Annu. Rev. Comput. Sci. 1990. 4:517–29

VISION

Takeo Kanade

School of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

Tom Binford

Department of Computer Science, Stanford University, Stanford, California **94305**

Tomaso Poggio

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138

Azriel Rosenfeld

Center for Automation Research, University **of** Maryland, College Park, Maryland **20742**

1. INTRODUCTION

Computer vision provides **a** major perceptual capability to autonomous or semiautonomous intelligent systems that operate in the physical real world by constructing scene descriptions (what is happening) from input image data (what is seen). Kinds and levels of descriptions depend largely on the tasks that the total systems are missioned to perform, but typically included are descriptions of the three-dimensional environment, of features detected, of objects identified, and of their relationships. The goal of computer vision research is to develop computational theories and technological means to realize an artificial vision system with at least the speed and capabilities of human vision in an unconstrained natural environment.

Currently, humans outperform machines by far in most tasks. Thus,

8756–7016/90/1115–0517$02.00

computer vision can learn much from human visual systems, but contrary to naive arguments, mimicking a human may not be the best way to meet our ultimate goals. Rather, to understand human vision as an information processing system is only one of the goals of computer vision research. In fact, machines may exceed human visual capacities in the future (and in some cases do so already), especially by using special sensors, collateral information, and absolute measurements. To bring this about we need a clear technical definition of the problems, along with development of rigorous computer vision algorithms.

Vision may be one of the hardest problems in the whole enterprise of understanding and replicating intelligence. A large number of factors are confounded in the intensity and color of image pixels: shape, illumination, surface properties, sensor characteristics, and so on. Most early-vision problems involving recovery of scene properties (such as shape) from images are underconstrained or ill posed. Correctly interpreting images and assessing situations from them require not only physical, geometrical. and optical knowledge. but also semantic domain knowledge. Moreover, interaction of the two types of knowledge is strong but remains undefined. The amount of information that must be processed is so huge that today's computers do not provide enough computational power to solve vision problems robustly in real time, even when we understand the problems well. Currently there **is** no uniform theory or approach for addressing **all** the issues in vision. In fact, developing such a "uniform" "general" theory may not be possible.

So important are the potential applications of computer vision (e.g. automatic target recognition, manufacturing. inspection, navigation. cartography. reconnaissance, and medical diagnosis) that even partial, short-term solutions have been useful. But we need both a fundamental understanding of vision processes and the advanced technology to map the algorithms onto hardware/software systems if we are to make computer vision work robustly in real time in natural unconstrained environments. The solution of these problems requires a concerted and steady long-term effort. It requires basic research and good engineering. We expect progress to continue and applications to increase in both availability and significance over the next decade.

2. BACKGROUND

2.1 *Progress to Date*

Automatic processing of imagery by computer (e.g. character recognition, image enhancement, and medical image processing) started in the 1950s, but PhD work by Roberts in ***1965*** is generally thought to be the first

computer vision research on understanding the three-dimensional world. In the **25** years since then substantial progress has been made. The approaches and emphases during this period are as follows:

Very Early (~ 1970)	Sequential: bottom-up knowledgeless pattern-recognition approach
Early (~ 1975)	Heterogeneous: top-down use of ad hoc domain knowledge
Middle (~ 1980)	Marr's paradigm: computational models of low-level vision modules
Recent (~ 1985)	Model or knowledge-based **3-D** vision: more systematic use of models
Most recent	Systems: vision as part of a larger system.

The last I5 years have seen major advances. Highlights of progress and the current state of the art are summarized below.

We have good foundations, in terms of mathematical techniques and analyses of constraints, for an understanding of early vision. Various optimal low-level image operators (edge detectors, among others) have been developed and are now considered standard. Reliable low-level segmentation of an image into meaningful parts is, however, still far off.

Many problems of **3-D** recovery from images have been formulated as "inverse optics" problems, and various "shape from x" techniques have been developed, such as shape from shading. shape from texture, and shape from contour. One of the major findings of this class of work is that ***generic*** natural constraints (i.e. general assumptions about the physical world that are correct in almost all situations) are often sufficient to solve the problems of early vision. Two main themes are therefore intertwined at the heart of the main achievement of early vision research: *(a)* the identification and characterization of generic constraints for each problem (e.g. epipolar constraint, rigidity) and ***(b)***their use in an algorithm to solve the problem (e.g. variational method, scale space, and regularization).

Various general mathematical tools have been developed and applied to vision, mostly early vision, to deal with uncertainty, prior models, posterior-model estimation, stabilization (of ill-posed problems), interpolation, and matching. They include regularization, weak continuity, Markov Random Field models, and Bayesian modeling and estimation.

Stereo techniques being developed recently will be capable of supporting DMA terrain elevation mapping (but not feature analysis, which is harder) in the near future. Practical near-real-time stereo systems are being studied in laboratories. Newer approaches to stereo, such as the use of more densely taken multiple frames, will provide further improvements in accuracy, detail and speed.

The topics of time-varying imagery, optical flow, and the structure-from-motion problem have attracted many researchers and resulted in many papers. Such work remains mostly academic, and the performances of most proposed algorithms with real imagery are not yet satisfactory for practical use.

Development of parallel vision algorithms has begun to have real impact, as parallel multiprocessor architectures (e.g. the Connection Machine, Warp, and Pipe) become available to the research and development community.

Many techniques for range sensing by active means have been developed. In addition to simple low-cost range sensing for industrial applications, medium-to-high resolution moderate-speed range sensors by scanning laser for outdoor navigation have been developed. Much more needs to be done to develop robust **3-D** range-analysis techniques.

Progress has been made on building capable model-based **3-D** interpretation systems. While early systems were ad hoc special-case-based systems for simpler problems like interpretation of line drawings, recent model-based vision systems are capable of symbolic reasoning on model vs image relations, incrementally building **3-D** scene descriptions from multiple images, and efficiently mapping an object model's geometrical constraints into search space. However, use of **3-D** constraints in a general manner is still greatly limited. We are just beginning to design interpretation schemes based on first principles and comprehensive models.

In the last several years, much progress has been made toward natural outdoor Scene understanding and its use for autonomous land vehicles, thanks to the Strategic Computing project. Some demonstrations of a whole system including perception, planning, and control have been given for following straight and curved paved and unpaved roads by using color vision. Considering the state of the art of vision when the Project started, these accomplishments are significant. However, the demonstrated capability is quite limited in terms of robustness, kinds of situations (types of road, lighting conditions, etc) it can handle, and use of other a priori information such as maps. Very little work has been done on open-terrain navigation, except for work on avoidance of obstacles such **as** rocks and bushes by using sonar or laser range finders. Research has just started on perception **for** very-rough-terrain navigation in the context of the Mars Rover.

2.2 *Related Fields*

Vision has relationships to many other fields. In its attempts to understand natural perception systems it is closely related to psychology, psychophysics, and neurophysiology. Though copying natural vision would be wrong,

these fields provide information about human and animal visual systems that can serve as motivation for developing computer vision systems.

Network analyses of all kinds are important in vision for representation and computation of a large-scale network of constraints. Neural network solutions, however, have made little impact on fundamental understanding of vision processes, except for the strong suggestion that fine-grained parallelism and possibly analog **VLSI** may play an important role in future real-time vision systems.

Vision requires powerful three-dimensional shape representations. Such representations are also used in CAD and 3-D graphics. However. the shape-modeling requirements for vision are different from those in commercial CAD systems, which are used for man-machine communication of shapes through production of drawings. In vision we require symbolic results for subsequent programs. Although geometric models are now better than the use we make of them, there is need for work on modeling and system building. Formal or informal systems that reason about shape. either symbolic or numerical, are of increasing importance in vision.

Signal-processing systems applicable to other types of real-world signals (e.g. acoustic, seismic) can provide vision researchers with techniques for efficient data processing. noise modeling. and optimal estimation.

Methodology and technology in sensor development have a crucial impact on vision-system performance. Some applications can be simplified considerably by developing application-dependent sensors. A prime example from biomedical research was simplification of chromosome analysis by development of staining techniques for chromosome banding. The use of optical techniques (e.g. interference, diffuse light, polarization) makes complicated industrial inspection problems trivial, or makes applications achievable that were out of reach by the current state of the art of general vision.

Computation is a major limiting factor for vision. Most vision problems today are computation-bound. Advances in computation technology. especially less expensive parallel computers, will have the largest impact on progress in vision, since they will allow researchers to develop and test sophisticated algorithms within a reasonable time. That massively parallel computers are becoming available to the research community is encouraging, but it must be emphasized that parallel computers are no substitute for a fundamental understanding of machine vision.

2.3 *Vision Research Centers*

Vision research is active in both academia and industry, in the **US** and abroad. In the US, CMU, Columbia, Illinois, Maryland, Massachusetts. **MIT,** NYU, Pennsylvania, Rochester, Stanford, SRI, and **USC** are the

leading groups in vision research. Most of these institutions have been supported by the DARPA Image Understanding program and Vision in Strategic Computing. In Japan, the Electrotechnical Laboratory and University of Osaka are good, and in Europe, INRIA (France), Karlsruhe (W. Germany), Oxford (Britain), and Utrecht (Netherlands) maintain significant activities.

3. RESEARCH OPPORTUNITIES

3.1 *Research Issues*

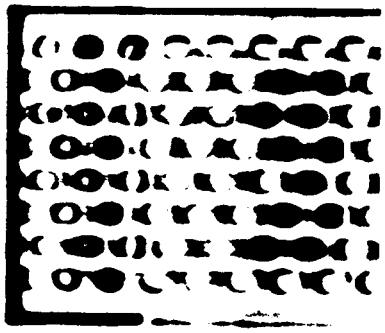
Vision is still fairly far from its ultimate goal, but important research opportunities exist for both the short and long terms.

MODULES OF EARLY VISION In the last decade we have seen major advances in understanding such aspects of early vision as color, texture. stereo, and motion. Continued work is required in these areas. Recent work on color and highlights based on physical and optical models of reflection will be likely to lead to deeper understanding of color image interpretation. Stereo appears ready for near-term significant applications in vision systems. Stereo reconstruction of depth has been motivated mainly by aerial mapping, which requires wide-angle stereo. Stereo techniques developed so far are now successful in integrating many constraints, not just the epipolar constraint, and are ready to merge with intelligent systems. At the same time. a new formulation of the stereo problem is also emerging that uses appropriate filtering techniques to integrate depth estimation from image sequences obtained by small motions of a camera. This formulation simplifies stereo matching problems because of the narrow baseline. and yet provides a good depth map because of the integration of many measurements. Implementation of algorithms on parallel machines and special hardware has also been progressing, and near-real-time depth recovery by stereo is expected in the near future.

Range-data analysis is promising. Sensor development is essential for range-data analysis techniques. Current active range sensors have severe limitations in cost, speed, dynamic range, and accuracy; they also have problems with specularity, glancing incidence, and narrow color spectrum. Substantial progress will be made in range sensing in the next decade, including **FM** (frequency modulation)-based active ranging. VLSI smart sensors, and real-time stereo.

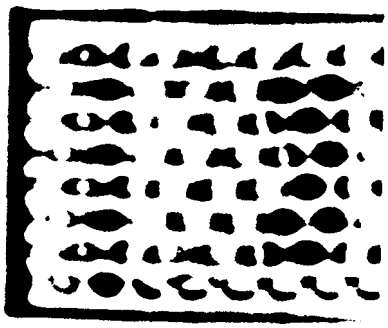
SEGMENTATION Segmentation of images into meaningful units and aggregation of primitive features into groups almost always limit the performance of current vision systems. The problem of segmenting textured surfaces is least understood. Texture is extremely common in military

problems, in infrared **(IR),** synthetic aperture radar **(SAR),** and visible imagery. Texture segmentation is closely related to perceptual organization, which is a difficult scientific problem. Debates about segmentation continue: Is it necessary or possible? To what degree is it task independent? How can a priori information be used? We expect steady, significant improvements in segmentation techniques, including answers to some of these questions, over a five-year period.



ROBUSTNESS BY QUALITATIVE OR ACTIVE VISION Many vision problems are ill conditioned, and in the presence of noise it is difficult to solve them robustly. For example, in image-sequence analysis, accurate quantitative measurements of the optical flow field appear to be very difficult to make. Trying to obtain qualitative solutions is sometimes more robust. Biological visual systems, for which motion analysis is ecologically of great importance, make use of relatively qualitative properties of the flow field which can be calculated more robustly. It has recently been demonstrated, for example, that the flow-field divergence, which provides adequate information for collision avoidance purposes, can be calculated quite robustly. Another way to obtain robust solutions is to formulate the problems from the viewpoint of an "active" observer—i.e. an observer who can control the geometric parameters of the sensory apparatus, for example by moving it. Problems such as shape from shading, texture, or contour and structure from motion are much easier to solve for an active observer than for a passive one.

INTEGRATION OF WEAK MODULES Computer algorithms developed for several early-vision processes (e.g. edge detection, stereopsis, motion, texture, and color) give separate cues to the distance from the viewer of three-dimensional surfaces. their shapes, and their material properties. Not surprisingly, biological vision systems still greatly outperform computer vision programs. One key to the reliability, flexibility, and robustness of biological vision systems is their ability to integrate several visual cues. Further development and application of mathematical techniques for integrating information processed by different (weak) modules will be critical, such as Bayesian reasoning and Markov Random Field models. A near-term payoff will reward techniques to integrate visual with nonvisual sensors—e.g. stereo and sonar, shape from motion and range, etc.



REAL-TIME PARALLEL ALGORITHMS Making computer vision programs work in real time is especially important for time-critical applications such as navigation, target recognition, and risk assessment. Now that parallel computers have become available, iconic retinotopic level algorithms will run in real or near-real time. This trend will accelerate in the future. In

addition to continuing development of parallel vision algorithms themselves, it is now crucial to develop a general programming model and environment for faster algorithm development and usage.

However, more difficult and important are computer architectures for higher-level nonretinotopic processing, since as low-level processing gets faster, intermediate to higher-level processing is becoming a bottleneck for real-time vision. *So* far, little or no work has been done in this area. There are several scientific questions. How can we make computers that are faster for total vision problems? High-level vision is the least understood and most difficult for parallel computation. What is the structure of vision algorithms? How can we partition problems on parallel machines? It is unlikely that we will have fundamental solutions to these questions in the near future.

In addition to digital **VLSI** technology, the use of analog or hybrid (digital and analog) circuits for some vision modules appears to be very promising for faster, time-continuous processing of data with more integration of sensors and processing. For focal plane processing, 3-D integrated circuitry is also useful.

MODEL-BASED RECOGNITION SYSTEMS Model-based recognition continues to be an important research issue. The ultimate goal is to achieve comprehensive interpretation of images based on geometric and semantic constraints and statistical uncertainty represented in the object models. This will require a broad array of research: representation of object models, effective use of domain knowledge for low- to intermediate-level processing. indexing (relating observables to object classes), verification of hypotheses, evidential reasoning with multiple sources of knowledge and multiple sensors, and strategy selection and resource allocation in hypothesis management. Some of these issues are long-range high-risk research areas, but improvements in any of them will lead to substantial improvement of performance of model-based vision systems. The applications for such systems range from target recognition in **SAR** images to industrial parts recognition. In the near future the first full, general, model-based system will be available—i.e. including significant segmentation and interpretation modules.

LEARNING AND KNOWLEDGE ACQUISITION Making vision algorithms learn has been a neglected area (the exception has been the training of statistical pattern classifiers), but its importance will grow in the next decade. Learning algorithms will be developed for edge detection, color constancy, and motion detection. **A** pragmatic application will be the automatic tuning of algorithms and their parameters from examples and from data. Back-

propagation algorithms of neural nets will have some impact in this level of learning.

More challenging and long-term research involves automatic generation of recognition algorithms or structural learning. This will require object modeling, sensor modeling, recognition strategy generation and automatic run-time program generation, and tuning of the resultant programs with real examples.

SYSTEMS We decompose vision problems into subproblems and seek to find modular solutions to those subproblems, and then we combine modules to solve many vision problems. This is an economical approach. System problems are often thought of as engineering problems, simply combining existing modules using known methods so that they work together. However, as recent development of the **ALV** system has proved, integration of modules is not a trivial undertaking. Problems of size, effort, and complexity characterize large systems. Combining modules requires ***(a)*** appropriate system architectures and tools that permit representation of a great variety of data types with complex relations among them and (6) smooth interactions among many modules that communicate using the representations. Working on systems integration also tells whether the decomposition into modules is apt.

3.2 *Measures of Progress*

Qualitative and quantitative measurement of progress in vision seems much more difficult than in other perception technologies. In speech, for example, one can use for evaluation several relatively well-defined measures, such as speaker independence vs dependence, vocabulary size, the types of grammars that limit permissible sentences, and the ratio of processing time to length of utterance. Comparing two speaker-independent voice recognition systems for numerals in terms of their correct recognition rates is very sensible. Comparison of the abilities of two vision systems to recognize office scenes containing only desks and chairs, however, is probably an ill-defined task. Only very low-level tasks, such as edge detection from grey-scale images, can be compared in terms of processing time, resolution, S/N ratio, etc. This is partly because vision can use so many different sensors (b/w, color, range, **IR,** etc) and a problem's difficulty varies with the choice **of** sensors. Another probable reason **is** that vision has not yet advanced to the level where the constraints involved in semantically meaningful (to humans) statements, like "office scene" and "chair," are well understood. In fact, easier vision problems, such as recognizing **2-D** flat pieces in an industrial setting, can be evaluated by the number of objects, special illumination, overlap vs nonoverlap, straight or curved boundaries, etc.

Yet vision has made significant progress in the past 30 years: Emphasis has shifted from binary to grey-scale images. from b/w to color, from planar to curved-surface objects, from knowledge-less to knowledge-based systems, and from the blocks world to the real natural world. These trends will continue.

Progress can be measured by the constantly increasing number of real-world vision problems that we are now in a position to solve. The breadth and level of usage of vision in real-world applications is the most appropriate measure of progress at this point. In each problem, one should consider the following indicators of system capabilities:

- Number of objects that can be recognized
- Constraints imposed on illumination, background, etc
- Precision in locating objects
- Recognition rate (correct-recognition and false-alarm rates)
- Static scenes or dynamic scenes
- Time of execution (scaled by processor speed).

4. IMPACTS

Applications of vision technologies to making artificial intelligence (AI) work in the real world are broad and significant.

- Manufacturing — inspection, assembly, transfer
- Assembly of sensitive devices
- Surveillance, spying, and verification — multi-sensor object recognition and counting. measurement using SAR, Inverse SAR, IR. visible spectrum and color
- Mapping and cartography — terrain, natural and cultural feature extraction
- Automatic target recognition — smart weapons
- Guidance and homing
- Autonomous navigation systems — Autonomous Land Vehicle, Autonomous Underwater Vehicle, Autonomous Air Vehicle
- Field robotics — maintenance, ammunition transfer and loading, runway repair, construction
- Space — space station, planetary exploration
- Transportation — driverless cars
- Exploration of hostile or hazardous environments — hazardous waste dump site cleaning, nuclear accidents
- Medicine — automatic disease detection, diagnosis, microsurgery.

Of these, military, space, and manufacturing are the areas in which

advancement of vision will have the largest impacts. It is not an exaggeration to say that in almost all intelligent machines conceived for Department of Defense (DoD) applications, such as ALV, AUV, smart weapons, surveillance, and field robotics, vision is the component technology most critical for their success. Vision is also important to increased productivity and quality in manufacturing. The broad field of inspection is one of the ripest for applications of vision: Printed circuit board inspection, IC wafer inspection, and size measurements of parts have been the most successful uses of machine vision so far. Advancement of model-based object recognition will greatly increase the use of vision in assembly processes. Vision has a big impact upon space robotics, especially in construction. inspection. and repair in space, and in unmanned or machine-aided planetary exploration or hazardous environment clean-up. Vision for flight telerobotics for the Space Station and the Mars Rover is already actively under study.

Progress in vision can be significant in an incremental way. Our assessment of how vision will be applied in the future follows:

- Short term (3-5 years)
  - more manufacturing applications — inspection. assembly
  - stereo mapping — elevation data; improved resolution and coverage
  - cruise missile mid-term guidance, terminal guidance to target
  - intelligent mobile robots in relatively well-structured environments
- Mid term (5–10 years)
  - mobile robots with simple tasks in hazardous or hostile environments
  - model-based surveillance with multi-sensor data
  - cartography
  - space or underwater inspection
- Long term (10 years or more)
  - completely autonomous planetary or underwater exploration.

Transition of newly developed technologies to the real world is a difficult issue. There is a huge quality gap between what can be achieved with available methods and what is done in industry. Because many US administrators and academics believe (incorrectly) that closing this gap is simply "engineering," and is thus less important and less intellectual, the task is delegated to the "less successful." This phenomenon may be why many successful industrial applications are achieved in Japan, where the "top quality" people often work in industrial research and development and on manufacturing technology. Directing a significant amount of the best forces to such problems is critical to technology transfer, and it could in turn create new funding resources because of new applications opened up.

In advanced military applications, too, transferring laboratory ideas to

the real world has not always gone smoothly. In general, the **DARPA** Image Understanding program has done well in technology transfer: Defense Mapping Agency, model-based vision, Automatic Target Recognition, later parts of ALV activity, etc. Team efforts (such as ALV) are a possible mechanism for a smooth transition. Still, there are problems. First, enforcing the teamwork is difficult: individual groups do whatever they want to do. Second, the demonstration of system integration, which was thought to be the mechanism of technology transfer, tends to become a goal in itself. For their initial series of ALV demos, for example, the integration contractor had to do almost everything themselves by shopping around and combining existing technologies rather than testing, advancing, and integrating newly developed ideas.

System building by a team effort of research laboratories and industries in a focused application area is still one of the best mechanisms for technology transfer. For it to work, however, careful planning is required. First of all, good choices must be made of problems and scenarios that have real-world relevance. The problem must be challenging and deep enough to demand the substantial advancement of technology for its solution, rather than to promote ad hoc short-term solutions. **At** the same time, scenarios must be developed in which progress can be evaluated and become visible. Second, the adequacy of technology should be subjected to both analytic evaluation and experimental evaluation. Mechanical evaluation with a so-called "standard set" of images is often inappropriate. Researchers should take part in the evaluation because they are painfully aware of their systems' strengths and limitations. Third, transfer of the component technology must not be attempted prematurely. Further development can often be done most efficiently by the original researchers, rather than by system-integration contractors.

One way to promote transfer of technology to the real world is to give researchers the primary responsibility for pushing the technology closer to system integration by providing incentives (funding) or facilities (or subcontractors) to do the necessary real-time implementation, extensive testing, and so on.

5. CONCLUSION

Intelligent machines need vision ***to*** deal with the real-world problems. Application areas in which advances in vision will have greatest impacts are the DoD domain (reconnaissance and guidance), space (construction and exploration), hazardous environments (nuclear, waste clean up), and manufacturing (manipulation and inspection).

At the same time, vision is extraordinarily difficult. We are unlikely to

see a single breakthrough in the near future that will completely solve vision problems or change our view of vision. There may be no such thing as "the principle of vision" from which we can derive solutions for all vision problems. Still, incremental progress in vision, in either basic or applied techniques, when combined with advances in tools, such as the availability of powerful parallel computers and special-purpose VLSI modules, may lead to dramatic leaps in demonstable capabilities.

Based on these observations, our recommendations are:

- Provide steady funding of basic research over an extended period. This should include support of necessary infrastructures.
- Push the technology to the level of system and implementation by researchers. Provide financial (funding) and perceptual (image) incentives for scientific engineering work.
- Set focused, long-term, challenging. and visible application goals. such as the use of robot vision in outdoor, space, or hazardous environments. Careful planning is required to promote a more cooperative and progressive process in development and technology transfer. Development projects with contractors must be organized in smaller steps. including testing and evaluation.
- Promote communication between the vision research community and other closely related technology areas, including computer architecture, VLSI, and sensors.