

# Cognitive (Vision) Systems

by Henrik Christensen

'Cognitive Vision Systems' is a European project addressing issues related to categorisation, recognition, learning, interpretation and integration in relation to vision systems for intelligent embodied systems.

Over the last decade there has been significant progress in the fields of artificial intelligence, computer perception, machine learning and robotics. Yet there has not been major progress on truly cognitive systems. Cognition is here interpreted as 'generation of knowledge on the basis of perception, reasoning, learning and prior models'. Cognition is not a *passive process*, where an observer merely is monitoring an external environment. In addition the system has facilities for communication with the environment through which it also can articulate its knowledge. The system is embedded in the world and interacts with its environment to gather knowledge and perform its mission. Consequently, a *cognitive system needs to be embodied* and it has a number of tasks that define its mission objectives. The operation of the system is defined by its mission objectives and it acquires knowledge about the environment through its perception system. The perception is active in the sense that the system can use the embodiment to interact with the environment, which can be used to change the state of the environment,

which subsequently can be observed by the perception system. The sheer amount of information available in the external environment calls for methods to generate 'abstract' models.

One important part of generation of models is the ability not only to RE-cognize objects, but also to perform recognition by means of categorisation. Categories form an important component to allow management of an abundance of information, association of function, physical layout, etc. for objects, situations and events. Categorical perception is, however, a major challenge as is seen on in the adjacent image. A task could here be "count the number of chairs". This is a non-trivial task. Without understanding of the physical layout of the scene some objects may be recognized incorrectly and using contours only, some of the shadows might be confusing. In addition the pictures on the wall do not represent places to sit. Visual cognition is obviously only one of many potential modalities that are of interest to cognitive systems. In addition for some objects there might be a need to interact with the

environment to determine if an object qualifies for a particular task. For instance, is a chair stable enough to allow sitting down? Some of these qualities can only be determined by interaction.

Another fundamental component of any cognitive system is memory. Inherently memory is limited and there is a need to consider how memory can be utilized for different purposes: context information, spatial layout, abstraction, etc. Memory is, however, limited which call for efficient methods to manage this scarce resource. This requires use of attention mechanisms for select information of interest, and data-mining methods or machine learning to generate abstractions and derived representations. A basic quality of memory is also forgetting or intelligent garbage collection methods.

The relation between prior models or long(er)-term memory and working memory is another issue to be addressed. As part of the study of memory there is also a need to study pedagogic models for acquisition of information. What is

**Example image that illustrates the complexity of recognizing chairs as no single technique in terms of contours, appearance, components etc. is adequate to correctly allow 'counting of the number of chairs'. (Source: Bülthoff, Max Planck Institute for Biological Cybernetics (MPIK), Tübingen, Germany.)**



the best possible model to interact with a system to allow it to acquire a new skill, a task model or the representation of a new concept? Can traditional pedagogic models also be used to teach artefacts new concepts in an efficient manner? Cognitive science has proposed models for how infants acquire new models of the world; one such example is the learning models proposed by Piaget. Can such models be directly applied (can we endow the system with curiosity) or can be rephrase these methods so as to make them operational and applicable for artificial cognitive systems.

The concepts outlined above are fundamental questions addressed in the EU project 'cognitive vision systems' in which issues related to categorisation,

recognition, learning, interpretation and integration are addressed in relation to vision systems for intelligent embodied systems. For categorisation of objects a new hybrid model that integrates multi-models for recognition with different types of memory has been proposed. In addition the relation between spatial models (of objects) is being integrated with models for scene dynamics to capture episodic information and tie these to particular objects. The relation between reasoning, interpretation, recognition and processing of basic information cues is another fundamental problem studied. How does context allow control of the visual process to make it tractable, while at the same time allowing for enough richness to detect unexpected events? Traditional formal

models in AI have not had enough richness, but recent progress on reasoning under uncertainty shows promise in terms of richness and efficiency. Finally a Piaget inspired model of skill and task acquisition is being implemented as a basis for teaching robot like 'creatures' to interact with the environment. At present basic technologies are available and pair wise integration of techniques is performed to allow in-depth studies of the interaction between vision, AI, cognition and biology, psychology, computer science, and robotics.

**Link:**

<http://cogvis.nada.kth.se>

**Please contact:**

Henrik I. Christensen,  
KTH Royal Institute of Technology, Sweden  
E-mail: [hic@nada.kth.se](mailto:hic@nada.kth.se)

## Retrieval of Images based on Visual Content: A Biologically Inspired Multi-Agent Architecture

by Socrates Dimitriadis, Kostas Marias and Stelios Orphanoudakis

**Understanding the mechanisms of visual perception is important in a number of application areas, including content-based image retrieval (CBIR). An interdisciplinary approach to CBIR, in the framework of cognitive science, is expected to yield significant results where traditional approaches have failed to do so. In the Computational Vision and Robotics Laboratory (CVRL) of the Institute of Computer Science - FORTH, an ongoing research effort aims to develop and implement an experimental platform for the investigation of CBIR, based on a biologically inspired multi-agent architecture.**

In recent years, interdisciplinary research has produced significant results in a number of areas in which traditional approaches appeared to have reached their limit. Based on the joint effort of biological neuroscientists, computational neuroscientists, psychologists, and computer scientists, the interdisciplinary field of cognitive science has emerged as the field of research in which brain function is investigated with respect to behaviour and mental activity. Visual perception is a major brain activity and has attracted the attention of many cognitive scientists. Understanding the mechanisms of visual perception is important in a number of application areas, including content-based image retrieval (CBIR).

In the Computational Vision and Robotics Laboratory (CVRL) of ICS-FORTH, an ongoing research effort is aims to develop and implement a biologically inspired image retrieval methodology. Due to the fact that images represent a particularly large volume of information, the efficient and possibly intelligent browsing of images based on visual content is becoming increasingly important in application fields which make use of large image databases, such as diagnostic medical imaging, remote sensing, entertainment, etc. Surfing on the Web may also be facilitated and made more intelligent if visual browsing is exploited.

Humans undoubtedly possess the ability to process visual information efficiently and to identify images as being similar

based on their visual content. However, computational approaches currently fall short of matching this ability. The reasons for this are many and cannot be explored in this short article. It is our research objective at ICS-FORTH to develop and implement CBIR mechanisms that are perceptually relevant and based on a biologically inspired system architecture. We must therefore take into account what is currently known about how humans process visual stimuli and subsequently represent and store visual information, so that it may later be retrieved efficiently from visual memory.

It is known that the human cerebral cortex consists of many cortical areas, each one specialised and able to execute a specific task. The visual cortex consists