Machine Vision Fundamentals

Prasanthi Guda, Jin Cao, Jeannine Gailey and Ernest L. Hall
Center for Robotics Research
Department of Mechanical, Industrial and Nuclear Engineering
University of Cincinnati
Cincinnati, OH 45221

Abstract

The field of machine vision can be thought of as the application of computer vision techniques to industrial automation. Although computer vision programs produce the desired result, every successful vision application can be traced back to a thorough understanding of the fundamental science of imaging. The purpose of this paper is to introduce machine vision concepts and review several facts and techniques used in machine vision applications. A review of some properties of the human visual system is necessary to provide a basis for understanding machine vision capabilities. In addition, the important data processing techniques including histogram operations, convolution and edge detection, Fourier transform spectral methods, and segmentation for measurement of lines and regions are reviewed. Industrial applications may be divided into two categories: automatic inspection and robot control. Several applications are reviewed in each category. As the field of machine vision continues to grow rapidly, its positive effects on technology are increasing. For instance, 100% inspection during production promises to provide major improvements in manufacturing where high quality products, reduced waste, and lower production costs are essential.
1. Introduction

The new machine vision industry that is emerging is already generating millions of dollars per year in thousands of successful applications. Machine vision is becoming established as a useful tool for industrial automation where the goal of 100% inspection of manufactured parts during production is becoming a reality. The purpose of this paper is to present an overview of the fundamentals of machine vision. A review of human vision is presented first to provide an understanding of what can and cannot be easily done with a machine vision system.

Human Visual System

While human beings receive at least 75% of their total sensory input through visual stimuli, our vision processes are executed in billions of parallel neural networks at high speeds mainly in the visual cortex of the brain. Even with the fastest super-computers, it is still not possible for machines to duplicate all of the functions of human vision. However, an understanding of the fundamentals of human image formation and perception provides a starting point for developing machine vision applications.

The human visual system is comprised of three main organs: the eyes, the optic nerve bundle, and the visual cortex of the brain. This complex system processes a large amount of electro-chemical data and performs its tasks in a highly efficient manner that cannot yet be duplicated by machine vision. However, the human visual system can also be confused by illusions. The key element to vision is light, which is radiant energy in the narrow range of the electromagnetic spectrum, from about 350nm (violet) to 780nm.
This energy, upon stimulation of the retina of the human eye, produces a visual sensation we call visible light. Photometry and colorimetry are sciences that describe and quantify perceived brightness and color and objective measurements of radiant energy.

The visual cortex of the human brain as shown in Figure 1 is the central location for visual processing. We believe that the inner layer, or white matter, of the brain consists mainly of connections while the outer layer, or gray matter, contains most of the interconnections that provide neural processing. The eyes function as an optical system whose basic components consist of the lens, the iris, and the retina. The lens of the eye focuses the incoming light to form an inverted image on the retina at the rear wall of the eye. The amount of light entering the eye is controlled by a muscle group called the iris. The retina, as shown in Figure 2, consists of about 125 million light-sensitive receptors that, because of their many-to-one connections, have some processing capabilities. These receptors consist of color sensitive “cones” and brightness sensitive “rods.” The central part of the retina is called the fovea. It contains a dense cluster of between six and seven million cones that are sensitive to color and are connected directly with the brain via individual nerves. When an image is projected onto the retina, it is converted into electrical impulses by the cones and then transmitted by the optic nerves into the brain. The optic nerve has between one and two million neurons. Around the periphery and distributed across the surface of the retina are the rods. Unlike cones, rods share nerve endings and are sensitive to light and dark but are not involved in color vision. The rods in the human eye can adapt to a range of light intensities over several orders of magnitude.
magnitude. This permits humans to see not only outside in the bright sunlight but also in a darkened room.

Most of the neural processing of image impulses carried via the optic nerve bundle takes place in the visual cortex. Various theories have been suggested to attempt to describe visual cortical processing, including edge enhancement, computing correlation, Fourier transforms, and other higher level operations [1]. The basic architecture of our organic neural networks is now being used as a template for developing neural network computer algorithms. These artificial neural algorithms can perform tasks such as recognizing objects, making decisions, function approximations, and even system identification and discovery. Nonetheless, many capabilities of human beings, such as visual understanding and description, still present challenging problems.

An example of early visual system neural processing is shown in Figure 3. The neural network in this experiment, known as the \textit{backward-inhibition model}, consists of a linear recurrent network followed by a non-linear element followed by another network. The inhibition equation may be written as:

\begin{equation}
\mathbf{y} = \mathbf{w}^T \cdot \mathbf{x} - \mathbf{y}^T \cdot \mathbf{w}_2 \cdot \mathbf{y}
\end{equation}
where the output responses are $y_i$, the input is $x_i$ and the coefficients, $w_{ij}$, regulate the amount of inhibition. This model was used to demonstrate the nonlinear nature of the frequency sensitivity of the human visual system [2]. The human visual system was found to respond in a nonlinear manner to the contrast of an input signal as shown in Equation (1). By the combination of linear and nonlinear processing, a model was developed which showed similar characteristics, as shown in Figure 3(a). An original image and the result of processing by the nonlinear model are shown in Figure 3 (b) and (c). Note that the image is blurred slightly, but that a considerable edge enhancement is produced. As may be observed in Figure 3, the reduction of irrelevant detail and enhancement of important edges in the dark regions of the image was achieved. This effect could be analogous to night vision. Perhaps early humans, living in caves or hunting and foraging at night, had required this survival ability to discern the outlines of predators in the dark shadows.

$$y = f \Theta \left[ w^T x - b \right]$$

(2)

The neural network model may also be written to emphasize the fact that the response must be greater than a threshold, $b$, to produce an output, as seen in Equation 2.

In this case, $g$ represents the nonlinear, zero-one step function whose value is 1 when the threshold, $b$, is exceeded and 0, otherwise. Additionally, the function $f$ may be another function that determines frequency sensitivity or recognition selectivity. The overall
neural network function is a composite function of a basic linear decision element combined with nonlinear function mappings that are characteristic of modern multi-layer neural networks.

**Machine Vision Hardware Components**

A machine vision system consists of hardware and software components. The basic hardware components are a light source, a solid state camera and lens, and a vision processor. The usual desired output is data that is used to make an inspection decision or to permit a comparison with other data. The key considerations for image formation are lighting and optics.

One of the first considerations in a machine vision application is the type illumination to be used. Natural, or ambient, lighting is always available but rarely sufficient. Point, line, or area lighting sources may be used as an improvement over ambient light. Spectral considerations should be taken into account in order to provide a sufficiently high contrast between the objects and background. Additionally, polarizing filters may be required to reduce glare or undesirable spectral reflections. If a moving object is involved, a rapid shutter or *strobe* illumination can be used to capture an image without motion blur. To obtain an excellent outline of an object’s boundary, back lighting can provide an orthogonal projection used to silhouette an object. Line illumination, produced with a cylindrical lens, has proven useful in many vision systems. Laser
illumination must be used with proper safety precautions, since high intensity point illumination of the retina can cause permanent damage.

Another key consideration in imaging is selecting the appropriate camera and optics. High quality lenses must be selected for proper field of view and depth of field; automatic focus and zoom controls are available. Cameras should be selected based on scanning format, geometric precision, stability, bandwidth, spectral response, signal-to-noise ratio, automatic gain control, gain and offset stability, and response time. A shutter speed or frame rate greater than one-thirtieth or one-sixtieth of a second should be used. In fact, the image capture or digitization unit should have the capability of capturing an image in one frame time. In addition, for camera positioning, the position, pan and tilt angles can be servo controlled. Robot-mounted cameras are used in some applications. Fortunately, with recent advances in solid state technology, solid state cameras are now available at a relatively low cost.

Since the advent of the Internet and the World Wide Web (WWW), a great variety of images are now available to anyone. This has also led to an increase in the variety of formats for image data interchange [3]. Some of the most common image formats now are bitmap (BMP), data compressed JPEG (JPG), and GIF87a (GIF). The Graphics Interchange Format (GIF), shown in Table 1, was developed by CompuServe, and is used to store multiple bitmap images in a single file for exchange between platforms. The image data is stored in a bitmap format in which numbers represent the values of the picture elements or pixels. The bit depth determines the number of colors a pixel can
represent. For example, a 1-bit pixel can be one of two colors, whereas an 8-bit pixel can be one of 256 colors. The maximum image size with the GIF format is 64,000 by 64,000 pixels. The image data stored in a GIF file is always compressed using the Lempel-Ziv-Welch (LZW) technique. The GIF data can also be interlaced up to 4:1 to permit images to display progressively instead of top down.

There are literally hundreds of various image file formats. Table 2 lists some common formats as well as the extension, creator and conversion filter(s) for each format.

**Machine Vision Algorithms and Techniques**

**Image Functions and Characteristics**

As mentioned by Wagner [4], in manufacturing, human operators have traditionally performed the task of visual inspection. Machine vision for automatic inspection provides relief to workers from the monotony of routine visual inspection, alleviates the problems due to lack of attentiveness and diligence, and in some cases improves overall safety. Machine vision can even expand the range of human vision in the following ways:

- Improving resolution from optical to microscopic or electron microscopic
- Extending the useful spectrum from the visible to the x-ray and infrared or the entire electromagnetic range; improving sensitivity to the level of individual photons
- Enhancing color detection from just red, green and blue spectral bands to detecting individual frequencies
• Improving time response from about 30 frames per second to motion stopping strobe-lighted frame rates or very slow time lapse rates

• Modifying the point of view from the limited perspective of a person's head to locations like Mars, the top of a fixture, under a conveyor or inside a running engine.

Another strength of machine vision systems is the ability to operate consistently, repetitively, and at a high rate of speed. In addition, machine vision system components with proper packaging, especially solid state cameras, can be used in hostile environments, such as outer space or in a high radiation hot cell. They can even measure locations in three dimensions or make absolute black and white or color measurements, while humans can only estimate relative values.

The limitations of machine vision are most apparent when attempting to do a task that is either not fully defined or that requires visual learning and adaptation. Clearly it is not possible to duplicate all the capabilities of the human with a machine vision system. Each process and component of the machine vision system must be carefully selected, designed, interfaced and tested. Therefore, tasks requiring flexibility, adaptability, and years of training in visual inspection are still best left to humans.

Machine vision refers to the science, hardware and software designed to measure, record, process and display spatial information. In the simplest two-dimensional case, a digital black and white image function, $I$, as shown in Figure 4, is defined as:

$$ I = I_f (x, y): x = 0, 1, \ldots, N - 1; y = 0, 1, \ldots, N - 1 $$

(3)
Each element of the image \( f(x, y) \) may be called a picture element or \textit{pixel}. The value of the function \( f(x, y) \) is its \textit{gray level} value, and the points where it is defined are called its domain, window or \textit{mask}.

The computer image is always \textit{quantized} in both spatial and gray scale coordinates. The effects of spatial resolution are shown in Figure 5.

The gray level function values are \textit{quantized} to a discrete range so that they can be stored in computer memory. A common set of gray values might range from zero to 255 so that the value may be stored in an 8-bit byte. Usually zero corresponds to dark and 255 to white. The effects of gray level quantization are shown in Figure 6, which shows the same image displayed at 1, 2, 4 and 8 bits per pixel. The 1 bit, or binary, image shows only two shades of gray, black for 0 and white for 1. The binary image is used to display the silhouette of an object if the projection is nearly orthographic. Another characteristic of such an image is the occurrence of false contouring. In this case, contours may be produced by coarse quantization that is not actually present in the original image. As the number of gray shades is increased, we reach a point where the differences are indistinguishable. A conservative estimate for the number of gray shades distinguishable by the normal human viewer is about 64. However, changing the viewing illumination can significantly increase this range.

In order to illustrate a simple example of a digital image, consider a case where a digitizing device, like an optical scanner, is required to capture an image of 8 by 8 pixels
that represent the letter ‘V’. If the lower intensity value is 0 and higher intensity value is 9, then the digitized image we expect should look like Figure 7. This example illustrates how numbers can be assigned to represent specific characters.

On the other hand, in a color image, the image function is a vector function with color components such as red, green, and blue defined at each point. In this case, we can assign a particular color to every gray level value of the image. Assume that the primary colors, red, green and blue, are each scaled between 0 and 1 and that, for a given gray level, a proportion of each of the primary color components can be appropriately assigned. In this case, the three primary colors comprise the axes of a unit cube where the diagonal of the cube represents the range of gray level intensities, the origin of the cube corresponds to black with values (0, 0, 0) and the opposite end of the diagonal (1,1,1) represents white. The color cube is shown in Figure 8.

In general, a three dimensional color image function at a fixed time and with a given spectral illumination may be written as a three dimensional vector in which each component is a function of space, time and spectrum:

\[
\text{Colorimage} = \begin{bmatrix}
R(x, y, z, t, \lambda) \\
G(x, y, z, t, \lambda) \\
B(x, y, z, t, \lambda)
\end{bmatrix}
\] (4)

Images may be formed and processed through a continuous spatial range using optical techniques. However, machine vision refers to digital or computer processing of the
spatial information. Therefore, the range of the spatial coordinates will be discrete values. The necessary transformation from a continuous range to the discrete range is called *sampling*, and it is usually performed with a discrete array camera or the discrete array of photoreceptors of the human eye.

The viewing geometry, shown in Figure 9, is also an important factor. Surface appearance can vary from diffuse to specular with quite different images. Lambert's law for diffuse surface reflection surfaces demonstrates that the angle between the light source location and the surface normal is most important in determining the amount of reflected light, as shown in Figure 10. For specular reflection, shown in Figure 11, both the angle between the light source and surface normal and the angle between the viewer and surface normal are important in determining the appearance of the reflected image.

**Frequency Space Analysis**

Since an image can be described as a spatial distribution of density in a space of one, two or three dimensions, it may be transformed and represented in a different way in another space. One of the most important transformations is the Fourier transform. Historically, computer vision may be considered a new branch of signal processing, an area where Fourier analysis has had one of its most important applications. Fourier analysis gives us a useful representation of a signal because signal properties and basic operations, like linear filtering and modulations, are easily described in the Fourier domain. A common example of Fourier transforms can be seen in the appearance of stars. A star looks like a small point of twinkling light. However, the small point of light we observe is actually
the far field Fraunhoffer diffraction pattern or Fourier transform of the image of the star. The twinkling is due to the motion of our eyes. The moon image looks quite different since we are close enough to view the near field or Fresnel diffraction pattern.

While the most common transform is the Fourier transform, there are also several closely related transforms. The Hadamard, Walsh, and discrete cosine transforms are used in the area of image compression. The Hough transform is used to find straight lines in a binary image. The Hotelling transform is commonly used to find the orientation of the maximum dimension of an object [5].

**Fourier transform**

The one-dimensional Fourier transform may be written as:

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-2\pi iux} dx$$  \hspace{1cm} (5)

In the two-dimensional case, the Fourier transform and its corresponding inverse representation are:
The discrete two-dimensional Fourier transform and corresponding inverse relationship may be written as:

\[
F(u, v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-i2\pi \frac{(ux + vy)}{N}}
\]

\[
f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u, v) e^{i2\pi \frac{(ux + vy)}{N}}
\]

for \(x=0,1,\ldots,N-1; y=0,1,\ldots,N-1\)

and \(u=0,1,\ldots,N-1; v=0,1,\ldots,N-1\)

**Convolution Algorithm**

The convolution theorem, that the input and output of a linear, position-invariant system are related by a convolution, is an important principle. The basic idea of convolution is that if we have two images, for example pictures A and B, then the convolution of A and B means repeating the whole of A at every point in B, or vice versa. An example of the convolution theorem is shown in Figure 12. The convolution theorem enables us to do many important things. During the Apollo 13 space flight, the astronauts took a photograph of their damaged spacecraft, but it was out of focus. Image processing methods allowed such an out of focus picture to be put back into focus and clarified.
Image Enhancement

Image enhancement techniques are designed to improve the quality of an image as perceived by a human [1]. Some typical image enhancement techniques include gray scale conversion, histogram, color composition, etc. The aim of image enhancement is to improve the interpretability or perception of information in images for human viewers, or to provide “better” input for other, automated image processing techniques.

Histograms

The simplest types of image operations are point operations, which are performed identically on each point in an image. One of the most useful point operations is based on the histogram of an image.

Histogram Processing

A histogram of the frequency that a pixel with a particular gray-level occurs within an image provides us with a useful statistical representation of the image. Consider the image shown in Figure 13 as an example. It represents a square on a light background. The object is represented by gray levels greater than 4. Figure 14 shows its histogram, which consists of two peaks.

In the case of complex images like satellite or medical images that may consist of up to 256 gray levels and 3,000 by 3,000 pixels, the resulting histograms will have many peaks.
The distribution of those peaks and their magnitude can reveal significant information about the information content of the image.

**Histogram Equalization**

Although it is not generally the case in practice, ideally the image histogram should be distributed across the range of gray scale values as a uniform distribution. The distribution, as the example in Figure 15 shows, can be dominated by a few values spanning only a limited range. Statistical theory shows that using a transformation function equal to the cumulative distribution of the gray level intensities in the image enables us to generate another image with a gray level distribution having a uniform density.

This transformation can be implemented by a three-step process:

1. Compute the histogram of the image
2. Compute the cumulative distribution of the gray levels
3. Replace the original gray level intensities using the mapping determined in (2).

After these processes, the original image shown in Figure 13 can be transformed and scaled and viewed as shown in Figure 16. The new gray level value set, $S_k$, which represents the cumulative sum, is:

$$S_k = \left( \frac{1}{7}, \frac{2}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{6}{7}, \frac{6}{7}, \frac{7}{7} \right) \quad \text{for } k=0,1,\ldots,7 \quad (8)$$

**Histogram Specification**

Even after the equalization process, certain levels may still dominate the image so that the eye cannot interpret the contribution of the other levels. One way to solve this
problem is to specify a histogram distribution that enhances selected gray levels relative to others and then reconstitutes the original image in terms of the new distribution. For example, we may decide to reduce the levels between 0 and 2, the background levels, and increase the levels between 5 and 7 correspondingly. After the similar step in histogram equalization, we can get the new gray levels set $S'_k$

$$S'_k = \left( \frac{1}{7}, \frac{5}{7}, \frac{6}{7}, \frac{6}{7}, \frac{6}{7}, \frac{7}{7}, \frac{7}{7} \right) \quad \text{for } k=0,1,\ldots,7$$

By placing these values into the image, we can get the new histogram specified image shown in Figure 17.

*Image Thresholding*

Thresholding is the process of separating an image into different regions. This may be based upon its gray level distribution. Figure 18 shows how an image looks after thresholding. The percentage threshold is the percentage level between the maximum and minimum intensity of the initial image.

*Image Analysis and Segmentation*

An important area of electronic image processing is the segmentation of an image into various regions in order to separate objects from the background. These regions may roughly correspond to objects, parts of objects, or groups of objects in the scene represented by that image. It can also be viewed as the process of identifying edges that correspond to boundaries between objects and regions that correspond to surfaces of objects in the image. Segmentation of an image typically precedes semantic analysis of the image. Their purposes are [6]:
• Data reduction: often the number of important features, i.e., regions and edges, is much smaller than the number of pixels.

• Feature extraction: the features extracted by segmentation are usually “building blocks” from which object recognition begins. These features are subsequently analyzed based on their characteristics.

A region in an image can be seen as a significant change in the gray level distribution in a specified direction. As a simple example, consider the single line of gray levels below:

0 0 0 0 0 1 0 0 0 0 0

The background is represented by gray levels with a zero value. Since the sixth pixel from the left has a different level that may also characterize a single point. This sixth point represents a discontinuity among all the other levels. The process of recognizing such discontinuities may be extended to the detection of lines within an image when they occur in groups.

**Edge Detection**

In recent years, a considerable number of edge and line detecting algorithms have been proposed, each being demonstrated to have particular merits for particular types of images [7]. One popular technique is called the parallel processing, template-matching method, which involves a particular set of windows being swept over the input image in an attempt to isolate specific edge features. Another widely used technique is called sequential scanning, which involves an ordered heuristic search to locate a particular feature.
Consider the example of a convolution mask or matrix, given below.

\[
\begin{array}{ccc}
  a_1 & a_2 & a_3 \\
  a_4 & a_5 & a_6 \\
  a_7 & a_8 & a_9 \\
\end{array}
\]  

(10)

It consists of a 3 by 3 set of values. This matrix may be convolved with the image. That is, the matrix is first located at the top left corner of the image. If we denote the gray levels in the picture corresponding to the matrix values \(a_1\) to \(a_9\) by \(v_1\) to \(v_9\), then the product is formed:

\[
T = a_1*v_1+a_2*v_2+\ldots+a_9*v_9
\]

(11)

Next, we shift the window one pixel to the right and repeat the calculation. After calculating the all pixels in the line, we then reposition the matrix one pixel down and repeat this procedure. At the end of the entire process, we have a set of \(T\) values, which enables us to determine the existence of the edge. Depending on the values used in the mask template, various effects such as smoothing or edge detection will result.

Since edges correspond to areas in the image where the image varies greatly in brightness, one idea would be to differentiate the image, looking for places where the magnitude of the derivative is large. The only drawback to this approach is that

\textit{differentiation enhances noise}. Thus, it needs to be combined with \textit{smoothing}. 

Smoothing Using Gaussians

One form of smoothing the image is to convolve the image intensity with a Gaussian function. Let us suppose that the image is of infinite extent and that the image intensity is \( I(x, y) \). The Gaussian is a function of the form

\[
G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}
\]  

(12)

The result of convolving the image with this function is equivalent to low pass filtering the image. The higher the sigma, the greater the low pass filter’s effect. The filtered image is

\[
I_\sigma(x, y) = I(x, y) * G_\sigma(x, y)
\]  

(13)

One effect of smoothing with a Gaussian function is a reduction in the amount of noise because of the low pass characteristic of the Gaussian function. Figure 20 shows the image with noise added to the original, shown in Figure 19.

Figure 21 shows the image filtered by a low pass Gaussian function with \( \sigma=3 \).

Vertical Edges

To detect vertical edges, we first convolve with a Gaussian function and then differentiate the resultant image in the \( x \) direction using
This is the same as convolving the image with the derivative of the Gaussian function in the x-direction; that is:

\begin{equation}
I_o(x, y) = I(x, y)^* G_o(x, y)
\end{equation}

One then marks the peaks in the resultant images that are above a prescribed threshold as edges (the threshold is chosen so that the effects of noise are minimized). The effect of doing this on the image shown in Figure 21 is shown in Figure 22.

**Horizontal Edges**

To detect horizontal edges, we first convolve the image with a Gaussian function, then differentiate the resultant image in the x-direction. This is the same as convolving the image with the derivative of the Gaussian function in the y-direction, that is

\begin{equation}
-\frac{x}{2\pi\sigma^2} e^{\frac{-x^2 + y^2}{2\sigma^2}}
\end{equation}

Then, the peaks in the resultant image that are above a prescribed threshold are marked as edges. The effect of this operation is shown in Figure 23.

**Canny edges detector**
To detect edges at an arbitrary orientation, one convolves the image with the convolution kernels of vertical edges and horizontal edges. Call the resultant images \( R_1(x,y) \) and \( R_2(x,y) \). Then form the square root of the sum of the squares:

\[
R = \sqrt{R_1^2 + R_2^2}
\]

This edge detector is known as the **Canny edge detector**, as shown in Figure 24, which was proposed by Canny [8]. Now set the thresholds in this image to mark the peaks as shown in Figure 25. The result of this operation is shown in Figure 26.

**Three Dimensional—Stereo**

Two-dimensional digital images can be thought of as having gray levels that are a function of two spatial variables. The most straightforward generalization to three dimensions would have us deal with images having gray levels that are a function of three spatial variables. The more common examples are the three-dimensional images of transparent microscope specimens or larger objects viewed with X-ray illumination. In these images, the gray level represents some local property, such as optical density per millimeter of path length.

Most humans experience the world as three-dimensional. In fact, most of the two-dimensional images we see have been derived from this three-dimensional world by camera systems that employ a perspective projection to reduce the dimensionality from three to two [9].
Spatially Three-dimensional Image
Consider a three-dimensional object that is not perfectly transparent, but allows light to pass through it. We can think of a local property that is distributed throughout the object in three dimensions. This property is the local optical density.

CAT Scanners
Computerized Axial Tomography (CAT) is an X-ray technique that produces three-dimensional images of a solid object.

Stereometry
Stereometry is the technique of deriving a range image from a stereo pair of brightness images. It has long been used as a manual technique for creating elevation maps of the earth’s surface.

Stereoscopic Display
If it is possible to compute a range image from a stereo pair, then it should be possible to generate a stereo pair given a single brightness image and a range image. In fact, this technique makes it possible to generate stereoscopic displays that give the viewer a sensation of depth.

Shaded Surface Display
By modeling the imaging system, one can compute the digital image that would result if the object existed and if it were digitized by conventional means. Shaded surface display
grew out of the domain of computer graphics and has developed rapidly in the past few years.

**Image Recognition and Decisions**

**Neural Networks**

Artificial neural networks can be used in image processing applications. Initially inspired by biological nervous systems, the development of artificial neural networks has more recently been motivated by their applicability to certain types of problem and their potential for parallel-processing implementations [10].

**Biological Neurons**

There are about one hundred billion neurons in the brain, and they come in many different varieties, with a highly complicated internal structure. Since we are more interested in large networks of such units, we will avoid a great level of detail, focusing instead on their salient computational features. A schematic diagram of a single biological neuron is shown in Figure 27.

The cells at the neuron connections, or synapses, receive information in the form of electrical pulses from the other neurons. The synapses connect to the cell inputs, or dendrites, and form an electrical signal output of the neuron is carried by the axon. An electrical pulse is sent down the axon, or the neuron “fires,” when the total input stimuli from all of the dendrites exceed a certain threshold. Interestingly, this local processing of interconnected neurons results in self-organized emergent behavior.
Artificial Neuron Model

The most commonly used neuron model, depicted in Figure 28, is based on the model proposed by McCulloch and Pitts in 1943 [11]. In this model, each neuron's input, $a_1-a_n$, is weighted by the values $w_{i1}-w_{in}$. A bias, or offset, in the node is characterized by an additional constant input $w_0$. The output, $a_i$, is obtained in terms of the equation:

$$a_i = f\left(\sum_{j=1}^{N} a_j w_{ij} + w_0\right)$$

(17)

Feed-forward and Feed-back Networks

Figure 29 shows a feed-forward network in which the neurons are organized into an input layer, hidden layer or layers, and an output layer. The values for the input layer are set by the environment, while the output layer values, analogous to a control signal, are returned to the environment. The hidden layers have no external connections; they only have connections with other layers in the network. In a feed-forward network, a weight $w_{ij}$ is only nonzero if neuron i is in one layer and neuron j is in the previous layer. This ensures that information flows forward through the network, from the input layer to the hidden layer(s) to the output layer. More complicated forms for neural networks exist and can be found in standard textbooks. Training a neural network involves determining the weights $w_{ij}$ such that an input layer is presented with information results in the output layer having a correct response. This training is the fundamental concern when attempting to construct a useful network.
Feed-back networks are more general than feed-forward networks and may exhibit different kinds of behavior. A feed-forward network will normally settle into a state that is dependent on its input state, but a feed-back network may proceed through a sequence of states, even though there is no change in the external inputs to the network.

**Supervised Learning and Unsupervised Learning**

Image recognition and decision-making is a process of discovering, identifying and understanding patterns that are relevant to the performance of an image-based task. One of the principal goals of image recognition by computer is to endow a machine with the capability to approximate, in some sense, a similar capability in human beings. For example, in a system that automatically reads images of typed documents, the patterns of interest are alphanumeric characters, and the goal is to achieve character recognition accuracy that is as close as possible to the superb capability exhibited by human beings for performing such tasks.

Image recognition systems can be designed and implemented for limited operational environments. Research in biological and computational systems is continually discovering new and promising theories to explain human visual cognition. However, we do not yet know how to endow these theory and applications with a level of performance that even comes close to emulating human capabilities in performing general image decision functionality. For example, some machines are capable of reading printed, properly formatted documents at speeds that are orders of magnitude faster that the speed that the most skilled human reader could achieve. However, systems of this type are
highly specialized and thus have little extendibility. That means that current theoretical
and implementation limitations in the field of image analysis and decision-making imply
solutions that are highly problem dependent.

Different formulations of learning from an environment provide different amounts and
forms of information about the individual and the goal of learning. We will discuss two
different classes of such formulations of learning.

Supervised Learning

For supervised learning, a “training set” of inputs and outputs is provided. The weights
must then be determined to provide the correct output for each input. During the training
process, the weights are adjusted to minimize the difference between the desired and
actual outputs for each input pattern.

If the association is completely predefined, it is easy to define an error metric, for
example mean-squared error, of the associated response. This in turn gives us the
possibility of comparing the performance with the predefined responses (the
“supervision”), changing the learning system in the direction in which the error
diminishes.

Unsupervised Learning

The network is able to discover statistical regularities in its input space and can
automatically develop different modes of behavior to represent different classes of inputs.
In practical applications, some “labeling” is required after training, since it is not known at the outset which mode of behavior will be associated with a given input class. Since the system is given no information about the goal of learning, all that is learned is a consequence of the learning rule selected, together with the individual training data. This type of learning is frequently referred to as self-organization.

A particular class of unsupervised learning rule that has been extremely influential is Hebbian Learning [12]. The Hebb rule acts to strengthen often-used pathways in a network, and was used by Hebb to account for some of the phenomena of classical conditioning.

Primarily, some type of regularity of data can be learned by this learning system. The associations found by unsupervised learning define representations optimized for their information content. Since one of the problems of intelligent information processing deals with selecting and compressing information, the role of unsupervised learning principles is crucial for the efficiency of such intelligent systems.

**Image Processing Applications**

Artificial neural networks can be used in image processing applications. Many of the techniques used are variants of other commonly used methods of pattern recognition. However, other approaches of image processing may require modeling of the objects to be found within an image, while neural network models often work by a training process. Such models also need attention devices, or invariant properties, as it is usually
unfeasible to train a network to recognize instances of a particular object class in all orientations, sizes and locations within an image.

One method commonly used is to train a network using a relatively small window for the recognition of objects to be classified, then to pass the window over the image data in order to locate the sought object, which can then be classified once located. In some engineering applications, this process can be performed by image preprocessing operations, since it is possible to capture the image of objects in a restricted range of orientations with predetermined locations and appropriate magnification.

Before the recognition stage, a system has to be determined, such as which image transform to use. These transformations include Fourier transforms or polar coordinates or other specialized coding schemes such as the chain code. One interesting neural network model is the neocognition model of Fukushima and Miyake [13], which is capable of recognizing characters in arbitrary locations, sizes and orientations using a multi-layered network.

For machine vision, particular operations include setting the quantization levels for the image, normalizing the image size, rotating the image into a standard orientation, filtering out background detail, contrast enhancement and edge direction. Standard techniques are available for these and it may be more effective to use these before presenting the transformed data to a neural network.
Setting Up an Application

The main steps in setting up an application are:

- Physical setup: light source, camera placement, focus, field of view
- Software setup: window placement, threshold, image map
- Feature extraction: region, shape features, gray scale values, edge detection
- Decision processing: decision function, training, testing

Future Development of Machine Vision

Although image processing has been successfully applied to many industrial applications, there are still many definitive differences and gaps between machine vision and human vision [14]. Past successful applications have not always been attained easily. Many difficult problems have been solved one by one, sometimes by simplifying the background and redesigning the objects. Machine vision requirements are sure to increase in the future, as the ultimate goal of machine vision research is obviously to approach the capability of the human eye. Although it seems extremely difficult to attain, it remains a challenge to achieve highly functional vision systems.

The narrow dynamic range of detectable brightness causes a number of difficulties in image processing. A novel sensor with a wide detection range would drastically change the aspect of image processing. As microelectronics technology progresses, three-dimensional compound sensors, large scale integrated circuits, (LSI), are anticipated, to which at least preprocessing capability should be provided.
As to image processors themselves, the local-parallel pipelined processor may be further improved to provide higher processing speeds. At the same time, the multiprocessor image processor may be applied in industry when the key-processing element becomes more widely available. The image processor is likely to become smaller and faster, with new functions, in response to the advancement of semiconductor technology such as progress in system-on-chip configurations and wafer-scale integration. It may also be possible to realize 1-chip intelligent processors for high-level processing, and to combine these with 1-chip rather low-level image processors to achieve intelligent processing, such as knowledge-based or model-based processing. Based on these new developments, image processing and the resulting machine vision improvements can be expected to generate new values not merely for industry but for all aspects of human life.

**Machine Vision Applications**

Machine vision applications are so numerous that only a list is given below.

**Inspection**
- Hole location and verification
- Dimensional measurements
- Part thickness
- Component measurements
- Defect location
- Surface contour accuracy

**Part identification and sorting**
- Sorting
- Shape recognition
- Character recognition
- Inventory monitoring
• Conveyor picking - non overlapping parts
• Conveyor picking - overlapping parts
• Bin picking

**Industrial robot control**

• Tracking
• Seam welding guidance
• Part positioning and location determination
• Collision avoidance
• Machining monitoring

**Mobile robot applications**

• Navigation
• Guidance
• Tracking
• Hazard determination
• Obstacle avoidance

**Overview**

High-speed production lines, such as stamping lines, use machine vision to meet on-line, real time inspection needs. Quality inspection involves deciding whether parts are acceptable or defective, then directing motion control equipment to reject or accept them. Machine-guidance applications improve the accuracy and speed of robots and automated material handling equipment. Advanced systems enable a robot to locate a part or an assembly regardless of rotation or size. In gauging applications, a vision system works quickly to measure a variety of critical dimensions. The reliability and accuracy achieved with these methods surpasses anything possible with manual methods.
In the machine tool industry, applications for machine vision include sensing tool offset and breakage, verifying part placement and fixturing, and monitoring surface finish. A high-speed processor that once cost $80,000 now uses digital signal processing chip technology and costs less than $10,000. The rapid growth of machine vision usage in electronics, assembly systems and continuous process monitoring created an experience base and tools not available even a few years ago.

**Inspection**

The ability of an automated vision system to recognize well-defined patterns and determine if these patterns match those stored in the system’s CPU memory makes it ideal for the inspection of parts, assemblies, containers and labels. Two types of inspection can be performed by vision systems: quantitative and qualitative. Quantitative inspection is the verification that measurable quantities fall within desired ranges of tolerance, such as dimensional measurements and the number of holes. Qualitative inspection is the verification that certain components or properties are present and in a certain position, such as defects, missing parts, extraneous components or misaligned parts.

Many inspection tasks involve comparing the given object with a reference standard and verifying that there are no discrepancies. One method of inspection is called template matching. An image of the object is compared with a reference image, pixel by pixel. A discrepancy will generate a region of high differences. On the other hand, if the observed image and the reference are slightly out of registration, differences will be found along
the borders between light and dark regions in the image. This is because a slight misalignment can lead to dark pixels being compared with light pixels.

A more flexible approach involves measuring a set of the image’s properties and comparing the measured values with the corresponding expected values. An example of this approach is the use of width measurements to detect flaws in printed circuits. Here the expected width values were relatively high; narrow ones indicated possible defects.

**Edge-Based Systems**

Machine vision systems, which operate on edge descriptions of objects, have been developed for a number of defense applications. Commercial edge based systems with pattern recognition capabilities have reached markets now. The goal of edge detection is to find the boundaries of objects by marking points of rapid change in intensity. Sometimes, the systems operate on edge descriptions of images as “gray-level” image systems. These systems are not sensitive to the individual intensities of patterns, only to changes in pixel intensity.

**Component or Attribute Measurements**

An attribute measurement system calculates specific qualities associated with known object images. Attributes can be geometric patterns, area, length of perimeter, or length of straight lines. Such systems analyze a given scene for known images with predefined attributes. Attributes are constructed from previously scanned objects and can be rotated to match an object at any given orientation. This technique can be applied with minimal preparation. However, orienting and matching are used most efficiently in applications
permitting standardized orientations, since they consume significant processing time.

Attribute measurement is effective in the segregating or sorting of parts, counting parts,
flaw detection, and recognition decisions.

**Hole Location**

Machine vision is ideally suited for determining if a well-defined object is in the correct
location relative to some other well-defined object. Machined objects typically consist of
a variety of holes that are drilled, punched, or cut at specified locations on the part. Holes
may be in the shape of circular openings, slits, squares or shapes that are more complex.

Machine vision systems can verify that the correct holes are in the correct locations, and
they can perform this operation at high speeds. A window is formed around the hole to
be inspected. If the hole is not too close to another hole or to the edge of the workpiece,
only the image of the hole will appear in the window and the measurement process will
simply consist of counting pixels. Hole inspection is a straightforward application for
machine vision. It requires a two-dimensional binary image and the ability to locate
dges, create image segments, and analyze basic features. For groups of closely located
holes, it may also require the ability to analyze the general organization of the image and
the position of the holes relative to each other.

**Dimensional Measurements**

A wide range of industries and potential applications require that specific dimensional
accuracy for the finished products be maintained within the tolerance limits. Machine
vision systems are ideal for performing 100% accurate inspections of items which are
moving at high speeds or which have features which are difficult to measure.

Dimensions are typically inspected using image windowing to reduce the data processing
requirements. A simple linear length measurement might be performed by positioning a long width window along the edge. The length of the edge could then be determined by counting the number of pixels in the window and translating into inches or millimeters.

The output of this dimensional measurement process is a “pass-fail” signal received by a human operator or by a robot. In the case of a continuous process, a signal that the critical dimension being monitored is outside the tolerance limits may cause the operation to stop, or it may cause the forming machine to automatically alter the process.

**Defect location**

In spite of the component being present and in the correct position, it may still be unacceptable because of some defect in its construction. The two types of possible defects are functional and cosmetic.

A functional defect is a physical error, such as a broken part, which can prevent the finished product from performing as intended. A cosmetic defect is a flaw in the appearance of an object, which will not interfere with the product’s performance, but may decrease the product’s value when perceived by the user. Gray scale systems are ideal for detecting subtle differences in contrast between various regions on the surface of the parts, which may indicate the presence of defects. Some examples of defect inspection include the inspection of:

- label position on bottles
- deformations on metal cans
- deterioration of dies
- glass tubing for bubbles
- cap seals for bottles
Surface Contour Accuracy

The determination of whether a three-dimensional curved surface has the correct shape or not is an important area of surface inspection. Complex manufactured parts such as engine block castings or aircraft frames have very irregular three-dimensional shapes. However, these complex shapes must meet a large number of dimensional tolerance specifications. Manual inspection of these shapes may require several hours for each item. A vision system may be used for mapping the surface of these three-dimensional objects.

Part Identification and Sorting

The recognition of an object from its image is the most fundamental use of a machine vision system. Inspection deals with the examination of objects without necessarily requiring that the objects be identified. In part recognition however, it is necessary to make a positive identification of an object and then make the decision from that knowledge. This is used for categorization of the objects into one of several groups. The process of part identification generally requires strong geometric feature interpretation capabilities. The applications considered often require an interface capability with some sort of part handling equipment. An industrial robot provides this capability.

There are manufacturing situations that require that a group of varying parts be categorized into common groups and sorted. In general, parts can be sorted based on several characteristics, such as shape, size, labeling, surface markings, color and other
criteria, depending on the nature of the application and the capabilities of the vision system.

**Character Recognition**

Usually in manufacturing situations, an item can be identified solely based on the identification of an alphanumeric character or a set of characters. Serial numbers on labels identify separate batches in which products are manufactured. Alphanumeric characters may be printed, etched, embossed or inscribed on consumer and industrial products. Recent developments have provided certain vision systems with the capability of reading these characters.

**Inventory Monitoring**

Categories of inventories, which can be monitored for control purposes, need to be created. The sorting process of parts or finished products is then based on these categories. Vision system part identification capabilities make them compatible with inventory control systems for keeping track of raw material, work in process, and finished goods inventories. Vision system interfacing capability allows them to command industrial robots to place sorted parts in inventory storage areas. Inventory level data can then be transmitted to a host computer for use in making inventory level decisions.

**Conveyor Picking – Overlap**

One problem encountered during conveyor picking is overlapping parts. This problem is complicated by the fact that certain image features, such as area, lose meaning when the
images are joined together. In cases of a machined part with an irregular shape, analysis of the overlap may require more sophisticated discrimination capabilities, such as the ability to evaluate surface characteristics or to read surface markings.

**No Overlap**

In manufacturing environments with high volume mass production, workpieces are typically positioned and oriented in a highly precise manner. Flexible automation, such as robotics, is designed for use in the relatively unstructured environments of most factories. However, flexible automation is limited without the addition of the feedback capability that allows it to locate parts. Machine vision systems have begun to provide the capability. The presentation of parts in a random manner, as on a conveyor belt, is common in flexible automation in batch production. A batch of the same type of parts will be presented to the robot in a random distribution along the conveyor belt. The robot must first determine the location of the part and then the orientation so that the gripper can be properly aligned to grip the part.

**Bin Picking**

The most common form of part representation is a bin of parts that have no order. While a conveyor belt insures a rough form of organization in a two-dimensional plane, a bin is a three-dimensional assortment of parts oriented randomly through space. This is one of the most difficult tasks for a robot to perform. Machine vision is the most likely tool that will enable robots to perform this important task. Machine vision can be used to locate a part, identify orientation, and direct a robot to grasp the part.
Industrial Robot Control

Tracking

In some applications like machining, welding, assembly, or other process-oriented applications, there is a need for the parts to be continuously monitored and positioned relative to other parts with a high degree of precision. A vision system can be a powerful tool for controlling production operations. The ability to measure the geometric shape and the orientation of the object coupled with the ability to measure distance is important. A high degree of image resolution is also needed.

Seam Welding Guidance

Vision systems used for this application need more features than the systems used to perform continuous welding operations. They must have the capability to maintain the weld torch, electrode and arc in the proper positions relative to the weld joint. They must also be capable of detecting weld joints details, such as widths, angles, depths, mismatches, root openings, tack welds and locations of previous weld passes. The capacity to perform under conditions of smoke, heat, dirt and operator mistreatment is also necessary.

Part Positioning and Location Determination

Machine vision systems have the ability to direct a part to a precise position so that a particular machining operation may be performed on it. As in guidance and control applications, the physical positioning is performed by a flexible automation device such
as a robot. The vision system insures that the object is correctly aligned. This facilitates the elimination of expensive fixturing. The main concern here would be how to achieve a high level of image resolution so that the position can be measured accurately. In cases in which one part would have to touch another part, a touch sensor might also be needed.

**Collision Avoidance**

Occasionally, there is a case in industry where robots are being used with flexible manufacturing equipment and the manipulator arm can come into contact with another piece of equipment, a worker, or other obstacles, causing an accident. Vision systems can be effectively used to prevent this. This application would need the capability of sensing and measuring relative motion as well as spatial relationships among objects. A real-time processing capability would be required in order to make rapid decisions and prevent contact before any damage would be done.

**Machining Monitoring**

The popular machining operations like drilling, cutting, deburring, gluing and others that can be programmed off-line have employed robots successfully. Machine vision can greatly expand these capabilities in applications requiring visual feedback. The advantage of using a vision system with a robot is that the vision system can guide the robot to a more accurate position by compensating for errors in the robot’s positioning accuracy. Human errors, such as incorrect positioning and undetected defects, can be overcome by using a vision system.
Mobile Robot Applications

This is an active research area. Some of the topics are listed below.

- Navigation
- Guidance
- Tracking
- Hazard determination
- Obstacle avoidance

Conclusions and Recommendations

Machine vision, even in its short history, has been applied to practically every type of imagery with various degrees of success. Machine vision is a multidisciplinary field. It covers diverse aspects of optics, mechanics, electronics, mathematics, photography, and computer technology. This paper attempts to collect the fundamental concepts of machine vision for a relatively easy introduction to this field.

The declining cost of both processing devices and required computer equipment make continued growth for the field likely. Several new technological trends promise to stimulate further growth of computer visioning systems. Among these are:

- Parallel processing, made practical by low-cost microprocessors
- Inexpensive charge-coupled devices (CCDs) for digitizing images
- New memory technologies for large, low-cost image storage arrays
- Inexpensive, high-resolution color display systems.
Machine vision systems can be applied to many manufacturing operations where human vision is traditional. These systems are best for applications in which their speed and accuracy over long time periods enable them to outperform humans. Some manufacturing operations depend on human vision as part of the manufacturing process. Machine vision can accomplish tasks that humans cannot perform due to hazardous conditions and carry out these tasks at a higher confidence level than humans. Beyond inspecting products, the human eye is also valued for its ability to make measurement judgements or to perform calibration. This will be one of the most fruitful areas for using machine vision to replace labor. The benefits involved include:

- Better quality products
- Labor replacement
- Warranty reduction
- Rework Reduction
- Higher machine productivity
References


**List of Illustrations**

Figure 1. The visual cortex is the center for high level processing in the human brain.

Figure 2. Anatomy of the eye showing location of retina.

Figure 3. Example of neural processing in the early stages of vision.

Figure 4. Black and white image function.

Figure 5. Image at various spatial resolutions.

Figure 6. Images at various gray scale quantization ranges.

Figure 7. Digitized image.

Figure 8. Color cube shows the three dimensional nature of color.

Figure 9. Image surface and viewing geometry effects.

Figure 10. Diffuse surface reflection.

Figure 11. Specular reflection.

Figure 12. An example of the convolution theorem.

Figure 13. A square on a light background.

Figure 14. Histogram with bi-modal distribution containing two peaks.

Figure 15. An example of histogram equalization.

Figure 16. Original image before histogram equalization.
Figure 17. New image after histogram equalization.
Figure 18. Image after thresholding.
Figure 19. A digital image from a camera.
Figure 20. Image with noise added.
Figure 21. Filtered image by a Gaussian filter with $\sigma = 3$.
Figure 22. Derivative of Gaussian filter.
Figure 23. Horizontal edges detected by of the Gaussian in the y-direction.
Figure 24. Canny edge operator.
Figure 25. Thresholded image to mark the peaks.
Figure 26. Results of the Canny edge detector.
Figure 27. A schematic diagram of a single biological neuron.
Figure 28. ANN model proposed by McCulloch and Pitts in 1943.
Figure 29. A feed-forward network

Table 1. GIF image format characteristics.
Table 2. Other image formats.
<table>
<thead>
<tr>
<th>Header and Color Table Information</th>
<th>Header</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Screen Descriptor</td>
<td></td>
</tr>
<tr>
<td>Global Color Table</td>
<td></td>
</tr>
<tr>
<td>Image 1</td>
<td>Local Image Descriptor</td>
</tr>
<tr>
<td></td>
<td>Local Color Table</td>
</tr>
<tr>
<td></td>
<td>Image Data</td>
</tr>
<tr>
<td>Image 2</td>
<td>Local Image Descriptor</td>
</tr>
<tr>
<td></td>
<td>Local Color Table</td>
</tr>
<tr>
<td></td>
<td>Image Data</td>
</tr>
<tr>
<td>Image N</td>
<td>Local Image Descriptor</td>
</tr>
<tr>
<td></td>
<td>Local Color Table</td>
</tr>
<tr>
<td></td>
<td>Image Data</td>
</tr>
<tr>
<td></td>
<td>Trailer</td>
</tr>
</tbody>
</table>

Table 1. GIF image format characteristics.
<table>
<thead>
<tr>
<th>Image Type</th>
<th>Extensions</th>
<th>Creator</th>
<th>Filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmp (Microsoft Windows bitmap image file)</td>
<td>bmp, dib, vga, bga, rle, rl4, rl8</td>
<td>Microsoft</td>
<td>Imconv, imagemagick, xv, PhotoShop, pbmplus</td>
</tr>
<tr>
<td>Cgm (Computer Graphics Metafile)</td>
<td>cgm</td>
<td>American National Standards Institute, Inc</td>
<td></td>
</tr>
<tr>
<td>ddiff (DEC DDIF file)</td>
<td>ddiff</td>
<td>Digital Equipment Co. (DEC)</td>
<td>pbmplus</td>
</tr>
<tr>
<td>gif (Graphics Interchange Format file)</td>
<td>gif, giff</td>
<td>CompuServe Information Service</td>
<td>Imconv, imagemagick, xv, PhotoShop, pbmplus, Utah Rasterfile Toolkit</td>
</tr>
<tr>
<td>ico (Microsoft Windows icon image file)</td>
<td>ico</td>
<td>Microsoft</td>
<td>imconv</td>
</tr>
<tr>
<td>img (Img-whatnot file)</td>
<td>img</td>
<td></td>
<td></td>
</tr>
<tr>
<td>jpeg (Joint Photographic Experts Group compressed file)</td>
<td>jpeg, jpg, jfif</td>
<td>The Independent JPEG Group</td>
<td>Imagemagick, xv, JPEGsoftware, Photoshop,DeBabelizer</td>
</tr>
<tr>
<td>mpnt (Apple Macintosh MacPaint file)</td>
<td>mpnt, macp, pntg, mac, paint</td>
<td>Apple Computer, Inc.</td>
<td>Imconv, sgitools, pbmplus, PhotoShop, Utah Rasterfile Toolkit</td>
</tr>
<tr>
<td>pcd (Photo-CD file)</td>
<td>pcd</td>
<td>Kodak</td>
<td>Imagemagick, pbmplus, PhotoShop, DeBabelizer</td>
</tr>
<tr>
<td>pcx (ZSoft IBM PC Paintbrush file)</td>
<td>pcx, pcc</td>
<td>ZSoft Corp. (PC Paintbrush)</td>
<td>Imagemagick, xv, Photoshop,DeBabelizer</td>
</tr>
<tr>
<td>pdf (Adobe Acrobat PDF file)</td>
<td>pdf</td>
<td>ADEX Corp.</td>
<td></td>
</tr>
<tr>
<td>photoshop (Adobe PhotoShop file)</td>
<td>photoshop</td>
<td>Adobe</td>
<td>PhotoShop, DeBabelizer</td>
</tr>
<tr>
<td>pix (Alias image file)</td>
<td>pix, alias</td>
<td>Alias Research, Inc.</td>
<td>Imconv, sgitools, DeBabelizer</td>
</tr>
<tr>
<td>ps (Adobe PostScript file)</td>
<td>ps, ps2, postscript, psid</td>
<td>Adobe</td>
<td>Imconv, sgitools, xv, pbmplus, PhotoShop, Utah Rasterfile Toolkit</td>
</tr>
<tr>
<td>tiff (Tagged-Image File Format)</td>
<td>tiff, tif</td>
<td>Aldus, MicroSoft, and NeXT</td>
<td>Imconv, sgitools, xv, pbmplus, PhotoShop, Utah Rasterfile Toolkit</td>
</tr>
<tr>
<td>xwd (X Window System window dump image file)</td>
<td>xwd, x11</td>
<td>X Consortium / MIT</td>
<td>Imconv, sgitools, pbmplus, Utah Rasterfile Toolkit</td>
</tr>
</tbody>
</table>

Table 2. Other image formats.
Figure 1. The visual cortex is the center for high level processing in the human brain.
Figure 2. Anatomy of the eye showing location of retina

Figure 3. Example of neural processing in the early stages of vision.

Figure 4. Black and white image function.
Figure 5  Image at various spatial resolutions.
Figure 6. Images at various gray scale quantization ranges.
Figure 7.

Figure 8. Color cube shows the three dimensional nature of color
Figure 9. Image surface and viewing geometry effects.

Figure 10. Diffuse surface reflection
Figure 11. Specular reflection

![Original image](image1.png) ![Image convolved with Roberts kernel](image2.png)

a) Original image; b) Image convolved with Roberts kernel

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Robert's kernel**

Figure 12. An example of the convolution theorem
Figure 13. A square on a light background

Figure 14. Histogram with bimodal distribution containing two peaks.

Figure 15. An example of histogram equalization. (a) original image, (b) histogram of original image (c) equalized histogram (d) equalized image.
Figure 16. Original image before histogram equalization

Figure 17. New image after histogram equalization.
Figure 18. Image thresholding

Figure 19. A digitized image from a camera

Figure 20. The original image corrupted with noise
Figure 21. The noisy image filtered by a Gaussian of variance 3
Figure 22. The vertical edges of the original image

Figure 23. The horizontal edges of the original image
Figure 24. The magnitude of the gradient

Figure 25. Thresholding of the peaks of the magnitude of the gradient
Figure 26. Edges of the original image
Figure 27. Model of a biological neuron.

Figure 28. McCulloch-Pitts model of a neuron.
Figure 29. Feed-forward neural network