Chapter 1

* Introduction

1.1 I - Vision classics: automation

Computer Vision is a relatively new field. It covers a wide range of subtopics, but their unifying feature is that computers are dealing with information in the form of an image. Besides offering the authors a living, Computer Vision has a few other things going for it. After all, vision is the sense we depend on most. More than any other kind of sensory information, vision provides the capability to emulate human abilities like learning, understanding, and adapting to the task environment and its unpredictable variations.

The first real market of computer vision was automation. In particular, vision systems were used for visual inspection. Vision systems also guide robots and production equipment. Here are some good reasons why computer vision quickly became successful in this arena:

increased productivity , as computer vision systems can operate 24 hours a day, without fatigue and with better predictability in their grading criteria and processing time, which in turn helps to optimize system control and reduce cycle times. The vision system can also send its information directly to a central computer for process steering,

increased reliability , again because of the better constancy of grading criteria, but also because computer vision often allows 100% rather than sampled inspection and because of the system's unabated 'attention' even at the dullest of jobs, in contrast to the boredom quickly felt by humans performing such tasks. Monday mornings and friday afternoons are periods the quality control department especially fears when working with human inspectors.

increased flexibility , since "seeing" machines are able to adapt to a wider range of changes in their environment. This capacity is of particular importance in batch production where part processing can automatically adapt to the type of part that passes by. The vision system can recognize the part, locate it, and load the correct

programme. Visual feedback can also make up for a lack of precision of the manipulation or processing equipment. If, for instance, a robot is not very precise in its repetitive motion, a vision system can correct for the deviations from the target position.

reduced cost According to the last point, the visual feedback allows one to work with e.g. a cheaper robot, and cheaper positioning and manipulation equipment in general. Moreover. automated visual inspection can reduce labour cost as well as required floor space, as the vision system can be integrated into the machine. Also some equipment typical of 'hard automation' can be eliminated or simplified, such as pallets, jigs, and feeders. A vision system may detect a workpiece's position and orientation on a conveyor, so that these parameters need not be controlled. Computer vision also helps to reduce waste material through on-line optimisation of tool trajectories.

Computer Vision applications related to automation (mainly visual inspection) are often referred to as 'Machine Vision' applications. In 1992 the European market for machine vision systems was estimated at more than 7500 systems per year. This figure exceeded 14000 systems in 1996. Here a quote from the sept. '99 issue of the Journal 'Image Processing': 'The world market for machine vision in 1997 reached nearly 4 B$, of which North America held a share of 4025to grow by some 5

1.2 I - A bright future

In the meantime, computer vision has found applications far beyond visual inpsection and automation. Medical image processing has become a major field of its own. The possibility to look inside the body with different types of scanners, has opened a whole gamut of possibilities. Computer vision is used to integrate the information coming from different sources and to ease the interpretation of the data. Computer vision also is much more precise in taking quantitative measurements than human vision. Humans, on the other hand, are far better in making qualitative judgements, like finding a region of interest in an X-ray. The physician can then interactively indicate such region, e.g. an artery that shows dangerous constriction, whereupon the comnputer measures the exact width.

Image processing also became a crucial tool for remote sensing applications such as gathering land use / land cover statistics, assessing the risk of environmental hazards, selecting prospection areas for oil, diamond, and gold mining, etc. As ever more satellite are launched, with each an increasing number of sensors on board, which in turn are capturing data in more and more spectral bands and under more polarisation states, the only way to keep track of all these data is automated storage, retrieval, and processing.

But by far the largest upcoming area of computer vision applications are those related to cummunications in a broad sense. These areas include video-conferencing, vision for games and special effects in the movies, collaboration over large distances via virtual and augmented reality, intelligent human-machine interfaces, the creation of avatars for the Internet, etc. Currently, much work is going into computers that recognize people's gestures

and facial expressions; into programs that go out on the Internet and come back with the kind of images you have asked for; into 3D reconstructions from simple video sequences; into...

These developments will be propelled by an important development that is bound to happen over the coming decade. The 80s saw the widespread use of small and cheap processors, with the PC as only one of the spin-off products. With a bit of exaggeration one could claim that it cause the proliferation of more 'intelligence' into all kinds of machine, including household appliances. The 90s witnessed a similar development, as small and cheap lasers became available. Through optical transmission, this caused boom in high bandwith communication and laid a necessary foundation for the development of the Internet. The first decade of the 3rd millenium is poised to bring small and cheap sensors. Cameras will be among them. In fact this trend is already visible, think of webcams. But this is only the start, cameras will be everywhere! And it goes without saying that the images will need to be analysed and interpreted by computers to make senseful use of them. And just as the Internet is difficult to predict as the outcome of new laser technology, we may only guess what kind of mindblowing ideas will emerge.

1.3 I - A day out of the life of Computer Vision

Throughout all these exciting applications, the basic tasks vision systems have to solve boil down to a combination of one or more of the following:

1. Inspection
2. Positioning, registration, metrology
3. Assembly, navigation, and visual control
4. OCR, document processing
5. Scene reconstruction, visualization, and editing
6. Image compression
7. Image enhancement and restoration
8. Pattern, object, and event recognition
9. Human motion / gesture / face recognition and interpretation

Inspection has been dealt with in section 1.1, so we can keep this discussion short. As said, this is the oldest application of computer vision, but still equally relevant. As the speed of hardware has kept on increasing, new applications continuously become feasible.

Positioning, registration, metrology is an item that bundles applications where the precise localisation of objects or their parts is involved. This includes precise mensuration of

metric dimensions of objects or their parts. In this context the art of 'sub-picel' precision measurements has evolved, i.e. getting these dimensions with a better precision than the size of the spatial samples taken by the camera. Another issue covered by this class is the calibration of the camera itself, i.e. the determination where the camera is with respect to the scene and what its internal parameters such as focal length and pixel size are.

Assembly, navigation, and visual control are applications where a vision system guides equipment in order to position itself correctly with respect to workpieces or the environment. Examples are visually guided grasping – in a stable grip! – for assembly operations or delivering the necessary input for path planning in moblie robot navigation. Tracking of surrounding cars and road markings has in the meantime lead to the first demonstrations of autonomous cars, albeit only on highways.

OCR and document processing have the goal of automatically reading and analysing documents. OCR or 'Optical Character Recognition' is about recognising the different characters in a text on paper, from an image. Obviously, just as with speech recognition OCR systems exploit context information, i.e. characters as part of words, words as part of sentences, etc. Especially the recognition of handwritten texts remains a challenge. A somewhat separate domain is signature recognition and verification. Document processing also aims at the automated extraction of a text's layout: how many columns does it have, where are the different lines, what are figures and are these photographs, tables or graphics, where is the title, the author's name, the publisher, the abstract or in the case of forms where the different fields are that have to be filled out. Alos bar and area code detection and reading belong to this class of applications.

Scene reconstruction, visualization, and editing quickly gain importance as vision applications. Games, movies, training simulators, virtual reality demonstrators, e-commerce sites all require high quality models of scenes and the objects therein. Often these are three-dimensional models. Computer vision has developed several methods for the extraction of these 3D model, with and without the surface texture. Also, tools have been developed for the automatic registration – i.e. fitting – of multiple, partial 2D and 3D models, as well as for the fitting of different kinds of parametric models, such as wireframes, splines, superquadrics, etc. Also have texture model been generated that can simulate surface structures, so-called textures, and that are learned from example images.

Image compression aims at representing image information using as few bits as possible, yet with a minimal loss of information or image quality. It is important for image transmission, as well as storage in and fast retrieval from large image databases. With new technologies such as HDTV and videotelephony ever increasing streams of image data have to be sent at high rates. Channels have limited bandwith that should be used as efficiently as possible. Compression for storage is e.g. important in hospitals, where large amounts of images from all sorts of scans are to be stored for later diagnosis. In typical medical applications, loss of information is even considered unacceptable and the compression should allow the reconstruction of the original image up to the last bit.

Image enhancement and restoration deal with the improvement of image quality. Restoration implies undoing a detoriation that an image has undergone, e.g. of a surveillance image that is obtained from a worn out video tape or the removal of scratches on an old movie.

Enhancement is a more general term in that it includes the improvement of the visual quality of images, but also of the saliency of the aspects or features in an image that are of particular importance. As an example, if a physician wants to interpret a radiography of the lungs, he would like to have good contrast, clear edges (e.g. the ribs), etc. Image processing techniques will help to make the information that is available in the image more accessible to human vision. Surveillance and forensics are other domains where image enhancement and restoration are key in making good use of the available data. If you want to identify an offendor, it is often necessary to enhance the images in order to better notice details on his/her cloothing. Similar requirements hold in astronomy, geology, non-destructive inspection, etc.

Pattern, object, and event recognition are essential components of many applications. The goal is to identify specific objects or object classes, or similarly, specific events or event classes. Think of the distinction between recognizing a specific building vs. detecting *a* building. It has to be admitted that visual recognition still is in its infancy, and all the more so if not a specific instance but the class to which it belongs has to be recognised. The class recognition problem is usually referred to as 'categorisation' rather than 'recognition'. The main problem is the variation in appearance that objects or events may have in the image. One may look at an object from different directions and under different illumination conditions. The corresponding images may be very different. In the case of categorisation the variations between the shapes and textures of the members of the class add to the problem.

Human motion / gesture / face recognition and interpretation could be considered a subclass from the previous item. Nevertheless, analysing the shape and motions of the human body is of particular importance. One can for instance learn how the realistically animate virtual characters, still one of the big bottlenecks in the production of animation movies. Also, by interpreting the actions and gestures of users, human-machne interfaces can be made more intelligent. There also is a keen interest in face recognition, for verification purposes for instance. Similarly, the corresponding categorisation problems are interesting, e.g. to detect faces / people in images for surveillance applications.

As will have become clear from this diverse list of vision topics, an interesting feature of computer vision is that it builds on the latest in various areas of technology, mathematics and physics. There is the use of state-of-the-art processors and processor configuations as vision still is one of the computationally most demanding applications. Likewise there are demanding applications that demand the utmost from cameras and scanners, and from optics. Moreover, vision is one of the most demanding applications of signal processing theory, geometry, statistics, optimization theory, etc. This interesting mix has lead computer vision to be a major force behind several theoretical and practical innovations. Think of wavelets, fractal, neural nets, support vector machines, independen component analysis, invariants, regularisation, etc., etc. To be good for vision, methods need not only be powerful and sophisticated, but also robust against noise and all kinds of unpredictable disturbances such as occlusions, changes in illumination, etc.

1.4 I - Challenges

Half our brain is devoted to the processing and interpretation of visual data. No wonder then that human vision still is no match for computer vision. There still are many visual tasks that we can easily do, but that are beyond the reach of computer vision systems. Here are a few nuts to track to move forward:

- *Cue integration:* So far, the emphasis in conmputer vision research has been on the development of powerful and robust algorithms for the extraction of single cues. Such cues are motion, stereo depth, colour, texture, etc. It is time now to integrate the extraction of different cues, such that further processing can exploit a richer set of basic features. For the moment, vision systems can mostly deal with scenes that are controlled in terms of their content or illumination. The combined use of several cues seems a prerequisite to deal with cluttered scenes.

- *Dynamic 3D:* Excellent 3D acquisition are available by now, but the world is 4D rather than 3D, in the senee that many objects also change their shapes over time. Think of a tree blowing in the wind, a steel plate being extruded, or the expressions on a human face. Ideally, such changes could be measured in 3D. The impact on top notch animation will be huge. Additional inspection tasks can be solved. Revalidation of patients with musculatory problems will be better supported. ...

- *Recognition of object categories:* This is the point already touched upon, namely that vision systems increasingly can recognise individual objects under a wide variation of circumstances, but are still largely unable to deal with categories of objects. There are a few exceptions to this, where lots of dedicated effort has gone into. Examples are face detection and character recognition. But the general categorisation problem has hardly been touched upon.

- *Shape and scene representations:* There are intriguing questions to be solved relating to how models of objects and scenes ought to look like. In how far do we need explicit 3D information, for wbich parts can we do with interpolation between images or a few planar depth layers, where does texture suffice and where do we need more geometric detail, can texture models be improved to also include effedts of viewpoint and illumination, etc. This is an area with a strong confluence of computer vision and graphics. Undoubtedly, the models should depend on context and tasks.

- *Grouping and segmentation:* There is a kind of chicken-and-egg problem underlying much of vision. When presented a scene, the first problem could be to distinguish the different objects, i.e. to delineate the regions in the image that correspond to different objects. We call this the segmentation problem. Once this problem has been solved, one can set out to recognize the objects. However, how to delineate such regions in the first place? It would be much easier if one had a clew what the objects where, but this is ecactly the problem we want to solve. Cue integration will be vital here. Grouping refers to the collection of special pattern configurations, like

e.g. mirro symmetric or periodic regions or regions of a homogeneous colour, in the hope that such regularities are non-accidental, i.e. indicate that such regions indeed belongs to a single object.

- *More self-learning and self-diagnosis:* Most vision systems don't adapt and don't learn once in operation. Future systems will have to work in far less structured environments. This implies that the manufacturer will no longer be able to foresee the situations the vision system can find itself in. Vision systems will have to learn to recognise new objects and new classes of objects from their own 'experience'. Maybe even more importantly, vision systems will have to understand what their own limitations are, and report back not only a result but also how confident they 'feel' about it.

1.5 I/II - Organisation of the course

The course comprises the following chapters:

1. Introduction - I
2. Light - I
3. Perception - II
4. Image Acquisition - I/II
5. Sampling and Quantization - I
6. Preprocessing and Enhancement - I/II
7. Basic Feature Extraction - I
8. Edge Detection - I/II
9. Motion Extraction - I/II
10. 3D Data Extraction - I/II
11. Shape Representations - II
12. Surface Features - II
13. Feature Selection - I/II
14. Object Recognition - II

The 'I' and 'II' symbols indicate whether these chapters are part of the first or the advanced course on computer vision, completely or partially.

These chapters have been ordered more or less according to the typical stages of a vision algorithm, starting with considerations on the input step (chapters 2 - 5), followed by processing preceding actual interpretation of the image (chapters 7 - 13). Chapters 14-16 deal with issues already closer to semantic interpretation.

Chapter 2 deals with the physics of light and its interaction with matter. These principles are the basis of vision and should therefore be understood to some extent. They are also relevant for the design of optical instruments and illumination setups.
Chapter 3 discusses human perception, i.e. it describes the way light and patterns is perceived by humans. Relevant aspects of psychophysics and neurophysiology are touched upon very briefly. The take-home message is that there is far more to perception than simple recording of incoming light.
Chapter 4 discusses devices to acquire images, mostly cameras. In particular, simple geometric and photometric camera models are introduced. It also discusses illumination techniques that may dramatically ease vision tasks.
Chapter 5 discusses the problem of turning the continuous input into an appropriate discrete representation, both with respect to the choice of spatial sampling schemes and the choice of intensity levels. The chapter also introduces the reader to Fourier analysis of images.
Chapter 6 discusses alternatives to the Fourier decomposition of images, that may be more appropriate for certain applications. Examples are unitary transforms in general, the Karhunen-Loeve transform, Independent Component Analysis, and wavelets.
Chapter 7 describes methods to sharpen images, to reduce noise, and to remove motion blur. The chapter also introduces the concept of convolution filters.
Chapter 8 discusses the detection of pixels that correspond to interesting intensity patterns, such as positions that correspond to edges, stripes, corners, or pixels that line up to straight edges.
Chapter 9 gives an overview of some techiques to extract edge strings from local edge pixels. Such edge yield compact descriptions for further processing, as they typically correspond to the boundaries of objects or surface patches, which are very important for scene analysis, tracking, stereo, etc.
Chapter 10 discusses the processing of video data to extract the motion of objects. Motion detection, motion vector extraction (optical flow), and tracking are different examples of motion related tasks.
Chapter 11 gives an overview of methods for the extraction of 3D information – range data – about a scene. Active and passive as well as uni- and multi-directional methods are illustrated.
Chapter 12 describes methods to represent shapes. Some are more global, other local, and some focus on decomposition into parts of some prescribed types.
Chapter 13 discusses the extraction and description of features that characterise surfaces: colour and texture. Such features can be seen as complementary to more shape oriented

features discussed elsehwere in the course.
Chapter 14 discusses the crucial step of 'segmentation'. Indeed, after the initial processing described in the earlier chapters, the scene should be divided into its meaningful entities.
Chapter 15 focuses on features that can characterise shapes. It discusses the principles and practical use of invariance. It also discusses how decisions can be made based on features, from a pattern recognition point of view.
Chapter 16 concludes the course by discussing how several of the foregoing results can be combined into algorithms that recognize objects. Model-based, appearance-based, and hybrid techniques are described and compared.

1.6 Where to get more information

Of course, computer vision is covered by several specialized journals. The following list gives an overview of some of the best-known ones:

Artificial Intelligence
Biological Cybernetics
Computer Vision, Graphics, and Image Processing: Image Understanding
IEEE Transactions on Geoscience and Remote Sensing
IEEE Transactions on Medical Imaging
IEEE Transactions on Image Processing
IEEE Transactions on Pattern Analysis and Machine Intelligence
Image Processing
Image and Vision Computing
International Journal of Computer Vision
International Journal of Robotics Research
Journal of Mathematical Imaging and Vision
Journal of the Optical Society of America A
Journal of Visual Communication and Image Representation
Machine Vision & Applications
Pattern Recognition
Pattern Recognition Letters

Psychophysical research can be found in

International Journal on Experimental Psychology: Human Perception and Performance
Journal of the Optical Society of America A
Perception
Perception & Psychophysics
Spatial Vision

Neurophysiological research is reported regularly in

Biological Cybernetics
Journal of Neurophysiology
Journal of the Optical Society of America A
Vision Research
Visual Neuroscience

The interested reader will also find references to books in the "Further reading" sections at the end of the chapters.

Finally, a number of conferences are devoted to computer vision. The list is growing quickly, but the following ones certainly are a good starting point to get the flavour of computer vision research:

Asian Conference on Computer Vision
British Machine Vision Conference
Conference on Computer Vision and Pattern Recognition
European Conference on Computer Vision
Image Understanding Workshop
International Conference on Computer Vision
International Conference on Pattern Recognition
Scandinavian Conference on Image Analysis
SPIE Conference on Visual Communications and Image Processing