Machine vision: the basics

By Valerie C. Bolhouse

Machine vision can be described as the replacement of human visual sense and judgment capabilities with a video camera and computer to perform a task. The camera is used to acquire an image of a scene, replacing the function of the human eye. The computer must have a means of receiving this information, and uses a predefined set of instructions (program) to analyze the information to make a decision.

Video cameras and computers have been available for several decades, but industrial machine vision has only recently become a viable solution for manufacturing. This is due in part to technological advances that make the price/performance of the systems reasonable, and also to the fact that increased world-wide competition has made American industry pursue every possible avenue to improve both quality and productivity.

Where can machine vision be used?

There are three major application areas for machine vision: inspection, process control, and robot guidance. The motivation for using vision for inspection is readily seen. Human operators simply do not make good inspectors, regardless of how conscientious they may be. It is estimated that they are about 85% effective in sorting defective parts. This rate is higher on the start of the shift, but decreases due to boredom and fatigue. Inspection is even more difficult when the rate of defects approaches several parts per million. Machine vision can be programmed to be better than 99% effective, and can be set to insure that no bad parts leave the factory by setting the parameters such that machine vision errors occur by rejecting good parts.

Despite the advantages of machine vision over human operators in inspection, the real payback for vision is in process control. Many of the vision applications in electronics (automatic wire bonding, solder paste inspection, and surface mounted device assembly) take advantage of the early feedback from the process to alter processing parameters and prevent mass production of bad parts.

Human operators just do not make good inspectors, regardless of how conscientious they may be.

Robot guidance accounts for less than 10% of all vision applications, but holds much potential for future robot installations. Some of the early robot installations (such as spot welding of automotive body panels) did not require vision since the parts had to be precisely fixtured to carry out the subsequent process. These "easy" applications have been done, and the ones left will require either expensive fixturing to locate the parts precisely, or sensory feedback from vision to tell the robot the location of the parts.

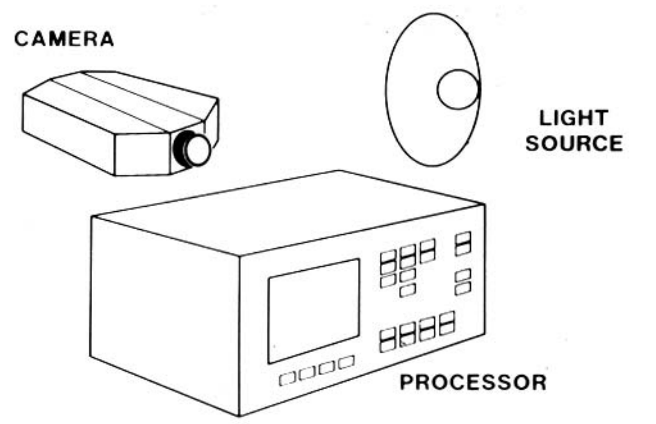

FIGURE 1. Hardware components of any vision system: camera, light source, and processor.

VALERIE C. BOLHOUSE is an automation systems specialist with the corporate robotics laboratory of the Ford Motor Co. in Dearborn, Mich.

Machine vision can be thought of as a process: image acquisition, image analysis, decision making, and finally, output to external devices. Although commercially available equipment varies considerably in appearance and capability, all systems perform this process and are comprised of similar hardware and software building blocks (Fig. 1). By analyzing the building blocks of any system, the performance limitations imposed by the individual components can be better understood.

Image acquisition

The objective of image acquisition is to obtain information about the appearance of the part, and present it to the vision computer in a format that can be analyzed and understood. Components that need to be considered in image acquisition include the lighting to highlight the features of interest on the part, the camera and optics to take the picture, and the electronics to transfer image information (intensity) into image data (numbers) that the computer can analyze.

■ **Lighting**

The importance of lighting is stressed most by experienced users of vision, but is rarely seen in product brochures from the vision vendors. This is because while the lighting design can make or break the vision application, most vendors purchase off-the-shelf lighting products. In fact, some of the best lighting designs use everyday components such as fluorescent or flood lights. They are inexpensive, readily available, and easy to maintain.

Machine vision applications are usually concerned with reflected light. The surface finish and the color of an object will determine how light will be reflected from that object. The cameras used in vision are monochrome (black and white) and respond to light similar to a light meter that measures the amount of light irrespective of color. For this reason, vision systems "see" colors differently than people do. Blue and red are the same intensity to the camera; the same can be true for brass, copper, and steel depending on the surface finish of the metal.

Surface finish quality plays an important role in how light gets reflected. Very smooth, shiny surfaces will merely bend the light rays in a new direction (angle of reflection will equal angle of incidence) without changing the spatial coherence of the rays. This is termed specular reflection. Dull or diffuse surfaces will scatter the light rays in all directions. Typically, diffuse surfaces are the easiest to illuminate.

In designing a lighting scheme, one must first select a light source, then decide where to put the light relative to both the part and the camera. The following characteristics are important in choosing a light source:

Spectral properties. The color of the light and the color of the part will determine how much energy gets returned to the camera sensor. The best color light source will depend on the contrast between the features to be looked at by the camera.

Efficiency. Some light sources such as fluorescent lights are very efficient and will emit a lot of light relative to heat generated. Others such as tungsten generate considerable heat along with the light. The lights are generally left on continuously, and inefficient light sources can cause localized heating problems and will be expensive to operate.

Life and aging characteristics. Light sources are usually rated in number of hours of continuous operation. A 1000-hour light source would only last several weeks for a two-shift operation and maintenance would be required to continuously change bulbs. As the light ages the output of light energy decreases. Depending on the type of light used, the decline could be gradual and small, or immediate and significant. Changes in light output will require parameter changes to the vision system, again resulting in much maintenance to the total system.

Cost and availability. Light bulbs will burn out during the life of the vision system. Generally spare parts will be stocked. If the parts are expensive, there is a tendency to stock fewer spares, and if they are not readily obtained from usual sources they will not be easily replenished.

Once the light source is selected, the position of the light relative to the part and the camera must be determined. There are three basic light approaches: back light, front light, and structured light.

Back lighting is used when part silhouette information is important (Fig. 2). The part is positioned between the light and the camera to provide the highest contrast image. However, this is not always possible or practical to implement. When the parts are in fixtures or moving down a conveyor, special handling would be required to position the part for back lighting.

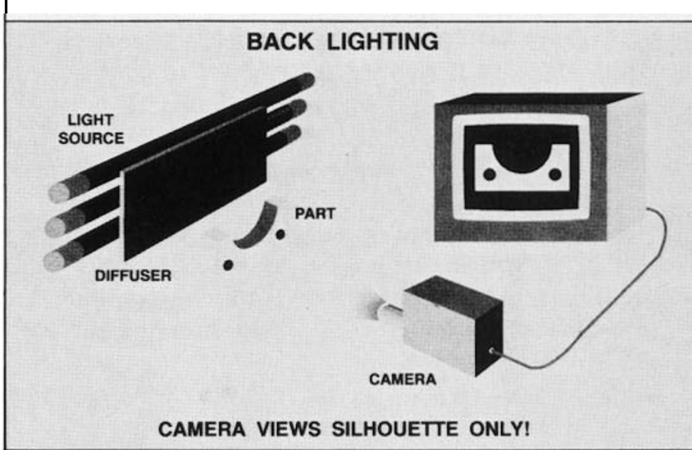

FIGURE 2. Set-up for back lighting.

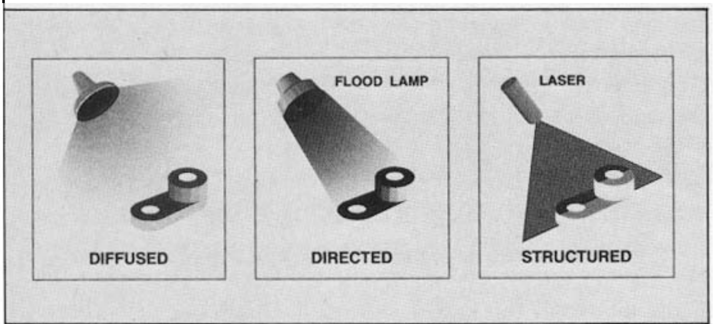

FIGURE 3. Three types of front lighting.

Front lighting is generally used due to process constraints (where it is impractical to position a light behind the part) or inspection requirements. If the application has to use surface information, front lighting is mandatory. Three different types of front lights are used (Fig. 3). The most common is diffused light—evenly distributed light that results in no shadows or reflections off the part. Diffused light works quite well for normally high contrast parts and for grayscale processing.

Directed light is angled to purposely create reflections and/or shadows to highlight critical features. Typical directed light sources are flood lamps and fiber optic lights.

Structured light is the third type of front lighting. A line of light is directed at the part, and only the reflected light pattern is analyzed by the system. A laser is generally used since it can be easily focused due to its spatial coherence. When the angular relationship between the light and the camera is known, geometric techniques known as triangulation can be used to determine height (or the z-dimension) of the features even though all information from the camera is two dimensional.

The lighting solution for each application has to be determined individually. After the key features are identified, the best way to highlight these features has to be decided—the type of lighting necessary for a person to see the detail is usually sufficient. Then the process constraints have to be considered. Certain obstructions on the line may preclude some lighting approaches. Back lighting may be desirable, but might not be practical.

The next step is to set up the light source and view the part on a video monitor. Parts with varying colors, surface finish, and defect modes should be analyzed. If the contrast is still there, the lighting design is probably appropriate. The important thing is to start simple. This will result in the lowest cost and easiest to maintain solution to the problem.

■ Optics

For the purpose of this discussion on lenses, think of the sensor element as being much like photographic film. The response of the sensor to light is very similar to 100 ASA film (the most common speed of film used for general purpose photography), and its size is 6.6 mm by 8.8 mm, about one fourth the size of a 35 mm film negative. Selecting a lens for vision is very similar to choosing a lens for home photography. Lenses are specified by their focal length—the distance between the optical center of the lens and the image plane when the lens is focused at infinity. Typical focal lengths used in industrial vision are 25 mm (wide angle) to 200 mm (telephoto).

The required lens focal length for an application is determined by the size of the area to be imaged and the camera position; in other words, by the magnification required and the standoff distance from the lens to the object. Magnification is defined to be the image size divided by the object size. The image size is the size of the sensor, which is almost always a standard 6.6 mm by 8.8 mm. The object size is the field of view, or that size area the camera will be viewing.

For the majority of vision applications outside of electronics, the parts viewed are much larger than 8 mm, so the magnification is much less than one. This greatly simplifies lens selection, since depth of focus and distortion are not much of a problem at low magnifications. Generally just buying a good quality lens ($200 to $500) will guarantee satisfactory performance.

Most video cameras used in machine vision have a screw thread C-mount. Standard off-the-shelf C-mount lenses were designed for closed circuit surveillance imaging and are not well suited for machine vision because they are optimized for large area viewing at very long standoff distances. Minimum focusing distances are usually four feet, while industrial applications require mounting the camera about 18″ to 24″ above the line. Extension rings are used to focus the lenses at these shorter distances. Unfortunately, the extension rings cause distortions and reduce the amount of light. An alternative to custom optics is to purchase standard off-the-shelf 35 mm camera lenses and use an adaptor to C-mounting. However, these lenses are quite large relative to C-mount lenses and can be too bulky if the camera will be moving on a robot arm.

■ Camera

The camera consists of the image sensor and the electronics required to read out the intensity information focused onto the sensor. There are two basic sensor technologies: vacuum tube sensors known as vidicon and solid state sensors known as Charge Coupled Devices (CCDs). Vacuum tube technology dominated the video camera market until the early 1980s. Most closed circuit surveillance cameras and home video cameras still use this technology, which is inexpensive and works well under controlled environmental conditions. However, the industrial environment can be quite harsh and the fragile sensor tube does not survive well. Vacuum tube sensors also experience linearity and stability problems, which make them inappropriate for applications which require high accuracies such as gauging.

With advances in semiconductor technology, the solid state sensor has replaced the vidicon sensor for most machine vision applications. The solid state sensor consists of an array of discrete photosensitive elements. Common sized include 320 × 240 and 512 × 512; resolutions greater than 1000 × 1000 are available, but not yet standard. Sensors with photo sites arranged in a two-dimensional format are called array sensors. Sometimes a single line of picture elements called line scan sensors are used. Common sizes are 1024 × 1,2048 × 1, and 4096 × 1.

The information stored in each individual photo site is the total amount of charge developed at that site as light strikes the sensor surface. There is no detail about the distribution of light within the individual picture element—the smallest unit of detail available in the image. The photo site is known as a pixel (the acronym for *pic*ture and *el*ement) and the number of pixels used in a vision system is often called the resolution of the system.

■ Camera electronics

The image acquired on the sensor consists of an electrical charge representative of the total amount of light striking each picture element. The camera electronics provides a means of transferring this information to the computer system to analyze one line at a time. The video format is called raster scan, with the read starting at the upper left corner of the sensor, moving left to right and top to bottom. This is the format used to standardize the broadcast television industry and is used by machine vision cameras and monitors. It is not ideally suited for high speed, high data rate machine vision applications, but does allow the low volume machine vision industry to reap the economic advantages of the very high volume television industry.

The electrical standard RS170 or RS330 describes the voltage (1 volt peak-to-peak) levels and timing synchronization used. The standard picture consists of 525 lines of video information (Fig. 4). Because of broadcast television requirements, the 525 lines are used for two pictures taken 1/60 seconds apart, interlaced every other line. Thus, for low resolution (320 × 240) images, the camera can take 60 pictures a second, whereas medium resolution (512 × 512) is limited to only 30 pictures a second.

The video information leaves the camera as an analog voltage signal across a coaxial cable. The machine vision hardware must now convert the analog video signal into an array of numbers that a computer can analyze. This happens in the frame grabber portion of the vision system.

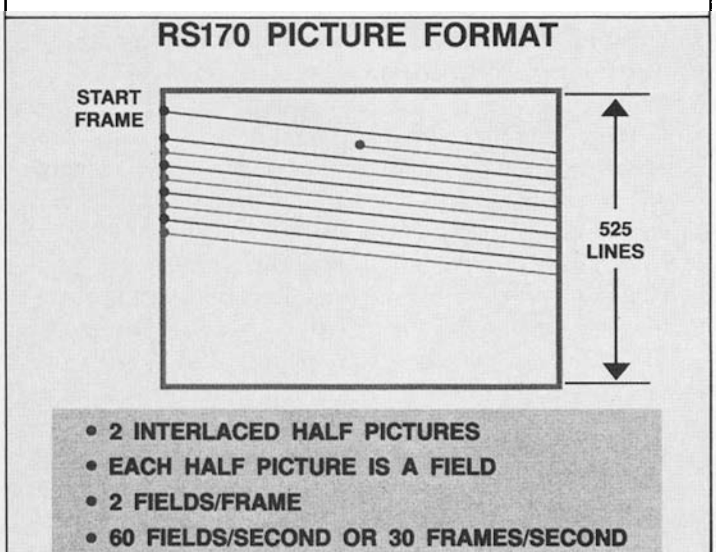

FIGURE 4. Video picture format used in machine vision cameras and monitors.

■ Frame grabber

Most vision systems can use more than one camera at a time (four or eight are typical). The video signal from each of the cameras is input to a multiplexer. The multiplexer is simply a gate, which will only allow one signal through at a time. This signal then goes to an Analog to Digital (A to D) converter, which, as its name implies, will convert the analog voltage signal into numbers. Typically, low light will result in a low voltage signal on the sensor, which gets converted to low numbers. Zero is black. Bright light results in higher voltage levels from the camera and higher numbers from the A to D converter. The maximum light level, or the largest number, is termed the grayscale resolution of the system. The most common resolution is 64 "shades of gray;" higher grayscale resolution systems use 8 bits for 256 gray levels.

The numbers from the A to D converter get stored in computer memory referred to as the image buffer. The vision computer operates on data from the image memory. Most systems can store more than one image in memory.

What the computer does with image data

Machine vision is asked to do a variety of things: verify presence/absence of features, locate a feature, measure objects or distance between objects, count or sort objects, and detect anomalies in parts that make them rejectable. The vision system makes these decisions by analyzing the information content in the image data.

The calculations performed on image data can be classified as either image enhancement or image analysis In image enhancement the vision computer alters the values of the pixels based on some pre-programmed steps called a vision algorithm. The output is another image. Images are enhanced to highlight features that are important, such as edges, or to filter out noise. The objective is to make decision analysis easier.

In image analysis, the computer analyzes image data and extracts "feature vectors" from the two-dimensional image data. The key is data reduction. The feature vectors are often only one number to represent the information in thousands of pixels. One example would be the number to represent the average brightness in the image. The value of the feature vector can be passed to a host computer, which can then decide if the part is acceptable or rejectable based on measured brightness. Image analysis is really image understanding. The vision computer attempts to make sense of the ambiguous image data by extracting feature vectors with a predefined set of calculations.

Image understanding is a difficult task for any computer system, and this job is further complicated by the speed required for real time application of machine vision in production. In food processing, data must be analyzed as fast as it is acquired (60 times a second). Electronic applications will range in the 100 millisecond to 1 second timeframe, and automotive is usually 3 to 10 seconds. There is a considerable amount of data to be analyzed in an image (262,144 numbers in a 512 × 512 array) in a very short time. Two data reduction techniques are employed by machine vision to improve processing time: thresholding and windowing.

■ Threshold to binary

Pixels in grayscale systems contain 6 or 8 bits of intensity information: black, white, and shades of gray in between. These are the numbers generated by the Analog to Digital Converter in the frame grabber portion of the hardware to represent intensity. This information is often reduced to two levels—black and white—by selecting an intensity level to label the darker gray pixels black and the lighter gray pixels white. This intensity level is called the threshold, and the resulting processing is called binary image processing. Binary processing is used in many applications because it is the fastest and most straightforward way to analyze images.

■ Windowing

Windows are used to reduce the number of pixels analyzed by the system. A window is a region of pixels in a continuous area that is processed by the computer while all pixels outside that area are ignored. Windows are typically rectangular in an orthogonal coordinate system, but can also be circular, annular, rhomboidal, or non-orthogonal. Windows are almost always used in inspection applications, where we can constrain the part location, and almost never in robot guidance, where the part typically can be anywhere in the field of view and it is the job of the vision system to find it.

Vision can work well in an industrial environment because we have a priori information about part appearance and location.

■ Information in images

Vision can work well in an industrial environment because we have *a priori* information about part appearance and location. In high volume production, parts are almost identical in appearance and are typically fixtured. A computer can be programmed to follow some sequence of operations to repeatedly extract information from the image from which to make decisions.

This information can be classified as either spectral or spatial. Spectral properties give information on the color or intensity of pixels, whereas spatial information gives the relationship of pixels relative to each other in space. Spatial information is used to identify edges and shapes of objects in the image. When vision is asked to locate or measure features, this requires spatial information be used. However, verification of presence or absence can often be done with spectral analysis only.

There are three types of spectral analysis typically implemented in machine vision: binary pixel counting, grayscale average intensity, and grayscale intensity distribution analysis. In binary pixel counting, the number of white or black pixels in a window is counted. This value is compared to a preset accept/reject number to make a decision. Binary pixel counting works well for presence/absence decisions and where high speed is critical. It does not work for measurement, location, or counting since spatial information within the window is not considered. In grayscale average intensity, the grayscale value of every pixel in the window is summed, and the total is divided by the number of pixels in the window. Again, this average value is compared to the accept threshold. Average intensity calculations are usually used to measure brightness of areas of the image. Intensity histograms are used to analyze the distribution of intensity values, and give useful information for texture analysis.

Spatial information is used to locate or measure objects. The computer looks not only at the brightness of the pixels, but also at the location of pixels relative to their neighbors. Finding the edges of an object is key to spatial analysis. An edge in an image is defined by a rapid change in grayscale intensity in neighboring pixels. There are numerous ways to find edges in an image (gradient analysis, linear and nonlinear convolution operators). Once edges are determined, the vision computer can count edge pixels, measure between edges, or correlate edge templates to known good parts. Pixels within edge boundaries are classified as objects. Geometric features of the objects can be calculated to give information on the size, shape, and location of the feature in pixel space.

■ Machine vision implementation

The type of vision processing performed and the feature vectors extracted are known as the vision algorithm. The resulting information is usually passed off to a host computer to make some decision to be output to external devices. The systems are usually programmed in a standard higher level language. "C" has become the language used most often.

The operator interface for application programming of the vision system is usually menu driven. The more flexible and powerful systems will use a programming language to allow for user definable functions. Design and execution of the operator interface is the point where systems become unique. Beginning users feel most comfortable using a menu; sophisticated users will want functions not available on simple menus and will require a language.

Interface to external devices

Once the host computer has made an accept/reject decision or located or measured the part, this information must be passed to the production processing equipment. Most sytems have both serial and parallel interfaces available. The serial port will generally be used to send statistical data on part features to an external computer to collect or to send location information to a robot. Parallel input/output is used to indicate part presence and to output the status of the inspection.

The study of machine vision requires knowledge from many disciplines: physics of light, optics, electronics, and computer processing.

A marriage of technologies

The study of machine vision requires knowledge from many disciplines: physics of light, optics, electronics, and computer processing. The marriage of all these technologies into one has made it feasible to emulate the human visual sense with mechanical hardware. It is not possible to replace human vision with machine vision across the board, yet in well defined applications the machine can be programmed to function better than the human. Computers work well at repetitive tasks, while people get bored. There are several thousand systems across the United States today performing this artificial vision task day in and day out. This technology is continuing to prove itself as a viable solution to many production problems.