Trends in Computer Vision

Introduction

Computer vision is the task of making computers understand what is happening in the environment they "see". This has been a fascinating subject of research for many decades. Behaviourial vision, active vision, machine vision and stereo vision are some examples of the same. David Marr defined 'vision' as achieving a particular visual task and this task-oriented, computational approach has been the basis of machine vision. Research on human vision has revealed that 'understanding the environment' is extremely complex, as such, one is yet to exploit the same in the realm of computer vision.

Whereas computer technology has progressed rapidly from programs written in procedural languages like Assembler, Fortran, Cobol, C to object-oriented languages like C + +, Prolog, Lisp, Sybase, Ingress etc., from Von Neuman's sequential machine to parallel processing work stations to supercomputers, from stand-alone nodes to client-server architectures, the progress on computer vision has been rather slow. Pavlidis (1992) notes that we are yet to realize a good 'edge detector' or a good 'segmentation' algorithm despite the best efforts from dedicated researchers over the decades.

However, the situation is not all that bleak. Computer vision is being increasingly applied successfully in a wide variety of areas.

(1) Motorola Inc. uses machine vision for quality control wherein billions of semiconductors are processed in a very short time span – a humanly impossible task.
(2) The US Postal Service has been using OCR machines since the 1970's. This system has now evolved into an automated sorting process that automatically locates the address block and identifies the ZIP code and sorts accordingly – a typical commercial exploitation of document image processing technology. Bar code reader is another typical example.
(3) Advances in 'image acquisition' hardware, i.e. sensor technology from high resolution vidicon cameras to precision opto-mechanical-electronic scanners and charge coupled device arrays of high sensitivity, resolution and precision, coupled with advances in software like preprocessing, segmentation, feature extraction, recognition and description, have directly contributed to commercial exploitation of remote sensing.
(4) Similarly, defence-aided research has contributed significantly to 'real-time vision' applications including robotics and autonomous land vehicles. These applications are computer-intensive, demand predictable performance and capability to interact with the environment (as in active vision) through appropriate interfaces with the sensors. Parallel processing is generally accepted as necessary to support real-time image understanding applications. However, 'which type of parallelism' is still an open relevant question.

The recent advances in hardware are being exploited to incorporate reasoning between different vision levels. University of Massachusettes and Hughes Research Laboratories have worked together to develop a hardware architecture called image understanding architecture (IUA) using VLSI technology, that uses content addressable array parallel processor (CAAPP) at low level, intermediate communication associative processor (ICAP) for intermediate level and symbolic processing array (SPA) at the higher level of vision. In the Indian context, parallel processing is currently restricted to low-level vision problems like edge detection, classification, transforms of images etc. Refresh memories of $1\,K \times 1\,K \times 1\,K$ with 4 bytes/voxel ($=4$ gigabytes) are expected to be available in the near future, with fast raster manipulation algorithms. Further, there is also a trend to differentiate between image supercomputers and computational supercomputers – with the human visual cortex and the brain as an analogy.

Machine recognition systems consist of two parts – a recognition engine and a model library (Besl & Jain 1986; Chin & Dyer 1986). Since these systems are required to identify the objects in the scene, understand the environment, and act accordingly, 'data' description plays a critical role and is based on (a) generalized cylinders, (b) multiple 2-D projections, (c) skeletons, (d) generalized blobs (volumes), (e) spherical harmonics, and (f) overlapping spheres. The recognition engines are similarly classified as being based on (a) local feature focus methods, (b) global feature focus methods, (c) interpretation trees, (d) pose clustering, (e) alignment and (f) geometric hashing.

The following are some recent interesting, encouraging developments that are expected to play a critical role in the areas of object recognition and 'information' retrieval. Stein & Medioni (1992) use 3-D scene data as an alternative to 2-D aspect graphs. They describe 3-D objects based on combinations of surface patches and curves of discontinuities along with methods to build and recognise them from specific instances. Swain *et al* (1991) use colour histograms of multicolored objects to provide robust, efficient use for indexing into a large database of models. These colour histograms are stable object representations even in the presence of occlusion, change in view etc. Hung *et al* (1991) use statistical decision theory to drastically reduce the storage and associated processing with sequential Bayesian estimators for surface parameters wherein information extracted from previous images is summarised in a quadratic form. Bobick & Bolles (1992) present a hybrid representation scheme that supports the evolution of descriptions progressing from 2-D blobs to complete semantic models such as bushes, rocks and trees. The order of representing these descriptions of an object is as a lattice of representation of space, after the object has been described a multiple number of times and is termed stable in the statistical sense. The same is enhanced by sets of explanations for deviations from expected measurements.

Steven Zuker (1993) describes a two-stage approach to get global curves i.e., boundaries, by exploiting information from local edge detectors, its tangents and curvature. He uses these global curves to illustrate that the problem of deformation in shape can be tackled. Eric Saund (1992) exploits many geometric properties and spatial relationships at many redundant levels of abstractions with a two-level methodology. At the first level the shapes are symbolically manipulated with information about relative locations and sizes while the pictorial organization inherent to shape's spatial geometry is preserved. At the second level, 'dimensionality reduction' devices are used to interpret shape tokens in terms of their 'deformation classes' with flexibility to account for deformations in their forms. Dickinson *et al* (1992) use complex primitives instead of simple 2-D primitives like lines, corners, inflexion etc., so that the process of searching to recognize a model becomes efficient. Grimson (1992) has shown that a tree search to locate objects in a cluttered environment is a quadratic with number of models and data features, if all the data are known to come from a single object, a cubic if one can group the data into subsets likely to have come from a single object, and exponential otherwise. This implies that naive approaches to library indexing are likely to carry expensive overheads.

In navigation applications, 3-D scene interpretation is important. Dickmanns & Mysliwetz (1992) have presented a method for autonomous road-following to reduce the death toll and economise losses in road traffic, by computing the local differential geometry of a road from motion information with an autonomous vehicle driven at speeds of roughly 100 km per hour on highways and at somewhat slower speeds on hilly rural roads.

Similarly, Zhang & Fangeras (1992) have computed 3-D motion from a long sequence of stereo frames with a two-level bottom up approach. At the first level, they use parallel processing to track 3-D tokens (lines, curves, segments etc.) from different frames and estimate the associated kinematics. At the second level, tokens are grouped into objects based on their kinematics to overcome problems of occlusion, appearance and disappearance of tokens. A prediction-matched approach is used to track the objects in which the problem of multiple matches is handled.

Artificial neural networks are being increasingly used to emulate the human visual system for handling a wide variety of vexing image processing problems associated with rotation, scaling, occlusion and restoration. Amongst the recent advances (Katsuji *et al* 1991) in this area are architectures that include plural networks and movement of the view point. Some recent salient functions for extraction of patterns include incorporation of filter networks to extract rectangular regions where a noticeable pattern in contained, position networks and size networks to detect the position and size (vertical and horizontal lengths) of the target pattern and framework networks to fix the target network exactly. Similarly the function for pattern recognition includes categorising networks that account for errors in position and size.

The role of genetic algorithms in vision is still not clear, but these appear to have some potential in the near future. Here, a problem solving system is defined to consist of genes, a performance function and a genetic algorithm (Nie & Surkan 1991). A string of bits (i.e. patterns) function as a gene (the bits are initialised with random numbers) and correspond to the processing elements, and a genetic algorithm is directed by the performance function tc evolve better offspring from the parent genes. Nie & Surkan (1991) illustrated this approach for evaluating the square root of a number and quote from Charles T Walbridge (1989) who has suggested that "chip" chromosomes for VLSI designs could be evolved and surviving organisms would mate to produce offspring having chromosomes for the most efficient chips. These result from a combination of different parts of different chromosomes in the offspring to preserve the characteristics derived from parents. Another very important application area for genetic algorithms is to build an explanation mechanism into neural networks (Mathew & Ping 1991) that is similar to those available in expert systems. One could then ascertain how a classification was obtained or ask about the differential between the current and another classification – currently a serious limitation.

Amongst the other recent trends in vision, the most notable is the incorporation of gaze (eye movements) (Olson & Coombs 1991) and vergence control (process of adjusting angle between cameras) (Rimey & Brown 1991), very useful strategies for binocular robots for fixating the object first and then moving the arms of the robots. Lately, the realisation of "virtual reality" for robotics applications has a great future.

In this special issue, a set of twelve papers have been chosen, broadly representing the current research areas in the Indian context. The paper by Ram Chellappa and Rosenfeld addresses the current issues in computer vision by noting that, in general, vision problems are ill-defined, ill-posed and computationally intractable. They suggest that useful solutions can still be obtained by limiting the domain of application, carefully choosing the task, using redundant data and by applying adequate computing power. Methods of designing such solutions are the subject of the emerging discipline 'vision engineering' and can often serve as computational models for biological visual processes as well. Rajan describes the frequently used operations in image processing like thinning, edge detection, segmentation, erosion, dilation etc., by treating a digital image as a cellular automata configuration and an operation as an evolution of the automaton – an approach of recent origin.

Das & Chatterji present a survey of distances in the geometry of digitized spaces of arbitrary dimensions and related issues in digital picture processing. Chattopadyay & Das develop a unified approach to quantize and reconstruct curves with two parameters that is useful in solving the domain finding problems as well. Venkatesh *et al* use Hermite polynomials for decomposing images into a multilayered representation of image called 'wavelet arrays' and also give a procedure to extract zero crossings at different scales. Shankar Pal & Lui Wang address the problem of image representation and retrieval for shape analysis and template matching, uncertainty management in recognition and creating new images of various poses by improving the compact representation of fuzzy medial axis transformation (FMAT). Arun K Pujari presents an active vision volume intersection based approach to recover shape from 2-D images. Krishnan & Kiron K Rao use time series techniques for shape recognition.

Ashfaq Khokar & Viktor Prasanna present parallel implementation of stereo matching and image matching with linear features as matching primitives. Sengupta & Sahasrabudhe address the problem of reducing the sparsity of depth points by orienting the epipolar line of the cameras in a direction that maximizes the number of feature points. Subhodev Das & Narendra Ahuja address the problem of surface reconstruction for stereo images for scenes with large depths using a four-step process. Arun Agarwal describes the organization, interaction of different modules of document image processing along with algorithms to extract higher level representations from visual sketches, physical and logical layouts and block primitives.

A list of recent references is included to serve as a ready reckoner for the 'state of art' in this line.

June 1993

B L DEEKSHATULU
Guest Editor

References*

Adelsomn E H, Wang J Y A 1992 Single lens with a plenoptic camera. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 99–106

* This list also includes suggestions for further reading

Ballard D H, Brown C M 1992 Principles of animate vision. *Comput. Vision Graph. Image Process. Image Understanding* 56: 3–21

Baribeau R, Rioux M, Godin G 1992 Color reflectance modelling using polychromatic laser range sensor. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 263–268

Besl P, Jain R C 1986 Three dimensional object recognition. *Comput. Surv.* 17: 75–145

Bobick A F, Bolles R C 1992 The representation space paradigm of concurrent evolving object descriptions. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 146–156

Brooks M J, Chojnacki W, Kozera R 1992 Impossible and ambiguous shading patterns. *Int. J. Comput. Vision* 7: 119–126

Campani M, Verri A 1992 Motion analysis from first-order properties of optical flow. *Comput. Vision Graph. Image Process. Image Understanding* 56: 90–107

Chen C-H, Mulgaonkar P G 1992 Automatic vision programming. *Comput. Vision Graph. Image Process. Image Understanding* 55: 170–183

Chin R T, Dyer C R 1986 Model based recognition in robotics. *Comput. Surv.* 18: 67–107

Choudhury A, Rauka S 1992 Parallel processing for computer vision and image understanding. *IEEE Comput.* 25: 7–10

Dickinson S J, Pentland A P, Rosenfeld A 1992 3-D Shape recovery using distributed aspect matching. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 175–198

Dickmanns E D, Mysliwetz B D 1992 Recursive 3-D road and relative ego-state recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 199–213

Eberhart R C, Dobbins R W 1991 Designing neural network explanation facilities using genetic algorithms. *Proc. Int. Joint Conf. Neural Networks*, pp. 1758–1763

Funt B V, Drew M S, Jian H 1991 Color constancy from mutual reflection. *Int. J. Comput. Vision* 6: 5–24

Gennery D B 1992 Visual tracing of known three dimensional objects. *Int. J. Comput. Vision* 7: 243–270

Grimson W E L 1992 The cost of choosing the wrong model in object recognition by constrained search. *Int. J. Comput. Vision* 7: 195–210

Grimson W E L, Huttenlocher D P 1992 Introduction to the special issue on interpretation of 3-D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 97–98

Heeger D J, Jepson A D 1992 Subspace methods for recovering rigid motion I: Algorithm and implementation. *Int. J. Comput. Vision* 7: 95–177

Hung Y-P, Cooper D B, Cernuschi-Frias B 1991 Asymptotic Bayesian surface estimation using an image sequence. *Int. J. Comput. Vision* 6: 105–132

Ishiguro H, Yamamoto M, Tsuji S 1992 Omni-directional stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 257–262

Katsuji I, Kazutoshi G, Yoshiki U 1991 Pattern extraction and recognition for noisy images using three layer BP model. *Proc. Int. Joint Conf. on Neural Networks*, pp. 262–267

Kweon I S, Kanade T 1992 High resolution terrain map from multiple sensor data. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 278–292

Lu Y, Jain R C 1992 Reasoning about edges in scale space. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 450–455

Marsh B, Brown C, Le Blanc T, Scott M, Becker T, Quiroz C, Das P, Karlson J 1992 The Rochester checkers player: Multimodel parallel programming for animatic vision. *IEEE Comput.* 25: 12–19

Mathew B I, Ping T C 1991 Some experiments on use of genetic algorithms in a Boltzman machine. *Proc. Int. Joint Conf. on Neural Networks*, pp. 2645–2652

Nelson R C 1991 Introduction: Vision as intelligent behaviour – An introduction to machine vision at the University of Rochester. *Int. J. Comput. Vision* 7: 5–9

Nelson R C, Polana R 1992 Qualitative recognition of motion using temporal texture. *Comput. Vision Graph. Image Process. Image Understanding* 56: 78–79

Nie X, Surkan A J 1991 Genetic algorithms: Hints from biological science. *Proc. Int. Joint Conf. on Neural Networks*, pp 426–429

Olson T J, Coombs D J 1991 Real-time vergence control for binocular robots. *Int. J. Comput. Vision* 7: 67–89

Pavlidis T 1992 Why progress in machine vision is so slow. *Pattern Recogn. Lett.* 13: 221–225

Rimey R D, Brown C M 1991 Controlling eye movements with hidden Markov models. *Int. J. Comput. Vision* 7: 47–65

Saund E 1992 Putting knowledge into a visual shape representation. *Artif. Intell.* 54: 71–119
Seibert M, Waxman A M 1992 Adaptive 3-D object recognition from multiple views. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 107–124
Shakunaga T 1992 3-D corridor scene modelling from a single view under natural lighting conditions. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 293–297
Stein F, Medioni G 1992 Structural indexing: Efficient 3-D object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 125–145
Straforini M, Coelho C, Campani M, Torre V 1992 The recovery and understanding of a line drawing from indoor scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 298–302
Swain M J, Ballard D H 1991 Color indexing. *Int. J. Comput. Visior* 7: 11–32
Tistarelli M, Sandini G 1992 Dynamic aspects in active vision. *Comput. Vision Graph. Image Process. Image Understanding* 56: 108–129
Vaillant R, Faugeras O D 1992 Using external boundaries for 3-D object modelling. *IEEE Trans. Pattern Anal. Mach. Intell.* 14: 157–173
Walbridge C T 1989 Genetic algorithms: What computers can learn from Darwin. *Technol. Rev.* 92: 46–53
Weems C C, Riseman E M, Hanson A R 1992 Image understanding architecture: Exploiting potential parallelism in machine vision. *IEEE Comput.* 25: 65–68
Zhang Z, Faugeras O D 1992 Three dimensional motion computation and object segmentation in a long sequence of stereo frames. *Int. J. Comput. Vision* 7: 211–241
Zhao F 1991 Machine recognition as representation and search – A survey. *Int. J. Pattern Recogn. Artif. Intell.* 5: 715–747
Zuker S W 1993 Endowing AI with vision: A biological and computational perspective. *Curr. Sci.* 64: 407–418