Introduction to the Special Section on Learning in Computer Vision

The goal of computer vision (CV) research is to provide computers with human-like perception capabilities so that they can sense the environment, understand the sensed data, take appropriate actions, and learn from this experience in order to enhance future performance. The field has evolved from the application of classical pattern recognition and image processing methods to advanced techniques in image understanding like model-based and knowledge-based vision.

In recent years, there has been an increased demand for computer vision systems to address "real-world" problems. As noted in the Final Report of a recently-held NSF Workshop on the Challenges in Computer Vision Research: Future Directions of Research [1], "... much of our current models and methodologies do not seem to scale out of limited "toy" domains." Therefore, the current state-of-the-art in computer vision needs significant advancements to deal with real-world applications, such as navigation, target recognition, manufacturing, photointerpretation, remote sensing, etc. It is widely understood that many of these applications require vision algorithms and systems to work under partial occlusion, possibly under high clutter, low contrast and changing environmental conditions. This requires that the vision techniques be robust and flexible to optimize performance in a given scenario.

The field of machine learning (ML) is driven by the idea that computer algorithms and systems can improve their own performance with time. Machine learning has evolved from the relatively "knowledge-free" general purpose learning system, the "perceptron" [2], and decision-theoretic approaches to learning [3] to symbolic learning of high-level knowledge [4], artificial neural networks [5], and genetic algorithms [6]. With the recent advances in hardware and software, a variety of practical applications of the machine learning research is emerging [7].

Vision provides interesting and challenging problems and a rich environment to advance the state-of-the art in machine learning. Machine learning technology has a strong potential to contribute to the development of flexible and robust vision algorithms, thus improving the performance of practical vision systems. Learning-based vision systems are expected to provide a higher level of competence and greater generality. Learning may allow us to use the experience gained in creating a vision system for one application domain to a vision system for another domain by developing systems that acquire and maintain knowledge. We claim that learning represents the next challenging frontier for computer vision research.

More specifically, machine learning offers effective methods for computer vision for automating the model/concept acquisition and updating processes, adapting task parameters and representations, and using experience for generating, verifying, and modifying hypotheses. Expanding this list of computer vision problems, we find that some of the applications of machine learning in computer vision are: segmentation and feature extraction; learning rules, relations, features, discriminant functions, and evaluation strategies; learning and refining visual models; indexing and recognition strategies; integration of vision modules and task-level learning; learning shape representation and surface reconstruction strategies; self-organizing algorithms for pattern learning; biologically motivated modeling of vision systems that learn; and parameter adaptation, and self-calibration of vision systems. As an eventual goal, machine learning may provide the necessary tools for synthesizing vision algorithms starting from adaptation of control parameters of vision algorithms and systems.

An innovative combination of CV and ML techniques has the promise of advancing the field of computer vision, which will contribute to better understanding of complex images of real-world scenes. There is another benefit of incorporating a learning paradigm in the computational vision framework. To mature the laboratory-grown vision systems into real-world working systems, it is necessary to evaluate the performance characteristics of these systems using a variety of real, calibrated data. Learning offers this evaluation tool, since no learning can take place without appropriate evaluation of the results.

Past research in applying ML techniques to CV problems has been limited. One of the reasons for this was the lack of understanding and availability of tools for low-level image analysis. However, in the last decade, good progress has been achieved at this level. Still, most current vision algorithms continue to be ineffective in dealing with multi-scenario, real-world situations. Solving the signal-to-symbol transition problem remains one of the key challenges in the application of symbolic learning to vision.

Generally, learning requires large amounts of data and fast computational resources for its practical use. However, all learning does not have to be on-line. Some of the learning can be done off-line, e.g., optimizing parameters, features and sensors during training to improve performance. Depending upon the domain of application, the large number of training samples needed for inductive learning techniques may not be available. Thus, learning techniques should be able to work with varying amounts of a priori knowledge and data.

The effective usage of machine learning technology in real-world computer vision problems requires understanding the domain of application. Abstraction of a learning problem from a given CV task and the selection of appropriate representations...
for the learnable (input) and learned (internal) entities of the system. To succeed in selecting the most appropriate ML technique(s) for the given CV task, an adequate understanding of the different machine learning paradigms is necessary. The five key machine learning paradigms currently available are: inductive learning, analytic learning, case-based learning, genetic algorithms, and connectionist learning. Mltistrategy learning techniques are also being applied to solve vision problems [8].

Theoretical and practical advances have begun to be made in the field of computer vision by new techniques and processes of learning, adaptation and representation. There have been several Ph.D. theses in applying adaptive and learning techniques to computer vision problems. Since learning in vision is a new area of research, there are many unexplored issues and there are potentially many different ways in which learning can be applied to solve vision problems, and to optimize the resources needed by a vision system. Learning needs to be carefully applied to selected problems where it makes sense to do so.

A learning system has to clearly demonstrate and answer the questions like what is being learned, how it is learned, what data is used to learn, how to represent what has been learned, how well and how efficient is the learning taking place and what are the evaluation criteria for the task at hand. Experimental details are essential for demonstrating the learning behavior of algorithms and systems. These experiments need to include scientific experimental design methodology for training/testing, parametric studies, and measures of performance improvement with experience. Experiments that exhibit scalability of learning-based vision systems are also very important.

I. PAPERS IN THIS SPECIAL SECTION

The goal of this special section is to provide the reader with samples of recent developments that use adaptive and learning approaches in computer vision. The special section has one paper and seven correspondences. They use a variety of techniques for adaptation and learning.

In the first paper on learning and feature selection in stereo matching, Lew et al. address the problem of how to learn which of the seven image features (like gradient magnitude, gradient direction, Laplacian, curvature, etc.) to match in a binocular stereo system. The learning involves the use of local context to determine which features are most effective for matching. The paper integrates learning-based feature selection, feature set refinement, and surface reconstruction. It describes a noise-tolerant instance-based learning algorithm for finding optimal feature sets. an approach for adaptive maximization of the class separation criterion for refined feature selection. and a strategy for determining the need for a priori surface information to resolve correspondences. The complete stereo system has been tested on a variety of real images and the technique shows improved matching results compared to a conventional pyramid-based nonlearning algorithm.

In the first correspondence, Cho and Dunn develop a technique to recognize and learn 2-D shape concepts in the presence of affine transformations, occlusion and articulation. The technique learns from examples using a representation, called "conjunction of local shape properties" (CLP) that characterizes a shape by a list of "robust" local properties that are defined at overlapping regions. The learning method, called "property based learning" (PBL), is an incremental learning approach, where concepts are defined relative to each other. It is used to learn conjunctions of local properties important for classification (indexing and matching) of shapes. The learning includes a garbage-collection algorithm to remove insignificant properties and a reinforcement algorithm to acquire new properties or adjust weights of properties for each class when a classification error is made. Results are presented using images of articulated tool objects and hand gestures. The application of the technique to real images will require sophisticated low-level image analysis.

The focus of the second correspondence by Draper et al. is the development of goal-directed classifiers for pixel labeling. The authors use Linear Machine Decision Tree (LMDT) algorithm that is a combination of a linear machine with a decision tree. The LMDT builds a multivariate decision tree that trains a linear machine to classify the initial training set. The LMDT algorithm uses the "thermal" training procedure to find the coefficients of the linear machine. The authors show how LMDT can be altered to induce decision trees that minimize arbitrary misclassification cost functions. The classifier is applied to the red, green and, blue components of a color image to find "roads." A comparison of the algorithm with well known decision tree algorithms in the machine learning literature would be desirable to understand the improvement provided by the LMDT algorithm for pixel classification.

In the third correspondence, Greenspan et al. present the application of learning techniques from the pattern recognition literature to the problem of structural and unstructural texture discrimination. Texture characteristics are learned from examples during training, rather than specified a priori as in some model-based statistical and structural techniques. A multistage approach is used, where at the first stage multi-channel filtering with a bank of orientation and spatial frequency tuned filters is used to extract a set of image features. At the second stage, unsupervised learning using the K-means algorithm is used to obtain a more compact description. At the third stage, supervised learning of rules and associated estimates of probabilities characterize the different texture classes. The rules are obtained by a technique that maximizes an information-theoretic measure. Empirical results are shown on a set of outdoor images.

The next two correspondences use genetic algorithms for geometric primitive extraction and target recognition.

The correspondence by Roth and Levine extracts geometric primitives like curves, circles and ellipses in edge images by representing these shapes by their minimal subset representation. The minimal subset representation is the smallest number of points necessary to define a unique instance of a geometric primitive. This is a nonstandard representation for
genetic algorithms that normally use bit-strings of parameters. The problem is formulated as an optimization problem and the genetic algorithm is used to find the optimal primitives. The evaluation function is based on matching a fixed-size template to the geometric data and finding the total number of points inside the template. The authors also compare their approach with the Hough transform.

The correspondence by Katz and Thrift present a genetic algorithm in a two-layered approach for automatic target recognition (ATR) in simulated and infrared imagery. The first layer, called "screener," selects regions-of-interest (ROI’s) from the original image. The second layer, called "classifier," takes the ROI’s generated by the first layer and performs actual classification. Genetic algorithms, applied to population of filters based on fitness criteria, are used at both layers for generating suitable filters for screening and classification. Fitness values are determined in each case by ground-truth information supplied during the training process. The genetic algorithm at the screen layer uses a fitness function for evaluation that is based on how close the peaks (obtained by applying the screen filter to the image) are to actual target locations. The genetic algorithm for the classifier layer is driven by the performance of a statistical classifier (threshold-based, Parzen, decision tree) to distinguish between target and clutter. There are as many filters as the number of potential targets in a image. A nonstandard representation is used so that the crossover operator retains a geometric meaning. The mutation operator works on each element of a filter by changing an element to a random real number in [-1, +1]. The recognition results are comparable to those obtained by a human observer.

Probabilistic Hough transforms are methods that are based on polling instead of voting to detect geometric features in an image. In the correspondence on the circular Hough transform, Yla-Jääski and Kiryati present a technique for adaptively setting the poll size (rather than keeping it fixed), for terminating the processing in a probabilistic Hough transform. The adaptive algorithm monitors the peaks in the parameter space as the input data from an edge detector is processed, stopping when a particular set of stable peaks have emerged. The stopping rules call for polls with an average size that is lower than the fixed poll size that would lead to the same error rate. The process terminates voting as soon as any number of objects seem to have been reliably detected even though the existence of other objects is not ruled out. The experimental results show how the termination criterion can be used with a Hough transform technique for circle finding. This is an adaptive algorithm, not a learning algorithm where the parameters are learned over time. It would be desirable to apply this work for detection of other geometric features and develop concrete analytical results.

The last correspondence by Smith et al. demonstrates that a simple statistical pattern recognition technique like K-NN and a large training database can be automatically optimized for parameters and weights to produce very high classification performance in the domain of handwritten digits. The K-NN algorithm uses three different metrics. They are the Hamming metric, the "squeeze" metric based on conventional distance maps, and the "penstroke" metric based on approximating the input pattern by arcs and matching against corresponding arcs of a template. The correspondence also presents a confidence metric for estimating the reliability of classification decisions. There are very few scalability examples for learning in vision and this correspondence is one that provides comparative experimental results and performs many large-scale experiments on a massively parallel supercomputer. The performance of these systems scale consistently with the size of the training database, where the error rate is cut by half for every tenfold increase in the size of the training set from 10 to 100,000 examples.

II. THE FUTURE

The field of machine learning in computer vision is just emerging. The number of CV systems that incorporate some learning component is expected to increase as vision researchers seriously attempt to fulfill the need for adaptation and learning in their vision systems. The abilities to reason and learn are the two highest-level capabilities associated with adaptive and learning-based systems for perception and manipulation.

There will be significant advances as the field evolves and there is increased interaction between machine learning and computer vision researchers. An even more likely development is that modern statistical and function approximation techniques—which are becoming widely attractive because of the neural network epidemics—will be applied to, and further developed for, vision problems by researchers in computer vision. There is little doubt in our minds that—despite the content of the current issue of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE—the next frontier in computer vision is indeed learning.

Bir Bhanu
University of California at Riverside
Tomaso Poggio
Massachusetts Institute of Technology
Guest Editors

ACKNOWLEDGMENT

We thank the authors who submitted and contributed their work to this special section and worked under tight schedule. We thank the referees who spent their valuable time in reviewing the manuscripts. We would like to thank A. K. Jain, Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, for providing useful suggestions during various stages of the special section. We wish to thank O. Firschein for his useful comments. Finally, we thank S. McKe and L. Highleyman for secretarial support.

REFERENCES


Bir Bhanu received the B.S. degree (with Honors) in electronics engineering from the Institute of Technology, BHU, Varanasi, India; the M.E. degree (with Distinction) in electronics engineering from Birla Institute of Technology & Science, Pilani, India; the S.M. and E.E. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology; the Ph.D. degree in electrical engineering from the Image Processing Institute, University of Southern California and the M.B.A. degree from the University of California, Irvine.

Since 1991, he has been a Professor of Electrical Engineering and Computer Science and Director of Visualization and Intelligent Systems Laboratory at the University of California, Riverside. Prior to that, he was a Senior Honeywell Fellow at Honeywell Systems & Research Center in Minneapolis. He has also worked with the University of Utah, IBM San Jose Research Laboratory, INRIA-France and Ford Aerospace & Communications Corporation. He has been the principal investigator of various programs from ARPA, NASA, NSF and other agencies and industries. Currently, he is the principal investigator on grants from ARPA and others in the areas of learning and vision, target recognition, and machine vision applications.


Tomaso A. Poggio was born in Genoa, Italy in 1947. He received the Ph.D. degree in theoretical physics from the University of Genoa in 1970.

From 1971 to 1982, he was Wissenschaftlicher Assistant at the max Planck Institut für Biologische Kybernetic in Tübingen, Germany. Since 1982, he has been a Professor at the Massachusetts Institute of Technology affiliated with the Department of Brain and Cognitive Sciences and the Artificial Intelligence Laboratory. In 1988, he was named to the Uncas and Helen Whitaker Professorship, and in 1992, he became a founding co-director of the Center for Biological and Computational Learning.

Dr. Poggio is the author of over 250 papers in areas ranging from psychophysics and biophysics to information processing in man and machine, artificial intelligence, machine vision and learning. He is on the editorial board of several interdisciplinary journals, a Fellow of the American Association of Artificial Intelligence, and a Corporate Fellow of Thinking Machines Corporation.
SPECIAL SECTION ON LEARNING IN COMPUTER VISION

Introduction to the Special Section on Learning in Computer Vision ....................... B. Bhau and T. Poggio 865

PAPERS

Learning and Feature Selection in Stereo Matching ......................... M. S. Lew, T. S. Huang, and K. Wong 869

CORRESPONDENCE

Learning Shape Classes ........................................ K. Cho and S. M. Dunn 882
Goal-Directed Classification using Linear Machine Decision Trees .......... B. A. Draper, C. E. Brodley, and P. E. Utgoff 888
Geometric Primitive Extraction Using a Genetic Algorithm ...................... G. Roth and M. D. Levine 901
Generating Image Filters for Target Recognition by Genetic Learning ........... A. J. Katz and P. R. Thrift 906
Handwritten Character Classification Using Nearest Neighbor in Large Databases ........................................ S. J. Smith, M. O. Bourgoin, K. Sims, and H. L. Voorhees 915

REGULAR SECTION

PAPERS

Computer Vision and Image Analysis

A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment .................... T. Kanade and M. Okutomi 920
Learning Compatibility Coefficients for Relaxation Labeling Processes .......... M. Pelillo and M. Refice 933

CORRESPONDENCE

Computer Vision and Image Analysis

Vector Field Analysis for Oriented Patterns ........................................ C.-F. Shu and R. C. Jain 946
Fast Parzen Density Estimation Using Clustering-Based Branch and Bound ................. B. Jeon and D. A. Landgrebe 950
A Least Biased Fuzzy Clustering Method ........................................ G. Beni and X. Liu 954