INFORMATION SOCIETY TECHNOLOGIES
(IST)
PROGRAMME

ECVision

European Research Network for
Cognitive AI-enabled Computer Vision Systems

Thematic Network:
Research Planning (WP1)
Ta1.7.6 Research Roadmap

IST 2001 Workprogramme
Section IV.2.1 – Real-Time Distributed Systems
Sub-section (ii) – Cognitive Vision systems

Project Acronym:	ECVision
Project Full Title:	European Research Network for Cognitive AI-enabled Computer Vision Systems
Date of Preparation:	13 February 2003

Table of Contents

Overview and summary

(Jim - Last thing to write)

The key things to address here are:

The objective of the document;
The reasons it is needed;
The people it is directed at;
How it is organized (and why it is organized this way);
The basis on which it was founded.
These will be written one the document is stable. However, here is a start.

Cognitive computer vision is an important emerging science that has the potential to (and, indeed, is probably a prerequisite for) the creation of a whole new class of technological and application development in fields where interaction between man and machine, and between machine and its environment, are needed. The goal of this Research road map is to provide the potential for impact of cognitive vision and to then identify the best way in which this potential can be realized. In particular, the document will set out an ontology of the emerging discipline of cognitive vision so that we can unambiguously discuss its constituent parts. It will then describe a set of application scenarios that illustrate the potential technologies that can be enabled by appropriate breakthroughs in cognitive vision.

1. Mapping the potential for innovation in Cognitive Vision

1.1 The emergence and growth of new technologies

Predictions about the emergence of new technologies are notoriously inaccurate. For one thing, innovation is a cumulative process, resulting in exponential changes. Humans are surprisingly poor at predicting exponential processes [Gladstone 00]. Human intuition tends to rely on linear extrapolations, resulting in predictions that are often over optimistic on the time scale of a year or two while pessimistic for scales beyond 5 years. More-over, the events that trigger a cumulative exponential process are unpredictable, difficult to discern and difficult to estimate.

A number of theories about the innovation process have emerged over the last few decades. These theories draw from the roots in econometrics and philosophy. Authors such as Mowery [Mowery and Rosenberg 98] and Utterback [Utterback96] have documented the emergence of a diverse array of technologies and synthesized theories that lead to general phenomena and predictable phases. This section reviews some of these phenomena and their possible sequences in order to establish an objective foundation for rational policies for triggering or accelerating innovation in cognitive vision.

1.1.1 The nature of innovation

In order to make statements about innovation we must have some objective properties with which to characterize innovation. The primary objective properties used to characterize innovation are performance metrics. These are properties that can be measured in an objective and reproducible manner independent of a specific technological implementation. An essential step in defining a technology is elucidating the performance metrics that can be used to characterize the technology. Without performance metrics it is impossible to make statements about a technology.

Innovations are ideas that permit improvements in performance. Such ideas may have many forms. In the following, we will describe several forms of innovations. We illustrate these forms using examples from the domain of electricity in the 19th century.

Innovations may have the form of an ontology of concepts that enable analysis. For example, statements about electricity required an ontology composed of such concepts as voltage, current, resistance, capacitance, and inductance. Each of these concepts is based (that is, grounded) on observable measurements (performance metrics). The emergence of this ontology, and the accompanying metrics, made it possible to reason and communicate about electric circuits.

Innovations may also have the form of theories that predict and explain phenomena. For example, Faraday discovered a relation between magnetism and electric charge. This relation made it possible to design a plethora of electrical devices. However, the theory could only be stated formally once the ontology and its measures were established.

Innovation may have the form of methods of analysis. For example, analysis of electrical circuits were made possible by Kirchoff's laws for current and voltage.

Analysis of communication systems was made possible by the introduction of Fourier domain analysis by Nyquist, Shannon, and others.

Innovations may have the form of the design for a device. Faraday's law made is possible to design devices that converted between kinetic energy and electric potential (motors and generators). Over the years, many new forms of electric motors and electric generators have been invented, refined and exploited. Each motor can be compared to others based on a common set of measurable properties such as efficiency or torque.

Innovations may have the form of systems architectures composed of multiple independent devices. Thomas Edison's creation of an electrical power distribution system was a revolutionary innovation that can be measured in terms of number of electrical devices constructed or total electric power generated .

Innovations often have the form of improvements to existing devices or architectures. George Westinghouse surplanted Thomas Edison by transmitting power using alternating current. The improvement in transmission efficiency led to important economic advantages.

In summary innovations may have the form of a proposal for or an improvement to an ontology, a theory, a method, a device, or a system. In order to map the domain of Cognitive Computer Vision, we must identify the ontology for the domain, and catalog the existing theories, methods, devices and systems.

1.1.2 The virtuous cycle of innovation.

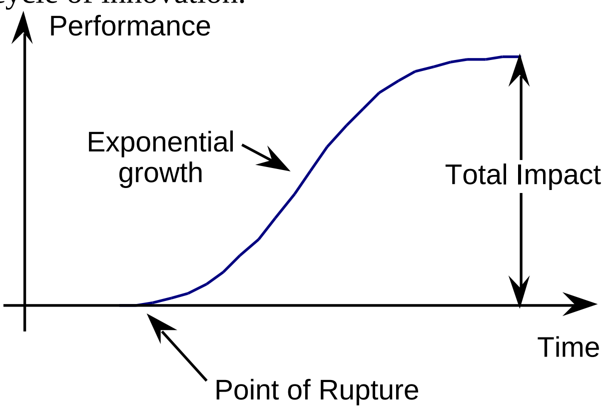

Figure 1. The S Curve can be described by three parameters: rupture , exponential growth and total impact.

The interaction of innovations tends to multiply performance. As a result, the accumulation of innovations leads to exponential growth in performance. The exponential nature of such growth in peformance has been widely documented throughout the history of human technology, and is widely known as the "S" curve, shown in figure 1 [Rogers 95], [Dent 93], .

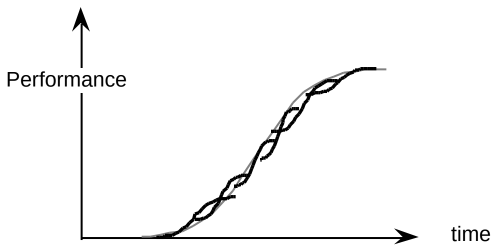

Figure 2. An S curve is the result of the composition of smaller S-curves.

The "S" curve is, in fact, a cumulative Gaussian distribution resulting for the accumulation of a multitude of interacting innovations. Each of these interacting innovations provides its own smaller S - curve, as shown in figure 2. The exponential growth in performance is the result of a virtuous cycle composed of the interaction of research, innovation and economic impact. In this spiral, the promise of economic gain leads to allocation of resources for research. Research produces innovation. Innovation triggers generation of wealth. The resulting wealth provides resources for research. All three components are fundamental. As long as the three components (or phases) of the cycle continue, innovations will accumulate and performance grows exponentially.

The virtuous spiral breaks when any of its three components attains a limit. For example, saturating the potential for generating economic benefit will result in a decline in the economic impact of innovations, resulting in a decline in resources for research, resulting in a decline in innovation. In some cases, the decline is due to a physical limit. For example, commercial air travel is currently restrained by the cost surpassing the sound barrier. While technologies exist for flying faster than sound, there is little economic incentive. As a result EADS, Boeing and their competitors concentrate on decreasing the cost of air travel.

The impact of a line of research can be described by the parameters of its accompanying S curve. For measuring past innovation the three most convenient parameters might be the growth rate, impact, and date.

The growth rate is the exponential coefficient that characterises the growth. For example, Gordon Moore has observed that transistor density in Integrated Circuits doubles every 12 months, leading to a double in the computing power on a chip every 18 months. An important measure of the impact of innovation is the discontinuous change in the exponential growth rate. Such a change is called "rupture". The increment in growth rate is an important measure of the impact of innovation or of a public policy decision.

The incremental growth rate is a direct result of the economic return on investment in research. Such growth rate is conditioned by a number of factors, including the propensity of the economic and administrative climate towards innovation. One property of innovation that interests us here might be called the "power" of the innovation. By this we means the effect that the innovation has on the efficiency of research. The term "power" corresponds to the coefficient of exponential growth. A "powerful" innovation enhances the efficiency of research on the design of devices and systems, thus leading to more rapid return on investment and a greater exponential growth.

The impact is the total growth in performance that can be achieved. For example, the introduction of jet engines led to improvements in the cruise speed of commercial aircraft from 300 to 600 knots. The duration of an S-curve is determined by the exponential coefficient and the impact. For a Cumulative Guassian, this is measured by the second moment (or standard deviation) of the derivative (the Gaussian density function). The impact is an important measure for choosing between competing public policy decisions.

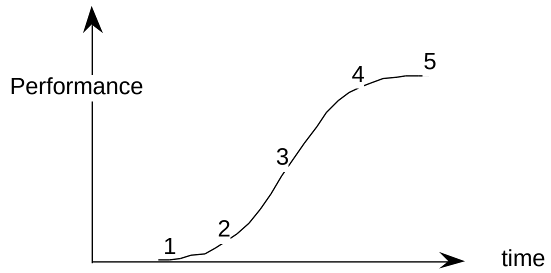

Figure 3. Phases of technology innovation: (1) rupture, (2) early development, (3) expansion, (4) maturation and (5) saturation

<u>1.1.3 The phases of innovation</u>

The S curve permits the process of technology innovation to be described in terms of 5 phases, as shown in figure 3. The phases tend to determine the nature of the innovation. For example, rupture tends to result from a new theory or method, often made possible by a new ontology or the emergence of an enabling technologies. Early development tends to be characterised by refinement to ontologies, theories and methods, and by innovation in the form of new devices or systems. Expansion is generally characterised by the emergence of a dominant design for devices or systems, and refinements to component devices or technologies. Maturation and saturation mark a period of diminishing returns on investment in refinements to devices or systems.

The above concepts allow us to make some simple recommendations about public policy on research funding. These recommendations are derived directly from the role that this can have in generating innovation.

a) Sustained exponential growth requires economic benefit. Except at times of national crisis (such as a military emergency), public funding can only maintain an S curve when accompanied by exponential growth in tax revenue. Investment in refinements to devices or system are rarely an effective use of public funds, and should be left to commercial investment. This implies that public funding should not be used in phases 3, 4, and 5 (of the S curve) of innovation.

b) An S curve requires perceived economic impact. Investors use research as a tool to multiply money. Such multiplication is only possible when investors perceive potential for economic impact. Demonstrating potential for economic impact in order to attract investment is a reasonable use of public funds. Investment in new devices or systems can help bring a technology from rupture to early expansion. Funding for work in phase 2 has impact when it when is serves to document, develop or demonstrate an innovation in phase 1.

c) The most powerful innovations are often accidental and a nearly impossible to predict. They are also hard to recognize As a result, innovations with the greatest potential for impact are in what might be called "fundamental science". Hence, public funding should be concentrated in phase 1. However, not all fundamental sciences have the same potential for impact and rupture.

We note that these prescriptions are perhaps too dogmatic. For example, attempting to build a demonstration of a new technology (phase 2) can often trigger the creation of new enabling innovation (phase 1) out of desperation.

The most powerful innovations are those that result in new theories and methods. Theories and methods enable the development of new more effective and more efficient ways of doing existing things and, more importantly, they enable the development of ways of doing new things. Such innovations require a proper ontological foundation. Public funding of the creation of an ontological foundation and the discovery of theories and invention of analytical methods is easily justified when then these activities enable a rupture in performance in a domain with substantial economic impact. The difficulty is that the time of rupture, and the increment in exponential growth rate are hard to predict.

1.2 The emergence cognitive vision as a subdomain of computer vision.

Computer vision is the science of machines that see. The goal of computer vision is to endow a machine with the ability to understand the structure, composition, and behaviour of its physical surroundings through the use of visual data. Over the last few years, exponential growth in computing power has provided inexpensive computing devices capable of processing images in real time. In parallel, rapid progress has been made in the development of theories and methods for using visual information to reconstruct and model the spatio-temporal world. The convergence of these two innovations has triggered significantly increased growth in the use of visual sensing. However, current growth rates are a small fraction of the potential growth rates for a technology of machines that see. Further more, the diversity and economic scale of potential applications offers the promise of a substantial impact.

This research roadmap documents innovations required to advance from a science of visual reconstruction to a science of visual perception where the machine can develop an understanding of its environment and thus behave adaptively, where it can reason about what it sees, anticipate events, and learn from its behaviour. Possibly, it can also interact with its environment and learn from that interaction. The question that arises is: what does it take to make this advance. In Section 2, we argue that the key missing component is a science of cognitive vision. We define the domain for this science from the viewpoints of cognitive science and of machine intelligence. From the view-points of these domains, we synthesize an ontology of fundamental concepts, and a glossary defining the technical vocabulary for cognitive vision. We then place cognitive vision within the structure of established scientific disciplines and delineate boundaries and relations with related domains.

Section 3 describes the technological and economic phenomena driving the emergence of cognitive vision. We survey the emergence of enabling technologies, and we describe applications domains that are currently attracting resources for science and technology.

Section 4 surveys some applications scenarios. For each survey, we identify fundamental scientific or technological problem whose solution would enable advancement. The production of these surveys is [has been] used as a tool to verify and complete the ontology as well as the list of enabling technologies. When [Now that] this analysis is complete, these surveys illustrate the ontology and need for the enabling technologies.

Section 5 summarizes the open fundamental problems identified in section 4 and describes the potential impact of progress.

The key question this document seeks to answer is: what should we be doing to cause the right ruptures (innovation phase 1) which will lead to strong take-up and exponential growth (innovation phases 2 and 3)? The answer is a set of fundamental research issues that complement existing computation vision in order for it to achieve the required innovative functionality. Ultimately, it is about creating new theories and methods, what Marr called computational theories, algorithmic formulations, and implementation realizations, that will allow us to do existing things in new more effective and more efficient and, more importantly, that will allow us to do new things altogether.

2. The Domain of Cognitive Vision

Some words from Goesta -

2.1 Cognitive Computer Vision

(Cognitive Computer Vision is the study of the acquisition and use of knowledge and reasoning in vision

1. Choose actions to accomplish goals
2. Use and generate explicit descriptions in terms that human use to communicate and reason.)

Cognitive vision is the study of the acquisition and use of knowledge through reasoning for computer vision. As mentioned above, cognitive vision is a subfield of computer vision. The aim of the Cognitive vision is to evolve computer vision from the science of visual reconstruction to a science of machines that see. Seeing requires abilities to use visual perception to know, understand and learn about an environment and, possibly, to facilitate adaptive and/or anticipatory interaction with that environment by the perceptual system. Thus, cognitive vision implies capabilities or functionalities for:

- **Knowledge representation** of events and structures. Ideally, the representation should exhibit some form of object or category invariance with respect to events and/or vision system behaviour. Many of the representations in the past have either been too application-specific to be adaptable to general (*i.e.* unanticipated) scenarios or environments or they have been too abstract to be applied in any meaningful way.

- **Learning**. There are two aspects to learning. First, there is *learning to see* (*i.e.* learning about the perceived visual environment.) Second, there is *learning to do* (*i.e.* learning to act and/or learning to achieve goals). It is an open question as to whether both are necessary in a fully-functional vision system.

- **Reasoning** about events and about structures (or some unified combination of both). One might distinguish three types of reasoning: reasoning to facilitate learning, reasoning to facilitate recognition/categorization, reasoning to facilitate hypothesis formation (*e.g.* 'what-if' scenario evaluation to aid system planning).

- **Recognition and categorization**, *i.e.*, mapping to previously-learned or *a priori* embedded knowledge bases. The majority of vision systems today rely on recognition or classification to effect their goals. This makes them inherently application-specific and quite difficult to re-use. Systems are able to recognize instances of particular objects or events (a car or type of car, a table or type of table) rather than being able to identify objects or events as instances of a meta-level concept of car (or road-vehicle) or table (or object-supporting platform). Cognitive vision systems would ideally have this categorizaton capability. It is, however, a difficult issue because objects of the same category can have completely different visual appearances.

- **Goal specification**, *i.e.*, identification of required system behaviour (this is the very essence of application development). Goal specification does not mean simply identifying the required information transformation from sense data to symbolic description – it may well include this but this in itself is inherently insufficient for a cognitive vision system which will typically have a number of often conflicting goals.

- **Focus of Attention** The abilities..
- **Context Awareness:** The ability to rec
- **Support Management:** Efficient retrieval from large spatio-temporal data sets. ??

Cognitive vision may also require some form of embodiment, autonomy, articulation, or agency but these are somewhat open questions of exactly the kind that ECVision will address.

Cognitive vision is a multi-disciplinary area, drawing for example on the disciplines of computer vision, artificial intelligence, cognitive science, and control and systems engineering. It is ultimately intended as a complement to conventional computational vision to facilitate a overall system functionality that is adaptive, anticipatory, and, possibly, interactive.

Cognitive vision draws on knowledge in many domains such as computer vision, perceptual psychology, psychophysics & neurobiology, artificial intelligence, cognitive science, autonomous systems, cybernetics. Concepts and capabilities from all of these fields are imported into the ontology that will be presented subsequently'.

2.2 The role of Ontology in defining a scientific domai

Ontology is the branch of philosophy that deals with being [Morris 69]. In philosophy of science, ontology refers to the irreducible conceptual elements in a logical construction of reality [Cao 97]. In other word, an ontology provides the set of logical elements from which theories and models are constructed. For example, in physics, rival ontologies propose particles, forces, or space-time as the fundamental elements of theories and models.

The theories and analytical methods of a domain are expressed in terms of an ontology. *An ontology provides a formal definition of the domain of discourse for a scientific discipline.* Without an ontology, there are no theories or analytical methods.

The possibility of rival ontologies is pivotal for a developing science. Unless a domain is mature, no ontology is perfect. Allowing rival ontologies in a given domain of discourse allows competing approachs to develop and to cross fertilize.

A research roadmap must allow for this multiplicity of ontologies and, hence, the roadmap must be presented at the level of a meta-ontology: not a single ontology of a cognitive vision system per se, but in the context of a class of ontological entitities or concepts. Thus we develop our ontology as a hierarchical structured glossary of concepts, theories, methods and representations.

Fundamental concepts

Jim rewrite and adapt from results of meeting.

The following provides an ontological foundation for cognitive computer vision. This ontology is a logical construction of concepts that allows us to make well defined statements about the problems and methods that define cognitive vision.

One may view these constructions as a hierarchy in which success levels are defined from earlier layers. It is possible to provide an even more formal definition of the fundamental terms, however, this would take us too far from the subject of this document. Thus we start with what we believe are terms that are unambiguous for any experienced engineer or scientist in informatics.

The most basic concepts are fundamental to any measurement system. From these we progressively move to concepts that are specific to vision and artificial intelligence. We begin with an assertion from Emmanuel Kant [critique of pure reason]. Their exists an unknowable and unique reality. Kant calls this reality "Noumea". Noumea is external and infinitely complex. It can not be completely "known", but only partially observed. Kant refers to the elementary perceptions of noumea as "phenomena". The task of any agent or machine that attempts to see is to estimate a partial description of noumea from partial and incomplete perceptions.

In modern terms we say that reality is "sensed" by a sensor (or transducer). A sensor produces a <u>measurement</u> based on reality. A <u>measurement</u> from the sensor is the basic input into the sensing system. The sensor samples a probability density function (noumea) to provide a value (a measurement).

A measurement is a kind of "observation". Observations are the outcomes of an "observational process" performed by an "observer". The observation is the "answer to a question to nature" that is posed by the sensor. The observation process can be based on measurements or on observations derived from measurements.

Predicates are conclusions from observations. They are the result of an observation process that produces a truth value. The truth value may be binary(t/f) or probabilistic [0,1] or any other form of belief function. The predicate is the result of the observation process (a function of observables, including measurements).

The concept of "State" poses problems because of different usages in neighboring domains. We would like to use "State" in the AI Planning sense as a logical proposition
defined over a set of observations. However, in estimation theory, a state is a characteristic vector of estimated observations.

The concept we need is an abstract metaclass that models a situation during which some invariant condition holds. We could refer to this as a "situation" to avoid the use of overloaded terms. A situation is a predicate defined over observations.

Time is a unique variable whose values are derived from an ordered directed sequence.
Time is what a clock measures. Time is a measurement provided by special sensor called a clock. Any observation may be associated with a time.

A point is an entity described by a vector of N observations (N is an integer). One of the variables may be time. A space is a set of points. A space can be ordered along each of the N component variables.

An entity (or model element) is an abstraction drawn from the system being modeled. Entities are associations (or correlations or agregates) of observations. Entities may represent objects, or concepts, or classes, or instances of classes. Entities may represent physical objects (This-pen) or phenomena (night). Entities allow the system to make propositions about the state of reality. Entities are defined by predicates (Boolean or probabilistic). The predicate may be defined by extension (by explicitly listing members of the class) or by intention (by an acceptance test).

A property of an entity is an observation associated with an entity. Relations are predicates (Boolean or probabilistic) defined over entities and their properties.

A Structure is a set of entities and their relations. A structure is a kind of situation.

An identity is an entity is value that is unique for each entity. An identity variable is produced by an identification function. For example, an identification function may return the same identity variable for an entity that has the same position in successive images. Identification functions may also be constructed based on complex predicate expressions. (The blue box with the red spot).

A category is an observable variable whose value is unique for a set of entities. The set may be closed (entities A, B, C) or open (green entities).

Recognition is the ability to determine the identity of an entity. Classification (or categorization) is the ability to determine if an entity belongs to a category.

Knowledge is the ability to predict or specify situations. Learning is the ability to acquire knowledge. Reasoning is the ability to generate knowledge from descriptions of other knowledge.

3. Technology Drivers

3.1 Enabling Technologies

(to be contributed by Rebecca)

Computing Power, Communications, Devices, Displays
-> Increasing demand and increasing offer.
Computing devices
Signal Processing
Optics
Networks

3.2 Recent Scientific Progress

(solicit contributes from Jon Olof Eklundh and bernd Neumann)

Evolutions (last decade) in Comptuer Vision (Eklundh?) and in AI. (Bernd Neumann) (Semantic Web), software architectures, ...
This should eventually lead to a survey paper for CVIU or another journal.

4. Applications and Potential Markets

This sequence of application scenarios are used both to develop and to illustrate the ontological concepts set out in Section 2. These scenarios highlight the theories and methods that must be developed. Mapping these scenarios onto the ontology defines the roadmap. These then are the fundamental research problems. Within each problem, one can identify (or focus on) the key issues, be they theoretical, algorithmic, or implementation/technological support.

(Each of the research dreamers should do the mapping onto the (meta-)ontology, identifying omissions in the ontology in the process. They should also then identify the key issues (theory, algorithm, implementation).)

For each section: up to 4 pages
What exists Today? (with motivations).
What is currently demanded?
What is the long term dream?
What are the fundamental research issues?

4.1 Robotics and Autonomous Systems

Markus Vincze (and colleagues)

A person's "best friend": It can perceive the mental state of a human to support her/him

Envisioned results of open research problems:
- recognise objects under all lighting and cluttered settings
- form relations between objects, reason about relations and consequences
- learn representations, learn to select a good representation, evolve representations
- learn expectations and strategies:
- automated learning of concepts ....

A system to aid society, a system that has a "broad understanding" of humanity,

4.2 Video Surveillance

(Rebecca, James, David?? - with jim)

The Vigilant Environment - A Research Dream
David Vernon

4.3 Man-machine interaction

Jim and Paolo

Interactive Cognitive Vision System

P.Bottoni, D. Fogli , C. Garbay, P. Mussio, P.Rizzo
an Interactive Cognitive Vision System which collaborates with its users in image interpretation tasks,
…adapting to cultural, social, and task-dependent contexts of usage.
1) Vision as a situated constructive activity
2) Vision as a grounding process

4.4 Smart environments and ambient intelligence

(David, and Jim)
What exists Today
What is currently demanded?
What are the fundamental research issues?

What are the future dreams

4.5 Mapping on demand

(Wolfgang)

User specifies a map and in real time gets a map drawn from multiple sources of information. Problem - How does the system specify what he needs.
What exists
Slow
Specification requires an experienced user. (Lack of interaction language)
No interpretation capabilities (neither GIS nor Image Sequ).
What is current demand
Demand will increase based on location based services (from GPS). and widespread introduction of information technology.
Long term perspective
Technology will be made ubiquitous (Homes, cars, public spaces, ...

4.6 Indexing Photo databases and Content analysis of images

Jan Mark Geusebroek

what exists today
Hand labeled annotation
indexing based on image appearance.

What is current demand
Subject and Category recognition (Automatic annotation) Image to text description.
Text to images. (Show me a picture of a tree).
Personalized indexing
Semantic Content based labeling and indexing.

Long term dreams

4.7 Film, TV and Entertainment

Adrian Hilton

what exists today
Hand labeled annotation
indexing based on image appearance.
Image sequence - Keyframe extgraction, shot segmentation, domain specific interpretation
What is current demand
Subject and Category recognition (Automatic annotation) Image to text description.
Text to image sequences. (Show me a video of a storm).
Personalized indexing
Semantic Content based indexing.
acticity learning and recognition
People tracking
Semantic labeling and segmentation for MPEG compression
Commercial Suppression
recognizing nice landscapes.
Long term dreams
Fundamental research issues

- action understanding
- Defining, learning, recognizing function from image sequences, Function based category definition. Learning the mapping from appearance to functional
- Defining, learning, recognizing human culture.
- Subject and Category recognition (Automatic annotation) Image to text description.
- Text to image sequences. (Show me a video of a storm).
- Learning, Representing, recognizing emotion in images and video scenes.
- Semantic Content based indexing. Representing and acquiring meaning.
- Visual cues for semantic content.

4.8 Aerial and Satellite Image Analysis

(James Ferryman).

4.9 Medical imaging and life sciences

(Patrick Courtney)

4.10 Industrial Inspection

?? Evaluate text from Patrick. May be deleted.

5. Fundamental Research Problems

The fundamental research problems should be drawn from the applications described in the previous section. However, the Cognitive Vision ontology group has done an excellent first cut at defining these issues. We take their description of Cognitive Vision Research Problems as a starting point for this section. To the extent possible, the research issues proposed in the previous section will be clustered around these problems. If necessary, we will define new categories of research problems.

5.1 Model Learning

Todays systems require the system designer to specify the model of reality that the system will use. This si a fundamental limitation (closed universe). To operate inan open universe, a system must be able to automatically acquire knowledge of 1 to 7 below. (Hillary or Gerhard)

1. Activity/Behaviors/Processes/Dynamics
2. Classification/Category (Approaches,Applications,General)
3. Context/Scenes/Situations
4. Function
5. Objects/Parts
6. Parameters(Approaches,Applications,General)
7. Task Control

Issues

1. Learning Control
2. Validation
3. Types of Learning

1.Case-based
2.Reinforcement
3.Supervised
4.Unsupervised

5.2 Knowledge Representation

5.3 Recognition, Categorization and Estimation

5.4 Reasoning about Structures and Events

5.5 Architecture and Visual Process Control

5.6 Interaction with the environment

5.7 Performance Evaluation

5.7.1 Evaluation of Methods and Algorithms

(James, Patrick Courtney, Adrian Clark...
Benchmarks
Performance Metrics

5.7.2 Self-monitoring, auto-description, auto-criticism, self diagnosis,

(Wolfgang Foerstner, Jim)
used for systems control

Towards conscious general vision systems

Jan-Mark Geusebroek and Arnold W.M. Smeulders

This research dream sketches the problem of content awareness in general vision systems.
We consider cognition as a highly distributed task…
..visual content … is laid down in invariant representations…
1) what kind of knowledge can be extracted from specific visual modules.
2) how to use knowledge to steer the focus of attention?

Towards an Intelligent Cognitive Vision Platform
Authors: Monique Thonnat and Francois Bremond

An intelligent cognitive vision platform is a scene understanding system with a full degree of autonomy… the platform …

1) must adapt itself in real-time. This feature is mandatory for robustness …
2) …must have scalable visualization tools…

6. Recommendations

What are the enabling technologies that may stimulate progress.

Annexes

Annex 1. A Glossary of terms for cognitive computer vision.

2.3 The ECVision Cognitive Vision Ontology

The contents of this section move to fundamental research issues.

(People directly involved in its development are: Bob Fisher, Wolfgang Förstner, Annett Faber and Hanns-Florian Schuster.)

Top level - Corresponds to different theory sets.

1. Model Learning (Survey Result)
2. Knowledge Representation (Survey Result)
3. Recognition, Categorization and Estimation (Survey Result)
4. Reasoning about Structures and Events (Survey Result)
5. Visual Process Control (Survey Result) & interaction with the environment
6. Emerging Topics'
7. Case Studies

The Hierarchy. (Need to find a more synthetic way to present this).

1.Model Learning (Survey Result)
  1.Specific approaches to learning these different types of content (See also Knowledge Representation->Content for "what" things that are learned and Recognition, Categorization and Estimation->Specific Approaches for "how" things might be recognized.)
    1.Activity/Behaviors/Processes/Dynamics
    2.Classification/Category (Approaches,Applications,General)
    3.Context/Scenes/Situations
    4.Function
    5.Objects/Parts
    6.Parameters(Approaches,Applications,General)
    7.Task Control
  2.Issues
    1.Learning Control
    2.Validation
  3.Types of Learning
    1.Case-based
    2.Reinforcement
    3.Supervised
    4.Unsupervised
2.Knowledge Representation (Survey Result)
  1.Content (See also Model Learning->Specific Approaches for learning different types of content and Recognition, Categorization and Estimation->Specific Approaches for "how" things might be recognized.)
    1.Activity/Behavior/Processes/Dynamics

        2.Classification/Category (Approaches,Applications,General)
        3.Context/Scene/Situations
        4.Function
        5.Objects/Parts
        6.Ontologies
        7.Parameters
        8.Task Control
    2.Issues
        1.Indexing
        2.Storage
    3.Style
        1.Appearance-based
        2.Embodied
        3.Generative
        4.Geometric
        5.Logical
        6.Ontological
        7.Probabilistic
        8.Procedural
        9.Relational/Graphical
3.Recognition, Categorization and Estimation (Survey Result)
    1.General Approaches
        1. Appearance
        2. Feature Sampling
        3. Geometric/Structural
        4. Physical Models
        5. Property
        6. Temporal (discrete or continuous)
    2.Issues
        1.Accuracy
        2.Generic Classes
        3.Labeling/Localization
    3.General Techniques
        1.Alignment
        2.Attention
        3.Search(Approaches,Applications,General)
        4.Figure/Ground
        5.Grouping/Perceptual Organization(Approaches,Applications,General)
        6.Labeling(Approaches,Applications,General)
        7.Parameter Estimation and Optimization
    4.Specific Approaches to recognizing things (See also Knowledge Representation->Content for "things" that are learned and Model Learning->Specific Approaches for "learning" different types of content).
        1.Activity/Behaviors/Processes/Dynamics (Approaches,Applications,General)
        2.Class/Category(Approaches,Applications,General)
        3.Context/Scenes/Situations
        4.Functions
        5.Objects/Parts(Approaches,Applications,General)
        6.Parameters

4. Reasoning about Structures and Events (Survey Result)
    1. Content
        1. Appearance/Visibility
        2. Objects & Spatial structures and their organistion
        3. Tasks/Goals
        4. Events & temporal structures and their organisation
    2. Issues
        1. Performance
        2. Prediction
        3. Self-analysis
        4. Uncertainty
    3. Methods
        1. Constraint Satisfaction
        2. Hypothesize and Verify
        3. Logical
        4. Model-based
        5. Rule-Based
        6. Statistical
5. Visual Process Control (Survey Result)
    1. Decision Making
        1. Probabilistic
        2. Rule Based
        3. Soft Control
    2. Issues
        1. Active Sensing
        2. Goal Specification
        3. Planning
        4. Process Control & Monitoring
        5. Speed of Response
    3. Paradigms
        1. Central/Distributed
        2. Covert Control
        3. Reactive
6. Emerging Topics (Survey Result)
    1. Vision & Language Fusion
7. Case Studies

A.2. Principal Research Groups in Cognitive Vision

<u>A2.1 European Union member states</u>

<u>A2.2 European Union associated states</u>

<u>A2.3 European Union associated states</u>

References

[Schumpeter 42] J. A. Schumpeter, Capitalism, Socialism and Democracy. New York: Harper & Brothers, 1942. Revised 2nd Edition, 1947. Enlarged 3rd edition, 1950.

[Utterback 94] J. M. Utterback , Mastering the Dynamics of Innovation, Harvard Business School Press, 1994.

[Mowery 98] D. C. Mowery and N. Rosenberg, "Paths of Innovation", Cambridge University Press, 1998.

Gladstone 00] M. Gladstone, "The tipping point: Or how little things can make a big difference.", New York: Little Brown and Company, 2000.

[Rogers 95] Rogers, E. M. Diffusion of innovations (4 th Edition). New York: The Free Press, 1995.

[Dent 93] Dent, H.. The Great Boom Ahead. Hyperion Press, Nov. 1993.

[Morris 69] American Heritage Dictionary, W Morris, Ed., Houghton Mifflin Co., 1969

[Cao 97] Cao, T. Y., Conceptual Developments in 20th centure field theories, Cambridge Univ.Press, 1997

D. C. Van Essen, C. H. Anderson, and D. J. Felleman, Information processing in the primate visual system: An integrated systems perspective, Science 255, 419-423, 1992.

L. J. M. Florack, Image Structure, PhD Thesis Utrecht University, 1991.

D. H. Foster and S. M. C. Nascimento. Relational colour constancy from invariant cone-excitation ratios.
Proc. R. Soc. London B 257, 115-121, 1994.

J. M. Geusebroek et al., Color invariance, IEEE Trans. Pattern Anal. Machine Intell. 23, 1338-1350, 2001.

L. Van Gool et al., Vision and Lie's approach to invariance, Image Vision Comput.. 13, 259-277, 1995.

D. Knill, W. T. Freeman, and W. S. Geisler (editors), Special issue JOSA-A on Bayesian and Statistical Approaches to Vision, American Optical Society, to appear, 2003.

J. J. Koenderink, The structure of images, Biol. Cybern. 50, 363-370, 1984.

D. Marr, Vision, Freeman and Co., 1983.

J. Mundy and A. Zisserman (editors), Geometric Invariance in Computer Vision, Springer-Verlag, 1992.

R. Nozick, Invariances: The Structure of the Objective World, Harvard University Press, 2002.

B. O'Shaughnessy, Consiousness and the World, Oxford University Press, 2002.

C. Schmid et al., Evaluation of interest point detectors, Int. J. Comput. Vision.. 37, 151-172, 2000.

A. W. M. Smeulders et al., Content based image retrieval at the end of the early years, IEEE Trans. Pattern Anal. Machine Intell. 22, 1349-1379, 2000.