Applied Artificial Intelligence, 15:693–705, 2001

0883-9514/01 $12.00 +.00

EDITORIAL: MACHINE LEARNING IN COMPUTER VISION

FLORIANA ESPOSITO
DONATO MALERBA
Dipartimento di Informatica
Università degli Studi
Bari, Italy

In this editorial we briefly discuss interaction between two important areas of artificial intelligence: computer vision (CV) and machine learning (ML). Although the two fields have a long-standing tradition and can be considered technologically mature, past research in applying ML techniques to CV problems has been limited. After a short introduction in the fields of computer vision and machine learning, we highlight some important issues in the intersection of the two areas and sketch both current achievements and future research directions. Our goal is to help the reader put the six contributions of this special issue into the proper context.

In recent years, artificial intelligence has been viewed as the study and construction of rational agents, where the term agent simply denotes something that perceives its environment through sensors and acts upon the environment through effectors (Russel & Norvig, 1995). Many sensory modalities have been made available to artificial agents, the most useful of which is certainly vision. In vision the sensory stimulus processed is the light scattered from objects in a scene and projected in two-dimensional images. In computer vision (CV), an image is represented as an array of pixels, which is a two-dimensional function $f(x, y)$ that returns the intensity recorded by a light sensor at all points (x, y) of an image plane. The task of a computer vision system is to understand the scene that an image depicts.

The development of computer vision systems requires the solution to several problems, since the information to be processed, which is images, is

We gratefully acknowledge the work of J. M. Buhmann, T. Caelli, M. Petrou, P. Perner, T. A. Poggio, A. Verri, and T. Zrimec, who acted as program committee of the ECAI '2000 Workshop on "Machine Learning in Computer Vision" and helped us in reviewing the papers of this special issue. We also thank additional reviewers V. Cantoni, A. Chella, E. R. Davies, L. Hermes, F. A. Lisi, L. Saitta, and M. Vento for their effort during the paper selection process. We also wish to thank Dr. L. Rudd for her help in rereading the article.

Address correspondence to Donato Malerba, Dipartimento di Informatica, Università degli Studi, via Orabona, 4 70125 Bari, Italy. E-mail: malerba@di.uniba.it

obtained from a variable physical environment and is subject to several forms of alteration. *Image processing* and *signal processing* are the disciplines that investigate how it is possible to transform an input image into another image which has some desirable properties, such as a better signal-to-noise ratio or a corrected orientation. The classification of either an image or part of it requires the development of segmentation techniques, the extraction of a proper set of features from images, the selection of some of these features, and the building of a proper classification model. All these aspects are investigated in the field of *pattern recognition*. The description of the image itself is only an intermediate step towards the understanding of the scene that it depicts. The study of methods and techniques to recover information on three-dimensional scenes from two-dimensional images is the main goal of *image understanding* and *scene analysis*. They base this process of reconstruction on some cues available in the visual stimulus, such as binocular stereopsis, shading, texture, contour, and motion.

Understanding an image requires a large amount of knowledge, which can bias the transformation of the input signal (e.g., salt-and-pepper noise is possible), the segmentation of an image (e.g., the background is black), the extraction/selection of features (e.g., textural features are relevant to distinguish between green belt and buildings), the recognition of objects (e.g., only big objects are expected in the surrounding environment), and the reconstruction of the scene (e.g., only some vertex types are possible in the perceived world). Although humans are very good vision experts, they have great difficulty explaining how they see. Generally speaking, their introspection capabilities are much better developed for problem-solving tasks than perception tasks. Therefore, providing computer vision systems with all the necessary knowledge at the appropriate level of abstraction is practically impossible, unless relatively simple worlds are considered, such as the blocks world.

Each agent has a task to perform as best it can. Given a performance measure, some percepts of the surrounding environment and a background knowledge of the perceived world, a *rational* agent will try to optimize the performance measure while carrying out its own task. The agent may also try to *improve* its performance on the basis of information coming from the environment itself. Machine learning (ML) is the subfield of AI that aims at studying any process by which an agent improves its performance with time (Simon, 1983). To date one can identify four major ML *paradigms*, each of which has its own collection of conceptual models that altogether determine the way in which learning machines should be designed. The four paradigms are: statistical, logical or conceptual, connectionist or neural, genetic or evolutionary. They emerged from quite different scientific roots and employ different computational methods, although they all share the common goal of building machines which can perform new tasks that could not be performed before or perform old tasks better (e.g., faster or more accurately). The type of information provided by the environment determines the different *strategies* that learning systems must employ in order to improve their performance. The most important strategies are: *inductive* (e.g., learning a general concept description from a set of instances of the concept), *deductive* (e.g., knowledge reformulation, knowledge compilation, and creation of macro-operators to be used in planning activities) and *analogical* (e.g., case-based learning for problem-solving). Most research in ML has focused on a single paradigm and strategy, although the interest in multiparadigm and/or multistrategy approaches has grown during the past decade with the understanding of limitations of monoparadigm and/or monostrategy learning methods in real world problems (Michalski & Tecuci, 1994; Esposito, Michalski, & Saitta, 2000).

Constructing computer programs that are simply too difficult to program by hand is one of the application niches for ML (Mitchell, 1997). Since computer vision systems are difficult to program fully, because of both the complexity of real-world modeling and the large amount of required knowledge, they can greatly benefit from the application of ML techniques. Machine learning is currently the only known way to develop computer vision systems that are robust and easily reusable in different environments. However, the application of ML techniques to CV is neither straightforward nor well understood. Several problems are still unsolved and researchers have great difficulty in transferring competence acquired in one application domain to a new one. In the following, some of the current research issues will be presented.

RESEARCH ISSUES ON LEARNING IN COMPUTER VISION

From the standpoint of computer vision systems, machine learning can offer effective methods for automating the acquisition of visual models, adapting task parameters and representation, transforming signals to symbols, building trainable image processing systems, and focusing attention on target objects. To develop successful applications, however, the following issues have to be dealt with.

What are those Models used by Computer Vision Systems that might be Learned Rather than Handcrafted by the Designer?

In many CV applications, handcrafting the visual model of an object is neither easy nor practical. For instance, humans can detect and identify faces in a scene with little or no effort. This skill is quite robust, despite large changes in the visual stimulus. Nevertheless, providing computer vision systems with models of facial landmarks or facial expressions is very difficult (Rowley, Baluja, & Kanade, 1988). Even when models have been handcrafted, as in the case of page layout descriptions used by some document image processing systems (Nagy, Seth, & Stoddard, 1992), it has been observed that they limit the use of the system to a specific class of images, which is subject to change in a relatively short time.

How is ML used in Computer Vision Systems?

Machine-learning algorithms can be applied in at least two different ways in computer vision systems: first, to improve perception of the surrounding environment, which improves the transformation of sensed signals into internal representations; second, to bridge the gap between the internal representations of the environment and the representation of the knowledge needed by the system to perform its task (Esposito, Malerba, & Semeraro, 1996). The latter is by far the most common use. For instance, in an application to flaw detection, rules are generated by means of a decision-tree learning system so that flaws can be detected in pipe images once these are described by a set of features (internal representation of the environment) (Perner, 2001). An exception is the work by Saitta and Zucker (2001), in which machine learning is applied to the task of building abstract representations of the perceived world. A possible explanation of this marginal attention given to learning internal representations of the perceived environment is that feature extraction has received very little attention in the machine-learning community, because it has been considered application-dependent and works on this issue are not of general interest. The identification of required data and domain knowledge requires the collaboration with a domain expert and is an important step in the process of applying ML to real-world problems (Saitta & Neri, 1998). Only recently, the related issues of feature selection and, more generally, data preprocessing have been more systematically investigated in ML. Data preprocessing is still considered a step of the knowledge discovery process (Fayyad, Piatesky-Shapiro, & Smyth, 1996b) and confined to data cleaning, simple data transformations (e.g., summarization), and validation. On the contrary, many studies in computer vision and pattern recognition focused on the problems of feature extraction and selection. Hough transform, Fast Fourier transform (FFT) and textural features, just to cite some, are all examples of features widely applied in image classification and scene understanding tasks. Their properties have been well investigated and available tools make their use simple and efficient.

How do we Represent Visual Information?

In many computer vision applications, feature vectors are used to represent the perceived environment. However, relational descriptions are deemed to be of crucial importance in high-level vision. Since relations cannot be represented by feature vectors, pattern recognition researchers use graphs to capture the structure of both objects and scenes, while people working in the field of machine learning prefer to use first-order logic formalisms. By mapping one formalism into another, it is possible to find some similarities between works done in pattern recognition and machine learning. An example is the spatio-temporal decision tree proposed by Bischof and Caelli (2001), which can be related to logical decision trees induced by some general-purpose inductive learning systems (Blockeel & De Raedt, 1998).

What Machine-Learning Paradigms and Strategies are Appropriate to the Computer Vision Domain?

In this special issue, inductive learning, both supervised and unsupervised, emerges as the most important learning strategy, while the favorite paradigms are conceptual (decision trees, graph-induction), statistical (support-vector machines) and neural networks (Kohonen maps and similar autoorganizing systems). No application of genetic algorithms, evolutionary learning, or case-based learning is reported, although this could be mainly attributed to the limited number of contributions. A common factor to all applications is that learning algorithms should handle numeric attributes and relations extracted from images or detected by sensors.

From the standpoint of machine-learning systems, computer vision raises interesting and challenging problems. Many studies in ML assume that a careful trainer provides internal representations of the observed environment, thus paying little attention to the problems of perception. Unfortunately, this assumption leads to the development of brittle systems with noisy, excessively detailed, or quite coarse descriptions of the perceived environment. Some specific research issues raised by the computer vision domain are the following.

What are the Criteria for Evaluating the Quality of the Learning Processes in Computer Vision Systems?

In benchmarking computer vision systems, estimates of the predictive accuracy, recall, and precision are considered the main parameters to evaluate the success of a learning algorithm. However, the comprehensibility of learned models is also deemed an important criterion, especially when domain experts have strong expectations on the properties of visual models (e.g., the size or shape of a piece on a conveyer belt) or when an understanding of system failures is important (e.g., a logical component in a document image is not recognized because its corresponding layout component is a few pixels wider than the expected maximum (Malerba, Esposito, Lisi, & Altamura, 2001)). Comprehensibility is needed by the expert to easily and reliably verify the inductive assertions and relate them to their own domain knowledge, facilitate debugging, and improve the inductive programs themselves (Giboin, 1995). When comprehensibility is an important issue, the conceptual learning paradigm is usually preferred, since it is based on the comprehensibility postulate stated by Michalski (1983): "The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities. Components of these descriptions should be comprehensible as single 'chunks' information, directly interpretable in natural language, and should relate quantitative and qualitative concepts in an integrated fashion."

When is it Useful to Adopt Several Representations of the Perceived Environment with Different Levels of Abstraction?

In complex real-world applications, multirepresentations of the perceived environment prove very useful. For instance, a low-resolution document image is suitable for the efficient separation of text from graphics, while a finer resolution is required for the subsequent step of interpreting the symbols in a text block (OCR). Analogously, the representation of an aerial view of a cultivated area by means of a vector of textural features can be appropriate to recognize the type of vegetation, but it is too coarse for the recognition of a particular geomorphology. By applying abstraction principles in computer programming software, engineers have managed to develop complex software systems. Similarly, the systematic application of abstraction principles in knowledge representation is the keystone for a long-term solution to many problems encountered in computer vision tasks.

How Can Mutual Dependency of Visual Concepts be Dealt With?

In scene labeling problems, image segments have to be associated with a class name or label, the number of distinct labels depending on the different types of objects allowed in the perceived world. Typically, image segments cannot be labeled independently of each other, since the interpretation of a part of a scene depends on the understanding of the whole scene (holistic view). Context-dependent labeling rules will take such concept dependencies into account, so as to guarantee that the final result is globally (and not only locally) consistent (Haralick & Schapiro, 1979). Learning context-dependent labeling rules is another research issue, since most learning algorithms rely on the independence assumption, according to which the solution to a multiclass or multiple concept learning problem is simply the sum of independent solutions to single-class or single-concept learning problems (Malerba, Semeraro, & Esposito, 1997).

Obviously, the above list cannot be considered complete. Other equally relevant research issues might be proposed, such as the development of *noise-tolerant learning techniques,* the effective use of *large sets of unannotated images,* and the identification of suitable *criteria for starting/stopping the learning process and/or revising acquired visual models.* We hope that papers selected for this special issue of *Applied Artificial Intelligence* will give the reader an insight into at least some of the research issues reported above.

RELATED ACTIVITIES AND ACHIEVEMENTS

In recent years, there has been a boom in initiatives and research projects on the specific theme of learning in vision. Several workshops have been organized in Europe with the aim of bringing together researchers from different communities, such as machine learning, computer vision, and pattern recognition. The most recent are the following:

- European Conference on Computer Vision (ECCV) Workshop on "Learning in Computer Vision," Freiburg, Germany, June 1998, http://www-dbv.informatik.uni-bonn.de/learning.html.
- International Conference on Machine Learning (ICML) Workshop on "Machine Learning in Computer Vision," Bled, Slovenia, July 1999, http://ai.fri.uni-lj.si/icml-99-ws-mlcv/, (Zrimec, 1999).
- Workshop on "Machine Learning and Data Mining in Pattern Recognition," Leipzig, Germany, September 1999, http://www.ibai-research.de, (Perner & Petrou, 1999).
- ECAI '2000 Workshop on "Machine Learning in Computer Vision," Berlin, Germany, August 2000, http://www.di.uniba.it/~malerba/ws-ecai2000/, (Esposito & Malerba, 2000).

Research on this topic is also active overseas, where, for instance, the Laboratory of Computer Vision and Computational Learning has been founded at the Massachusetts Institute of Technology (MIT) (Poggio & Shelton, 1999; Poggio & Verri, 2000).

Among the fundamental achievements in the *learning theory and algorithms* suitable for computer vision, four deserve special mention. First, the development of *sparse function approximators*, such as the support vector machines (Vapnik, 1995), which can manage very high-dimensional feature vectors. In fact, for some low-level vision problems, it is quite difficult to define a small set of good features to be extracted from images, and it is better to generate a high number of features for each image and let the learning algorithm work with them in order to produce reliable classifiers. Second, the design of algorithms that can effectively exploit various forms of *domain* *knowledge* already available in the learning system. This is the case in robots that have some expectations of the environment, or model-based vision systems that support radiologists in the task of x-ray interpretation, for which several specialized books already exist. Third, the investigation of methods that can learn *multiple concepts*, which are defined relative to each other. There are many model-based high-level vision problems (e.g., scene-labeling problems) where many objects to be detected in the same scene interact with each other. Ignoring such interactions, due to functional dependencies and reflected in spatial location, might lead to inducing object models that are both more complex and less accurate. The fourth important achievement concerns the advances of studies on *abstraction* in learning. Abstraction is a fundamental computation tool to apply in complex vision problems, where some details are purposely ignored while building multiple representations of the same visual object.

All these theoretical and methodological achievements permitted the engineering of many CV applications. Some of them have a great potential impact on everyday life, such as the *detection of regions of interests* (ROIs), for example, face (Sung & Poggio, 1998), car and pedestrian detection (Papageorgiou & Poggio, 2000) in camera images, the recognition of faces or gestures (Essa, 1999), the *automated annotation* of images (Demšar & Solina, 1996; Belongie, Carson, Greenspan, & Malik, 1998) and videos aiming at subsequent information retrieval, the *remote-sensed data interpretation* supporting data acquisition in geographical information systems (Huang, Jensen, & Mackey, 1995; Jung, Jedynak, & Geman, 1997), the *extraction of both graphical and textual information from document images* (Esposito, Malerba, & Lisi, 2000) available either in workflow management systems or in digital libraries.

FUTURE RESEARCH TRENDS

Detecting trends in research on learning in computer vision is difficult because of the intrinsic multidisciplinary nature of the subject.

Regarding the *learning paradigm*, the mainstream of current research is oriented towards probabilistic or statistical approaches, which are characterized by sound mathematical foundations. Regularization networks (Girosi, Poggio, & Jones, 1995), support vector machines (Vapnik, 1995), and Bayesian methods (Knill & Richards, 1996) are now employed in many vision problems, both at high and low levels. When efficiency of the recognition phase is a strong requirement, as in the case of online handwriting recognition in pen-based computers (Tappert, Suen, & Wakahara, 1990), artificial neural networks are preferred because of their inherent computational model that can be easily implemented on VLSI chips.

The main *learning strategy* applied in CV is inductive, possibly supervised, learning. Reinforcement learning is confined to some robotic vision tasks. Recent applications of support vector machines to text categorization problems (Joachims, 1999), where unclassified observations abound, are likely to provide some insights into the issue of using large sets of mostly unannotated images, by means of a new strategy named *transductive learning* (Gammerman, Vapnik, & Vovk, 1998).

An attempt to bypass the limits of the simple feature-vector *representation* adopted in statistical learning theory has also been made by working on similarity matrices computed on graphs (Myers, Wilson, & Hancock, 2000). This approach seems in some way to be related to the current trend in machine learning to propositionalize structural representations of objects (Kramer, 1999); nevertheless, the motivations are quite different. In both CV and PR communities, researchers aim at dealing with structural representation without leaving the statistical approach which provides a solution to the problem of noisy data. On the contrary, researchers in ML consider propositionalization a way to bypass computational problems raised by first-order logic formalism. A different perspective is offered by abstraction mechanisms, which try to overcome problems of computational complexity by changing the representation of objects to keep all relevant information for the learning application and to discard useless details.

Another emerging trend is the focus on *scalability* in order to work on large volumes of images and videos. In handwritten character recognition, it is usual to find training benchmarks for more than 100,000 examples (e.g., the database provided by the National Institute of Standards and Technologies (NIST)). Another prominent example is the sky survey conducted at the California Institute of Technology (Caltech), which is collecting about 3,000 digital images of $23{,}040 \times 23{,}040$ 16-bit pixels each (Fayyad, Djorgovski, & Weir 1996). In these applications, and more generally, in multimedia data mining, computational complexity is such an important issue that only simple, fast numerical methods can be effectively employed.

ARTICLES IN THIS ISSUE

A recent MLnet workshop, held in Berlin before the 14th European Conference on Artificial Intelligence, focused on the discussion of the potential contribution of ML to CV (Esposito & Malerba, 2000).[1] The main goal was to bring together researchers from different communities, such as machine learning, computer vision, and pattern recognition to promote discussion and the development of new ideas and methods to deal with applications of machine learning to computer vision. The workshop emphasized the great potentialities and difficulties of all multidisciplinary events that include people with different backgrounds, experience, and terminology. In order to enlarge the audience interested in the discussion it was decided to solicit papers for this special issue on "Machine Learning in Computer Vision" of *Applied Artificial Intelligence*. Topics of interest include, but are not limited to:

- learning to recognize shapes
- supervised learning of visual models
- unsupervised learning for structure detection in images
- multistrategy learning in vision
- learning and refining visual models
- multilevel learning and reuse of learned concepts
- learning important features for image analysis
- relational learning in vision
- context in visual learning
- mining from large collections of images and videos
- interpretation of discovered visual models
- image segmentation via learning
- probabilistic model estimation and selection
- applications, such as medical imaging, object recognition, remote sensing, digital maps, document image analysis and recognition, spatial reasoning.

From the submitted papers, six have been selected. They are really representative of different areas involved in the topics of this special issue, which is computer vision, machine learning, and pattern recognition.

The first article in this special issue deals with two important problems for computer vision applications: how to represent complex objects so as to take relations among subparts into account and how to handle both numeric and symbolic data. Bischof and Caelli extend their previous work on learning from graph-based representations of objects in order to deal with numeric/symbolic spatio-temporal data. Learned classification rules are in the form of conditional cluster trees, which can be easily interpreted and validated. An application to the problem of learning actions is also presented.

The second article deals with an artificial high-level vision agent for the symbolic interpretation of data coming from a video camera. Perceived images are processed to extract snakes, which are deformable curves that move in the image under the influence of forces related to the local distribution of the gray levels. Then snakes are processed and the positions of interesting points, characterizing the posture of the arm or the posture of a person, are extracted from derived skeletons. A neural network is subsequently used to classify the temporal sequences of the movements of the interesting points into meaningful prototypical predicates.

The third article describes an interesting application to the problem of estimating the weight of fish sliding through a transparent channel. By applying a skeletonization technique, 13 features are computed from silhouettes extracted from side and top views of the fish. Support vector machines are then used to learn the regression relation between fish parameters and fish weight.

The fourth article concerns the problem of feature selection. The author compares two techniques and their effect on the classification accuracy of rules generated by a decision-tree learning system for a flaw detection problem in welds. Thirty-six features are extracted for each region of interest in a set of gray-level images. Features are computed by means of morphological edge detection operators, a Gaussian filter, a one-dimensional FFT filter, and a wavelet filter.

Saitta and Zucker describe how abstraction can be used in concept representation and propose a model of abstraction, which allows perception and categorization to be closely related. The authors found the model useful for distinguishing different aspects and roles of knowledge, as well as for automating the choice of the granularity level when solving a task. An application of the model to robot learning in a real-world environment is also reported.

The last article in this issue, by Song and Chiesielski, describes another automated visual inspection system applied to mashing assessment in the brewing process. Developing automated visual inspection systems is a time-consuming process, because there is no one set of preprocessing methods or set of features that suits all visual inspection tasks. The literature on image features is quite vast and several libraries of feature extraction programs are available. However, determining the best set of features for a new application is a time-consuming process involving much trial and error. The authors debate the use of a large library of features and machine learning methods for removing redundant features, so that development time for vision systems can be significantly reduced.

NOTES

1. A report on the workshop can be found in the MLnet information system http://www.mlnet.org/.

REFERENCES

Belongie, S., C. Carson, H. Greenspan, and J. Malik. 1998. Color- and texture-based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval. In *Proceedings of the International Conference on Computer Vision,* Bombay, India.

Bischof, W. F., and T. Caelli. 2001. Learning spatio-temporal relational structures. *Applied Artificial Intelligence* 15(8):707–722.

Blockeel, H., and L. De Raedt. 1998. Top-down Induction of first-order logical decision trees. *Artificial Intelligence* 101(1–2):285–297.

Demšar, J., and F. Solina. 1996. Using machine learning for content-based image retrieving. In *Proceedings of the 13th International Conference on Pattern Recognition,* Vol. IV, pp. 138–142. Los Alamitos, CA: IEEE Computer Society Press.

Esposito, F., D. Malerba, and G. Semeraro. 1996. From machine perception to machine learning: Generating models of visual objects. In *Machine Learning and Perception,* eds. G. Tascini, F. Esposito, V. Roberto, and P. Zingaretti, Series in Machine Perception and Artificial Intelligence, 23, 139–148. Singapore: World Scientific.

Esposito, F., and D. Malerba (Eds.) 2000. *Proceedings of the ECAI '2000 Workshop on Machine Learning in Computer Vision,* Berlin, Germany, http://www.di.uniba.it/~malerba/publications.mlcv.pdf

Esposlto, F., D. Malerba, and F. A. Lisi, 2000. Machine learning for intelligent processing of printed documents. *Journal of Intelligent Information Systems* 14(2/3):175–198.

Esposito, F., R. Michalski, and L. Saitta, (Eds.) 2000. Special Issue on Multistrategy Learning. *Machine Learning* 38(1/2).

Essa, I. A. 1999. Computers seeing people. *AI Magazine* 20(2):69–82.

Fayyad, U., S. G. Djorgovski, and N. Wier. 1996. Automating the analysis and cataloging of sky surveys. In *Advances in Knowledge Discovery and Data Mining,* eds. U. Fayyad, G. Piatesky-Shapiro, P. Smyth, R. Uthurusamy, 471–493. Menlo Park, CA: AAAI/MIT Press.

Fayyad, U., G. Piatesky-Shapiro, and P. Smyth. 1996. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, eds. U. Fayyad, G. Piatesky-Shapiro, P. Smyth, R. Uthurusamy, 1–35. Menlo Park, CA: AAAI/MIT Press.

Gammerman, A., V. Vapnik, and V. Vovk. 1998. Learning by transduction. In *Proceedings of the Fourteenth International Conference on Uncertainty in Artificial Intelligence,* pp. 148–155. San Francisco, CA: Morgan Kaufmann.

Giboin, A. 1995. ML comprehensibility and KBS explanation: Stating the problem collaboratively. In *Proceedings of the IJCAI'95 Workshop on "Machine Learning and Comprehensibity,"* pp. 1–11. Montreal, Canada.

Girosi, F., T. Poggio, and M. J. Jones. 1995. Regularization theory and neural network architectures. *Neural Computation* 7:219–269.

Haralick, R. M., and L. Shapiro, 1979. The consistent labelling problem: Part I. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1(2):173–184.

Huang, X., J. Jensen, and H. Mackey. 1995. A machine learning approach to automating construction of knowledge bases of expert systems for remote sensing image analysis with GIS data. In *Proceedings of GIS/LIS '95 Annual Conference and Exposition,* Vol. 1, pp. 456–465.

Joachims, T. 1999. Transductive inference for text classification using support vector machines. In *Proceedings of the Sixteenth International Conference of Machine Learning,* pp. 200–209. San Francisco, CA: Morgan Kaufmann.

Jung, F., B. Jedynak, and D. Geman. 1997. Recognising buildings in aerial images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images II,* eds. A. Gruen, E. P. Baltsavias, and O. Hericsson, 173–182. Basel: Birkhauser Verlag.

Knill, D. C., and W. Richards (Eds.) 1996. *Perception as Bayesian Inference.* London, Cambridge University Press.

Kramer, S. 1999. Relational learning vs. propositionalization: Investigations in inductive logic programming and propositional machine learning, Ph.D. Thesis, Department of Cybernetics and Artificial Intelligence, University of Vienna, Austria.

Malerba, D., G. Semeraro, and F. Esposito. 1997. A multistrategy approach to learning multiple dependent concepts. In *Machine Learning and Statistics: The Interface,* eds. C. Taylor and R. Nakhaeizadeh, 84–103. London: Wiley.

Malerba, D., F. Esposito, F. A. Lisi, and O. Altamura. 2001. Automated discovery of dependencies between logical components in document image understanding. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition.* Los Vaqueros, CA: IEEE Computer Society Press (in press).

Michalski, R. S. 1983. A theory and methodology of inductive learning. In *Machine Learning: An Artificial Intelligence Approach,* eds. R. S. Michalski, J. G. Carbonell and T. M. Mitchell, 83–134. Berlin: Springer-Verlag.

Michalski, R. S., and G. Tecuci. 1994. *Machine Learning: A Multistrategy Approach,* Vol. 4. San Mateo, CA: Morgan Kaufmann.

Mitchell, T. 1997. Does machine learning really work? *AI Magazine* 18(3):11–20.

Myers, R., R. C. Wilson, and E. R. Hancock. 2000. Bayesian graph edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22:628–635.

Nagy, G., S. C. Seth, and S. D. Stoddard. 1992. A prototype document image analysis system for technical journals. *IEEE Computer* 25(7):10–22.

Papageorgiou, C., and T. Poggio. 2000. A trainable system for object detection. *International Journal of Computer Vision* 38(1):15–33.

Perner, P., and M. Petrou. (Eds.) 1999. *Machine Learning and Data Mining in Pattern Recognition*. Lecture Notes in Artificial Intelligence, 1715. Berlin: Springer.

Perner, P. 2001. Improving the accuracy of decision tree induction feature preselection. *Applied Artificial Intelligence* 15(8):747–760.

Poggio, T., and C. Shelton. 1999. Machine learning, machine vision, and the brain. *AI Magazine* 20(3):37–55.

Poggio, T., and A. Verri. 2000. Learning and vision at CBL. Special Issue of the *International Journal of Computer Vision* 38(1):5–7.

Rowley, H. A., S. Baluja, and T. Kanade. 1998. Neural network-based face detection. *IEEE Trans on Pattern Analysis and Machine Intelligence* 20(1):23–38.

Russel, S., and P. Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice–Hall.

Saitta, L., and F. Neri. 1998. Learning in the "real world." *Machine Learning* 30(2/3):133–163.

Saitta, L., and J. D. Zucker. 2001. A model of abstraction in visual perception. *Applied Artificial Intelligence* 15(8):761–776.

Simon, H. 1983. Why should machines learn? In *Machine Learning: An Artificial Intelligence Approach,* eds. R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, 25–37. Berlin: Springer-Verlag.

Sung, K. K., and T. Poggio. 1998. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(1):39–51.

Tappert, C. C., C. Y. Suen, and T. Wakahara. 1990. The state of the art in on-line handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(8):787–808.

Vapnik, V. 1995. *The Nature of Statistical Learning Theory.* New York: Springer-Verlag.

Zrimec, T. (Ed.) 1999. *Proceedings of the ICML'99 Workshop on Machine Learning in Computer Vision,* Bled, Slovenia.