ROBOT VISION AND VISUAL INSPECTION

T. Kasvand
National Research Council
Ottawa, Ont.

ABSTRACT

We take for granted our own ability to see and to understand what we see, resulting in a world around us consisting of scenes, objects, situations and activities. In robotics vision is also used to provide information about object identity, location, state, etc. A robot provided with an intelligent vision system does not need accurate parts positioning and special fixtures, since it can be made "aware of" what it is doing. A fair number of specialized robot vision systems are already used in industry and more are under development. However, the present robot vision systems can only "understand" very simplified scenes, especially if economic constraints are applied. At our present level of knowledge and technology, we do not know how to design electronic vision systems that can in general compete with our own vision.

The article describes the robot vision and visual inspection problem in narrative terms, to serve as an explanation of the present trends and may provide a philosophical basis from which to view these problems, in order not to become overwhelmed by the amount of literature available.

KEY WORDS: Robot vision, visual inspection, computer vision, image processing, pattern recognition.

INTRODUCTION

We ourselves and our cohabitant species are extremely well equipped with sensors for vision, touch, smell, hearing, taste, and internal sensors for hunger, thirst, pleasure, pain, and so on. These sensors are "backed up" by elaborate signal processing, enormous memory, and inherited as well as learnt responses. We "operate" on our environment by using our very dexterous hands, legs and body to carry out the appropriate actions.

Our eyes receive a multi-dimensional signal from the environment which contains spatial, frequency (colour), and time information. We understand these signals at a glance. Such signals manifest themselves as understandable composite scenes of objects and situations to which we attach meaning, dynamic behaviour, beauty, etc. We take our own ability to see for granted, do it effortlessly and we instinctively feel no reason to believe that the interpretation of such a signal may be a very complicated task. In fact, we take pleasure in the scenery around us, enjoy music, and so on. Consequently, those who have not tried to write computer programs for understanding speech, images and other such problems, do find it very hard to believe that problems even exist! Their reactions may be summarized by: "If it is so obvious and simple for me, why can't you write a program"? However, for those who have tried, the proverb "familiarity breeds contempt" is not true in this context, and should be replaced by "familiarity breeds respect".

Our visual information processing system (vision) is extremely elaborate and sophisticated (Polyak). About one billion (1,000,000,000) years of "R & D" (evolution) has gone into it. Vision and the other senses have been optimized to facilitate survial in our natural environment. Our vision together with the other senses forms an exceedingly elaborate information handling system backed up by large memory and many information processing and behavior strategies. This is all "below the surface" and we are not directly aware of the complexities, psychologists excepted.

Research in image processing, pattern recognition, scene analysis, etc., has been carried on for more than twenty-five years. However, even though much progress has been made in specific application areas, there is still no generalized methodology or theory of how to extract information from and to "understand" general visual scenes by mechanized means. The problem has proved to be surprisingly difficult.

For better insight, one may compare the capabilities of mechanical or computerized image processing systems with our own visual abilities. The result is that if our own innate abilities are totally lacking, as for example in image reconstructions, the mechanized results are considered to be excellent. If our own abilities are meager, as for example in image restoration and enhancement, the computer results are judged as very good. If our own and the mechanized capabilities are about the same, as for example in optical character reading, we are not greatly impressed except possibly by the speed of the mechanical device. In the understanding of complex scenes, recognition of objects in such scenes, etc., where our own abilities are excellent, comparatively speaking, the computer results are so far negligible.

There is thus a profound difference between the abilities of our own vision and that of the mechanized or computer vision systems. Understanding of complex scenes is easy for ourselves but nearly impracticable for computer vision systems. The environment of our production systems, machines, assembly lines, etc., are all designed to suit our own visual (and tactile) systems. The visual scenes generally are extremely complex. Even bins or buckets full of machine parts, which are a typical sight in a small volume production system, constitute a problem which has not yet been solved satisfactorily in a general sense. The ultimate nightmare of an intelligent robot system designer is likely to be a disorganized repair shop, for fixing cars, for example. One has to look for tools to disassemble a mechanism, replace defective pieces, inspect and clean the remaining ones and then put the mechanism back together again by using whatever tools are handy and at least semi-appropriate. We attach no particular intellectual requirements to the performer of such tasks. However, the required scene analysis, image understanding and object recognition, the decision strategies and planning, the required dexterity and control of two manipulators (hands), etc., is far beyond our present intellectual and technological know-how in computer programming, hardware design, etc.

CATCH 22

Our production and assembly systems, repair shops, etc., are designed for use by ourselves. In these designs normal human abilities are taken for granted. Thus, the information processing required to find the objects (tools, parts), and so on, is assumed to be available and included in the cost of the worker. The popular belief, created by movies, television, numerous articles and, of course, science fiction, replaces the worker with an extremely versatile general purpose robot. This "all singing and dancing" robot is "just around the corner". Such a belief actually requires but little imagination. From a purely mechanistic point of view, we ourselves are built according to a "standard design". We extend our own sensors by a variety of instruments and signal transmission and processing techniques. We amplify our muscle power by using machines, extend our range of tolerable environments by appropriate protective clothing, and so on. Obviously, the "all singing and dancing" robot should behave likewise! A robot, however, need not be limited to two eyes, two arms, etc., but should rather be designed to suit a particular process and environment. The robot need not only use visible light for its vision, but also radar, x-rays, particle beams, accoustic signals, etc. The possibilities are limitless, being only bounded by our imagination and scientific and technical know-how.

In realistic applications of robots, however, we are limited by speed, reliability, and cost effectiveness (Heer). At the present level of science and technology the general purpose robot with abilities resembling our own is only a dream, which we neither can build nor afford. However, we can design many types of very effective special purpose devices ("robots") which will take over jobs presently performed by humans. Our everyday tasks are actually extremely complex, even though we call them "routine", and unless we can build some sort of a self-learning robot which learns the intricasies of various tasks on its own, we have to study the details of each task and program the machinery ("robot") accordingly.

The rather unstructured environment of a smaller car repair shop or a hobby workshop are typical examples of situations where one worker tackles a task involving a multiplicity of problems. The worker is actually happy in such an environment, especially if it is his hobby shop. If many workers are involved in a task then confusion results, since they get into each other's way, the tools are never in the place where one worker left them, etc. Even the Romans knew that for mass production a structured environment is needed where each worker repetitively performs the same small subtask.

A modern assembly line, where even each motion of the worker is optimized via time and motion studies, reduces the human to an automaton (robot). The resultant semiconscious state during work leads to boredom, carelessness and frustration. Movies such as "The Modern Times" and "Metropolis" may be worth seeing again, even though it would be much better for the prospective "roboticist" to actually work for a couple of weeks at the particular task, in the same environment, at the same speed and on the same pay as the worker! An interesting comparison is likely to result: The tasks on a modern assembly line are simplified to a few motions. The human finds them boring while the same tasks are still frequently too complex for the present day robots. There are also a very large number of exceptionl situations which the worker handles as "a matter of course", such as a broken tool, a defective component, a missing item, etc., each of which has to be anticipated for a robot. There is also a large information waste.

During the machining of a piece, for example, its position and orientation, as well as its identity are known. All this information is lost when this piece is thrown into a bin. To recover the information about the location, orientation and identity of a piece can be a very costly computational exercise, if done by computers. In robotics this cost has to be considered also and an optimal strategy chosen.

It should be noted that in the so-called "hard automation", typified for example by a screw cutting machine or a transfer line, the object being produced is kept in known positions and orientations throughout the process, otherwise this form of automation would be impossible. There are, of course, many ingenious devices (feeders, shaking tables, etc.) for retrieving by mechanical means the positions and orientations of simple objects. All this adds to the cost. The extremely high cost of hard automation is thus at least partially due to the retention and/or recovery of the information about the position, orientation and identity of the object being produced.

The so-called "robotics" or "flexible automation" is expected to decrease the cost of automation, besides contributing additional flexibility to the production process. However, it should not be forgotten that our ability to design and build sensory systems for robots and to improve their dexterity is both limited and also costly. Large information processing tasks also tend to be rather slow on general purpose computers. Special hardware can increase the speed significantly but, naturally, at much added cost. Thus judicious compromises in the applications of robots are mandatory if successful economical results are to be achieved. One step towards flexible automation is to redesign both the products and the methods of handling and assembling the parts. Another step is the construction of "flexible manufacturing cells" where several numerically controlled cutting and assembly machines are linked by conveyors, robot trolleys or robot arms for parts transport. The "cell" operates as a unit for the production of a whole range of objects and is easily reconfigurable and reprogrammable. The least unorthodox method is to surround existing production machines with robot arms and conveyors. The future automated factory, however, is unlikely to look like the present one with the workers replaced by robot arms.

The "catch 22" is that there is no "royal road" to the use of robots. Given the reliability, speed and cost constraints in ordinary manufacturing situations, we have to build many special purpose systems, and anticipate and solve a myriad of "nitty-gritty" problems. This is now happening. Many special purpose systems are already on the market, and even more in the various laboratories. The idea that a general purpose intelligent robot will be here tomorrow is just a dream or nightmare. If, however, by some breakthrough a truly intelligent robot could be designed, why should it be interested in doing our dirty work?

THE ROBOT PROBLEM

Research and development of robot systems has a long history. They have been built for the entertainment of kings and queens. In the 1700's beautiful robots were built to play piano, write, walk, talk, and some were built as early as 500 BC (Reichardt). More recently robots with vision have been built for the study of psychological questions, such as the Gray Walter turtles, for the study of artificial intelligence problems, such as the "Shakey" at Stanford Research Institute, and numerous other so-called hand-eye systems. Some of the more ambitious ones were the two-armed and 8-eyed Hitachi robot for assembling vacuum cleaners, the tool-using robots build at the Electrotechnical Laboratory in Japan, which could use hammer and saw and put nuts onto bolts, the robots at Edinburgh that could assemble simple objects, given a pile of parts, etc. Industrial applications were started in earnest more than a decade ago. By now the number of laboratories and systems is very large.

Exactly like ourselves, a robot needs a whole variety of sensors for its internal and external environment. The signals from these sensors have to be processed in order to take appropriate actions. Even though it may be inappropriate to compare ourselves to a robot, such a comparison offers an easily comprehensible base of reference.

Our own vision, as well as the vision system for a robot, serves as a non-contact sensor for locating an object, i.e., to determine its identity, position and orientation in space. Identification requires pattern recognition, location in space requires distance measurements. Inspection requires knowledge of the ideal form of the object and the nature of the defects. Defects, however, can be very numerous and varied. These steps are usually followed by guidance instructions to the manipulator (hand) on how to grab the object. When the manipulator (hand) is close to the object, vision also provides an error signal indicating the difference in position between the manipulator gripper (fingers) and the object. Touch, i.e., force and friction (slip) sensors come into play as soon as contact is made with the object. A tactile form of pattern recognition also occurs. This is usually followed by further visual inspection to ensure that the object has been correctly grasped to bring it to the desired new location, and so on. It should be observed that the information processing and decision and control tasks are quite complex, and dependent on the objects, their state, unforeseen mishaps and complications, and so forth.

The speed, cost and reliability constraints on robot vision are most severe in the ordinary production environment where human workers, in a sense, directly compete with the prospective robot systems. In such situations the cost of the entire robot system, including the manipulator, its feeding devices, etc., cannot cost much more than about two years' wages and benefits of the workers it replaces. Even though robots can work for 24 hours per day, need no lunch breaks and holidays, etc., the two-year figure has frequently been quoted as the approximate price of a robot that a manufacturer may be willing to pay. The robot vision system in such situations cannot cost much more than about one tenth of the robot system, say $5000 to $20000. Furthermore, the vision system has to operate at least as fast as the human, say in one tenth of a second to a second. These are the main reasons why the present commercially available "robot vision systems" are so simple and specialized.

The cost and speed constraints, however, are far less important for environments which are more or less hostile to humans, or where eye strain, boredom, and so forth, become significant factors. Such hostile environments occur, for example:

1) In soldering of leads to micro-circuits, circuit inspection, and so on, where eye strain becomes significant.

2) In welding and spray painting, where the fumes become intolerable.

3) In mining applications where the environmental support for human life may be more costly than the use of "rather clever" robots.

4) In deep sea operations where environmental support may not be feasible.

5) In atomic reactors where radiation levels are intolerable.

6) In outer space for constuction, repair, exploration, etc.

All these areas and many more require fairly sophisticated robot vision systems and can be very fertile grounds where research and applications can go hand-in-hand. The image processing problems, manipulator control, etc., however, do not necessarily need to differ much from those required for the factory environment.

INDUSTRIAL REQUIREMENTS FOR ROBOT VISION

As may be deduced from the rather lengthy preliminaries, the intelligent all-purpose robot is not feasible at present. We are thus forced to analyse the various problems in detail and, at best, only hope to solve categories of robot vision problems, rather than each specific problem individually.

The tasks for robot vision may be divided into general representative categories, i.e.:

* Manipulation and control functions.
* Inspection for quality control.
* Overall control and safety.

A still rather general breakdown of the desired functions for visual sensing and interpretation in the industrial environment may be the following (Rosen):

* Recognition of workpieces and assemblies.
* Determination of position and orientation of workpieces or assemblies relative to given coordinates.
* Extraction and location of salient features of a workpiece or assembly to establish spatial references for manipulator control.
* In process inspection (verification that a process or assembly has been or is being satisfactorily completed).
* Safety to equipment and personnel, in case of malfunctions and to avoid accidents.

A further breakdown of these categories into task areas may be as follows (Rosen):

A) Manipulation and Control Functions.

A1: For use in acquisition of workpieces (picking up operations)
- A1.1: Workpieces on conveyors.
  - A1.1.1: Lying on belts, in stable positions, unobstructed view,

separated or lightly touching.

A1.1.2: Hung from hooks, partly constrained, i.e. swinging and slightly rotating.

A1.2: Bin picking, workpieces in a container.

A1.2.1: Random spatial order, a "bucket" full of jumbled and possibly interlocking pieces.

A1.2.2: Partly organized, a container of arranged pieces, not separated, "egg-crate" or "chocolate box" type packaging.

A2: For use in manufacturing processes, robot uses a tool or holds and moves the piece. Accurate path control from visual (tactile, etc.) information. Examples: deburring, cutting, finishing, flash removal, liquid casketing, process control, sealing.

A3: For use in assembly operations.

A3.1: Fastening operations. For example, arc welding, bolting, gluing, nailing, riveting, spot welding, stapling.

A3.2: Fitting operations. For example, mating of parts, parts presentation.

A3.3: Inspection during process.

B) Inspection for Quality Control

B1: Quantitative measurements (mensurations) of critical dimensions.

B1.1: Measurements of critical dimensions of workpieces to stay within tolerances given.

B1.2: Measurements of tool wear for adjustment or replacement of tools.

B2: Qualitative and semiquantitative measurements.

B2.1: Optical character reading of labels and bar codes, inspection of labels.

B2.2: Sorting. Selection and identification of workpieces for orderly packing, presentation and inventory control.

B2.3: Integrity and completeness of work pieces and assemblies. Are all parts present in an assembly? Are the parts undamaged? Are all parts correct and in correct positions?

B2.3.1: Overall integrity and completeness. Approximate size and location of key features.

B2.3.2: Nature of defects, warping, cracks, burrs, broken parts, pits, etc.

B2.4: Cosmetic and surface finish, stains, smears, surface blemishes, and discontinuities, colour inconsistencies, etc.

C: Overall Control and Safety

C1: Overall system control to ensure that the whole automated process works correctly. This is mainly a problem of programming the central computer(s), rather than vision alone.

C2: Safety to ensure that the machinery does not damage itself and the personnel present.

The subdivision of the tasks may be continued down to the "nitty-gritty" technical details of hardware and programming algorithms. However, very generally stated, only for categories A1.1.1 (for non-touching objects), B1.1 and B2.1 have the solutions reached some commercial feasibility. At the present level of technology most of the subareas need to be treated as individual problems in computer vision. The literature contains numerous articles on such special solutions to special problems. From the practical and economical point of view, this is the only feasible approach. The general problems of artificial intelligence and robotics are exceedingly interesting and important and will form the foundations for more generalized solutions.

INDUSTRIAL ROBOT VISION

Given the situation as outlined in the previous sections, what problems can present robot vision solve?

One attempt to answer this question could be based on reviewing the present commercially available devices and those in development laboratories. Numerous books, magazines, conference proceedings, and brochures by manufacturers are available. Even though there is a veritable "flood" of articles, one should not be too surprised, however, to find that the manufacturers are reluctant to reveal the "intimate" details of their systems. Thus, it may be difficult to predict how a particular device may work in a novel situation.

The spectrum of devices is quite varied, some of which one may be reluctant even to call "vision systems". As a trivial example, a door that opens automatically when we approach it, does the door have "vision" just because we walked through a beam of light, stepped on a contact, or triggered a proximity sensor? There are systems that compute the gray level histogram of the image (Banard), usually within a preset window, or the gray level profile along a path, or a silhouette, and compare the resultant one-dimensional signals against given thresholds or prestored values. Other systems compare a two-dimensional image, or parts of it, pixel by pixel against prestored templates (Hsieh, Kashioka). The more general commercially available vision systems are a compromise between well-known simple algorithms (mostly 15 to 25 years old, Dodd, Duda, Hall, Kasvand, Rosenfeld) for binary or black and white images and what the designer of the vision system thinks is needed by the industry and what can be built at modest cost. The "grand daddy" of many such systems is the SRI vision system (Bolles, Nitzan). It should be emphasized, however, that these are very realistic approaches to the computer vision problem, given the constraints of cost and speed. All that is needed is that they do the job for which they have been designed (Bitter, Boykin, Perkins, Shirai, Takeyasu, Tsuji, Ward).

Another way of trying to answer the question of what robot vision can do is to approach the problem from a conceptual point of view, without neglecting the constraints of speed, reliability and cost. However, this approach requires a fair amount of practical and theoretical knowledge about image processing, pattern recognition, etc. Futhermore, one should never forget that our own vision is exceedingly good at understanding a scene and the objects in it, which can easily lead us to erroneous conclusions about the difficulty of a computer vision problem.

One of the basic questions is: What is it in the scene-understanding problem that makes it so difficult? To prevent the discussion from becoming entirely philosophical, the simplest but by no means satisfactory explanation is that the scenes are too complicated. We have neither the knowledge nor the computational power to untangle the complexities of an ordinary three-dimensional scene. Such scenes contain a multitude of arbitrary objects of any size and in any configuration, illuminated somehow. The scene contains shadows and specular reflections, objects obscure each other unless they are transparent, the objects move, the observer moves, etc. Thus, for any practical application of computer vision, there are essentially three cardinal rules:

1) Reduce the complexity of the scene to the bare acceptable minimum.

2) Control the sources of illumination to further simplify the resultant image.

3) Observe the scene from the most advantageous angles, to yet further simplify the analysis.

This may be called "the triple S rule", Simplify, Simplify, Simplify!

From the opposite side the attack on the computer vision problem consists of improving the computational capabilities of the hardware, reducing its cost, and aiming for generality of the algorithms, i.e.:

* Design mass produceable hardware for vision systems thereby reducing hardware costs.
* Develope generalized and fast algorithms as hardware and/or firmware, to obtain faster operation.
* Write basic image processing programs for general microprocessors, hoping that the microcomputers can be configured as multi-processor and multi-resolution systems.

And if everything else fails, the practical engineer will

* Use whatever tricks that will do the job without any regard to generality.

APPLICATIONS OF THE TRIPLE S RULE

The industrial scenes are far simpler than our natural environment, and can be at least partially modified to simplify the scene even further, thereby reducing the computational requirements placed on computer vision. However, to emphasize again, the greatest constraints on the industrial vision systems are that they have to be fast, reliable and economical. There are many ways of simplifying the scene while satisfying the constraints, such as:

1) The scene is to contain preferably only one, or in general only very few objects.

An object is expected to be in one of its stable positions and presented to the vision system on a flat and hard surface. The total number of objects, including their stable positions, that a usual system can identify is relatively small, say 10 to 100, since the recognition accuracy drops and the processing time increases as a function of the number of objects. A system should be constructed such that it can be "taught" new objects and the old ones can be "forgotten".

2) The objects in the scene have to be easily detectable from the background.

The use of some basic techniques such as gray level slicing (thresholding), colour separation by filters, polarized light, fluorescence, structured light, etc., is common. Thus, in most cases the separating variable is a physical parameter which is characteristic of the object(s). Some flexibility is obtained by making the threshold for the separating parameter variable under computer control. The main reason for this restriction is that searching for known objects in a cluttered scene is computationally expensive, time-consuming and not too reliable. Furthermore, the analysis of gray-level images is complex and time-consuming and the sooner the image of an object can be converted to a binary (silhouette) form the better.

3) If there is more than one object in the scene, the objects should not touch or overlap.

The reason is that the objects are first detected only as unidentified "blobs". On these "blobs" certain measurements are made which are used for identification. Typical measurements are area, length of contour, longest and shortest dimensions, moments, etc. Obviously, any two (or more) objects that touch can form a near infinity of new composite "blob shapes", and the elementary methods used for identification fail.

If the contact area between the objects is relatively small compared to the size of the objects, the so-called shrinking and expanding operators may be used to resolve the objects. This, however, requires both additional time and extra computer storage. Another method is to perturb the overlapping objects (i.e., nudge the unidentified "blob" with the manipulator) thereby hoping to separate them before renewed visual analysis.

4) The desired object must be fully within the scene to be analyzed.

The reasons are similar to the overlapping case above (3), i.e., there is no definite shape to the "blob". This condition, however, is easy to detect since the unknown "blob" intersects the frame or edge of the picture. Thus, a feedback is available for either moving the camera or the object.

5) The objects in the scene are of known size and shape.

Normally the sizes and shapes are somewhat variable depending on the distances and the angles of view. These variations, however, are mainly known from the three-dimensional geometry of the visual environment, since the image is a two-dimensional projection of it. These variations may be obtained experimentally or from a CAD data base.

6) The objects are usually in rather well-known positions and possibly also orientations.

This constraint reduces the amount of search required to find the objects in the image and lessens the number of processing steps for the image. Mechanical feeders and palletizing leave the objects in relatively well-known positions, thereby recovering or preserving information which otherwise would have to be extracted from the image at additional cost and time. In this case also some charcteristic features, for example hole combinations, may be used immediately for recognition, since the objects themselves may touch each other, or the edges of a "chocolate box" packaging method are visible in the image.

In a "cluttered scene", as for example in the image of a chip onto which the bonding wires are to be soldered, the pads and other characteristic features may be detected by matching prestored templates, i.e., subimages of what is expected to be seen in a particular small area in the scene. Template matching is practical only when the orientation, position, and size of the objects to be located is fairly well known in advance. In general, if the position, orientation and size of the area to be inspected is well known in advance, many specific algorithms can be applied directly. This is the case in most inspection situations, for example GM's Keysight (Perkins), screw thread inspection, etc.

7) The objects are frequently made "two-dimensional" by illuminating them from underneath or from the side. The machine thus only sees the shadow of the object.

The signal-to-noise ratio can be made large, resulting in "clean" images and the image signal can be easily sliced (thresholded) to produce binary or black and white images for analysis. This creates black-and-white silhouettes of the objects, which are much easier to process than gray level images. With this method one avoids shadows, reflections, and surface texture. The techniques for analyzing binary images have been studied for more than twenty years. Consequently, many algorithms are available.

8) Each stable position of a three-dimensional object, when placed on a flat surface, after back lighting, produces its own silhouette and is treated as a separate "two-dimensional" object.

In the image analysis system a 3D object with say five stable positions is thus represented by five different binary images (some of which may be the same) linked to one common source object. Objects which are supported by soft underlays or hang from hooks may present too many different silhouettes for these methods to be feasible. Of course, flexible objects are also excluded, except in a few cases where a feature remains constant despite flexing, such as for example the length of a piece of string.

9) In some situations it is possible to mark the objects in various ways during earlier stages of processing.

This is particularly true for larger objects, in particular if the markings do not interfere with say the appearance of the assembly. The marks should be properly designed such that they can be located more easily and recognized instead of the object itself, at least as the first step in image processing. The use of a controllable light source and retro-reflectors produces an easily thresholdable image where only the marks remain visible. The marks can also be used to determine the position and orientation of the object, i.e., its 6 degrees of freedom. For details see photogrammetry (Kratky).

10) There are, of course, many other constraints and considerations depending on the image processing strategy chosen and the nature of the images that the computer vision system has to deal with. The above mentioned ones, however, are fairly typical. It should also now be fairly obvious why picking of objects from a pile (bin or container) where they are all jumbled up is a complex problem. The objects touch and overlap, they present no unique stable positions and cannot be back-lit. One shortcut to avoiding this problem is to pick up a random number of objects from the bin, say magnetically or by suction, and to scatter these on a flat, hard backlit surface.

Further simplifications of the image of the scene are obtained by selection and control of the sources of illumination. The number of light sources, their spectral characteristics, illumination patterns, and shapes are to a large degree under the control of the designer of the robot vision system. The care and effort expended for proper illumination of the scene may frequently be the decisive factor between success and failure to solve a given problem. Furthermore, the light sources should be under computer control to allow frequent recalibrations. A variety of possibilities exist (Jarvis, Mundy and Jarvis).

1) Careful selection and positioning of the light sources.

The light sources are used to reduce unwanted effects in the image and to enhance the desired ones. Diffuse light may be used to reduce specular reflections and shadows in one application, while in another highly oblique illumination or even specular reflections are generated intentionally in order to highlight small imperfections on smooth surfaces.

2) A flash of light to "stop" the motion of an object.

If an object is moving at known speed then, after image analysis and object recognition, its position can be predicted.

3) The use of structured light. The use of "structured" or specially shaped light sources offers many possibilities for generating special effects, resulting in images with special properties.

The simplest form of structured light is a "pencil beam" of light which produces a single illuminated dot in the image. The light source is normally a laser, since an intense monochromatic light can easily be distinguished from the surrounding ambient light by the use of proper filters. By observing the scene with two cameras, or a properly designed single camera, the distance to the dot of light is computed from the parallax or positional discrepancy between the bright dots in the two images. Depth cameras have been designed in many laboratories and are available.

Instead of using a single dot, one may use patterns of dots of light, controlled by computer to turn them on and off in predetermined sequences if necessary. The use of such dot patterns resembles the use of "local operators" in image processing, but in this case depth and surface orientation are computed.

If the illumination source is designed to produce a strip of light, the rest of the scene being in darkness, only the strip of light is visible to the camera. In the image the strip is straight only if it is falling onto a flat surface in the scene. The curvatures or bends, the slopes, kinks, and breaks in the bright strip are analyzed to estimate various object parameters, such as location, shape and edges.

The light may also be shaped in many other ways for specific applications, such as, patterns of dots, rasters, and multiple convergent strips. Methods based on Moire fringes and Fourier optics are additional possibilities. Obviously, the image processing techniques vary for each kind of structured light and problem environment.

4) The uniformly lit background to generate silhouettes or profiles. Side illumination is also used.

5) Several sources of light used in various sequences, in order to highlight specific aspects of the scene in a controlled sequence.

The above mentioned methods do by no means exhaust the possibilities. Thus, for example:

1) The vision system (eye) may be mobile where one robot arm is essentially an eye only, and brought to the position of best view in a given circumstance.

2) A light source may also be carried by a robot arm.

3) The manipulating arm may have an eye in its "palm".

4) Many light sources and "eyes" may be placed in strategic locations around the work area, and so on.

In a practical problem any "trick" that works is acceptable. Consequently, the possibilities are mainly limited by our imagination and the cost effectiveness of the solution. Thus, one may use special strip cameras, multiple robot eyes put into judiciously chosen locations, eyes on the manipulator, in its hand, and so on. There is a nearly endless number of combinations possible, specially when considering the interaction of other sensors such as touch, force, proximity, etc.

THE BIGGER PICTURE

A vision system cannot operate alone. In automatic visual inspection, the parts to be inspected have to be fed to the machine, positioned properly, inspected, and then removed from the inspection machine. If the inspection machine requires manual feeding (parts positioning), in many situations the operator might as well inspect the part also, making the inspection machine unnecessary. Furthermore, the more accurately the part can be positioned, the simpler the inspection technique is likely to be, since algorithms to compensate for errors in parts positioning may not be needed. Thus, the largest component of the cost of the inspection machine is in the equipment required for parts handling, and not in computer vision (Mundy and Jarvis).

In the application of vision to robots the same is true. The robot system contains many mechanical component for parts handling, i.e., for pickup, transport, and placement, besides the cutting and/or assembly machinery. The vision component is again a relatively small, but of course vital, part of the entire system. The integration of vision with other sensors, the commands to and the control of the actuators (motors on robot's joints, belts for transport, the starting, control and stopping of the cutting and assembly machines, and so on) requires total integrated control. Besides controlling the normal operation of the system, the integrated control must also be able to handle malfunctions, to prevent damage to the product, the machinery, and the presonnel. Consequently, the programming of robot systems is an extremely complicated task, which requires a functional breakdown and hierachical systems to be manageable. It has been stated that "at present there are about as many robot programming languages as there are robot types" (Bonner and Shin).

On the next level up is the control of the entire manufacturing operation within a factory, where the overall routing to meet production schedules has to work also in the presence of breakdowns of some of the machines involved. Some practically completely automated factories exist at present (Bylinsky).

The social implications of automation is an even "higher level" problem. It appears that only in Japan has a logical and systematic approach been taken (Kasvand).

CONCLUSIONS

The three constraints of speed, inexpensiveness and reliable operation put very severe limits on commercially feasible robot vision systems in the ordinary factory environment. The economic constraints are not so stringent, however, in more "exotic" situations where the human neither cares nor dares to work. In the factory environment these constraints can at present be met to a certain degree only by using binary images, low resolution, special hardware or firmware and relatively simple image processing algorithms, or the applications are restriced to very specialized cases.

Only a relatively small number of robot vision systems are successfully used in actual production operations. Many more systems are expected, since even the present methods can find many more economically feasible application areas. The introduction of robots with vision in industry, however, has been fairly slow, mainly owing to the novelty of such systems, the need to redesign manufacturing processes and their expense and general uncertainty.

The greatest need is to make the robot vision systems more versatile, robust and easier to program for specific applications. Modular construction, special hardware "building blocks", parallel and pipelined processing (Hwang and Fu, Preston), special programming languages, and so on, are all needed to extend the application areas and still remain economically viable. The amount of R&D needed for computer vision is enormous, and as already indicated, this challenge has been taken up in many places (Jarvis).

Generalized vision systems approaching human capabilities are at present neither scientifically nor technologically possible. Furthermore, to be economical, a robot system should be designed to meet the needs of a particular task environment rather than being modelled on the human worker. The multitude of possible sensors, algorithms, decision making and operational strategies, and so on, means that there is no single unique optimal solution to the robotics problem. There are likely to be only good solutions and less good ones.

The industrial requirements on robot vision may be classified into categories according to the nature of the task, as was done in this paper. Such categorization is understandable to the industrial user. From the point of view of solving computer vision problems, however, such a categorization is not necessarily optimal. The needs of robot vision should be met by formulating a set of task categories specified by image content rather than from where the images originate. This has not been very noticeable yet, but it will bring this problem to the general field of computer vision research, where robot vision simply becomes one application area among many.

ACKNOWLEDGEMENTS

The author is grateful to the Division of Electrical Engineering of NRC for encouraging the research and applications of image processing.

REFERENCES

Information retrieval give thousands of references on robots and the number of articles is rapidly increasing. Searches geared to "robot" do not find articles related to image processing in general and to special image processing devices such as for example the Quantimet, Leitz TAS and TAS 2, Zeiss IBAS, the Hamamatsu system, Toshiba system, the Cyto-Computer, Bauch and Lomb Omnicon, Leiden TAS, the Image 100, and many new image buffer based systems.

Numerous papers may be found in the conference proceedings such as ISIR (International Society for Industrial Robots), ICRP or IJCPR (International 'Joint' Conference on Pattern Recognition), PRIP (Pattern Recognition and Image Processing, Conference Proceedings of IEEE Computer Soc.), IJCAI (International Joint Conference on Artificial Intelligence), etc., as well as many special workshops, and IEEE publications. Magazines specializing on robotics are available, such as: Robotics Age, The International Journal of Robotics Research, Robotics Today, Robot Systems and Products, and Robots - Le Journal de la Robotique Industrielle. Bibliographies on robotics are available as ERB-935 and ERB-936 by N. Abdelmalek (NRC), and from IFS Publications Ltd., England, by R. Gomersall and P. Farmer.

The references below are in essence only scattered samples from the enormous amount of available literature.

Barnard S. T., Automated Inspection Using Gray-Scale Statistics, SRI TN 216, March 1980.

Bitter K.H., Application of Opto-Electronic Image Sensors, SPIE, vol. 392, Robot Vision and Sensory Controls, p. 93-116.

Bolles R. C., Part Acquisition Using the SRI Vision Module, SRI TN 193, Sept. 1979.

Bonner S., and Shin K.G., A Comparative Study of Robot Languages, Computer, vol. 15, Num. 12, Dec. 1982, p. 82-96.

Boykin W. R., Diaz R. G., The Application of Robotic Sensors - A Survey and Assessment. Advances in Computer Technology - 1980, vol. 1, pp. 160-165 (Dept. of Engineering Sciences, Univ. of Florida, USA).

Bylinsky, G., The Race to the Automatic Factory, Fortune, Feb 21, 1983, p. 52-64.

Dodd G. G., Rossol L., Computer Vision and Sensor-Based Robots, Plenum Press, 1979.

Duda R. O., Hart P. E., Pattern Classification and Scene Analysis, Wiley, 1973.

Hall E. L., Computer Image Processing and Recognition, Academic Press, 1979.

Heer E., Robotics and Manipulators, A Review of Current Technology, (Tutorial Summary), Advances in Computer Technology - 1980, vol. 1, p. 159. (Jet Propulsion Laboratory, Pasadena, Calif).

Hsieh Y. Y., Fu K, S., Automatic Inspection of IC Chips, Advances in Computer Technology - 1980, vol. 1 pp. 12-45. (School of Electrical Engineering, Purdue Univ., USA).

Hwang K., and Fu K.S., Integrated Computer Architectures for Image Processing and Data Base Management, Computer, vol. 16, num. 1, Jan. 1983, p. 51-60.

Jarvis J.F., Research Directions in Industrial Machine Vision: A Workshop Summary, Computer, vol. 15, Num. 12, Dec. 1982, p. 55-61.

Kashioka, S., Mazakazu E., Yusaburo S., A Transistor Wire-Bonding System Utilizing Multiple Local Pattern Matching Techniques, IEEE Trans., SMC-6, No. 8, Aug. 1976, pp. 562 - 570.

Kasvand T., The Robotics Scene in Japan, ERB-933, NRC.

Kasvand T., Vision, as Applied to Robots, ERB-932, NRC.

Kasvand T., Computer Simulation of Pattern Recognition by Using Cell-Assemblies and Contour Projection, IFAC Tokyo Symposium on Systems Engineering for Control System Design, Aug. 25 - 28, 1965, pp. V-11 to V-20.

Kasvand T., Some Thoughts on Picture Languages, in Graphic Languages, Nake F., Rosenfeld A., editors, North - Holland pub. Co.,

1972, pp. 163 - 186.
Kratky V., Real-Time Photogrammetric Support of Dynamic 3-D Control, Photogrammetric Engineering and Remote Sensing, vol. 45, 1979, pp. 1231 - 1242.
Mundy J.L, and Jarvis J.F, Automatic Visual Inspection, Chapter 6 in Applications of Pattern Recognition, K.S. Fu, editor, CRC Press, 1982.
Nitzan D., Assessment of Robot Sensors, NSF Robots Research Workshop, Newport, Rhode Island, Apr. 15 - 17, 1980.
Nitzan D., Rosen C., Agin G., Bolles R., Gleason G., Hill J., McGie D., Prajoux R., Park W., Sword A., Machine Intelligence Research Applied to Industrial Automation, Ninth Report, Aug. 1979, SRI International.
Perkins W. A., A Model-Based Vision System for Industrial Parts, IEEE Trans. on Computers, vol. C-27, no. 2, Feb. 78, pp. 126-143. (Computer Science Research Laboratories, General Motors Corp., Warren, Mich.)
Perkins W.A., Vision System that Learns how to Inspect Parts, Proc. of the International Society for Optical Engineering, vol. 336, Robot Vision, 1982, p. 50-58.
Polyak S.L., The Vertebrate Visual System, Univ. of Chicago Press, 1957.
Preston K. Jr., Duff M. J. B., Levialdi S., Norgren P. E., Toriwaki J., Basics of Cellular Logic with Some Applications in Medical Image Processing, Proc. IEEE, vol. 67, No. 5, May 1979, pp. 826 - 856.
Reichardt, J., Robots, Fact, Fiction, and Predictions, Penguin Books, 1978.
Rosen C. A., Machine Vision and Robotics: Industrial Requirements, SRI TN 174, Nov. 1978.
Rosenfeld A., Pfaltz J. L., Sequential Operations in Digital Picture Processing, Journal of ACM, Oct. 1966, pp. 417 - 494.
Rosenfeld A., Picture Processing by Computer, Academic Press, 1969.
Rosenfeld A., Kak A. C., Digital Picture Processing, Academic Press, 1976.
Shirai Y., On Application of 3-Dimensional Computer Vision, Bulletin of the Electrotechnical Laboratory, vol. 43, No. 6, 1979, pp. 8 - 27.
Takeyasu K., Masuo K., Shimomura R., Goto T., Matsumoto Y., An Approach to the Integrated Intelligent Robot with Multiple Sensory Feedback: Construction and Control Functions, (a manuscript obtained during a visit).
Tsuji S., Survey on Vision Systems for Advanced Automation in Japan, Proc., SPIE, vol. 182, Imaging Applications for Automated Industrial Inspection and Assembly, Apr. 19-20, Washington, D.C., pp. 2-5.
Tsuji S., et al., Future Directions of Industrial Applications, panel discussion, IJCPR-78, pp. 1144 - 1149.
Ward M. R., Rossol L., Holland S. H., Dewar R., Consight: A Practical Vision-Based Robot Guidance System, Proc. 9-th ISIR, Mar.13-15, 1979, Washington, D.C. pp. 195-211. (General Motors Research Laboratories and Manufacturing Development, USA).