII of *Lectures on Anglican Difficulties.*

Here we note that pattern-man is used as a model and described in terms of qualities or features. John Henry was originally an Anglican and his quotation took me back to the elementary school I attended as a child in India. St. Mark's School, run by two Welsh sisters, imparted a love of the English Language through tales from Shakespeare, English history, and translations from Celtic, Greek and Norse mythology. The pupils, Christians, Muslims, Hindus like myself, and others attended Bible class and on many Sundays we were expected to attend services at the Church of England next door to the school. No doubt there we heard about divine grace and sin and repentance. The quotation also triggered a memory of part of a rock-a-billy song-that's rock plus hillbilly[1]:

> "If you want to get to Heaven
> You've got to raise a little Hell."[1]

Obviously, natural virtue corresponds to local minima, not the global desired state and we know that to get out of the local minima you need to raise the temperature every now and then. Clearly a simulated annealing approach to salvation!

Some sketches from the current pattern recognition scene

Although use of simulated annealing in pattern recognition was a later development, many ideas and methodologies which V encountered when I entered the field of pattern recognition, have come back, fortunately in a more robust form, with the resurgence of artificial neural networks, dynamical systems and other "complex" systems. Some in the field have greeted this resurgence with dismay. To them we might well say what Guidillo said to Rocco in Ignazio Silone's novel *A Handful of Blackberries:*

> "I see you are tired son, and disappointed. You have the sadness of one who has traveled far and ends up finding himself where he began. Didn't they teach you at school that the world is round?"

Unlike the critics who are dismayed I am not sad but happy at this resurgence. Neural Nets, dynamical systems, search, optimization, mean-field annealing, graduated non-convexity optimization, Genetic Algorithms-I like these approaches because not only are they proving useful and interesting, they use concepts and language I learned years ago, and so I don't have to learn many new tricks to continue playing in this game. To present an overview of "this game" now is a lot harder than it Was when I wrote my article "Patterns in Pattern Recognition, 1968-1974" (1974). Here I will just mention a few items that my colleagues and I have found useful or that have struck my fancy.

Artificial neural networks

In 1961-62, my group at GD/E had filed two patents on a Perceptron type network and implemented it in a computer system called Adaptive Pattern Encoder, or APE. As I have written elsewhere, those were the days of catchy names and audacious claims (1972). The lack of a well-defined training algorithm for multilayer networks and the emerging understanding of the relationships between perceptron type training algorithms and statistical classification methods had by 1962, when I joined Philco-Ford, shifted my attention to training procedures derived from parametric and non-parametric statistics (Kanal (1962) and my foreword, in Sethi and Jain (1992)). By 1968 hybrid linguistic-statistical models seemed to me to be more appropriate for many pattern recognition problems than purely statistical or purely structural/linguistic models. (Part of our work on this topic was presented in Kanal (1970), Kanal and Chandrasekaran (1972)). Thanks to an invitation from Azriel Rosenfeld, in 1970 I became a Professor of Computer. Science at the University of Maryland. The research projects I started there were on interactive pattern analysis and statistical classification using decision trees (1972 and 1977), on generative and descriptive models for characterizing error patterns in communication channels with memory (1971 and 1978), on linguistic models for graphical formula translation (Underwood and Kanal (1973)) and on hybrid linguistic-statistical models for waveform and time-series analysis (Stockman et al. (1973) and (1974)). At Maryland I started teaching courses in Artificial Intelligence (Al) and became interested in applications of search and problem reduction methods of AI to problems in pattern recognition (Kulkarni and Kanal (1976) and (1978), Stockman et al. (1976), Kanal (1979), Stockman and Kanal (1983)). This also led to my interest in understanding

[1] Ozark Mountain Daredevils, “If You Want to Get to Heaven”, A&M Records, 1973

the exact relationships of the Al search procedures to the dynamic programming and branch and bound methods I had encountered when teaching Operations Research (OR) courses as a part-time visiting professor at the Wharton School of Business of the University of Pennsylvania (Kanal and Kumar (1981), Kumar and Kanal (1983), Kanal and Kumar (1988)). Although my IEEE-IT Trans. survey paper noted that a 1974 conference on Biologically Motivated Automata indicated a revival of interest in neural models and adaptive networks, and I was aware that Albus, Amari, Fukushima, and Kohonen worked in this area, -by and large I was not following that literature.

It was in the above context that a young man came by in the spring of 1975 to introduce himself as a new assistant professor in the Government and Politics department at the University. He wanted to talk about his Ph.D. dissertation which he had recently finished at Harvard. He said quite enthusiastically that it related to pattern recognition, learning machines, and intelligent systems. With the best of intentions I told him to lend me a copy and he lent me a report titled "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences" (Werbos (1974)). Soon the report got buried in the usual flood of paper that unfortunately still remains very much part of my life. I have often wondered in recent years if the title had been different, e.g., if it had mentioned something about a procedure for training multilayer neural networks, or if my desk had been less cluttered, would I have paid more attention to this report or was I so tuned away from artificial neural networks that it would have made no difference? Of course that young man was Paul Werbos and the method he wanted me to read about was the error back-propagation (EBP) method for training multilayer feedforward neural networks. I was not alone in neglecting his work. It was not until some years later, that the independent, brilliant work of David Rumelhart and his colleagues at the University of California, San Diego, on Parallel Distributed Processing (Rumelhart and McClelland (1986)) made EBP famous, and it was some time after that that Paul Werbos gained recognition for his pioneering work on EBP. Perhaps in science catchy names and titles play a positive role. At least they do get our attention.

Several deficiencies of EBP have been noted by researchers. Improvements continue to be proposed and alternative iterative and non-iterative procedures continue to be cited as being equivalent or better. However, EBP was one of the main reasons for the resurgence of artificial neural networks, at least in the United States. My guess is that currently it is probably the most widely used general procedure for training neural networks for pattern classification. A basic problem of statistical pattern recognition, viz, the dimensionality sample size problem, also arises in artificial neural systems. In the design of multilayer feedforward networks one question is how many hidden units to use. An experimental approach to this question is to make the activation of the hidden units visible to the designer and see which units don't have much activation during the classification process and suppress them. For determining the number of parameters, i.e., the order of a statistical model, Akaike (1974) introduced an information criterion now referred to as the Akaike Information Criterion (AIC). A generalization of the AIC, called the Network Information Criterion (NIC), for determining the number of hidden units for an artificial neural network, is presented in a report by Murata et al. (1992). The NIC is for model selection among a sequence of nested, hierarchical models, where a lower dimensional model is a submodel of the next higher dimensional model in the sequence. These studies represent some of the current efforts to develop a theoretical methodology for determining the number of hidden units in a multilayer back-propagation network.

The attention attracted by the energy function minimization approach introduced by J.J. Hopfield in 1981-82, for developing self-organizing, associative memory networks was possibly a primary reason for the resurgence of significant scientific activity in artificial neural networks in the early years of the last decade. The application of energy function minimization in the design of neural networks for combinatorially hard optimization problems, and for pattern classification and motion vision has been very fruitful and interesting for me and my colleagues. Hopfield and Tank (1985) introduced a technique for embedding optimization problems, such as the traveling salesman problem (TSP), in mean-field thermodynamic neural networks. One of the shortcomings of their proposed technique is that each discrete optimization problem must be reduced to the minimization of a 0-1 Hamiltonian. A sufficient but not necessary condition for network stability is that the network connection weights be symmetric. The technique yields fully connected networks of functionally homogeneous "visible" units with loworder symmetric connections. Theoretically, all NP-hard optimization problems can be reduced to the TSP and consequently, 0-1 Hamiltonian minimization problems. However, direct reductions have been found for only structurally simple problems. In Hellstrom and Kanal (1990), we showed that embedding the NP-hard knapsack optimization problem in neural networks required the use of functionally heterogeneous units. In Hellstrom and Kanal (1992a), a knapsack packing neural network with both low-order and conjunctive asymmetric synapses is derived from a non-Hamiltonian energy function. Extensive simulations showed that for the knapsack optimization problem, solution quality was consistently better than approximate solutions given by a simple greedy algorithm; often exact solutions result as verified by solutions obtained using a fast parallel enumerative algorithm. The underlying mechanism of the general method involves the decomposition of arbitrary problem energy gradients into piecewise linear functions which can be modeled as the outputs of small groups of hidden units. In Hellstrom and Kanal (1992b), we derived thermodynamic mean-field neural networks for multiprocessor scheduling. Simulations of networks of up to 2400 units gave very good and often exact solutions. Our general method for treating non-Hamiltonian energy functions also has broad applicability to problems in pattern recognition.

A related optimization method which we and others have found very useful is the Graduated Non-Convexity (GNC) technique developed by Blake and Zisserman (1987). In addition to using it for edge-preserved smoothing, we have used it in developing algorithms for discontinuity-preserved motion field computation in an application involving multiple-object tracking and motion parameter estimation from passive imagery (Raghavan et al. (1992a) and (1992b), Gupta and Kanal (1992)). It is interesting to note the essential equivalence among Mean-Field Annealing, GNC, and a feature extraction technique known as Variable Conductance Diffusion, which has been shown in some recent papers (Bilbro et al. (1992), Snyder et al. (1992)).

While the new generation of artificial neural networks excite us, we should keep in mind that: (1) as has been shown by us (1989) and others, often fairly simple statistical decision tree methods give equivalent or better results; (2) the various neural network paradigms for pattern classification introduced in recent years have close connections with stochastic approximation, estimation and classification procedures known in statistical pattern recognition; and (3) rather good algorithms have been developed in recent years for fairly large combinatorial optimization problems whereas neural networks have so far only been demonstrated on much smaller problems. It remains to be shown that combinatorial optimization is a good arena for artificial neural networks. However, there is no denying the positive results already achieved in recent years with neural networks, genetic algorithms, and other such newer techniques which I think are very intriguing and fertile areas for theoretical and experimental inquiry.

Hybrid systems

For the solution of complex problems in pattern recognition and more generally in machine intelligence, involving heterogeneous data sources of both numeric and symbolic information, the use of hybrid methodologies integrating multiple paradigms is becoming increasingly popular. The following statement appeared in Kanal (1972):

> "It is now recognized that the key to pattern recognition problems does not lie wholly in learning machines, statistical approaches, spatial filtering, heuristic programming, formal linguistic approaches, or in any other particular solution which has been vigorously advocated by one or another group during the last one and a half decades as the solution to the pattern recognition problem. No single model exists for all pattern recognition problems and no single technique is applicable to all problems. Rather what we have in pattern recognition is a bag of tools and a bag of problems."

Twenty years later I was pleased to see the following comments in a recent article by Marvin Minsky (1991):

> In the 1960's and 1970's students frequently asked, "Which kind of representation is best", and I usually replied that we'd need more research before answering. But now I would give a different reply: "To solve really hard problems, we'll have to use several different representations."

Later in the same article, Minsky continues:

> "It is time to stop arguing over which type of pattern classification technique is best because that depends on our context and goal. Instead we should work at a higher level of organization and discover how to build managerial systems to exploit the different virtues and evade the different limitations of each of these ways of comparing things."

I hope that with Minsky joining in what I had long thought should be apparent, fewer researchers amongst us will yield to the temptation to attempt to boost their current favorite technique or theory by knocking what are viewed as competing methodologies and areas of inquiry. In recent years we have heard such comments praising Neural Networks and criticizing AI, ignoring the many seminal contributions of Al, or denigrating neural networks and praising case-based reasoning, or comments about Neural Nets versus Decision Trees or Genetic Algorithms, etc.

In the 1960's we heard such comments about statistical versus, structural-linguistic techniques. What is incumbent on us is to attempt to understand the capabilities and applicability of the various tools and exploit the complementary advantages of the different paradigms. Examples of such attempts on combining linguistic and statistical models, combining heuristic search with statistical pattern recognition, and heuristic search with structural pattern recognition appear in Kanal and Chandrasekaran (1972), Kulkarni and Kanal (1976), Kanal (1979), Stockman and Kanal (1983), Kanal and Dattatreya (1986). A variety of combined approaches to the design of decision trees is surveyed in Dattatreya and Kanal (1985). Combining structural and probabilistic knowledge with hypergraph search to develop alternate scenarios for decision making in uncertain domains is presented in Bhatnagar and Kanal (1993a,b).

Figure 1 displays some of the tools currently available in the pattern recognition tool kit, and some of the problems to which these tools have been applied. In "Hybrid Systems-A Key to Intelligent Pattern Recognition" (Kanal and Raghavan (1992)), we discuss some of the key issues which need to be addressed in integrating

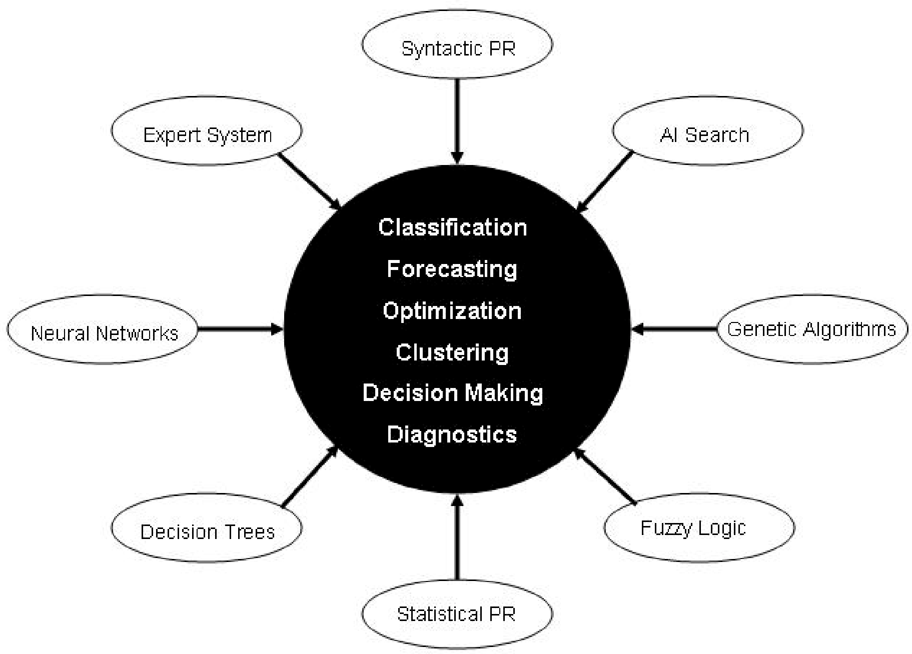


Figure 1 Some tools for a pattern recognition tool kit.

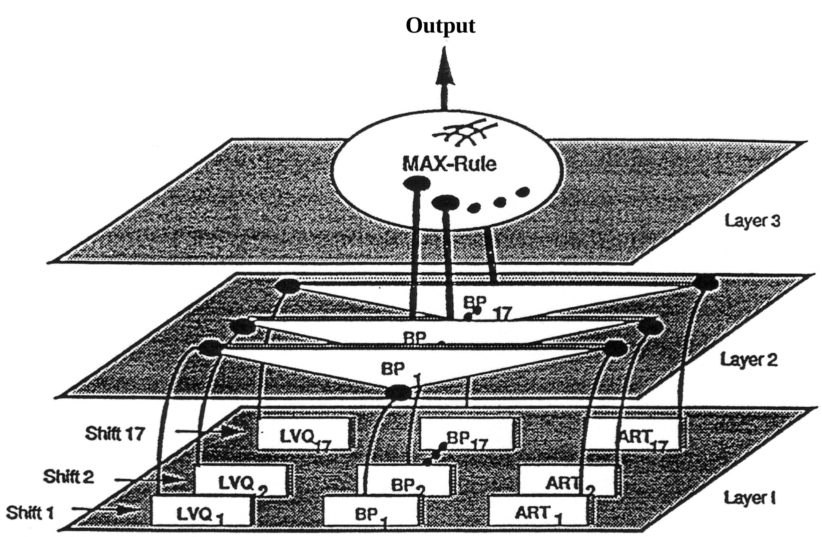


Figure 2. A hybrid of neural networks for RCS-based target recognition

heterogeneous methodologies for intelligent solution of pattern recognition problems and some of the design principles we have found useful. In that paper and in Raghavan and Kanal (1992), we describe some proof-of-concept systems developed by my associates at LNK Corporation. These include a multilevel hybrid of several types of neural networks for classification of Radar Cross Sections (Figure 2); and an integrated platform for intelligent fusing of 3 components-sensors, spatial databases, and maps, using neural networks, an expert system,

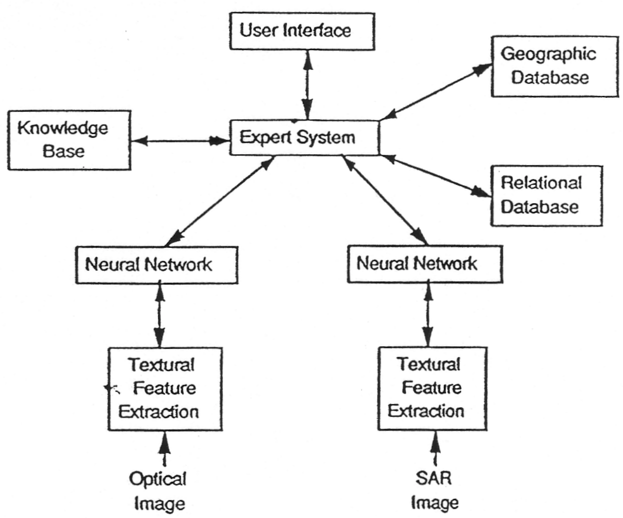


Figure 3. A hybrid system for feature extraction using sensor fusion.

and a fuzzy logic controller (Figure 3). From our experience, as outlined in the above mentioned papers, hybrid systems provide a key to achieving significantly better performance than individual component methodologies, in solving complex pattern recognition problems. A recently published book (Kandel and Langholz (1992)) appears to contain an excellent collection of papers on hybrid architectures and applications for intelligent systems.

A significant program based on hybrid pattern recognition methodologies, including neural nets, expert systems, and decision trees, is the Intelligent Data Management (IDM) project at NASA/Goddard Space Flight Center. The research being done for this project is described in a series of papers (see Campbell et al. (1989), Cromp (1991)). The Earth Observing System (EOS) is expected to generate massive quantities of satellite imagery and other data from a variety of instruments. While being rapidly archived, this data needs to be processed by fast algorithms which can characterize the data in such a way as to allow efficient querying and retrieval from object-oriented databases. An Intelligent Information Fusion System (IIFS) for the IDM is illustrated in Figure 4.

Applications such as the NASA IDM point to the problem of scalability which remains one of the basic concerns for employing various pattern recognition, parallel processing, and machine intelligence tools on real-world problems. Hierarchically organized hybrid systems are being explored as one alternative to meeting the challenge of scalability, i.e., having such systems work well on problems involving very large data sets and tough real time constraints.

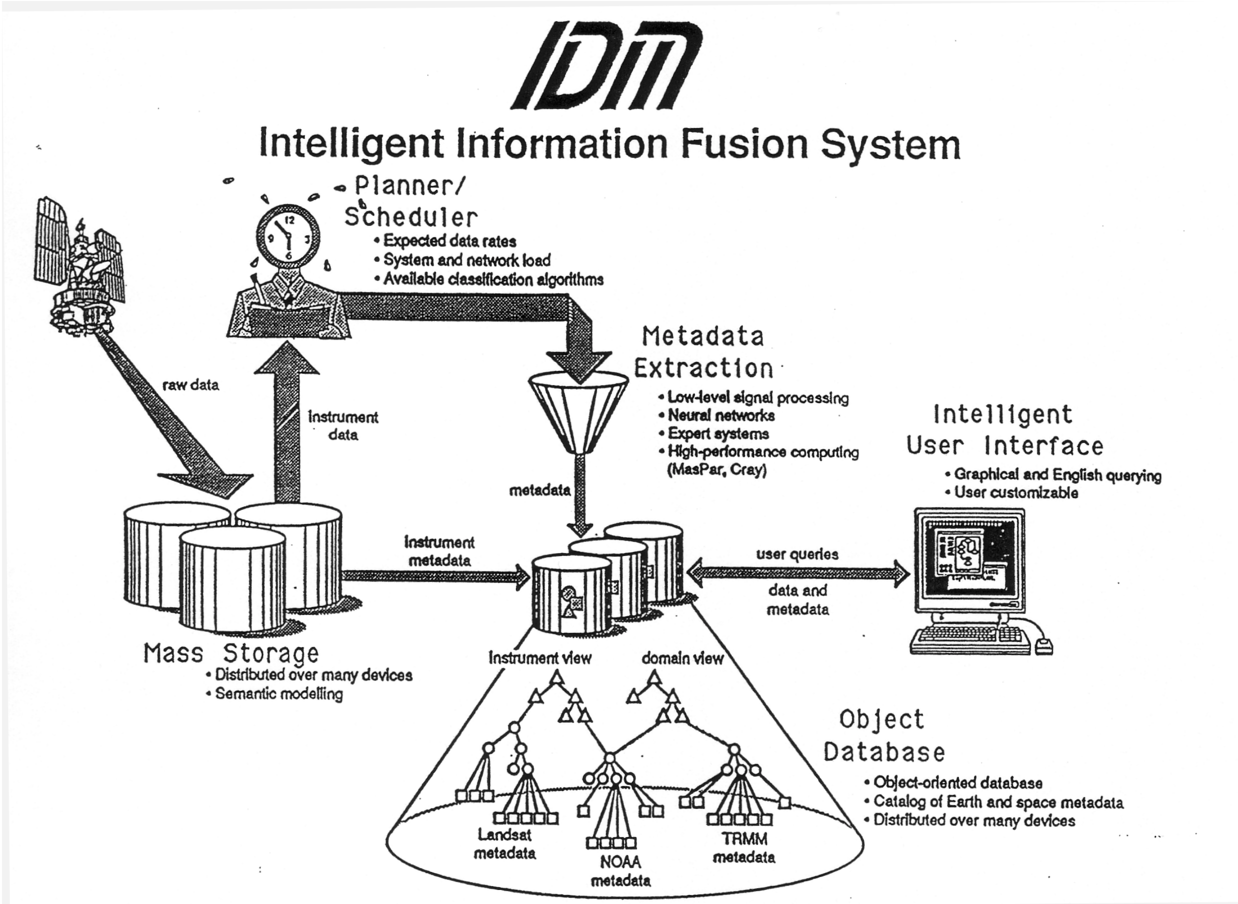


Figure 4. The NASA/Goddard Space Flight Center Intelligent Data Management IIFS (courtesy of R.F. Cromp).

“Where's the AI?"

The issue of scalability is a key point of an interesting article by Roger Schank in the winter 1991 issue of *AI Magazine* (Schank (1991)). Key position statements of the article are (1) "The answer to where's the AI? is that it's in the size"; and (2) “AI means a machine that learns and from this view point there is no AI, at least not yet". Commenting on the view that, "if no machine did it before, it must be AI." Shank says, "Two important types of programs to discuss within this conception of Al are optical character readers and chess-playing programs. Are these programs AI? Today, most people would say that they are not. Years ago, they were. What happened? The answer is that they worked. They were Al as long as it was unclear how to make them work."

This is also not an uncommon experience concerning reactions to human intelligence and problem solving as has been commented on by a number of people. However, in many areas, including OCR and speech recognition, the statement that something "works" must be highly qualified. While we are much further along in OCR than the technology used in the postal readers when I was with Philco-Ford Corp., the latest report on the status of postal address reading is that the system being developed at the Center of Excellence for Document Analysis and Recognition at the State University of New York at Buffalo is able to read the Zip Code about 75% of the time and can read an address and Zip Code about 30% of the time. The current failure of hand-held pen computers to catch on in the marketplace is attributed to poor performance of the recognition software in reading handwritten input. Possibly in another two years much better performance will be achieved. Currently automation of handwriting recognition is far from the state of "working".

At a recent voice recognition show, the systems being exhibited performed equally poorly except in highly constrained situations. However, recent reports indicate progress. My former student, P.S. Gopalakrishnan (Gopal) works in Fred Jelinek's group at the IBM Watson Research Center. According to their publications, that group has been following an approach to speech recognition based on hidden Markov modeling and statistical techniques. In a March 1992 paper (Bahl et al. (1992)) they have published results on continuous speech dictation. Their techniques produce an error rate of 4.8% in transcribing natural English sentences from a vocabulary of 5000 words, spoken as in a normal conversational mode, i.e., no forced pauses between words. If "perplexity" is viewed as the average branching factor at any point in a partially expanded sentence, this is a high perplexity task, i.e., it has a very large search space. This system is speaker-dependent requiring sufficient training data from each speaker. The Defense Advanced Research Projects Agency (DARPA) of the United States is sponsoring research on spoken language systems and the task chosen for this effort is an airline transport information system (ATIS). A February 1992 Proceedings of a DARPA workshop on speech and natural language processing contains a paper by F. Kubala et al. of BBN Corp. (Kubala et al. (1992)) on a task with a vocabulary of about 1500 words. The task has low perplexity. However the sentences are spoken spontaneously, not read from a prepared text; this leads to all types of hesitation and stammering. Also there is noise and the task is speaker-independent. The BBN system reported in the paper achieved a 6.2% error rate on a subset of the test data, with some noisy data removed. Certainly these achievements are significant when compared to past performance but the benchmark in such domains is human performance.

Suppose the above speech recognition systems scaled up and even learned, but we knew that they use hidden Markov modeling and certain statistical techniques and learning *procedures.* Would we call them intelligent? Perhaps some of us would and others would not. After all, we can search very large spaces using some search algorithms. Search is a very useful contribution of mathematics, OR and Al, and a key ingredient of Al systems. But even many within the field of Al are unwilling to view search as an important part of Al, and many in OR are unable to appreciate the value of heuristic search. It seems that the argument of what is or is not AI will continue even as the contributions from the field known as Al become an integral part of computer science, engineering and other disciplines.

Schank's article ends by suggesting case-based reasoning and case-based teaching as promising areas for Al. Currently case-based reasoning is a popular topic in Al. Relevant in this context is a theorem enunciated many years ago by Satosi Watanabe, a physicist-engineer and philosopher of pattern recognition.

Categorization

Assessing similarity and using near-neighbor classification underlies case-based reasoning. Watanabe's "Theorem of the Ugly Duckling" proved many years ago that if the resemblance or similarity between two objects is measured by the maximum number of predicates shared by them, then the degree of similarity between any pair of arbitrary objects is the same. Thus a swan and a duck, and two swans are equally similar. This situation arises because all predicates of the same rank are treated equally. As Watanabe discussed in his papers and books (Watanabe (1969) and (1985)), performing logical manipulation on raw data resulting from observation does not provide a grouping among the observed objects because unless some predicates are considered more important than others, i.e., weighted more heavily, the above theorem holds. Logic may be "a ballet dance of bloodless categories" (Bradley (1922)), but the preferential weighing of predicates has its origins in human values and in the objective of performing a classification. As Watanabe puts it, "What makes human cognition possible is the evaluative weighing whose origin is aesthetic and emotional in the broadest sense of the terms." In comments on non-similarity grouping and object-predicate inversion, he also points out the weakness of simple-minded similarity theory as a foundation for pattern recognition and gives examples where the relationship among elements of a group is different from similarity. Interesting examples include trilateral circular relations and multilateral relations. In the first several chapters of his 1985 book titled, *Pattern Recognition-Human and Mechanical,* he summarizes earlier papers of his that cover a variety of philosophical views on categorization, from the Greeks and Western philosophers to Brahmanism and Buddhism.

Many of the points about categorization touched on in Watanabe's papers and books are addressed at length in an excellent book by George Lakoff, which brings together thinking and empirical evidence from several disciplines, languages and cultures, on how humans form categories (Lakoff (1987)). The provocative[2] title *Women, Fire and Dangerous Things,* comes from the fact that these three items are placed in the same category in

[2] Lakoff's title invariably provokes a response. After my talk the following anonymous note showed up on bulletin boards in the convention center: "Women, Fire and Dangerous Things; Men Dust and Uninteresting Things."

an Australian aboriginal language, Dyirbal-but not on the basis that women are fiery and dangerous. Anything connected with water, the sun, and the stars is also placed in this category.

Alternate realities

One of the themes of the above mentioned books is a readiness to accept the existence of alternate perceptions of reality with alternate meanings and objectives. Consider the following. Sometimes, as in the goblet/two faces picture or drawings such as the young girl/old lady picture (Figure 5), or Wittgenstein's duck/ rabbit drawing, there are actually two different perceptions of reality simultaneously

Figure 5. Young woman/old woman drawing

represented by the same data. Other times, even if there be one reality known to an all-seeing God, it would inevitably be reported differently by observers based on their individual perceptions. For example, there is the classic Japanese movie, *Roshomon,* in which different witnesses to a crime give very different accounts of the events that transpired. That such multiple perceptions of reality are quite common in scientific affairs is brought home to us by the following observations concerning the book *Perceptrons* written by Marvin Minsky and Seymour Papert. In (Hecht-Nielsen (1990)), Robert Hecht-Nielsen quotes from the book *Perceptrons;* a part of the extract is presented here:

> "However, we were now involved in establishing at M.I.T. an artificial intelligence laboratory largely devoted to real `seeing machines,' and gave no attention whatsoever to perceptrons until we were jolted by attending an I.E.E.E. workshop on Pattern Recognition in Puerto Rico early in 1967. Appalled at the persistent influence of perceptrons (and similar ways of thinking) on practical pattern recognition, we determined to set out our work as a book..."

I mentioned this I.E.E.E. workshop at the beginning of this talk. I was one of the organizers and also edited the book *Pattern Recognition* (Kanal (1968)) which was the Proceedings of the workshop. As noted in the preface of the book, the workshop was held on October 24-26, 1966. As I have mentioned elsewhere, many workers had abandoned perceptron type learning machines and moved on to statistical and syntactic procedures. So only one session of five papers, chaired by Prof. Widrow of Stanford, was devoted to papers on adaptive networks for pattern recognition. Thus of the thirty papers in the book only five (5) are on this topic. Yet Minsky and Papert found themselves appalled at the persistent influence of perceptrons! I believe I attended most of the sessions. I had a different perception regarding the influence of perceptrons at the workshop. Whatever the reasons, their perception was different and motivated them to do some very interesting research on the single layer perceptron. I certainly would not say their book was the main reason for the demise of perceptron type networks. I think the main reason was inadequate technology and training algorithms for multilayer perceptrons and the unrealistic, premature promises made to funding agencies. Still, Hecht-Nielsen shares a widely held feeling that research on artificial neural networks was severely set back by Minsky and Papert's book.

In uncertain and ill-structured domains, i.e., in almost all areas of human activity, alternate possibilities in the form of different objectives, different perceptions and different partial information should be systematically considered. That is the motivation underlying the work presented in recent papers by Bhatnagar and Kanal (1993 a,b,c) and Bhatnagar's Ph.D. dissertation (1989). Our inspiration also derives from the Kantian and Hegelian models of enquiry developed in the writings of C.W. Churchman and his students ((Churchman (1971), Mitroff and Turoff (1973)). Instead of having a single domain model on which one performs Bayesian reasoning, we assume that the overall domain model is a forest which has embedded in it a multiplicity of situation models. The dissertation presents a hypergraph-search based methodology for hypothesizing alternative situation models for reasoning and planning. This produces inferences of interest based on different objectives of inquiry, rather than determining inferences in pre-determined causal models. The work develops what I think is a novel and useful method for representing qualitative knowledge about the known causal relations along with probabilistic knowledge about the domain. Formalisms for reasoning based entirely on qualitative cause effect relationships fail to capture the associated uncertainty. On the other hand, reasoning based on probabilistic methods alone takes into account statistical correlations and ignores the available -knowledge about the qualitative cause-effect relationships. This is not unlike the situation in pattern recognition years ago when it became apparent that both probabilistic and structural or linguistic aspects of patterns need to be employed. The problem of automated abductive reasoning which we address in Bhatnagar and Kanal (1993a,b), is to find interesting explanations, i.e.; situation models for a set of partially observed events, in terms of the known qualitative cause-effect relationships while using statistical correlations whenever the qualitative cause-effect knowledge is unavailable. We also use probabilistic and fuzzy (Bhatnagar and Kanal (1993c)) knowledge about the domain to determine the uncertainty associated with various parts of an explanation. The capability to hypothesize interesting, alternative causal structures is needed in several reasoning and planning domains, e.g., legal battles in a courtroom, military battles, medical diagnosis, and fault diagnosis.

A variety of approaches to reasoning in uncertain domains is surveyed in (Bhatnagar and Kanal (1992)) and developed in several volumes of *Uncertainty in AI* in the book series *Machine Intelligence and Pattern Recognition* published by Elsevier/North-Holland. An excellent talk by Prof. Edwina Rissland on AI and Legal Reasoning, given at the 1992 AAAI conference, covered a lot of recent Al research on the topic in a very interesting manner. A starting point for references on this topic is Rissland (1988). A recent book by Peng and Reggia (1990) develops abductive reasoning techniques for diagnostic problem solving.

Alternate realities are very much a part of the riddles being tackled by quantum physicists as is discussed in an interesting article (Horgan (1992)) in the July 1992 issue *of Scientific American.*

Recent workshops indicate that many in neuroscience, pattern recognition, neural networks and related fields are strongly attracted to theoretical approaches based on quantum physics. At some workshops there is much discussion of the mind-body problem, consciousness, and materialist versus dualist philosophies. All very interesting and lots of fun. Sometimes the temptation to make grand claims is hard to resist, even for Nobel prize-winners:

> "We are at the beginning of the neuroscientific revolution. At its end, we shall know how the mind works, what governs our nature, and how we know the world."
>
> Gerald M. Edelman (1992)

One should never say never, and perhaps Edelman will one day be proved right. But at least in this lifetime of mine, on this claim as on many other such claims, I think I will bet on the following words from Goethe:

> Faust: Wohin der Weg?
> Mephistopheles: Kein Weg! Ins Unbetretene.
>
> (Where lies the way? No way! Its Untrodden.)

Prospects

In the early 1960's, conferences titled Adaptive systems, Self-Organizing Systems, Bionics, and Pattern Recognition, attracted scientists from many different disciplines, including psychology, engineering and neuroscience. Artificial Intelligence and Pattern Recognition were viewed as a common enterprise. Part of the excitement was due to neural networks, especially perceptrons. After two decades of growth and splitting into narrower and narrower subfields, with many specialized names, there is once again a confluence of ideas and results from neuroscience, physics, psychology, computer science, engineering, mathematics, etc., and an appreciation of the need to have a more integrated or holistic approach to pattern recognition. Although the current round of excitement also started with neural nets, in practical applications hybrid methods, combining numeric, symbolic, statistical and structural, crisp and fuzzy representations and logics, are rapidly taking over center stage. The fast hardware, desk-top workstations with dazzling color displays, easy to use software tools and user-friendly interfaces all add to the renewed excitement about our work. But the main reason is that with the new hardware and software technology, and three decades of algorithm development, satisfactory solutions to some complex pattern recognition problems are being obtained.

Formalisms for developing algorithms and parallel implementations for many of the individual tools shown in Figure 1 have received much attention in recent years. But the integration of heterogeneous computational components, multiple sensors producing different types of data, and heterogeneous knowledge bases, is a significant systems design problem for which we currently have only ad-hoc techniques. Clearly more systematic methods and formalisms need to be developed for the design of complex multilevel systems consisting of heterogeneous modules performing specialized local computations while interacting with other modules at the same and different levels of a hierarchical organization. Such interaction involves information and decisions flowing back and forth, with competition, and cooperation, all in the context of global constraint satisfaction. The management of hybrid systems, not unlike the management of complex human organizations, requires significant attention to language and communication between different modules for reasoning and decision making (Kanal and Perlis (1989)). In the process of developing formalisms for hybrid systems, foundations underlying current approaches should be open to question. This occurred, for example, with Lotfi Zadeh's development of fuzzy set theory; that theory should become ‘fuzzier’ with interval based computation (Womg.et al. (1992)). It is likely that the ability to model complex cognitive tasks using digital systems will be more and more, in question. In Kanal and Tsao (1987) we have expressed some doubts regarding perception being modelable by a Turing machine. The need for and possibility of non-Turing machines are likely to be topics of increasing discussion and argument in the coming years.

Concluding remarks

In its broadest sense pattern recognition is at the heart of all scientific inquiry, including our attempt to understand ourselves and the world around us. That task remains a most challenging one, but- what could be more interesting? In the IAPR we should continue to take a broad interdisciplinary view of pattern recognition, be open to new ideas, keeping in mind the multicultural nature of our enterprise.

And now it appears that the time has gone. But here's an alternate reality:

> Le temps s'en va, le temps s'en va, madame!
> Las! le temps non: mais Nous nous en allons!
>
> Time goes, you say? Ah, no!
> Alas Time stays, we go;
>
> Pierre de Ronsard, The Paradox of Time
> (Austin Dobson, tr.)

This is a beautiful poem but ultimately a sad one and I don't want to end on a sad note. So I end with a final definition of "pattern". A "pattern" in Ireland is a feast or merriment in honor of a patron saint. Let us dedicate our banquet to the memory of King-Sun Fu. Let us have a "pattern"[3]

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19 (6), 716-723.
Bahl, L.P. de Souza et al. (1992). A fast match for continuous speech using allophonic models. *Proc. Internal. Conf. on*

[3] The banquet two nights later was indeed a feast of food and music.

Acoustics, Speech, and Signal Processing. San Francisco, CA, 1-17 to 1-20.
Bhatnagar, R. (1989). *Construction of Preferred Causal Hypotheses for Reasoning With Uncertain Knowledge.* Ph.D. Dissertation, Computer Sciences Dept., University of Maryland, College Park, MD.
Bhatnagar, R. and L.N. Kanal (1992). Reasoning in uncertain domains: A survey and commentary. In: A. Kent and J.G. Williams, Eds., *Encyclopedia of Computer Science and Technology, Vol.* 27, Supplement 12, Marcel Dekker, New York, 297-316.
Bhatnagar, R. and L.N. Kanal (1993a). Structural and probabilistic. knowledge for abductive reasoning, to appear in *IEEE Trans. Pattern Anal. Machine Intell.,* Special issue on Probabilistic Reasoning.
Bhatnagar, R. and L.N. Kanal (1993b). Hypothesizing situation models from domain Knowledge. Revised manuscript submitted for publication.
Bhatnagar, R. and L. N. Kanal (1993c). Hypothetical reasoning in the framework of possibility theory. *Proc. Internal. Symposium on Uncertainty Modeling and Analysis.* ISUMA 93, College Park, MD.
Bilbro, G.L., W.E. Snyder et al. (1992). Mean-field annealing: A formalism for constructing GNC-like algorithms. *IEEE Trans. on Neural Networks 3 (1),* 131-138.
Blake, A. and A. Zisserman (1987). *Visual Reconstruction.* MIT Press, Cambridge, MA.
Bradley, F.H. (1922). *Principles of Logic,* 2nd ed. Oxford Univ. Press, London.
Campbell, W.J., S.E. Hill and R.F. Cromp (1989). The utilization of neural nets in populating an object-oriented database. NASA Conf. Publication 3033. Goddard Space Flight Center, Greenbelt, MD.
Churchman, C.W. (1971). *The Design of Inquiring Systems: Basic Concepts of Systems and Organization.* Basic Books, New York.
Cromp, R. F. (1991). Automated extraction of metadata from remotely sensed satellite imagery. *Proc. 1991 ACSM/ ASPRS/A UTO-CAR TO 10 Annual Convention,* Baltimore, MD.
Dattatreya, G.R. and L.N. Kanal (1985). Decision trees in pattern recognition. In: L. N. Kanal and A. Rosenfeld, Eds., *Progress in Pattern Recognition 2.* North-Holland, Amsterdam, 189-239.
Edelman, G.M. (1992). *Bright Air, Brilliant Fire. On the Matter of the Mind.* Basic Books, New York.
Gupta, N. and L.N. Kanal (1992). *Recovering Shape & Motion from a Sequence of Images.* Technical Report UMIACSTR 9267; CS. TR 2916. Computer Science Dept., University of Maryland, College Park, MD.
Hecht-Nielsen, R. (1990). *Neurocomputing.* Addison-Wesley, Reading, MA.
Hellstrom, B.J. and L.N. Kanal (1990). *The Definition of Necessary Hidden Units in Neural Networks for Combinatorial Optimization.* Technical Report UMIACS-CSD-CSTR2388, Dept. of Computer Science, University of Maryland, College Park, MD.
Hellstrom, B.J. and L.N. Kanal (1992a). Knapsack packing networks. *IEEE Trans. Neural Networks 3 (2),* 302-307.
Hellstrom, B.J. and L.N. Kanal (1992b). Asymmetric mean-field neural networks for multiprocessor scheduling. *Neural Networks 5,* 671-686.
Hopfield, J.J. and D.W. Tank (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics* 52, 141-152.
Horgan, J. (1992). Quantum philosophy. *Scientific American* 267 (1), July, 94-104.
Kanal, L. (1962). Evaluation of a class of pattern recognition networks. In: Bernard and Kate, Eds., *Biological Prototypes and Synthetic Systems I.* Plenum Press, New York, 261-269.
Kanal, L. (1970). Generative, Descriptive, Formal and Heuristic Modeling in Pattern Analysis and Classification. Report to Aerospace Medical Division, Air Force Systems Command, WPAFB, OH, Contract F33615-69-C-1571.
Kanal, L. (1971). Modeling Communication Channels with Memory. *Proc. Second USSR Symposium on Information Theory,* Tsakhadzor, Armenia, USSR Academy of Sciences.
Kanal, L. (1972). Interactive pattern analysis and classification systems-A survey and commentary. *Proc.* IFFF 60 (10), 1200-1215.
Kanal, L. (1974). Patterns in pattern recognition: 1968-1974. *IEEE Trans. Information Theory 20 (6),* 697-722.
Kanal, L. (1977). Hierarchical classifier theory and interactive design. In: P. Krishnaiah, Ed., *Applications of Statistics.* North-Holland, Amsterdam.
Kanal, L. (1979). Problem-solving models and search strategies for pattern recognition. *IEEE Trans. Pattern Anal. Machine Intell. 1 (2),* 193-201.
Kanal, L. and S. Raghavan (1992). Hybrid systems-A key to intelligent pattern recognition. *Proc. 1992 Internal. Joint Conf. on Neural Networks. Vol. IV,* 177-183. IEEE Catalog No. 92 CH3114-6.
Kanal, L.N. and G.R. Dattatreya (1986). Problem solving methods for pattern recognition. In: T.Y. Young and K.S. Fu, Eds., *Handbook of Pattern Recognition and Image Processing.* Academic Press, New York, 143-165.
Kanal, L.N. and E.S. Gelsema et al. (1989). Comparing hierarchical statistical classifiers with error-back-propagation neural network. Presented at VISION 89, SME, Detroit.
Kanal, L. and T. Tsao (1987). Artificial intelligence and natural perception. In: L.O. Hertzberger and F.C.A. Groen, Eds., *Intelligent Autonomous Systems.* North-Holland, Amsterdam, 60-70.
Kanal, L. and B. Chandrasekaran (1972). On linguistic, statistical and mixed models for pattern recognition. In: S. Watanabe, Ed., *Frontiers of Pattern Recognition.* Academic Press, New York.
Kanal, L. and A.R.K. Sastry (1978). Models for channels with memory and their application to error control. *Proc. IEEE* 66

(7), 724-744.

Kanal, L. and V. Kumar (1981). A branch and bound formulation for sequential and parallel game tree searching. *Proc. 7th Internal. Joint Conf. on AI.* Vancouver.

Kanal, L. and V. Kumar (1988). *Search in Artificial Intelligence.* Springer, Berlin.

Kanal, L.N., Ed. (1968). *Pattern Recognition.* Thompson Book Co. (out of print), Library of Congress Catalog No. 68-31794.

Kanal. L. and D. Perlis (1988). Uniform accountability for multiple modes of reasoning. *Int. J. Approximate Reasoning 2,* 233-246.

Kandel, A. and G. Langholz (1992). *Hybrid Architectures for Intelligent Systems.* CRC Press, Boca Raton.

Kubala, F. and C. Barry et al. (1992). BBN BYBLOS and HARC February 1992 ATIS Benchmark Results. *Proc. DARPA Workshop on Speech and Natural Language,* Harriman, NY.

Kulkarni, A. V. and L.N. Kanal (1976). An optimization approach to hierarchical classifier design. *Proc. 3rd Internal. Joint Conf. on Pattern Recognition,* San Diego. IEEE. Cat. #76CH1140-3C, 459-466.

Kulkarni, A.V. and L.N. Kanal (1978). Admissible search strategies for parametric and non-parametric hierarchical classifiers. *Proc. 4th Internal. Joint Conf. on Pattern Recognition.* Kyoto, Japan.

Kumar, V. and L.N. Kanal (1983). The composite decision process: A unifying formulation for heuristic search, dynamic programming, and branch and bound procedures. *Proc. National Conf. on AI.* Washington, D.C., 220-224.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* University of Chicago Press.

Minsky, M. and S. Papert (1969). *Perceptrons: An Introduction to Computational Geometry.* MIT Press, Cambridge, MA.

Minsky, M. (1991). Logical versus analogical or symbolic versus connectionist or neat versus scruffy, *AI Mag.* 12 (2), 34-51.

Mitroff, I.I. and M. Turoff (1973). The whys behind the bows. *IEEE Spectrum* March, 62-71.

Murata, N., S. Yoshizawa and S. Amari (1992). *Network Information Criterion-Determining the Number of Hidden Units for an Artificial Neural Network Model.* Technical Report #METR 92-05. Dept. of Mathematical Engineering and Information Physics, Faculty of Engineering, Tokyo University, Tokyo, Japan.

Peng, Y. and J.A. Reggia (1990). *Abductive Inference Models for Diagnostic Problem-Solving.* Springer, Berlin.

Raghavan, S. and L. Kanal (1992). Artificial neural networks are indeed smart. Biology Inspired Research (SBIR) Tools. In: J.L. Davis, Ed., *Proc. Navy SBIR Conf.* Office of Naval Research, Washington, D.C., 85-108.

Raghavan, S., N. Gupta and L.N. Kanal (1992a). *Tracking and Recognition of Multiple Objects in Passive Imagery.* LNK Corp. Final Rept. SDIO Contract N 00014-89-C-0011 to Office of Naval Research, Feb. 26, 1992.

Raghavan, S., N. Gupta and L.N. Kanal (1992b). *Computing Discontinuity-Preserved Image Flow.* Proc. 11th ICPR, The Hague Netherlands. IEEE. Computer Society Catalog No. 0-8186-2910-X/92, 764-767.

Rissland, E. (1988). Al and legal reasoning. *AI Mag. 9 (3),* 45-55.

Rumelhart, D.E. and J.L. McClelland, Eds. (1986), *Parallel Distributed Processing.* The M.I.T. Press, Cambridge, MA.

Schank, R.C. (1991). Where's the AI? *AI Mag.* 12 (4), 38-49.

Sethi, I. and A.K. Jain (1992). *Neural Nets and Statistical Pattern Recognition-Old and New Connections.* North Holland, Amsterdam.

Snyder, W.E., Y-S. Han and G.L. Bilbro (1992). A unified theory of edge-preserving smoothing. Manuscript. Personal Communication.

Stockman, G.C., L.N. Kanal and M.C. Kyle (1974). An experimental waveform parsing system. *Proc. 2nd Internal. Joint Conf. on Pattern Recognition,* Copenhagen, Denmark. IEEE Spec. Publ. #74CH0885-4C.