# Integration of Machine Learning and Vision into an *Active Agent* Paradigm

**Peter W. Pachowicz**
Systems Engineering Department and Center for Artificial Intelligence
George Mason University, Fairfax, VA 22030
ppach@mason1.gmu.edu

## Abstract

The paper introduces a transition from the traditional *train-recognize* paradigm into a learning-based *active agent* paradigm for the object recognition task. The role of different learning techniques is pointed out. We justify that a progress in the integration of learning and vision is in the development of new paradigms rather than in a simple transition of developed learning programs into the vision domain. The development of such paradigms provides an opportunity to build systems capable to learn and behave in an active manner even after the initial training is finished. We suggest that learning-based vision systems should be capable to run learning processes within the close-loop with the recognition processes; i.e., running in a "*never stop learning*" mode. In this case, the traditional direct distinction between the training and recognition phases tends to disappear. We indicate that the learning technology will have a major influence in the development of such autonomous active behaviors for robust object recognition systems. Some of example issues are discussed in terms of the face identification problem.

## 1. Introduction

The integration of learning and vision has direct implications on both research areas; where the objective of this integration is to advance the state of the art in machine vision by bringing a learning component to the vision research and vision systems. This indicates that the research in both areas should be reconsidered and revolutionized in order to join them together. At this point, we do not believe that slight modifications to vision and learning research will be sufficient. But, the integration of both areas can be achieved by implementing major changes within both areas (redefining vision and learning). The goal of such an integration is to enhance vision systems in these points where (i) stand alone vision techniques fail to solve the problem efficiently, (ii) there is no solution to the problem so far, or (iii) the solution is more natural (sound) using a learning approach.

Researchers frequently ask themselves whether the state of the art in learning and in vision is sufficiently mature for their integration. In this paper, we dispute such an approach. We demonstrate that the integration of learning and vision is responsible for much more than the simple transition (application) of developed learning tools (programs) to the vision domain - what characterized the most of early research in learning and vision. We argue that the integration of learning and vision requires the development of new paradigms.

This paper focuses on the transformation of the traditional static *train-recognize* paradigm into a new paradigm, a so called *active agent* paradigm and a "*never stop learning*" approach. We answer the questions of why and how such a transformation is applied. The special emphasis is on the role of learning research in this transformation, and the necessity to develop new learning and data/model manipulation methods. We believe that this paradigm will help to create better, active, and robust vision systems. At several points, we use the face identification problem to illustrate the development of this paradigm and the implications for the learning and vision research.

## 2. Traditional *Train-Recognize* Paradigm

An architecture of the traditional *train-recognize* paradigm is shown in Figure 1. There are two completely separate phases; i.e., the training phase and the recognition phase. The processes of these phases do not cooperate between themselves.

This research was conducted in the Center for Artificial Intelligence at the George Mason University. The Center's research is supported in part by the Advanced Research Project Agency under the grant No. F49620-92-J-0549, administered by the Air Force Office of Scientific Research, and the grant No. N00014-91-J-1854, administered by the Office of Naval Research, in part by the National Science Foundation under grant No. IRI-9020266, and in part by the Office of Naval Research under grant No. N00014-91-J-1351.

During the training phase, the system acquires object models from provided preclassified data (training examples). The attribute extraction processes are domain specific and pre-programmed. This attribute data is used then to acquire object models. The majority of traditional methods used to acquire models belong to the class of pattern recognition methods. When the training process is finished, the models acquired are transferred to the recognition phase. During the recognition phase, unseen data is matched with the models.

In the face identification problem, for example, domain specific attributes are calculated as distances or other specific numerical factors characterizing (specializing) pre-defined *generic geometric model* of a face. These generic models, frequently called *templates*, help to extract attribute values both for local geometric substructures and global geometric relationships. A fixed-length vector of such attributes represents a given face. A training set of attribute vectors can be obtained by repeating the attribute extraction processes under slightly repositioned face or changed perceptual conditions. These changes influence the extraction of geometric elements and in the same way the deviation of parametric values of the generic model. The model acquisition process uses a parametric, non-parametric, or neural net methods to find the parameters of a predefined *mapping procedure;* e.g., parameters of the Bayes risk minimization equation, or weights of the net connections. Next, these tuned *mapping procedures* are used in the recognition phase to map input test sample into the classification decision.

## 3. Justification for a new paradigm

The traditional *train-recognize* paradigm has serious problems. These problems relate to the static behavior of (i) the attribute selection and computation, (ii) the tuning of mapping parameters, and (iii) the mapping process with unknown instances. There is also lack of cooperation between the modules of the system. However, we admire that this traditional approach, if well suited to the application domain, is providing a very simple and powerful object recognition tool. But for more complex problems, especially problems of ill-defined complexity boundaries, it has obvious shortcomings. Again, the major critique is caused by system's static behavior compared to the dynamic behavior of the majority of application problems.

The following is a list of specific issues of our primary concerns characterizing object recognition task performed by a human:

1) The data characterizing objects are hybrid and noisy. Faces, for example, can be described by attributes that are numeric, symbolic, structural, functional, and relational. The noise influence changes attribute value. Sometimes the attribute value cannot be computed at all due to the visibility and object positioning problems.

2) Attributes are frequently grouped according to the local concepts. A concept of a face is composed of several local concepts (e.g., mouth template, eye template, chin template). These local concepts combined together with other object characteristics derive a face template.

3) Key attributes are not known a-priori. We do not know which attributes (from a given set) are the most distinctive for the application domain. Some of these attributes can be combined into new attributes (not defined a-priori) of a higher distinctive power.

4) Model acquisition processes are task-specific. The learning process has frequently additional objectives to the general goal of learning distinctive class descriptions. For example, let us assume that a system must be able to distinguish all faces it is trained to recognize above a certain probability level. If so, we have to organize the learning process in such a way that the descriptions of weakly recognizable faces will be modified in order to increase their discrimination power. This modification can be done by decreasing the recognition effectiveness of the strong classes to pay for the improvement of the weaker classes.

5) Model acquisition is controlled by background knowledge and generic models. Generic models are used in vision to find the attribute values from an image data. In the same way, the learning and recognition processes should also use the background and generic models in learning face descriptions and in face identification.

6) Model acquisition frequently incorporates more than one learning strategy. A single strategy is frequently not sufficient to deal with hybrid, complex, and noisy data.

7) Model acquisition is an active process. The human learning process actively searches for key clues in building concept descriptions. It includes, for example, the generation of sensor management procedures and requests for object repositioning.

8) Model matching is based on pre-organization of the concept memory and requested model exploration. The efficiency of object recognition depends on the optimality (speed, accuracy, etc.) of the matching processes. Appropriate memory organization and the manipulation of learned concept descriptions is of a great help to achieve such an optimality of recognition processes.

9) Matching processes handle the "I do not know decision". It is desirable to have a system capable to

provide the recognition decision indicating the lack of confidence to what is seen; i.e., saying that I do not know or I am not sure about this face identity. Such feedback information can be very useful in further exploration of models, or a request to learn a new object autonomously.

10) Matching processes are active. Matching processes for the recognition of objects have to deal with the complexity of objects and concept descriptions, and partial information available. It is necessary to apply active and incremental search of concept descriptions for clues which can help to recognize an unknown object.

11) Object characteristics are evolving according to variability of object appearances. Evolution of object characteristics is yet another example relevant to the active behavior of a learning-recognition system. Object models must be modified on-line according to the change in object characteristics caused by variable perceptual conditions.

## 3. An *Active Agent* Paradigm

### 3.1. System architecture and characteristics

Employing the already developed learning techniques along with the new techniques requested within an object recognition system provides an opportunity to define a new architecture and to explore new challenges. Such an architecture proposed for the face identification system is presented in Figure 2. Since we stress the differences between the traditional paradigm and the new one, this architecture is based on the modules of the traditional architecture presented in Figure 1. However, special marks and shadowed links indicate changes to the traditional architecture. Numbers associated with these modules and links point to the list of specific issues of our concern listed in section 2, and to the proposed solutions/approaches listed in section 3.2.

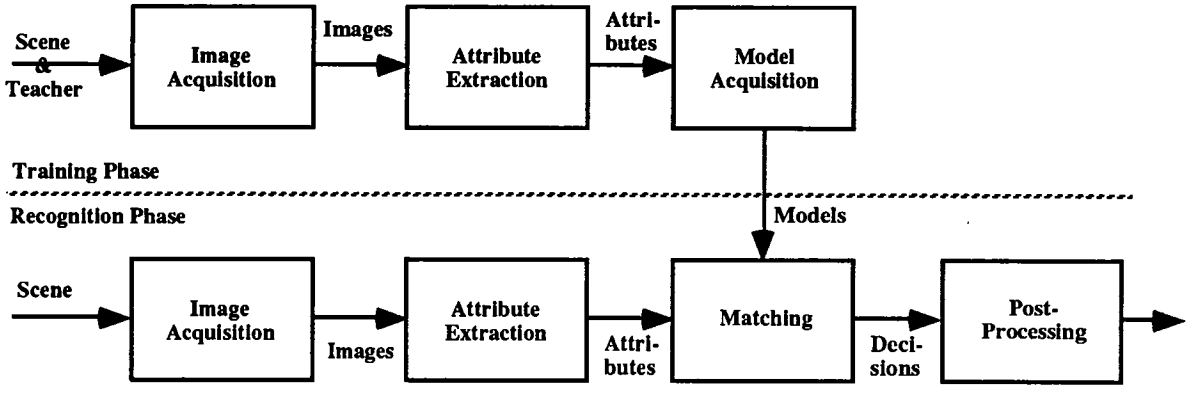


Figure 1. Architecture of the traditional *train-recognize* paradigm for object recognition task.

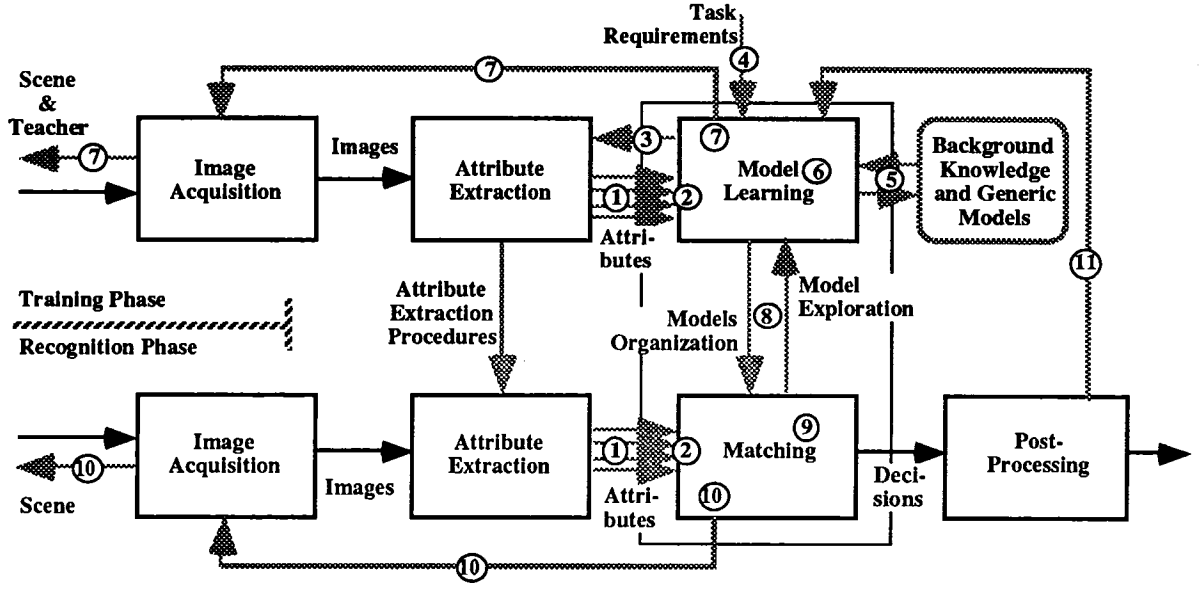


Figure 2. Transformation of the traditional *train-recognize* paradigm into an *active agent* paradigm. The closed loop links support a "*never stop learning*" approach to the autonomous object recognition. (The marks associated with modules and links correspond to the items listed in sections 2 and 3.2)

The major difference between the traditional architecture (Figure 1) and the new one (Figure 2) is that the processes of learning and recognition are no longer separated and they cooperate closely. So, the traditional distinction between the training and the recognition phases tends to disappear. The initial training phase is still needed, but the following system activity tends to be autonomous in memory organization, model reorganization, learning an optimal matching schedule, activation of learning processes, model verification, model evolution.

The *active agent* paradigm is also characterized by additional links supporting closed loop interactions between system modules. Adding system activity is a solution to system robustness and the increase in system autonomy (as a measure of system independence on human guidance, and the frequency of breakdowns). For example, the recognition experience can lead to the improvement of object models over time (see mark No. 11). Such an active behavior constitutes a "*never stop learning*" approach to the object recognition problem.

## 3.2. Challenging issues

The following is a list of selected solutions / recommendations to the issues listed in section 2. We indicate these issues in the context of the new paradigm for object recognition. (The items numbers refer to the labeled modules and links in Figure 2.)

1) Learning from hybrid data. Currently the solution to this problem is limited to quantization of continuous numeric attributes into subsymbolic intervals of the pre-defined (by a teacher) interval length. Further advances should go towards the self-organization of quantization processes; including the automatic selection of quantization parameters (e.g., guided by information measures) and reorganization of subsymbolic values into values under the other resolutions, if necessary (the memory organization problem). A qualitatively different but very difficult solution can be based on learning multiple representations and their fusion. Such representations and fusion schemas, however, are not known a-priori, and must be constructed by a system itself, for example, based on generic models, local concepts, etc.

2) Attribute grouping, hierarchical modeling, and distributed learning. For complex objects like faces, it seems more appropriate to organize / distribute the learning processes to acquire local concepts (mouth, nose, eye, etc.) first and then the concept of a face using learned local concepts, rather than to learn it from all available information at once. A preceding acquisition of local concepts can help in the exploration of differences and discriminating connectivities for learning a higher level description of a face. We strongly believe that performing information organization (attribute grouping) and subsequent exploration of learned local concepts for the most distinctive clues is necessary in learning complex concepts. The creation of hierarchical descriptions also gives an opportunity to use them by matching processes flexibly on different hierarchical levels according to available visual data (partial information, noisiness, different object appearances).

3) Discovery of new key attributes. Constructive induction can be very useful in the creation of new key attributes. A significant difficulty in finding these attributes is in the exponential character of attribute combination processes. Specific guidance is necessary to narrow the search for new attributes. Such guidance can be based on domain-dependent knowledge (e.g., functionality of different face structures, rules of the geometry), and representation-dependent knowledge (e.g., rules of simplicity, inter-relationships).

4) Designing control inputs to the learning and recognition system. Since learning and matching processes are a part of a task-specific application system, they are run according to the learning and matching control requirements defined by the global objectives. It is necessary to design and build an integrated learning and vision system controlled by the additional inputs which can dynamically influence and monitor the internal processes. These inputs can include such information as time restrictions, stability requirements, recognizability requirements, complexity/memory limitations, etc.

5 & 6) Multistrategy learning systems. For some very simple application problems, a single learning strategy can be sufficient to achieve good results. However, a single strategy is frequently not sufficient to deal with hybrid, complex, and noisy data and multiple learning / recognition requirements. For example, the face identification problem is complex enough that it does not seem to rely on one strategy only. Multiple strategies are needed to learn face descriptions based on (i) generic templates of faces' local concepts, (ii) hierarchical models that explore differences and similarities of local concepts, (iii) using domain-specific background knowledge (explanation), (iv) supporting noise tolerance and forgetting, etc.

7) Incremental and active search for key clues. We redefine the initial training phase where traditionally a teacher has passively supervised the training process and in majority of cases has given no feedback about the correctness and/or effectiveness of the descriptions learned. In the active approach, the process of learning concept descriptions should cooperate with a teacher. The goal for such a cooperation is to search for key information characterizing a given object (face). Frequently, a very distinctive clue can be found by

requesting object repositioning or by observing certain behavioral patterns of face local structures. Discovering and representing such clues provides a very powerful tool for (i) building very simple descriptions, and (ii) distinguishing thousands of objects (faces) among themselves based on simple features rather than on very complex and detailed (complete) models. A major need in such an activity is the system's capability to (i) actively search the available information and the descriptions being built, (ii) manage the sensor site, and (iii) incorporate the Focus of Attention (FOA) mechanisms.

8) Dynamic memory organization for matching processes. Appropriate organization of concept descriptions for the matching processes (different than for the learning processes) is needed to secure high recognizability and optimality of the object recognition task. Recognition of complex objects requires an active search through image data according to the descriptions learned. This is an active process supported by a dynamic organization of concept descriptions. Organization of concept descriptions must be modified according to new hypotheses generated and then confirmed / rejected through the matching process.

9) Complementing concept descriptions by the descriptions of their boundaries. Finding descriptions of concept boundaries will help to handle the "I am not sure" or "I do not know" decisions. For example, the descriptions of concept boundaries can be found and stored using a Fuzzy Set representation.

10) Active vision. Recognizing complex objects requires an active vision capability. These processes, however, are guided by an active search for partial hypotheses and by the feedback information about their confirmation. It is necessary to develop active matching techniques based on the exploration of hierarchical models, background knowledge, and generic models.

11) A "*never stop learning*" approach to object recognition. Certainly, we want to have systems that recognize objects very fast based on simple descriptions. But in many cases, very complex matching must be run to recognize an object. This occurs frequently since objects change characteristics under variable perceptual conditions, different object positioning, and changed expressions of the object's internal state (e.g., the face expression representing different state of human emotion, action, and the thinking process). Closing the loop between matching processes and learning processes supports the self-adaptation capabilities of the object recognition system. Model evolution implements the idea of closed-loop recognition and learning, and provides a solution to the model modification and an optimization of matching processes. Such an approach, called a "*never stop learning*" approach protects the system against stagnation in a similar way as a human does.

## 4. Evaluation criteria

Developing new paradigms for integrated vision and learning requires a discussion of benchmarking. We believe that a goal of minimizing the error rate at any price can be misleading in the evaluation of new paradigms. Frequently, the new methods created within these paradigms are not fully developed and require years to mature. Therefore, traditional benchmarking should be offset by the analysis of the situations which can be handled by new paradigms / methods when fully developed and deployed, in contrast to the traditional well-developed methods. We believe that benchmarking should emphasize the qualitative difference between new paradigms and the traditional paradigms. The analysis of methods' performance by means of the error rate should be employed later when newly developed paradigms have matured.

## 5. Summary

This paper has presented the author's position on the integration of learning and vision. It is justified that this research should be directed toward finding new qualitatively different paradigms, rather than looking for bridges (connections) between stand alone vision and learning modules.

An *active agent* paradigm has been introduced and compared to the traditional *train-recognize* paradigm. Some of the most distinctive problems have been listed, along with selected solutions.