European Research Network for
Cognitive Computer Vision Systems

ECVision White paper on industrial applications of cognitive vision

Authors: Patrick Courtney, Pia Böttcher

Date: 23. April 2003

Version: 1.2

Affiliation: ECVision Network of Excellence
contact: www.ecvision.info

Executive summary

A survey was carried out to determine existing and emerging applications of cognitive vision technology in a range of industry sectors: surveillance, industrial inspection, stock photo databases, industrial robotics, film TV and media, life science and aerospace. The relative role of a number of major characteristics of cognitive vision including learning, reasoning, recognition and goal orientation was outlined. Progress in a number of additional technical areas appear to be necessary to support development of new and improved applications within these sectors including user interfaces, system architectures, and design methodologies across the life cycle. There would appear to be a strong need for generic person detection.

Table of contents

1 Background

As part of the ECVision network of Excellence in Cognitive Vision, a workshop was held to identify existing and emerging applications of cognitive vision technology in a range of industry sectors. The objective was to collect this information as input into the research planning activities of the network. Eight individuals took part in the workshop held on 30th September 2002 covering aspects of surveillance, industrial inspection, stock photo databases, industrial robotics, Film TV and media, life science and aerospace. While this does not represent the full range of potential applications of Cognitive Vision, we believe that this white paper provides a useful industrial perspective.

2 Introduction

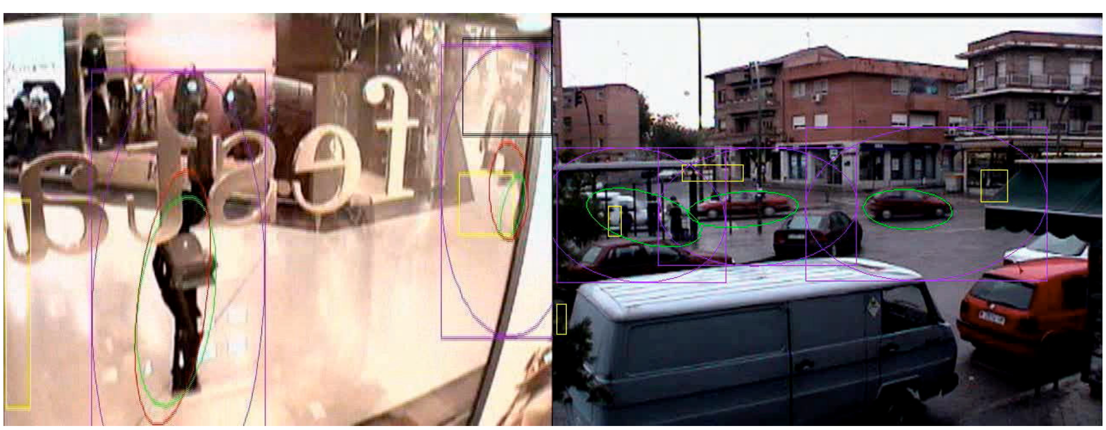
Image processing and analysis techniques have developed greatly in the past few years and are being successfully applied in a number of fields. Cognitive vision technology proposes to enhance the capabilities by the use of more sophisticated knowledge representations, learning, reasoning about events and structures, recognition and categorisation, goal specification and achievement, and so to create systems which are more robust, more flexible and generally smarter.

In the next section we survey emerging applications of Cognitive Vision as were considered in the workshop. In sections 4 and 5, we consider various trends and barriers that will have significant impact on the future application of Cognitive Vision technologies. Finally we make several recommendations about future development of these technologies.

3 Existing and emerging applications

3.1 Surveillance

Established applications in the surveillance and monitoring market include **object tracking**, **license plate recognition** and **video motion detection**, which are already widely used in the protection of property, traffic monitoring and similar. Face detection and recognition systems have also been installed and have been found to produce useful results in certain circumstances.



Picture1: Object tracking in people and traffic estimation

There is also a distinct need for surveillance systems that can collect and check business data in retail environments, which can be used for market research, to optimize operations, to investigate peak times and also assess suspicious situations. For this class of applications 100% automation is not required since competing techniques for counting people moving past a certain point or estimation of traffic (manual counting or magentic barriers) do not exceed 90% reliability and statistical data is sufficient.
Large numbers of cameras have already been installed for manual monitoring purposes. Watching the images directly is time consuming, and it is often difficult for the observer to concentrate over a long period of time. Alternatively these cameras could in principle form the basis of a cost-effective and robust solution. However the quality of the images provided is not always suitable. Changing illumination, occlusion, reflections, changing foreground and background conditions in the scene often make great demands on the image processing algorithms.

Emerging applications
The tasks still to be solved provide answers to questions such as: How many people are where, when? Where are they going? What are they doing? For these purposes it is also necessary not only to count people entering the scene but to **track them continuously**. They could then be tracked around the whole shop to analyse e.g. their buying behaviour. Further applications are related to understanding of the scene. An example would be the **automatic recognition of suspicious** or unusual behaviour (many people starting to move suddenly). Queue analysis can help schedule work.

One key requirement for a high volume market (5000-10000 units pa) is ease of installation. This would require some level of auto-configuration by learning of the background etc.

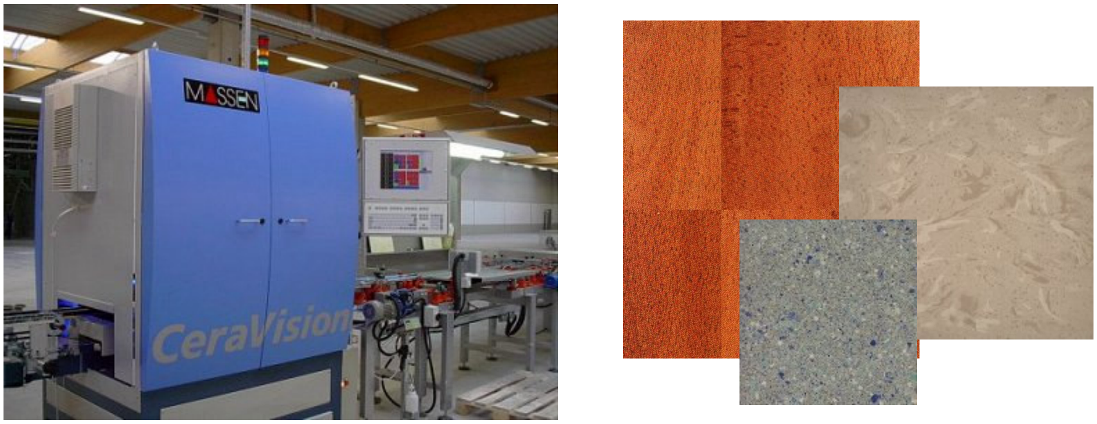
The lack of **baseline architectures** makes the design, construction and installation of robust system more time consuming than desired, which raises costs and reduces market take up. Some work has been carried out in this areas using CORBA in the Visorbase and Advisor projects. Multiple interacting modules could be used to improve performance, although simpler systems are often adequate and over-engineering is a potential problem.

3.2 Industrial Inspection

In recent years manufacturing industry has seen increasing levels of automation. Handling systems, automatic packaging and transportation systems are state of the art in many plants. One gap being closed is that of optical quality control. Some of the required tasks may be adequately solved by traditional image processing systems using physical measures such as size and shape. Sometimes however the quality being described is a more complex set of features such as colour stability or **aesthetic appearance**.

These problems occur for various kinds of products (textile, ceramic tiles, carpets, etc), in fact everywhere where a complex design has to be assessed. As in most other automation applications, the aim of this task is to simulate the human ability of adapting to a new job. This requires the emulation of the aesthetic assessment according to human pre-attentive perception.

The particular difficulties of the task are the tight time scales in the production lines and lean production, high production rates, limited number samples available to set up of the system, and unknown errors due to complex production methods. In addition to these more general challenges, surface inspection requires significant flexibility to cope with permanently changing designs due to demanding end markets. A typical system of this kind would be the Ceravision system.



Picture2: CeraVision® installation, sample material

Emerging applications

The ability to learn from a very small sample remains an important characteristic as batch sizes get small, requiring powerful **generalisation** ability.

An increasingly important issue is the 'extraction' of the knowledge of the quality expert who knows intuitively what is a good sample for a range of classification grades, and who knows the production context. New **interactive user interfaces** capable of capturing this knowledge have to be found and implemented.

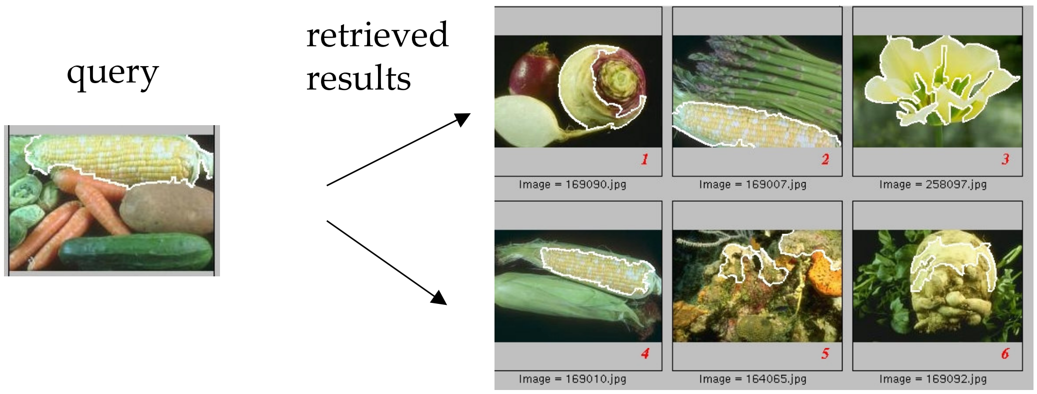
Finally the results should be used to establish a **closed loop control** to actually improve the quality of the production, where traditional PID-type control is inadequate for modelling complex non-linear effects. This is only possible on the condition that the images are fully understood in the context of the particular production process.

3.3 Stock Photo databases

The selection of photos for magazine and newspaper articles and advertisements is an important part of the work of many journalists and graphic designers. These photos principally serve to illustrate, or to draw attention to an article. Several independent studies have indicated that journalists usually (70%) require a picture with specific semantics (people, entities, events or locations) or occasionally (30%) general semantics (e.g. interesting or novel). A huge number of photos already exist in internal and external archives, and the stock is increasing all the time, but the main problem is to locate a suitable picture to meet a tight deadline.

Current retrieval technology is based on databases of manually labelled images. This entails numerous problems: it is time consuming to annotate photos (7-40 minutes per photo); the individuals performing the annotation are typically not the same individuals who benefit from enhancements to the annotation process; the metadata often provides incomplete or inconsistent descriptions since a single photo may have multiple interpretations and the annotation vocabulary changes over time.

Computer vision research has developed 'Content Based Image Retrieval' systems such as QBIC, Virage and Excalibur. These operate on presented example images to extract low-level image features such as colour, texture, shape descriptors and histograms. Typically these systems compute similarity measures in terms of these features, to images in the database, thus presenting ranked retrieval results. It was initially hoped that such systems could meet the requirements of typical image retrieval scenarios. However, to date, this approach has only provided satisfactory results in specific niche applications (textile patterns, logos, stamp collecting).



Picture3: Content based image retrieval system

Emerging applications

The next generation of image retrieval systems could benefit from **semantic annotation or search** options in order to better meet the requirements of realistic image retrieval settings. These would allow labeling of photos using special purpose categorizers: for instance, to detect specific people, locations, activities and events, or to categorize on the basis of emotive content. Recent progress on generic methods for detecting faces and people from a variety of views in a variety of poses and may be a significant step in this direction.

There is also significant industrial interest in addressing the broad range of user needs in the **management of consumer photo databases**. Emerging applications include automatic methods for "annotation propagation," album creation and image illustration of text.

Finally more realistic **benchmarks** that go beyond simple COREL datasets would be useful.

3.4 Industrial Robotics

Automation is now commonplace in nearly every industry, e.g. car manufacturing, food industry, packaging. An important concern is to achieve the benefits of automation without giving up too much flexibility. As a consequence hard automation is not appropriate because of variation of the objects and of changing tasks.

The primary goal of an increased level of automation is not simply to replace personnel with industrial robots. Practice has clearly shown that particularly monotonous and un-ergonomic tasks are error-prone when performed by humans. The substitution of these tasks by automated systems leads to increased process repeatability and product quality.

One can consider a '**Robot Zone'** that has been extending across a range from hard automation where robot use is optimal to situations requiring significant flexibility but where for example, a high degree of repeatability or cheap labour is not available. One way of improving flexibility is to provide robotic systems with a means of sensing their surrounding. Packaging of consumer goods is a significant area where the object shape cannot be controlled for the benefit of the robot, but fast and cost effective handling is required.

Existing systems using vision therefore include those employing fast (flyby) **2D recognition** capable of determining size and location of a known object type. Often a network of cameras is used to control the handling systems on a production line. The set-up of the tasks can be done by integrators with a graphical user interface where ease of use is an important factor for moderate volume markets (2000 units pa).

Picture 4: ABB PickMaster

Emerging applications
The main issues for further applications in robotics include improving the sensing ability of the systems by expanding 2D recognition to **2,5D or full 3D** capability to deal with more difficult and varied objects.

Complex non-linear closed loop control is important for improved flexibility, allowing tracking and correction of complex moving and misaligned components. Mechanisms that support **safe interaction** between human and robots are also becoming of increasing importance.

An especially challenging future application is the use of **mobile robots** with a further level of flexibility (moving) featuring real time task planning and decision making, knowledge acquisition, and easy to use expert systems.

3.5 Film, TV and Entertainment

3.5.1 Content analysis of images

The management of the huge amount of broadcast material currently being produced would benefit from **automatic indexing and retrieval** systems, able to analyse and identify the content of cinema film and TV. Typical applications needs include searching an interesting scene in a news program and filter out the 'uninteresting' sequences of the studio images. Existing systems are capable of detecting shots and extracting key frames using syntactic edge information and colour histograms, and limited face recognition. Players include Tecmath/Blue Order, Virage and a few research systems. They are useful in a limited range of situations in management and production.

Another important area of application is in sports event where a viewer may wish to **obtain further information** about the event - for example in a football match the distance of a player to the goal or the precise position of the offside line. Current systems like VIZ-RT (D) and Epis FAST from Symah Vision (F) offer such functionality on a manual basis needing a manual calibration step without awareness of the different actors (ball, players, referees) and or using footage from fixed cameras.

Picture5: Automatic player/ball tracking

Emerging applications

To provide a **more complex statistics** such as ball possession by a team or individual player, number of goals and fouls for fans and clubs, new approaches are necessary. Another very important aspect, especially in sport programs is **advertisement monitoring**. The question how often and how long a brand is visible and at which size and quality is of major interest for all advertising companies. In the DETECT project, both automatic ball and player tracking and brand monitoring is being realised.

Other needs emerge with **interactive television** where, a viewer may click on a player or actor and be provided with the name and additional information such as web-pages, advertising and products, which requires person and object tracking and identification.

3.5.2 Modification of images

Often there is the need or wish not only to analyse and interpret an image but to modify it. Again the technology is driven by the advertising industry. To show the right adverts to the right group of people is essential in this business. Targeted advertisements can be **inserted on-line** into broadcast sports program and this is already available using systems such as Symah Vision and Viz-RT.

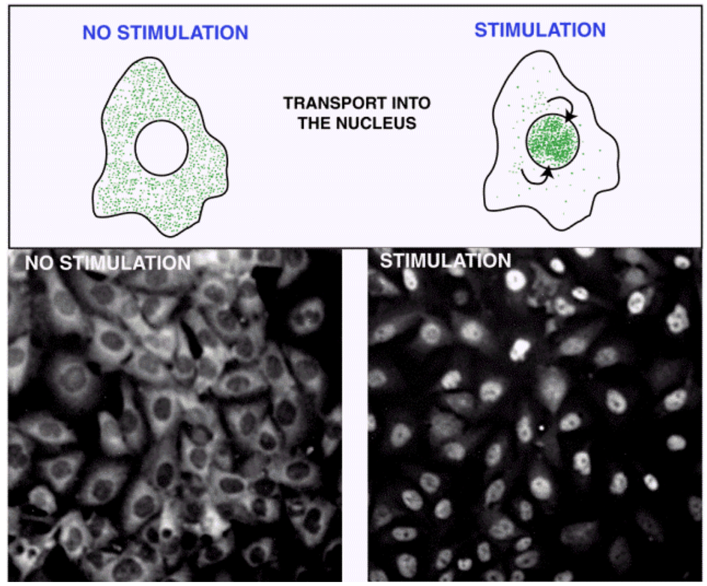
Well-established applications exist in the **digital restoration** of damaged film material. Image processing algorithms are used to remove scratches, flicker or damages caused by mould and dust. Most of these algorithms have to work motion-compensated to avoid introduction of digital artefacts. The quality is directly related to the quality of the estimated motion. Typical software based products are Diamant, Restore and MTI, hardware based products are e.g. Archangel.

Emerging applications

The idea of digital cinema restoration can be further developed. **Automatic film coloration** and **motion compensated** conversion into different formats (e.g. from interlaced to progressive or HD to SD) are desirable. Both of this applications need high-quality object based motion information also clearly indication occluded regions. **Virtual re-shoots** by digitally 'reconstructing' actors from partly destroyed material and combining them with three-dimensionally modelled backgrounds would be a very useful, albeit ambitious application.

3.6 Life sciences

The discovery and development of new drugs is a long and complicated process with various stages of tests. The pharmaceutical and biotechnology industries already make extensive use of images in the latter stages of clinical trials (MRI, X-ray/CT) that are examined by medically trained staff. Increasingly images are being analysed in the earlier stages of the process. Drug development frequently makes use of animal models of disease – for example in a study to examine the effects of an anti-anxiety drug, the **behaviour of rats** may be observed. Existing systems, such as those from Noldus, are able to track the animal in an image sequence and record its positions, even if it's shape varies considerably during the movement. Manual annotation is required to link the movement to specific behaviours. Another application area is in the **study of living cells**. Cells change their appearance over time and after being stimulated or exposed to substances. Workstations supporting specific application algorithms are now appearing eg Arrayscan and Kineticscan from Cellomics.
Another of the challenges of the task is to be able to link changes in cells and animal behaviour to information in the expanding **genetic and other databases** and thus obtain new insights.



Picture6: Cytoplasm to Nucleus Translocation

Emerging applications
Although image processing is used to extract data from a scene, the interpretation of the images is mostly still in the hands of a skilled human observer with clear limitations in terms of repeatability, experiment duration and cost where **skilled staff are limited**. To automate this time consuming task an understanding of the scene is needed. The movement of the animal, a rat for example, has to be **identified as posture** such rat grooming, as well as **novel postures** that might otherwise be missed. **Social interactions** are also difficult using simple tracking systems.

A final problem is the developed of **user interfaces** that allow interaction with skilled staff and to present data in a form which is understandable in biological terms.

3.7 Aerospace

One of the main goals in the area of aerospace is safer flying systems. Whilst the technical feasibility of vision guided control has been demonstrated, there are no operating systems in civil aviation. The reliability requirements of such systems are extremely high. The SLATS project (**Safer Landing and Taxiing**) underway with support from the DTI, in collaboration with the Civil Aviation Authority, aims to build the safety case for the use of a dual redundant (vision and GPS) system by providing better aircraft localisation, and so to improve civil aircraft approach, to decrease runway occupancies and to reduce incidents that damage landing gear.

Picture 7: Taxiing incidents

Emerging applications

In the military, long-term (10 years) development programmes exist to demonstrate these capabilities to meet the new defence needs of the 21st century. The direction of development is towards greater autonomy, decentralised intelligent systems and decentralised decision-making with sensing in an unknown environment.

The US Department of Defence has developed a four element reference architecture for data fusion which includes picture compilation (**what can I see?**), situation assessment (**where am I?**), threat classification (**what does it mean?**) and resource allocation (**what should I do about it?**). This may fulfil some of the requirement for cognitive vision. However the system-building aspects and the lack of well-characterised modules remain challenges at present. In particular the ontologies underlying module interaction and the subsequent negotiation have yet to be fully explored.

Research systems are in development to carry out constrained tasks with some civil applicability – for example exploring a single storey building and building a map – for disaster recovery.

4 Trends

The wide variety of existing and emerging applications covered in the examples above reflects the general situation. Trends of all aspects of our life influence the development of computer vision systems: political, environmental, social, technological and economic.

Technological

Probably the most obvious trend is emerging in the tremendous amount of new, smart, and affordable technologies in the consumer market. For example cheap cameras, mobile devices and the mass take-up of consumer digital photography creates new demands for handling and archiving the generated images (surveillance, film/TV, stock photo). New technology generates as well new challenges in industrial applications. Advanced systems, able to gather huge amounts of data force one to think of new ways to deal with the data. Automatic approaches have to be found to cope with the increasing flood of information. This trend can be observed in various situations such as the huge amount of data in archives for photo and film, as well as in the life sciences (increasing amount of genome data to be correlated)

Economic

The economic impact of new trends should not be underestimated. Companies are facing the pressure of price competition, which forces them to reduce overheads wherever possible. Companies tend to solve this problem by introducing new production methods. Lean production for example means no storage and no redundancy in their operations. This leads to higher demands in reliability for the technology employed. The time component is another constrain in this competitive arena; 'Just-in-Time' production allows no long gaps to install, set up or adjust automation systems. The situation is aggravated by the lack of skilled staff to operate complex and complicated machinery.

Environmental

On the other hand it is understood that employees work best, when they are feeling well. Hazardous, unergonomic or boring activities are being avoided or substituted by automatic systems as far as possible or feasible. The production times are increasingly 24 hours 7 days a week. Some jobs are critical especially at times humans are not at their peak of performance (e.g. during night time).

Social

One of the significant societal trends has been increased leisure time with a demand for sophisticated entertainment. This is where interactive television, the internet and mobile communications are all seen as playing a role. The question is what do people really want and what is the technology at the moment capable of offering. The internet especially provides a platform with unsuspected possibilities. The availability and distribution of online data needs new approaches. Increased leisure time also means more travel and increased demands of limited capacity within aerospace systems. Meanwhile an aging population places more demands on new medical approaches and pharmaceuticals.

Political / Organisational

The political response to changes in crime and security has included the deployment of huge numbers of surveillance cameras with substantial local support, at least in some countries. At a global level, changes in international relations have resulted in the need for completely new military systems, compared to 10-20 years ago.

5 Barriers and R&D Challenges

5.1 Demands of cognitive vision

An attempt is made here to tabulate some of the specific capabilities identified as being required, in each of the general cognitive vision functions. The function most in demand for each sector is represented by marking that **sector name** in bold.

Cognitive vision function	required capability	application sector
Learning	• Generalising from a small set (good example/no bad examples) • Linking low level features to higher level rules • Learning 'faces' or similar objects	**Surveillance** **Industrial inspection** Film/TV/entertainment **Life Science**
Reasoning about Events and structures	• Complex non-linear closed loop control • Understand high level 'rules of the game'	Industrial inspection Stock photo databases **Film/TV/entertainment** **Industrial robotics** **Aerospace**
Recognition and categorisation	• Categorisation for classes that are poorly defined at start • Data driven categories (unsupervised)	Surveillance **Stock photo databases** Film/TV/entertainment Life Science Aerospace
goal specification and achievement	• Capture user ontology • User interface • External knowledge sources • Relevance feedback	Industrial inspection Stock photo databases Life Science Industrial robotics Aerospace
Knowledge representation	Illumination/view normalisation	**All sectors**

Although several sectors are represented in each function, some appear more significant than others. In particular, learning and reasoning appear most in demand at this time.

5.2 Additional challenges

Having accepted that the capabilities that cognitive vision techniques would bring to the wide variety of application areas described earlier, some additional common issues emerge:

- better communication between the user and the systems
- people detection, human tracking and recognition
- system design
- clear system architectures

User Interfaces

The use of a fully automated system without human interaction is in many applications neither possible nor sensible at this time. The systems discussed in this paper are not dealing with easy tasks where it is enough to choose an appropriate threshold. Decisions often have to be made in a complex scenario according to a certain context in a certain situation. It is therefore often necessary to include a **human expert** in the loop of data acquisition and decision making.

Objects under study maybe the behaviour of an animal or a person, or the description of an aesthetic feature that cannot easily be described with traditional input elements, such as numbers or sliders. Complex relationships have to be presented in an easy, readily understandable manner using a language the users are familiar with and based on suitable onotologies and a common vocabulary. In addition, knowledge that may be 'extracted' from an operator will be essential in many applications. This aspect was found in all areas discussed.

To get all the additional information beside the image data, interfaces to **external information sources** may also be required. The user interfaces have to cope with such challenges.

People detection module

One common technical requirement that could be identified across the different applications was the observation and tracking of people. This requirement was found to be common to surveillance, stock photo, TV/Film/Media, robotics and (in modified form) to the life science areas. To understand the contents of an image it is necessary to detect people, what they are doing and what kind of activities are carried out. This is especially challenging in the analysis of still images. Such a module would have many forms depending on data sources (colour/monochrome, fixed/moving cameras, resolution range) and conditions of use (indoor/outdoor, etc).

System design

In the early years of image processing a lot of mistakes were made in particular with the installation of reliable systems. This led to a lack of confidence in this technology by many users for a considerable length of time. The challenge is to create a system that is easy to install, easy to use and easy to maintain. In particular the ability to Set Up/Teach the system for a new task in a very short time using only a few samples (surveillance, inspection, etc) is of great importance. This requires considering not only the system during normal operation, but the entire life cycle of the system – from design, installation, testing, to use and maintenance. The techniques to carry this out are still not well developed.

Architecture

A baseline system architecture of an abstract description of the task to be performed is still missing. The reuse of specific modules via standardised interfaces would reduce the cost of deployment and increase the potential market. This is especially important when trying to integrate off the shelf components and/or customising systems to particular needs. It goes hand in hand with suitable system design protocols. This aspect was found to be common to the surveillance, TV/film/media and aerospace areas.

6 Recommendations

In summary, the following recommendations appeared as a result of the discussions:

- ⇨ Cognitive vision techniques will bring benefits to broad range of application areas and in particular, learning and reasoning.
- ⇨ At the current state of development suitable end user interfaces must be considered
- ⇨ A generic person detection module would be of benefit in many areas
- ⇨ A baseline system architecture would be of benefit
- ⇨ System design methodologies need to be developed and the system life cycle considered

Many of the currently discussed cognitive vision applications focus on the consumer as the end user. It is worth pointing out that there are many business and professional applications that would benefit from the technology. They are likely to be less cost sensitive and more rational in their requirements compared to consumer needs. However they have significantly different ease of use and reliability requirements, and the organisational business processes remain as barriers to be understood.

7 Acknowledgements

The authors wish to thank Paco Gomez (Visual Tools), Chris Dance (Xerox), Henrik Knobel (ABB), Georg Thallinger (Joanneum Research), Walter Plaschzug (HS-ART) and Gareth Rees (BAE systems) for their contribution and for permission to include images. We gratefully acknowledge the support of David Vernon and the executive committee of the ECVision project, as well as the financial support from the IST programme of the Commission of the European Communities.