

ECAI - 2000 WORKSHOP

MACHINE LEARNING IN COMPUTER VISION

Final report

Organizing Committee:

Joachim M. Buhmann, University of Bonn, Germany
Terry Caelli, The University of Alberta, Alberta, Canada
Floriana Esposito, University of Bari, Italy (*cochair*)
Donato Malerba, University of Bari, Italy (*cochair*)
Petra Perner, Inst. of Computer Vision and Applied Computer Science, Leipzig, Germany
Maria Petrou, University of Surrey, UK
Tomaso A. Poggio, MIT, Boston, MA
Alessandro Verri, University of Genoa, Italy
Tatjana Zrimec, University of Ljubljana, Slovenia

Place and date: Berlin, August 22, 2000

1. Introduction

An *agent* is anything that perceives the surrounding environment through its sensors and performs actions upon it through its effectors. AI research aims to describe and build *rational* agents, which try to optimise their performance, given the information perceived from the environment and their background knowledge. Computer vision and machine learning investigate two important capabilities of rational agents: Human-like perception of the sensed data and improvement of agent performance with time. In recent years, there has been an increased interest in the synergetic contribution of these two fields to the development of agents that can solve “real-world” problems.

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, focusing attention on target objects. From the opposite side, computer vision can present interesting and challenging problems for people working in machine learning. Indeed, many studies in that field assume that a careful trainer provides internal representations of the observed environment, thus paying little attention to the problem of perception. Unfortunately, this assumption leads to the development of brittle systems with noisy, excessively detailed or quite coarse descriptions of the perceived environment.

The main goal of this workshop 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 such applications of machine learning to computer vision. Therefore, the workshop presented the great potentialities and difficulties of all multidisciplinary events that try to put in touch people with different background, experience and terminology.

2. Important issues in machine learning applications to computer vision

The different presentations at the workshop contributed to clarify what are the main issues to cope with when machine learning is applied to computer vision.

How is machine learning used in current computer vision systems?

Machine learning algorithms can be applied in at least two different ways in computer vision systems: firstly, to improve perception of the surrounding environment, that is to improve the transformation of signals captured by sensors into internal representations; secondly, to bridge the gap between the internal representations of the environment and the representation of the knowledge needed by the performance element to carry out its function.

The latter turned out to be common to almost all research activities presented in the workshop. Indeed, the way in which the perceived environment should be represented internally was generally defined a priori, although alternative representations were sometimes investigated. In particular, in his invited talk, Prof. Edwin Hancock assumed that graphs are extracted from training images in many high-level vision tasks. In their application to robotics, Bischof and Caelli selected specific unary attributes and binary relations to represent the movement of three parts of a robot arm, while in their application to intelligent document processing, Esposito et al. used a set of numeric/symbolic attributes and relations to describe the page layout of multi-page documents. In two applications to flaws detection, rules were generated by means of the system C4.5, and features were extracted by means of Gaussian, FFT and wavelet filters (see work by Perner and Apte) or by Hough transform and correlated Hough transform (see work by Cucchiara et al.). Eigenvalues of a covariance matrix computed from the pixels rows of a single ROI are used in a scene understanding application presented by Ardizzone et al., while the scene description component of a system for the interpretation of dynamic scenes presented by Chella et al. operates on the coordinates of some interesting points of some extracted skeletons.

The only work in which machine learning was applied to the task of building abstract representations of the perceived world is that presented by Prof. Lorenza Saitta in her invited talk.

A possible explanation of this marginal attention to learning internal representations of the perceived environment is that many studies in computer vision and pattern recognition focused on the problems of feature extraction and selection. Hough transform, FFT, 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. On the contrary, feature extraction has received very little attention in the machine learning community, because it has been considered application-dependent and works on this issue not of general interest. Only recently, the related issue of feature selection has been more systematically investigated in machine learning.

The application to cartographic generalization, that is compiling a map from larger scale source maps, which was illustrated by Lorenza Saitta, seems to be one of the rare cases in which machine learning can be profitably exploited to build suitable representations of the perceived environment.

What are the models of a computer vision system that might be learned rather than hand-crafted by the designer?

In almost all applications presented in the workshop, it was clear that handcrafting a model was neither easy nor practical. For instance, rules for the recognition of particular actions of a robot arm involve spatio-temporal conditions that are relatively easy to interpret once they are automatically generated, but which would be difficult to compile because of the many factors involved. Building the layout of a document image is the only application related to visual perception that was performed by means of hand-crafted rules encoding knowledge on typographical standards. This application is somehow similar to that of cartographic generalization, since both involve abstraction processes. Actually, some examples of systems that perform cartographic generalization by means of hand-coded rules are reported in the literature, but the main difficulty is that these rules depend on the particular map and cannot be easily reused. Therefore, the automated acquisition of cartographic knowledge is looked at with great interests by specialists.

What machine learning paradigms and strategies are appropriate to the computer vision domain?

In the workshop, inductive learning, both supervised and unsupervised, emerged as the most important learning strategy, while favourite paradigms are symbolic or conceptual learning (ILP, decision trees, graph-induction), statistical learning (discriminant analysis, support vector machines) and neural networks (Kohonen maps and similar auto-organizing systems). No application of genetic algorithms or evolutionary learning or case-based learning was reported in the workshop, although this should be mainly attributed to the limited number of presented papers. A common factor that emerged from all applications is that learning algorithms should handle numeric attributes and relations extracted from images or detected by sensors.

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 critical importance in high-level vision, as pointed out in the talk given by Hancock. Since relations cannot be represented by feature vectors, researchers working in pattern recognition use graphs to capture the structure of both objects and scenes, while people working in the field of machine learning prefer using first-order logical formalisms. By mapping a 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, that is related to the type of decision trees induced by some ILP systems (e.g., Tilde).

How does machine learning help to transfer the experience gained in creating a vision system in one application domain to a vision system for another domain?

Does experience gained in the domain of learning actions teach us something on handling spatio-temporal data in GIS applications? Does the abstraction model of visual perception developed by Lorenza Saitta fit well to the domain of document processing or object recognition? Is the work reported by Petra Perner on in-service inspection of welds in pipes useful for researchers interested in distinguishing between defective and non-defective industrial workpieces? These questions have not found an answer during the workshop but they certainly deserve great attention, if we want to prevent the occasional reinvention of the wheel.

How can mutual dependency of visual concepts be dealt with?

This issue was the main topic of the paper by Esposito et al. The solution sketched in this work is based on the parallel learning of clauses defining distinct concepts. The presented paper also described a new generalization model and a novel consistency recovery strategy to manage problems generated by the non-monotonicity property of the normal ILP setting.

What are the criteria for evaluating the quality of the learning processes in computer vision systems?

In almost all presented works, predictive accuracy, recall and precision are somehow estimated and are taken as main parameter to evaluate the success of a learning algorithm. In some papers, however, the capability of interpreting induced rules was also deemed an important criteria. This is the case of the application to document processing, where expected results in terms of rules were compared to actual output of the learning algorithms. This comparison was useful to discover current limits of the adopted representation as well as of the classification procedure.

When is it useful to adopt several representations of the perceived environment with different levels of abstraction?

In two applications of machine learning, multi-representations of the perceived environment proved very useful. One is cartographic generalization, where providing cartographers with some tool to support the abstraction process is the main requirement. The other domain is document understanding, where different concepts can be sometime attributed to layout components at different levels in the abstraction hierarchy.

3. The workshop in figures

Two invited talks were given in the workshop:

- Prof. Edwin Hancock from the Department of Computer Science, University of York, UK, presented a work on “A factorisation framework for structural pattern recognition” ;
- Prof. Lorenza Saitta from the Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, Italy talked about “An abstraction model of visual perception” .

Other regular papers were:

E. Ardizzone, A. Chella and R. Pirrone

Feature-based shape recognition by Support Vector Machines

W. Bishof and T. Caelli

Learning actions: induction over spatio-temporal relational structures

A. Chella, D. Guarino, I. Infantino and R. Pirrone

A high-level vision system for the symbolic interpretation of dynamic scenes by the ARSOM neural networks

R. Cucchiara, P. Mello, M. Piccardi and F. Riguzzi

An application of machine learning and statistics to defect detection

F. Esposito, D. Malerba and F. Lisi

Understanding multi-page printed documents: a multiple concepts learning problem

P. Perner and C. Apte

Improving the accuracy of C4.5 by feature pre-selection

The workshop was intended to be a genuinely interactive event and not a mini-conference. Thus, ample time was allotted for general discussion.

Acknowledgements

Thanks to ECAI 2000 Programme Committee and in particular to the Workshop Chair, Prof. Marie-Odile Cordier, for supporting the organisation of this workshop and the European Network of Excellence in Machine Learning (MLNET) for the economical support.