Mitteilung Nr. 124

VISION SYSTEMS: STATE OF THE ART AND PROSPECTS

BERND NEUMANN

IFI-HH-M-124/84

JUNI 1984

FACHBEREICH INFORMATIK
UNIVERSITÄT HAMBURG
Schlüterstr. 70
D-2000 Hamburg 13

This report is based on an invited lecture at the GDI/ECCAI Conference on "Artificial Intelligence", Zürich, April 12-13, 1984

Zusammenfassung

In diesem Bericht wird die Frage untersucht: Was kann die Künstliche Intelligenz zur Weiterentwicklung und Verbesserung industrieller Sichtsysteme beitragen? Die Ausführungen wenden sich auch an Leser, denen das Gebiet der maschinellen Bildverarbeitung bisher fremd ist. Nach einer einführenden Erläuterung des Fachgebiets "Bildverarbeitung" wird der Stand der Kunst im Bereich industrieller Sichtsysteme dargestellt. Es werden Gründe für die begrenzte Anwendbarkeit der gegenwärtig kommerziell erhältlichen Systeme aufgezeigt. Demgegenüber stehen Ansätze aus der KI-Forschung, die eine Verbesserung ermöglichen können. Als besonders vielversprechend werden eine "wissensbasierte" Verarbeitung und die Verwendung von Konfigurationsexperten hervorgehoben.

VISION SYSTEMS: STATE-OF-THE-ART AND PROSPECTS

Bernd Neumann

1. Introduction

Let me begin with a disclaimer: In this contribution I will not attempt a complete survey of existing vision systems. For this the reader is refered to the representative collection of articles edited by Pugh [PUGH 83] or any of the recent surveys [CHIEN and HARLOW 82, FU 83, KELLEY 83, PAU 84, PINSON 83, STIEHL and KRAFT 84]. Nor does this contribution present systems under development in the laboratories. An excellent paper of Binford [BINFORD 82] is recommended to the reader interested in a critical analysis of major laboratory systems.

Bearing in mind the theme of this volume I shall discuss vision systems with respect to the following question: What can AI contribute to industrial vision systems? This may sound somewhat strange presuming that vision systems are AI products to begin with. I shall show, however, that a typical industrial vision system does not exhibit typical AI architecture. Major shortcomings and limitations of present systems can be traced to a design which may be improved by adopting a "knowledge-based" approach. It is shown that this may lead to systems with extended applicability and predictable performance. Furthermore the process of adapting a vision system to a new task may be facilitated by "configuration experts", i.e. expert systems guiding the configuration process.

In order to provide the layman with some background, the discipline of "Computer Vision" will be briefly introduced in the next section. The reader familiar with this field may safely proceed to the discussion of industrial vision systems in section 3. First, a broad range of applications is presented to demonstrate the versatility of the existing vision technology. Then an exemplary vision task - object recognition - is discussed in some detail. This leads up to a critical evaluation of today's systems. Although many limitations are dictated by the need of fast processing, and consequently point to a hardware bottleneck, there is also the need for methodological improvements. This is where an AI approach will be helpful.

In section 4 the notion of a knowledge-based vision system is introduced. It is shown that the use of suitable object models and model-based recognition techniques holds many promises with respect to adaptability and predictability of vision systems. While it is unlikely that efficient knowledge-based vision systems of large generality will be available in the near future, the underlying principles can be brought to bear on practical systems to some extent by employing configuration experts.

2. What is Computer Vision?

Computer Vision had its beginning with the advent of fast digital computers after 1950. It was recognized early that visual data can be represented by numbers and thus be made the input of computer programs. The problem, of course, is how to process these numbers in order to obtain useful results. Right from the beginning the performance of the human visual system set the mark. Selfridge proposed "eyes and ears for the computer" [SELFRIDGE 55]. But for many years programs could only handle very restricted tasks, e.g. analysis of toy scenes composed of blocks [ROBERTS 65] or recognition of printed characters [FISCHER et al. 62]. Tasks like the latter motivated the paradigm of "pattern recognition". This is a particular abstraction of a recognition task where an unknown object is taken to be represented by a set of numbers - the features - and the main problem is to assign it to one of a given set of classes using the features as a decision basis. Today a very well developed theory of Pattern Recognition is available for all recognition tasks which can be cast into those terms.

Computer Vision, in general, is more than an assignment problem - it requires methods much different from Pattern Recognition. In particular, features and classes proved to be insufficient notions for representing complex intermediate or final results. Today, the leading paradigm for Computer Vision is a multilevel, knowledge-based process which reconstructs and describes a real-life scene from projections. This view is also spelled out by the authors of one of the leading textbooks of the field [BALLARD and BROWN 82] who define Computer Vision as

> "...the construction of explicit, meaningful
> descriptions of physical objects from images."

Computer Vision is a well established field with a large fundus of methods. Yet much research remains to be done before technical systems can compete with biological vision systems. Various related fields are relevant for progress in Computer Vision:

- Electrical Engineering
- Computer Science
- Artificial Intelligence
- Cognitive Science
- Psychology
- Neurophysiology
- Mathematics
- Pattern Recognition

Computer Science and Artificial Intelligence are usually considered the parent disciplines of Computer Vision. But strong ties also exist to Psychology and Neurophysiology which provide valuable data and insights concerning biological vision. Although most computer vision researchers do not seek detailed correspondence with the human visual system, there is probably none who would not happily adopt a process from biological systems.

3. Industrial Vision Systems

In this section I shall provide some information on existing industrial vision systems. Hopefully this will help the reader to form an opinion about what can possibly be done and what not. I shall begin by discussing major areas of applications, including examples selected from recent publications [ISIR 83, ROVISEC 83, PUGH 83].

Applications

By far the largest number of applications are in the area of <u>inspection</u> (quality control). Vision systems are used to check for

- completeness (missing or excessive parts)
- geometry (shape, alignment, tolerances)
- material defects

Here are some representative inspection tasks handled by machines (the letters C,G and M indicate the objective according to the three possibilities mentioned above):

- drillings in automotive parts (C,G)
- keycap assemblies (C)
- flaws in castings or forgings (M)
- printed circuits, chips (C,M)
- labels and fill levels of champagne bottles (C,G)
- correct reading of gauges and instruments (G)
- surface defects of hot steel slabs (M)
- cleanliness of Coca Cola bottles before filling (C)
- packages of birth control pills (C)
- gap tolerances in car door assemblies (G)

Inspection tasks are particularly suited for machine vision since human workers tend to perform poorly on such tasks: Only a small fraction of the samples exhibits interesting faults, the overwhelming majority is flawless and quite boring. Hence attention deteriorates rapidly and errors result.

The second major area of industrial applications is <u>robot position control</u>, where vision systems cooperate with a manipulator to perform tasks such as object recognition, automatic assembly, or seam tracking for welding. Here are some illustrative examples:

- pick-and-place from moving conveyor belts
- chip insertion into circuit boards
- iron coil recognition for crane control

- stacking and unstacking work pieces
- engine block location
- bolt position for car assembly
- collision avoidance
- pose refinement for electric motor assembly
- are welding of sheet steel

Most of these applications have been realized by adapting a general-purpose vision system to the particular task. The keycap assembly system [in PUGH 83] is a notable exception. It can only inspect keycap assemblies.

From the available reports it is difficult to assess the labour (and cost) which went into the adaptations. My personal experience suggests that, as a rule, more is required than setting certain parameters, for example: arranging the illumination by trial and error or even reprogramming the system software. This is one of the major shortcomings of present-day vision systems: They may require considerable efforts from human experts before they perform satisfactorily at a particular task. There are about fifty brands on the market from which to choose today. A recent survey can be found in [KINNUCAN 83].

Methods

A typical industrial vision system is shown schematically in Fig. 1. It consists of lighting equipment for controlled illumination of the object of interest, a sensor to pick up the image, a preprocessing unit for image transformation and fast feature extraction, a postprocessing unit for complex computations and decision making, and finally a user interface for interaction with a human operator. Using the task of object recognition as an example, I shall now briefly describe these components in some detail.

Illumination plays a critical part in most applications. It is the essential means to make a vision task manageable with present-day technology. In fact, industrial vision has been defined as "computer vision with controlled illumination". For object recognition, one typically tries to produce a clear contrast between object and background, e.g. by backlighting. But there are also more sophisticated illumination techniques, e.g. structured light which can be used to extract 3-dimensional object properties. To this end laser beam techniques have also been developed.

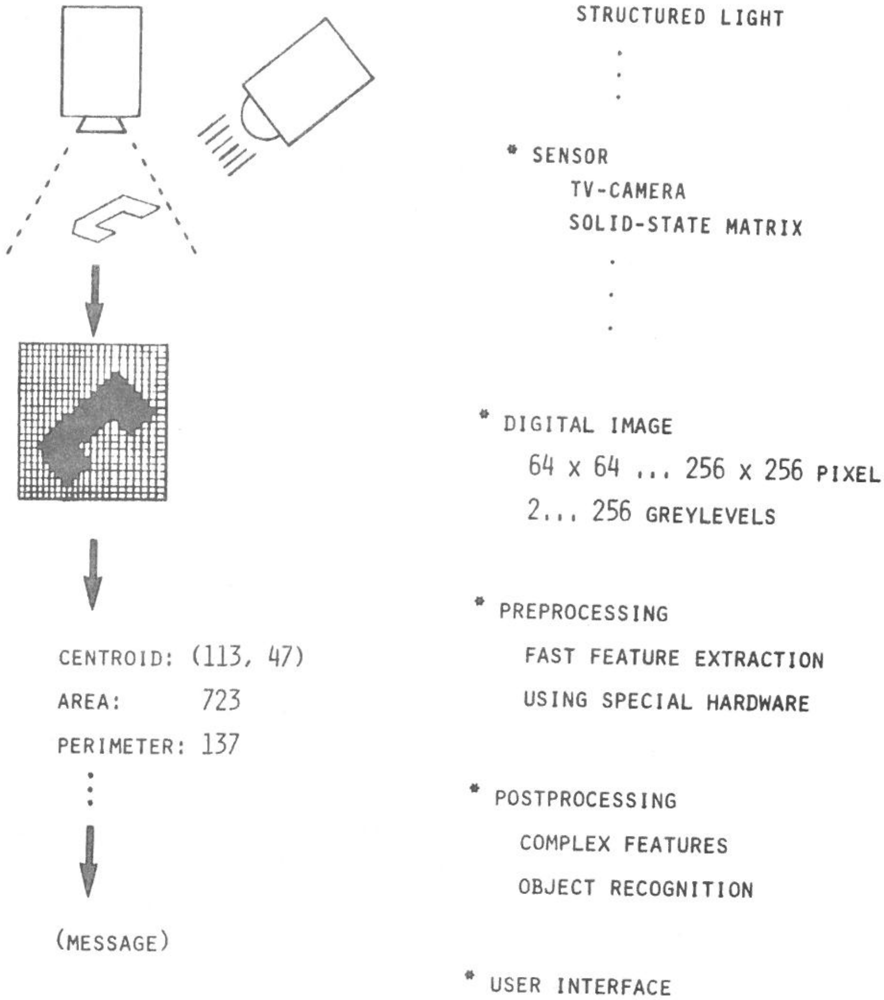

Fig. 1: Components of industrial vision system

There is a variety of sensors available, ranging from TV-cameras to solid-state devices. The latter ones are extremely rugged and can be built as small as a pack of cigarettes. After digitization the sensor output is a field of numbers, called the digital image. Each number represents the brightness ("greyvalue") of a picture element ("pixel"). Typically, industrial vision systems use digital images of a size between 64 x 64 and 256 x 256. Greyvalues are quantized into up to 256 greylevels. For ease of processing, however, most systems abandone greyvalue information very early by thresholding, thus all further processing is based on a binary (i.e. black-and-white) image. Threshold selection is a critical step if slight variations of object or illumination properties cause radically different binary images. This can often - but not always - be avoided by a judicial choice of illumination and background. It cannot be ruled out that binarization is alltogether inappropriate - these tasks call for greyvalue processing which is usually more complex and costly.

The preprocessing phase encompasses operations which can be carried out very fast, often using dedicated hardware. For binary images these operations usually belong to a standard set which is listed below. The output is a description of the black (or white) image regions in terms of feature values.

Binary image processing:

smoothing, shrinking, expanding

connectivity analysis

computation of features
- total area
- number of holes
- perimeter
- centroid
- longest and shortest diameter
- axes of equivalent ellipse
- compactness (area/perimeter^2)
- enclosing rectangle
- polar signature

If interesting image regions cannot be extracted by binarization, greyvalue processing is in order. Below are some of the operations used by the few existing greyvalue systems. (The repertoire of laboratory systems is much larger).

Greyvalue processing:

filtering
- averaging
- subsampling
- median filtering

edge analysis
- gradient value and direction
- local gradient maximum
- grouping of contour elements
- straight line fitting
- blob extraction

While preprocessing operations have to cope with large pixel arrays, the input data of all ensueing operations have small or moderate volume. Hence postprocessing is rarely time-critical and can be carried out by general-purpose microcomputers. The main task at this stage is classification of image regions by comparing the features with prototype descriptions. Most systems use the nearest-neighbour decision rule, i.e. decide for the prototype which is closest in feature space. Some advanced systems permit "relational" object descriptions involving more than a single image region. In this case classification is based on the relational match paradigm which calls for structural comparison of descriptions. In any case, the success of the classification step depends on the similarity of the unknown object - described by its image features - with prototype descriptions given to the system in a "learning" phase.

This is - in short - how a typical industrial vision system performs object recognition. In this simplified presentation I cannot do justice do all systems on the market, particularly not to those which are specialized for a narrow range of tasks. For example, systems performing quality control using Xray images may very well require processing steps which are quite different from the steps described above. Our "typical vision system" is exemplary in the sense that it reflects the architecture of the majority of systems dealing with the general task of object recognition.

Limitations

The performance of present-day industrial vision systems is quite limited in the sense that there are many potential applications which cannot yet be realized. In most cases the cause is a combination of task complexity, time requirements and costs. Hence cheaper and faster hardware will continue to push the limits and open up new fields of application. But this sort of progress will not cause the qualitative jump needed to overcome troublesome limitations arising from the currently employed methods. What are these methodological limitations?

First, there is the use of binary images. Low background contrast, variable illumination and variable object surface properties may cause highly varying binarization results. Hence all these factors have to be controlled to secure reliability. If anything changes, e.g. if the conveyor belt gets dirty, malfunctions are likely.

Second, let us examine the features which are used for recognition. Most of them are global with respect to an image region and thus are sensitive to local degradations, e.g. region deformations due to partial occlusion or imperfect binarization. Hence a correct classification may be jeopardized although a large part of the region is in perfect shape. This could be improved by employing descriptions based on local features, e.g. contour pieces embedded in relational structures. Such systems have the virtue of "graceful degradation", however at the cost of increased complexity. So far, few systems use relational descriptions.

Virtually all existing systems can only recognize planar shapes. This means that three-dimensional objects can only be recognized, if the image has features matching those of a prototype image. Since, in general, an object may have an unlimited number of different projections, there must be as many prototypes if it is to be recognized in all views. This is, of course, prohibitive and care must be taken that an object is always seen in one of a limited number of perspective views.

In summary, industrial vision systems of today have restricted applicability. As a consequence of the methods used, it is difficult to predict the performance when task parameters change. Expert knowledge and sometimes trial-and-error is required to adapt a system to a new task. In the next section it will be shown that these problems may be eased to some extent by using an AI approach to vision.

4. AI Approach

In order to assess future developments and potential improvements of existing vision systems we now take a look into the laboratories. There are several characteristics which distinguish laboratory systems from an industrial vision system. To a large extent these characteristics are typical for present-day AI research, hence it is appropriate to speak of an AI approach to vision.

First, most laboratory systems tackle tasks of considerable generality. (This does not mean that they actually achieve generality - see BINFORD 82). Many of the restrictions of industrial vision tasks mentioned earlier do not apply to experiments carried out in the laboratory. For example, there are systems dealing with 3-dimensional shapes, analyzing scenes with shadows, recognizing partially occluded objects, etc. Hence there exist methods which will improve the applicability of industrial vision systems.

If there is a single characteristic of AI systems from which industrial vision would benefit most, it is the knowledge-based architecture. This is, of course, a trade-mark of AI systems in general and I shall now briefly describe what this may mean for vision systems. Knowledge-based processing calls for an explicit representation of all static information pertinent to a given problem. In the context of computer vision this means primarily that all information must be made explicit which links the image to physical reality. This information is often called "image formation knowledge". Fig. 2 illustrates the image formation process. A ray of light leaves the light source, reflects from the object surface and is picked up by the sensor. From the physics of this process we know that the resulting image depends on each of the following components.

Object: shape
surface properties
position and orientation

Illumination: spectral characteristics
reflecting surfaces
position and orientation

Sensor: resolution, sensitivity, etc.
position and orientation

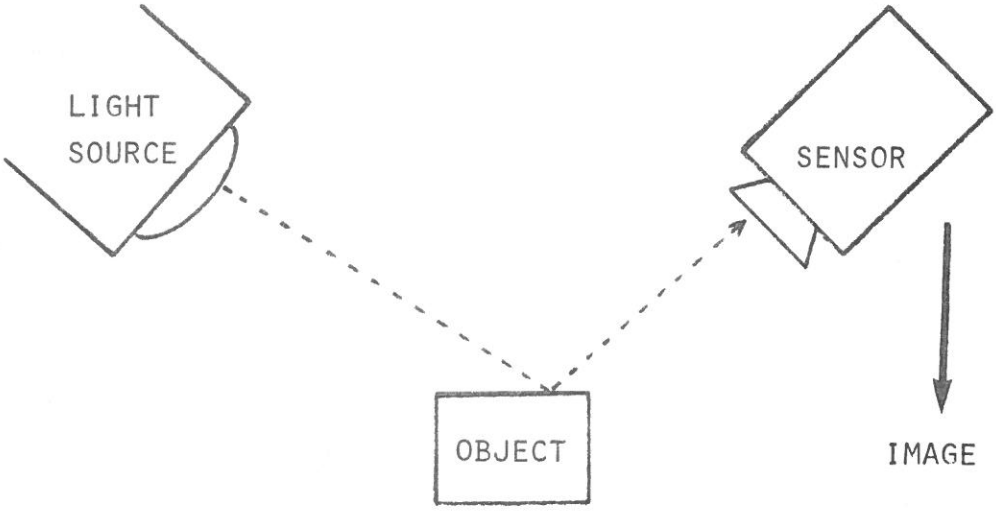

Fig. 2: Image formation process

Hence each image is an encoding of various properties of the physical reality, and image analysis processes must be set up accordingly. In a knowledge-based architecture this is achieved by using explicit representations of these influencing factors. In other words: Knowledge about illumination, sensor and object characteristics is kept in a data-base which can be easily accessed and modified. This has several consequences.

First, prototype objects are represented by strictly object-inherent properties:

- 3-dimensional shape (incl. tolerances)
- surfaces properties (reflectivity)
- range of possible positions and orientations

Hence, different from todays industrial vision systems, prototypes are not represented by image features which have the disadvantage of also encoding illumination and sensor properties. This makes object representations independent of other task components, improving the adaptability of vision systems to new tasks. It also paves the way to using CAD/CAM models for vision tasks, eliminating the process of "learning" prototypes.

Secondly, the effect of a particular choice of sensor or lighting equipment is coded in such a way that changes may be made easily without heavy reprogramming. This will certainly facilitate the adaptation to new tasks.

A third advantage concerns unwanted changes of sensor, illumination or object properties which are quite characteristic for an industrial environment. Using knowledge-based image processing the effect may be precalculated to a considerable accuracy. Hence the performance of vision systems becomes more predictable.

Unfortunately the knowledge-based approach sketched out above also adds considerable complexity. In view of the time-constraints omnipresent in industrial applications such a system will presently be too slow by orders of magnitude. However, some of the advantages of a knowledge-based architecture may be achieved without waiting for faster hardware. For this I suggest an expert system approach to vision system configuration.

Configuration experts are software tools which guide a human user in choosing system components, setting parameters and making decisions occuring in a system configuration task. A well-known example is XCON (formerly R1) developed for the configuration of VAX computer systems [MCDERMOTT 82]. For computer vision, a configuration expert has the objective of supporting vision system adaption to new tasks. Thus it should contain expertise of the kind which is presently required from human experts (and makes the configuration task a costly process). As for the knowledge-based approach discussed earlier, the idea is to provide the expert system with a high-level knowledge-based program which would be slow but could be used as a simulation tool and as a source from which fast solutions can be compiled. Simulation of a vision system is within reach: The hard part is generating realistic synthetic images based on image formation knowledge. This can in fact be realized with advanced Computer Graphics methods. The second feature - compilation of fast programs from high-level specifications - does require considerable research. The work of Bolles and Cain [in PUGH 82] and Goad [GOAD 83] are promising examples of this kind of approach. In my view, this is a feasible and necessary way to bring AI methodology to bear on industrial vision system performance.

5. Summary

In this contribution I have tried to answer the question: What can AI contribute to industrial vision systems? Reviewing the present capabilities of industrial vision systems I discussed several methodological weaknesses resulting from a simplistic approach dictated by time and cost constraints. In consequence, todays vision systems are limited in many ways and often lack good engineering virtues, e.g. a predictable performance at a new task.

The discussion of potential - and expected - improvements focussed on two major contributions of AI research: a knowledge-based approach to image analysis and an expert system approach to system configuration. It was shown that the development of such systems will lead to extended applicability, improved adaptability, predictable performance and a potential integration with CAD/CAM.

6. References

Ballard and Brown 82
Computer Vision
D.H. Ballard, C.M. Brown
Prentice Hall, 1983

Binford 82
Survey of Model Based Image Analysis Systems
T.O. Binford
International Journal of Robotics Research, Vol. 1, No. 1, 1982, 18-64

Chien and Harlow 82
Automated Visual Inspection: A Survey
R.T. Chien, C.A. Harlow
IEEE PAMI Vol. 4 No. 6 (1982) 557-573

Fischer et al. 62
Optical Character Recognition
G.L. Fischer, D.Y. Pollock, B. Radack, M.E. Stevens
Spartan, Washington D.C., 1962

Fu 83
Robot Vision for Machine Part Recognition
K.S. Fu
in: D. Casasent, E.L. Hall (eds.), Robotics and Robot Sensing Systems, Poc. SPIE Vol. 442, 1983, 2-14

Goad 83
Special Purpose Automatic Programming For 3D Model-Based Vision
C. Goad
Image Unserstanding Workshop, 1983, 94-104

ISIR 83
Proc. 13th International Symposium on Industrial Robots and Robots 7
North Holland, 1983

Kelley 83
Binary and Gray Scale Robot Vision
R.B. Kelley
in: D. Casasent, E.L. Hall (eds.), Robotics and robot Sensing Systems, Proc. SPIE Vol. 442, 1983, 27-37

Kinnucan 83
Machines That See
P. Kinnucan
High Technology Vol. 4 No. 4 (1883) 30-36

McDermott 82
R1: A Rule-Based Configurer of Computer Systems
J. McDermott
Artificial Intelligence 19 (1982) 39-88

Pau 84
Approaches to Industrial Image Processing and Their Limitations
L.F. Pau
Electronics & Power, February 1984

Pinson 83
Robot Vision: An Evaluation of Imaging Sensors
L.J. Pinson
in: D. Casasent, E.L. Hall (eds.), Robotics and Robot Sensing Systems, Proc. SPIE Vol. 442, 1983, 15-26

Pugh 83
Robot Vision
A. Pugh (ed.)
Springer Berlin/Heidelberg/New York 1983

Roberts 65
Machine Perception of Three-Dimensional Solids
L.G. Roberts
in: D.T. Tippett et al. (eds.), Optical and Electro-Optical Information Processing, MIT Press, Cambridge, Mass., 1965, 169-197

ROVISEC 83
Proc. 3rd International Conference on Robot Vision and Sensory Controls
IFS Publications, 1983

Selfridge 55
Pattern Recognition and Modern Computers
O.G. Selfridge
Western Joint Comp. Conf. 1955, 91-93

Stiehl and Kraft 84
Sichtsysteme fuer Industrieroboter
H.S. Stiehl, H.R. Kraft
HMD 115, 1984, 65-89