# IMAGE PROCESSING

## BACKGROUND

Image processing consists of a wide variety of techniques and mathematical tools to process an input image. An image is processed as soon as we start extracting data from it. The data of interest in object recognition systems are those related to the object under investigation. An image usually goes through some enhancement steps, in order to improve the extractability of interesting data and subside other data.

Extensive research has been carried out in the area of image processing over the last 30 years. Image processing has a wide area of applications. Some of the important areas of application are business, medicine, military, and automation.

Image processing has been defined as a wide variety of techniques that includes coding, filtering, enhancement, restoration registration, and analysis. In many applications, such as the recognition of three-dimensional objects, image processing and pattern recognition are not separate disciplines. Pattern recognition has been defined as a process of extracting features and classifying objects. In every three-dimensional (*3-D*) object recognition system there are units for image processing and there are others for pattern recognition.

There are two different approaches to image processing:

1. *Analog processing.* This approach is very fast since the time involved in analog-to-digital (*AD*) and a digital-to-analog (*DA*) conversion is saved. But this approach is not flexible since the manipulation of images is very hard.
2. *Digital processing.* This approach is slower than the analog approach but is very flexible, since manipulation is done very easily. The processing time of this approach is tremendously improved by the advent of parallel processing techniques.

### Digital Image Processing

Digital image processing is defined in Ref. 1 as "the processing of two dimensional images by a digital computer." A digital image is represented by an array of regularly spaced and very small quantized samples of the image. Two processes that are related to any digital system are sampling and quantization.

When a picture is digitized, it is represented by regularly spaced samples of this picture. These quantized samples are called *pixels.* The array of pixels that are processed in practice can be quite large. To represent an ordinary black and white television (*TV*) image digitally, an array of $512 \times 512$ pixels is required. Each pixel is represented by an 8 bit number to allow 256 gray levels. Hence a single TV picture needs about $2 \times 10^6$ bits.

Digital image processing encompasses a wide variety of techniques and mathematical tools. They have all been developed for use in one or the other of two basic activities that constitute digital image processing: image preprocessing and image analysis.

An approach called the state-space approach has been recently used in modeling image processors. These image processors are made of linear iterative circuits. The state-space model is used efficiently in image processing and image analysis. If the model of an image processor is known, the realization of a controllable and observable image processor is then very simple.

**Image Preprocessing.** Image preprocessing is an early stage activity in image processing that is used to prepare an input image for analysis to increase its usefulness. Image preprocessing includes image enhancement, restoration, and registration.

Image enhancement accepts a digital image as input and produces an enhanced image as an output; in this context, *enhanced* means *better* in some respects. This includes improving the contrast, removing geometric distortion, smoothing the edges, or altering the image to facilitate the interpretation of its information content.

In image restoration, the degradation is removed from the image to produce a picture that resembles the original undegraded picture.

In image registration, the effects of sensor movements are removed from the image or to combine different pictures received by different sensors of the same field.

**Image Analysis.** Image analysis accepts a digital image as input and produces data or a report of some type. The produced data may be the features that represent the object or objects in the input image. To produce such features, different processes must be performed that include segmentation, boundary extraction, silhouette extraction, and feature extraction. The produced features may be quantitative measures, such as moment invariants, and Fourier descriptors, or even symbols, such as regular geometrical primitives.

### Sampling and Quantization

Quantization is the process of representing a very large number (possibly infinite) of objects with a smaller, finite number of objects. The *representing* set of objects may be taken from the original set (e.g., the common number-rounding process) or may be completely different (e.g., the alphabetical grading system commonly used to represent test results).

In image processing systems, quantization is preceded by another step called *sampling.* The gray level of each pixel in an image is measured, and a voltage signal that is proportional to the light intensity at each pixel is generated. It is clear that the voltage signal can have any value from the voltages that are generated by the sensing device.

Sampling is the process of dividing this closed interval of a continuous voltage signal into a number of subintervals that are usually of equal length. In an 8 bit sampling and quantization process, for example, the interval of voltage signals is divided into 256 subintervals of equal length.

In the quantization process, each of the generated intervals from sampling is represented by a code word. In an

8-bit quantization process, each code word consists of an 8 bit binary number. An 8 bit analog-to-digital converter (*ADC*) can simply accomplish the tasks of sampling and quantization. The image data are now ready for further processes through use of digital computers.

For systems that involve dynamic processing of image signals [e.g., TV signals or video streams from charge-coupled device (*CCD*) cameras], the term *sampling* refers to a completely different process. In this context, sampling means taking measurements of the continuous image signal at different instants of time. Each measurement can be thought of as a single stationary image. A common problem associated with image digitization is aliasing.

The sampling theorem states that for a signal to be completely reconstructable, it must satisfy the following equation:

$$w_s = 2w \tag{1}$$

where $w_s$ is the sampling frequency and $w$ is the frequency of the sampled signal. Sampling, in this context, means taking measures of the analog signal at different instants separated by a fixed time interval $\Delta t$. This theorem is applicable on the sampling of stationary images as well, where sampling is carried through space instead of time. If the signal is band limited, sampling frequency is determined according to the frequency of its highest-frequency component. Image signals, however, are subjected to truncating, mainly because of the limitations in sensors and display devices. Sensors are capable of recognizing a limited range of gray levels. Real objects usually have wider ranges of gray levels, which means that both the gray levels higher and lower than the range of the sensor are truncated. Truncating is what causes the aliasing problem. To explain how this happens, consider the simple sinusoidal function given by $f(x) = \cos(x)$. Figure 1 shows a plot of this function and Fig. 2 shows a plot of its Fourier transform. Figure 3 shows a truncated version of that function, and Fig. 4 shows the equivalent Fourier transform. This function has infinite duration in the frequency domain. The Nyquist frequency is given by $w_n = w_s/2$. If we try to sample this signal with a sampling frequency of $w_s$, then all frequencies higher than the Nyquist frequency will have aliases within the range of the sampling frequency. In other words, aliasing causes high-frequency components of a signal to be seen as low frequencies.

This is also known as *folding*. A practical method to get red of aliasing is to prefilter the analog signal before sampling. Figure 4 shows that lower frequencies of the signal contain most of signal's power. A filter is designed so that filtered signals do not have frequencies above the Nyquist frequency. A standard analog filter transfer function may be given as

$$G_f(s) = \frac{w^2}{s^2 + 2\zeta ws + w^2} \tag{2}$$

Where $\zeta$ is the damping factor of the filter and $w$ is its natural frequency. By cascading second- and first-order filters, one can get higher-order systems that have higher performances. Three of the most commonly used filters are the Butterworth filter, ITAE filter, and Bessel filter. Bessel filters are commonly used for high-performance applications, mainly because of the following two factors:

1. The damping factors that may be obtained by a Bessel filter are generally higher than those obtained by other filters. A higher damping factor means a better cancellation of frequencies outside the desired bandwidth.
2. The Bessel filter has a linear phase curve, which means that the shape of the filtered signal is not much distorted.

To demonstrate how we can use a Bessel filter to eliminate high-frequency noise and aliasing, consider the square signal in Fig. 5. This has a frequency of 25 Hz. Another signal with a frequency of 450 Hz is superimposed on the square signal. If we try to sample the square signal with noise [Fig. 5(a)], we will get a very distorted signal [Fig. 5(b)]. Next, we prefilter this signal using a second-order Bessel filter with a bandwidth of 125 Hz and a damping factor of 0.93. The resultant signal is shown in Fig. 5(c). Figure 5(d) shows the new signal after sampling. It is clear that this signal is very close to the original square signal without noise.

## IMAGE RESTORATION

Image restoration refers to a group of techniques that are oriented toward modeling the degradation and applying the inverse process in order to recover the original image.

Each component in the imaging system contributes to the degrading of the image. Image restoration techniques try to model the degradation effect of each component and then perform operations to undo the model, to restore the original image (3).

There are two different modeling approaches for degradation: the a priori approach and the a posteriori approach. These two approaches differ in the manner in which information is gathered to describe the characteristics of the image degrading.

The a priori approach is to try to model each source of noise in the imagery system by measuring the system's responses to arbitrary noises. In many cases, deterministic models cannot be extracted and stochastic models are used instead.

The a posteriori approach is adopted when a great deal of information is known about the original image. We can develop a mathematical model for the original image and try to fit the model to the observed image. Figure 6 shows a simplified model for the image degradation and restoration processes. The original image signal $f(x, y)$ is subjected to the linear degrading function $h(x, y)$. An arbitrarily noise signal $\eta(x,y)$ is then added to create the degraded image signal $\eta(x, y)$. Reconstruction approaches try to estimate the original image signal $f(x, y)$ given the degradation signal $g(x, y)$ and some statistical knowledge of the noise signal $\eta(x, y)$.

We can broadly classify reconstruction techniques into two classes: the filtering reconstruction techniques and the algebraic techniques (5).

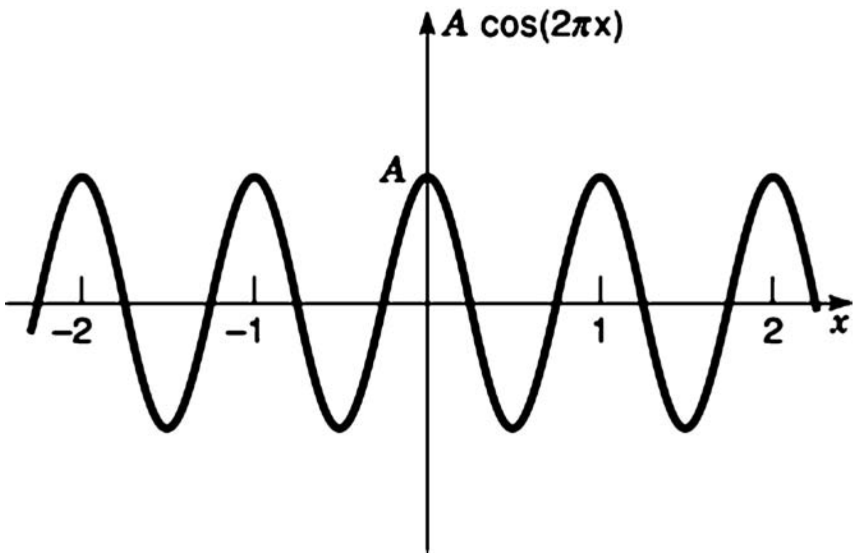


**Figure 1.** Cosine function with amplitude $A$ and frequency of 1 Hz.

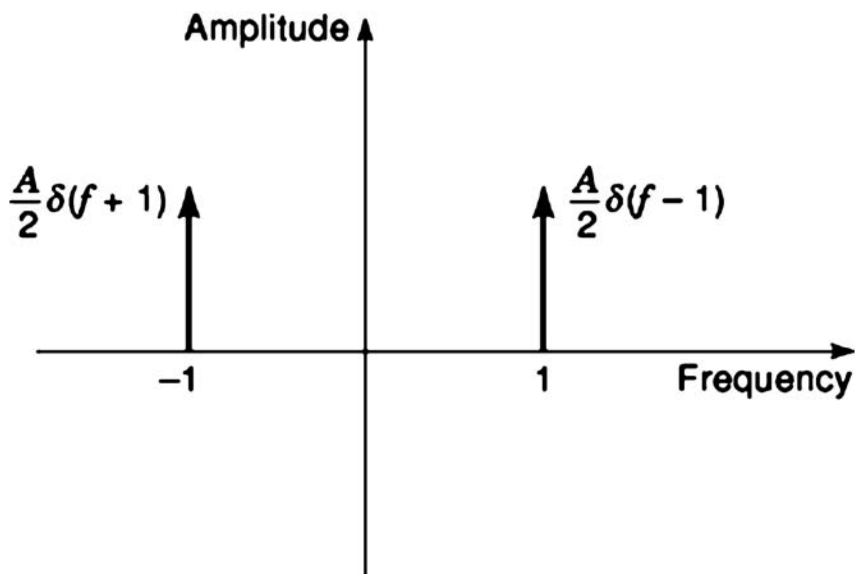


**Figure 2.** Power spectrum of the cosine function with amplitude $A$ and frequency of 1 Hz.

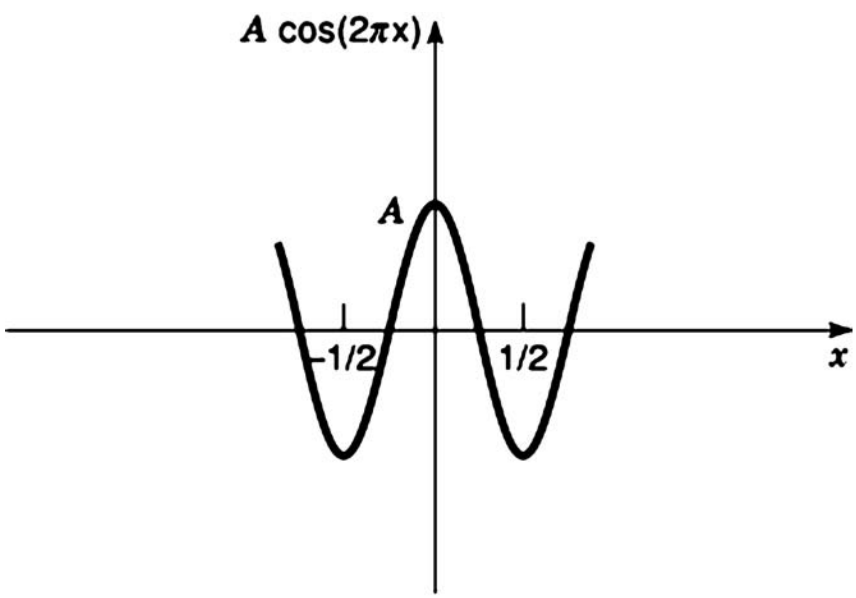


**Figure 3.** Truncated cosine function. The truncation is in the variable $x$ (e.g., time), not in the amplitude.

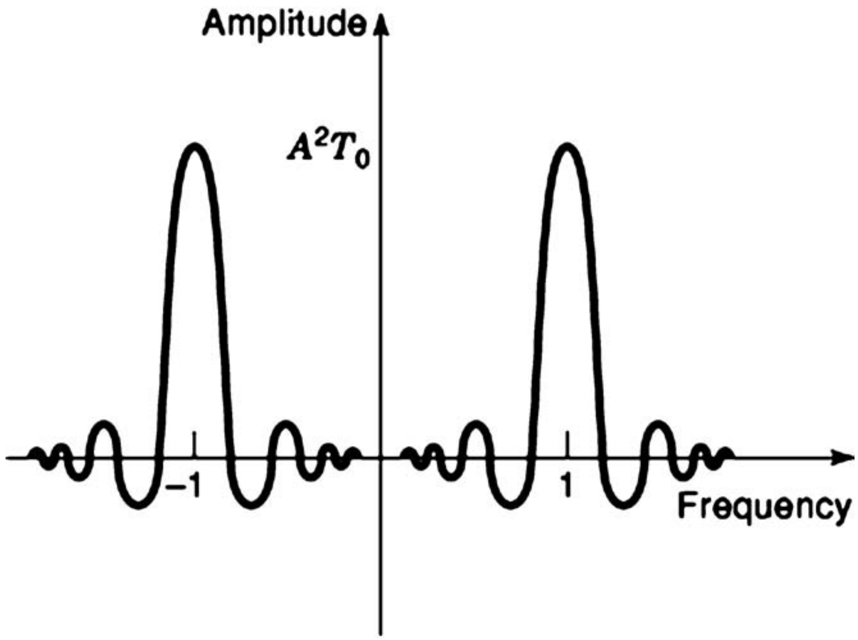


**Figure 4.** The power spectrum of the truncate cosine function is a continuous one, with maximum values at the same points, like the power spectrum of the continuous cosine function.

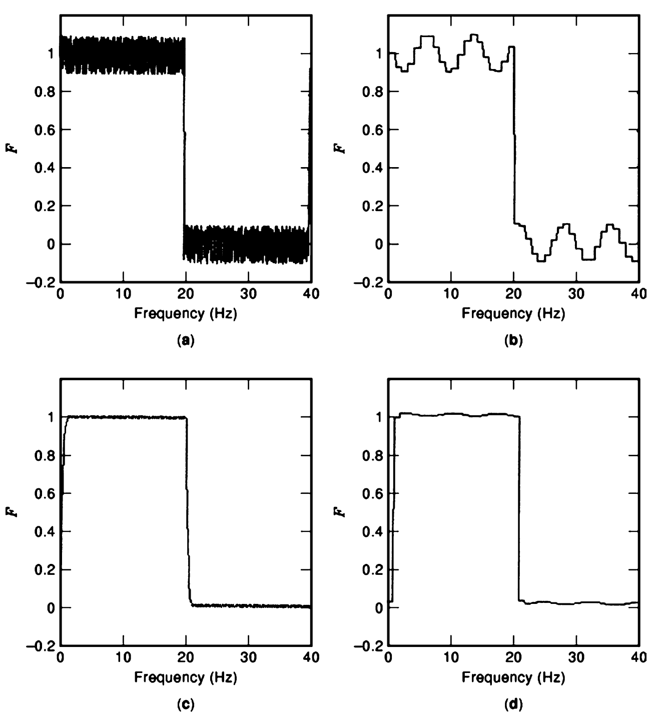


**Figure 5.** Antialiasing filtering: (a) square signal with higher-frequency noise, (b) digitized signal with noise, (c) continuous signal after using second-order Bessel filter, and (d) digitized filtered signal. Using the Bessel filter reduced noise and improved the digitization process substantially.

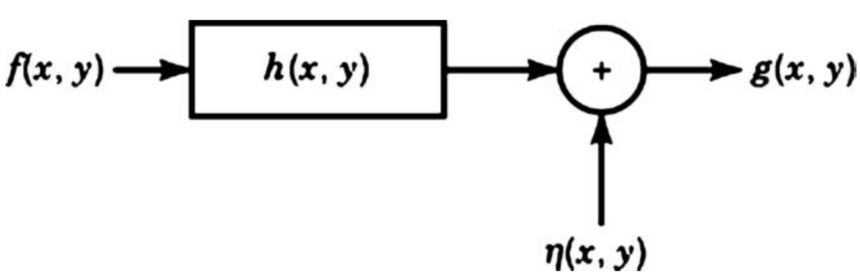


**Figure 6.** Simplified model for the degradation process. The image signal $f(x, y)$ is subjected to a linear degrading function $h(x, y)$ and an arbitrary noise $\eta(x, y)$ is added to produce the degraded signal $g(x, y)$.

### Filtering Reconstruction Techniques

These techniques are rather classical and they make use of the fact that noise signals usually have higher frequencies than image signals. This means that image signals die out faster than noise signals in high frequencies. By selecting the proper filter, one can get a good estimate of the original image signal, by reducing the effect of noise. Examples of the reconstruction filters are the deconvolution filter and the Wiener filter.

**Deconvolution Filter.** This filter is based on the concept of inverse filtering, in which the transfer function of the degraded system is inverted to produce a restored image. Figure 7 shows typical spectra of a deconvolution filtering image restoration system. If no noise is added to the system, the image is perfectly reconstructed. The presence of noise, however, will add a reconstruction error, the value of which can become quite large at spatial frequencies for which $f(x, y)$ is small. This reflects directly on regions of the image with a high density of details because they have higher spatial frequencies than other regions.

**Wiener Filter.** This filter uses the mean-squared error (*MSE*) criterion to minimize the error signal between the original and degraded image signals. The mean-squared restoration error is given by

$$\epsilon = E\{[f(x, y) - g(x, y)]^2\} \tag{3}$$

where $f(x, y)$ and $g(x, y)$ are as before. The Wiener filter equation is

$$W(m, n) = \frac{H^*(m, n)}{|H(m, n)|^2 + P_\mathrm{n}(m, n)/P_\mathrm{f}(m, n)} \tag{4}$$

where $P_\mathrm{n}(m, n)$ and $P_\mathrm{f}(m, n)$ are the power spectra of the signal and noise, respectively. The Wiener filter acts as a band-pass filter. At low spatial frequencies, it acts as an inverse filter, whereas at higher frequencies, it acts as a smooth rolloff low-pass filter.

This filter is not very suitable for use in cases in which images are investigated by the human eye. The MSE technique treats all errors equally, regardless of their spatial location in the image. The human eye, on the other hand, has high degree of tolerance to errors in darker areas on the image than elsewhere. Another limitation in this filter is that it cannot handle dynamically changing image and noise signals.

Another type of filter that is closely related to the Wiener filter is the power spectrum equalization (*PSE*) filter, the equation for which is

$$\mathrm{PSE}(m, n) = \left(\frac{1}{|H(m, n)|^2 + P_\mathrm{n}(m, n)/P_\mathrm{f}(m, n)}\right)^{1/2} \tag{5}$$

The power spectrum obtained by this filter for the reconstructed filter is identical to that of the original signal.

### Linear Algebraic Restoration Techniques

These techniques utilize matrix algebra and discrete mathematics for solving the problem of image restoration. To extract the discrete model for the restoration system, we make the following assumptions:

- The digitized original image $f(x, y)$ and the restored image $f_\mathrm{r}(x, y)$ are stored in the $M^2 \times 1$ column vectors $f$ and $f_\mathrm{r}$, respectively.
- The digitized degrading function $h(x, y)$ is stored in a square matrix $H$ ($M^2 \times M^2$).
- The degraded image $g(x, y)$ and noise $\eta(x, y)$ is stored in the $M^2 \times 1$ column vectors $g$ and $n$, respectively.

We can express the observed (degraded) image vector in the compact form:

$$\boldsymbol{g} = \boldsymbol{Hf} + \boldsymbol{n} \tag{6}$$

We use this model to derive some of the algebraic restoration techniques, namely, the unconstrained reconstruction technique, the constrained reconstruction technique, and the pseudoinverse filtering technique.

**Unconstrained Reconstruction.** If we know very little about the noise $n$, then we try to find an estimate image $f_\mathrm{r}$, such that $Hf_\mathrm{r}$ approximates $g$ in a least-squares manner. This can be accomplished by minimizing the norm of the noise $n$. Squaring the norm of both sides of Eq. (6) after substituting for $f$ by the estimate vector $f_\mathrm{r}$ and moving $Hf_\mathrm{r}$ to the other side of the equation, we get

$$\|\boldsymbol{n}\|^2 = \|\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r}\|^2 = (\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r})(\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r})' \tag{7}$$

where $\|a\|^2$ is the square of the norm of vector $a$ and is given by $\|a\|^2 = aa'$, where $a'$ is the transpose of the vector $a$. Consider the error function $E$, where

$$E(\boldsymbol{f}_\mathrm{r}) = \|\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r}\|^2 = (\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r})(\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r})' \tag{8}$$

Then our goal is to minimize $E$ with respect to $f_\mathrm{r}$. This can be accomplished by taking the derivative of $E$ with respect to $f_\mathrm{r}$, and equating the result to zero, that is

$$\frac{\partial E(\boldsymbol{f}_\mathrm{r})}{\boldsymbol{f}_\mathrm{r}} = -2\boldsymbol{H}'(\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r}) = 0 \tag{9}$$

Assuming that $H^{-1}$ exists, the solution for this equation is given as

$$\boldsymbol{f}_\mathrm{r} = \boldsymbol{H}^{-1}\boldsymbol{g} \tag{10}$$

**Constrained Reconstruction.** To account for the noise term in Eq. (6), we introduce the square matrix $Q$ ($M^2 \times M^2$) to represent some linear operator on $f$. By selecting different $Q$'s, we are able to set the goal of restoration as desired. Equation (8) is now modified to

$$E(\boldsymbol{f}_\mathrm{r}) = \|\boldsymbol{Qf}_\mathrm{r}\|^2 + \lambda(\|\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r}\|^2 - \|\boldsymbol{n}\|^2) \tag{11}$$

where $\lambda$ is a constant called the LaGrange multiplier. Again we try to minimize the error function $E$, by taking its derivative with respect to $f_\mathrm{r}$ and equating the result to zero, that is

$$\frac{\partial E(\boldsymbol{f}_\mathrm{r})}{\boldsymbol{f}_\mathrm{r}} = 2\boldsymbol{Q}'\boldsymbol{Qf}_\mathrm{r} - 2\lambda\boldsymbol{H}'(\boldsymbol{g} - \boldsymbol{Hf}_\mathrm{r}) = 0 \tag{12}$$

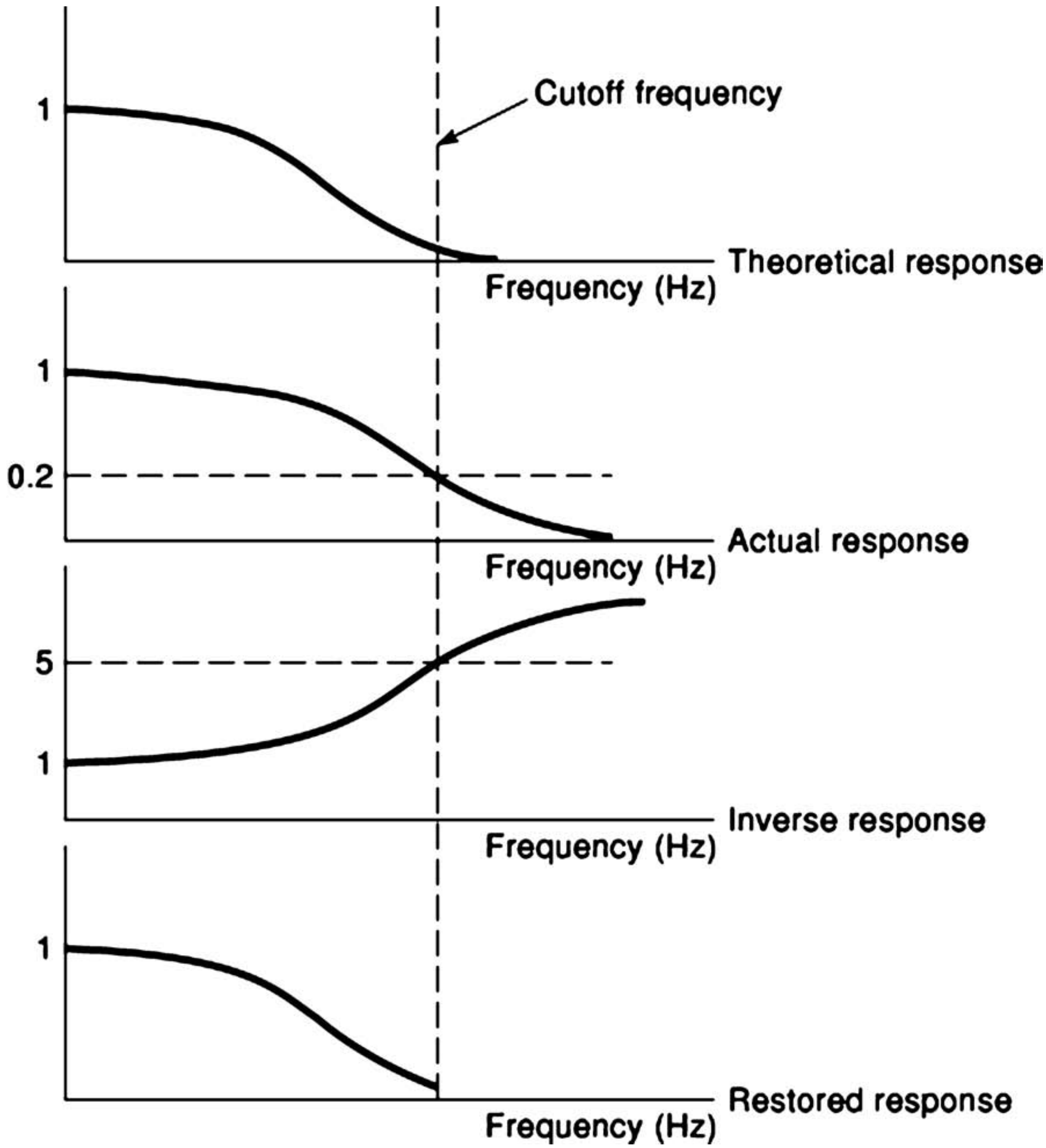


**Figure 7.** Typical spectra of a deconvolution filtering image-restoration system. The actual response of the restoration process is first inverted. Then signals with frequencies higher than the cutoff frequency are removed. The new response is finally inverted to result in a response very close to the theoretical one.

Solving for $f_r$, we get

$$\boldsymbol{f}_r = (\boldsymbol{H}'\boldsymbol{H} + \alpha\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{H}'\boldsymbol{g} \tag{13}$$

Where $\alpha = 1/\lambda$ is a constant that we adjust to satisfy the constraint.

**Pseudoinverse Filtering.** The pseudoinverse filtering is a special case of the constrained reconstruction techniques. In this technique, the constraint matrix $Q$ is the $M^2 \times M^2$ identity matrix. Equation (13) then becomes

$$\boldsymbol{f}_r = (\boldsymbol{H}'\boldsymbol{H} + \alpha\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g} \tag{14}$$

If $\alpha = 0$, this equation reduces to Eq. (10), which represents the unconstrained restoration technique.

## IMAGE ENHANCEMENT

Image enhancement refers to a group of processes that aim toward making the processed image much better than the unprocessed image. In this sense, image enhancement is closely related to image restoration. Enhancement techniques are basically heuristic procedures, that are designed to manipulate an image by taking advantage of the human vision system (2).

Image enhancement is used to remove the noise that may be added to the image from several sources, including electrical sensor noise and transmission channel errors. There are many enhancement techniques to improve the input image. Some of these techniques are gray-scale modification, point operators, contrast enhancement and filtering.

### Gray-Scale Modification

The gray-scale modification technique is very effective to modify contrast or dynamic range of an image. In this technique, the gray scale of an input image is changed to a new gray scale according to a specific transformation. The specific transformation is the relation between the intensity of the input image to the intensity of the output image. By properly choosing this transformation, one can modify the dynamic range or the contrast of an input image. For each input pixel the corresponding output pixel is obtained from the plot or the look-up table.

The transformation desired depends on the application. In many applications a good transformation is obtained by computing the histogram of the input image and studying its characteristics. The histogram of an image represents the number of pixels that has a specific gray value as a function of all gray values.

The gray-scale transformation is very simple and very efficient, but it has the following two limitations:

1. It is global—the same operation is applied to all the pixels that have the same gray level in the image.
2. The transformation that is good for one image may not be good for others.

### Point Operators

Point processing is a group of simple techniques used for image enhancement. Point operation is a direct mapping from an input image to an output image, where the pixel value in the enhanced output image depends only on the value of the corresponding pixel in the input image (5). This mapping can mathematically be given by

$$P_o = N(P_i) \tag{15}$$

where $P_o$ is the output enhanced pixel, $P_i$ is the corresponding input pixel and $N$ is the point operator. In case of discrete images, a point operator may be implemented by a look-up table (*LUT*), in which for each pixel in the output image, there is a corresponding pixel in the input image. It should be noted that point operators affect the gray level of pixels in an image but have no effect on the spatial relationships between pixels.

Point operators may be used in applications such as photometric calibration, contrast enhancement, display calibration, and contour lines.

**Types of Point Operations.** Point operators are generally classified as either linear or nonlinear operators.

***Linear Point Operations.*** The output gray level is a linear function of the input gray level. The mapping function is given by

$$G_o = N(G_i) = aG_i + b \tag{16}$$

where $G_o$ and $G_i$ are the gray levels of the output pixel and input pixel, respectively. By changing the values of the coefficients $a$ and $b$, different effects may be accomplished. For example, if $a = 1$ and $b = 0$, the gray levels of the image pixels do not change. A negative value of $a$ has the effect of complementing the image. That is, dark areas of the image become light and light areas become dark.

Figure 8 shows a plot of a linear point operator.

***Nonlinear Monotonic Point Operations.*** Nonlinear point operations are mainly used to modify the midrange gray level. One type of nonlinear point operation increases the gray level of midrange pixels, while leaving dark and light pixels little changed. This type of operator may be given by

$$G_o = G_i + AG_i(G_m - G_i) \tag{17}$$

where $G_m$ is the maximum gray level and $A$ is a parameter that determines the amount of increase ($A > 0$) or decrease ($A < 0$) in the midlevel gray range.

The *logarithmic* operator is a point operator that maps each pixel value to its logarithm.

The logarithmic mapping function is given by

$$G_o = c\ \log(G_i + 1) \tag{18}$$

In an 8 bit format, the value of $c$ is selected such that the maximum output value is 255.

### Contrast Enhancement

Contrast is defined as the range of brightness values present in an image. Contrast enhancement is a very important issue when the final output image has to be evaluated by a human observer. The human eye can distinguish up to 64 gray levels, compared to 256 different gray levels that can be achieved with an 8 bit imagery system.

The range of brightness values collected by a sensor may not match the capabilities of the output display medium. Contrast enhancement involves changing the range of brightness in an image in order to increase contrast.

Selecting a suitable method for contrast enhancement depends on two main factors: the state of the original image and the requirements from the final image. In many cases, more than one step of contrast enhancement is required to obtain the desired output. Two common contrast enhancement methods are the contrast-stretching technique and window-and-level technique.

**Contrast Stretching.** This process expands the range of brightness of an original digital photograph into a new distribution. As a result, the total range of sensitivity of the display device can be utilized. Contrast stretching makes minute variations within the image data more obvious.

Contrast stretching may be accomplished by both linear and nonlinear methods. Examples of linear contrast enhancement are the minimum-maximum linear contrast stretch, percentage linear contrast stretch, and piecewise

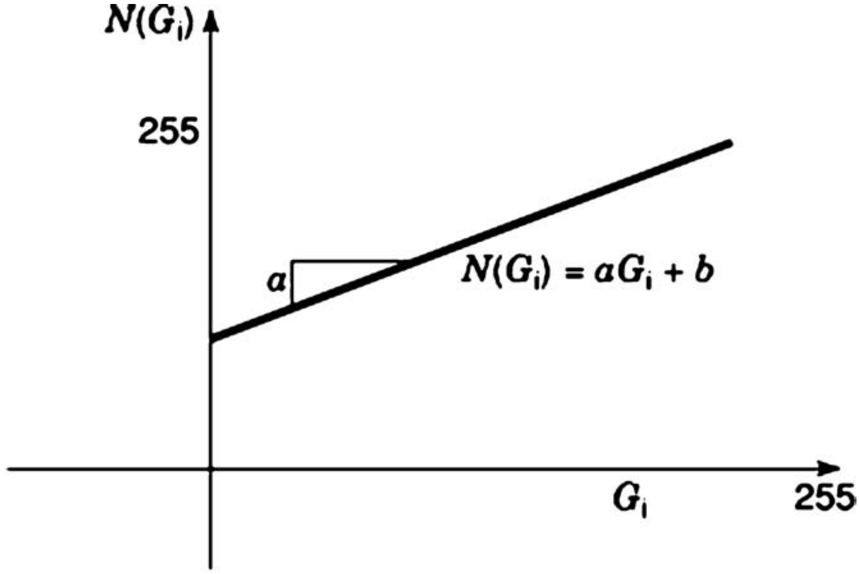


**Figure 8.** Linear point operator. By changing the slope $a$ and vertical intersection $b$, we can use point operators to produce different effects in the processed image.

linear contrast stretch. An example of nonlinear contrast enhancement is the histogram equalization.

**Window and Level.** If the brightness range of the element of interest in an image is limited, then a method known as Window and Level can be used to enhance its contrast on the expense of other elements. All signal levels below the desired range will be mapped onto the same low output level and all those with higher levels will be mapped onto the same high output level. Only pixels within the desired range will be represented with an acceptable range of contrast.

Two parameters define the desired range of contrast: the middle point Level and the width of the range Window. Figure 9 shows how the window-and-level technique affects the histogram of the image.

If the original image has a poor contrast level, then the window-and-level technique alone will not produce satisfactory results. A better technique would be to use one of the contrast-stretching methods to increase the contrast range first and then apply the window-and-level method on the desired element, taking into consideration its new range of contrast.

### Low-Pass Filtering

Low-pass filtering reduces the high-frequency components while retaining the low-frequency components. The low-frequency components of an image constitute most of the energy in an image. In other words, low-pass filtering removes a large amount of noise at the expense of removing a small amount of the actual signal.

Low-pass filtering can be used efficiently to remove additive random and multiplicative noise, but at the same time it blurs the processed image. Blurring is the primary limitation of low-pass filtering. Figure 10 shows a photograph of the famous Taj-Mahal, one of the Seven Wonders of the World. Figure 11 shows the effect of low-pass filtering on the same image.

### High-Pass Filtering

High-pass filtering reduces the low-frequency components while preserving the high-frequency components of an input image. Since the high-frequency components generally correspond to edges or fine details of an image, high-pass filtering increases the contrast and thus sharpens the edges of the image.

High-pass filtering can be used to remove image blur but at the same time it tends to increase background noise power. Figure 12 shows the effect of high-pass filtering on Fig. 10.

### Median Filtering

Median filtering is a nonlinear digital technique used for image enhancement. The basic idea behind median filtering is the removal of noise by finding the median of neighboring pixels and assigning it to the center point. It is used to reduce impulsive noise and to enhance edges while reducing random noise. Figure 13 shows the effect of filtering the photograph image with a median filter with a window of three pixels.

## OBJECT RECOGNITION

Object recognition includes the process of determining the object's identity or location in space. The problem of object or target recognition starts with the sensing of data with the help of sensors, such as video cameras and thermal sensors, and then interpreting these data in order to recognize an object or objects. We can divide the object-recognition problem into two categories: the modeling problem and the recognition problem.

### Image Segmentation

Image segmentation is the process of partitioning a digital image into disjoined, meaningful regions. The meaningful regions may represent objects in an image of three-dimensional scene, regions corresponding to industrial,

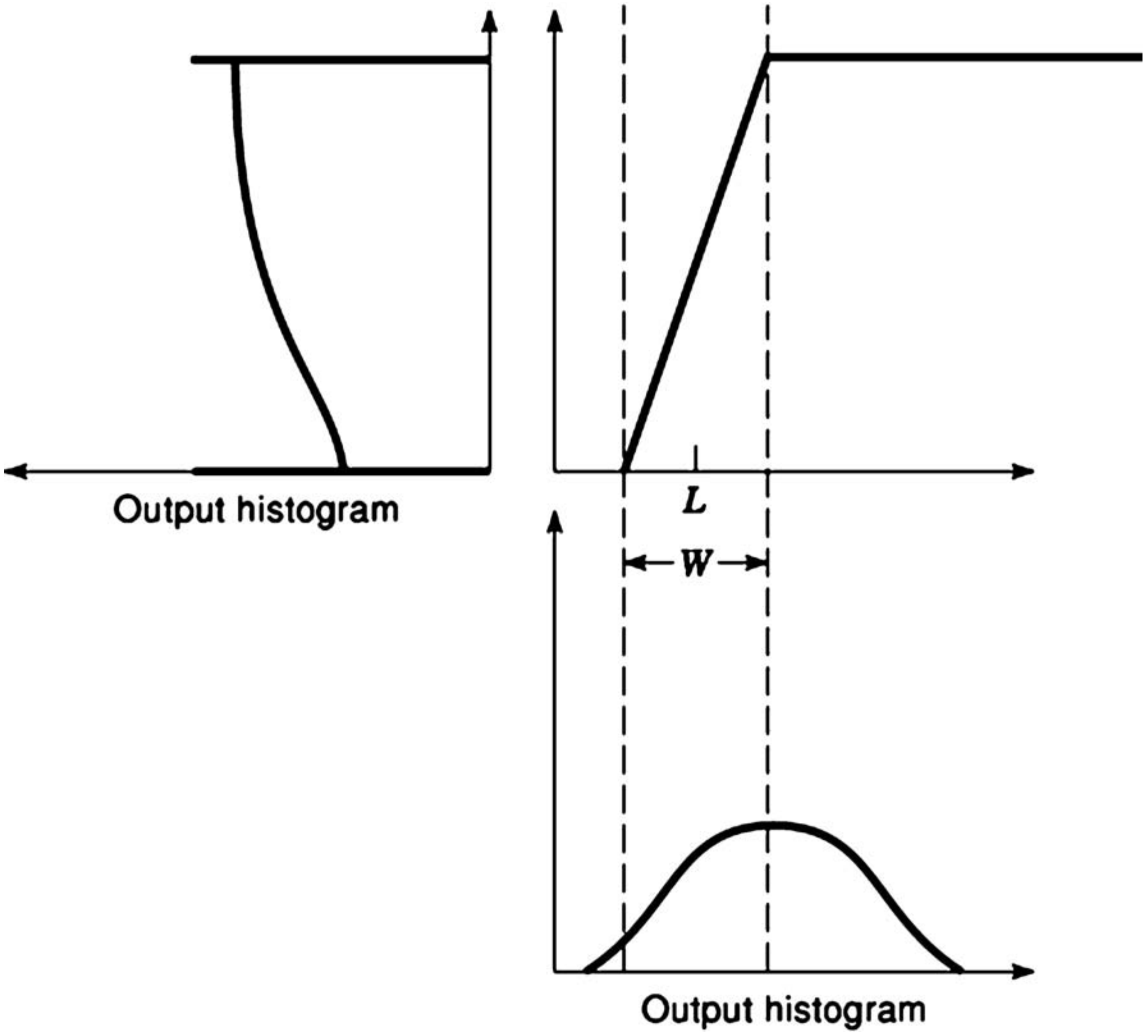


**Figure 9.** Effect of the window-and-level technique on the image histogram. By using this technique, we can select and enlarge any portion of the image histogram we need.

residential, agricultural, or natural terrain in an aerial recognizance application, and so on. A region is a connected set of pixels and the objects are considered either four-connected, if only laterally adjacent pixels are considered, or they can be eight-connected, if diagonally adjacent pixels are also considered to be connected. Image segmentation is an efficient and natural process for humans. A human eye (or rather, mind) sees not a complex scene, but rather a collection of objects. In contrast, image segmentation is not an easy task in digital image processing, and it may become a serious problem if the number of objects is large or unknown or if the boundaries between objects are not clear.

Three of the most commonly used techniques for digital image segmentation are the gray-level thresholding technique, gradient-based segmentation technique, and the region growing technique. A brief description of the three techniques is given below.

**Gray-Level Thresholding.** This technique is an example of the regional approach, which implies grouping pixels into distinct regions or objects. When using this technique for image segmentation, all pixels that have gray levels equal to or above a certain threshold are assigned to the object. The rest of the pixels are assigned to the object's background. If the image contains more than one object, then more threshold levels can be used to do the segmentation.

**Gradient-Based Segmentation.** This technique concentrates on the boundaries between different regions. Reference 5 describes three techniques that are based on the gradient technique. They are the boundary tracking technique, the gradient image thresholding, and Laplacian edge detection. We describe each of these techniques briefly.

***Boundary Tracking.*** This technique starts with scanning the image for the pixel with the highest gradient. This pixel is for sure on the object's boundary. Then a 3-by-3 pixel segment (with the original pixel in its center) is used as tracking probe for the next pixel with the highest gradient, in the neighbor of the original one. This pixel is considered part of the object's boundary and the probe is moved to the new pixel to search for another boundary pixel. The process is repeated until a closed contour is formed. If, at

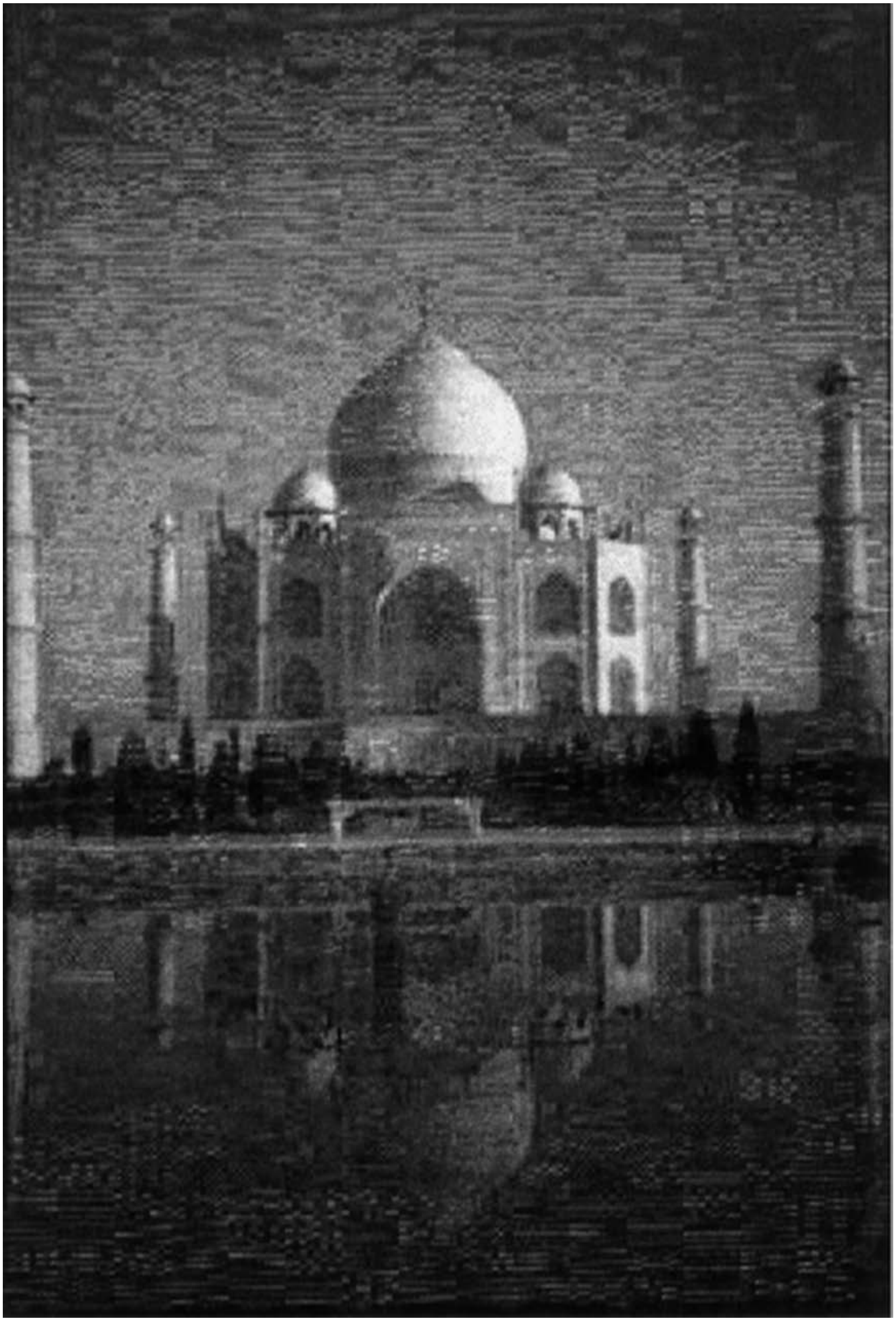

**Figure 10.** Photo of the Taj-Mahal with arbitrary noise.

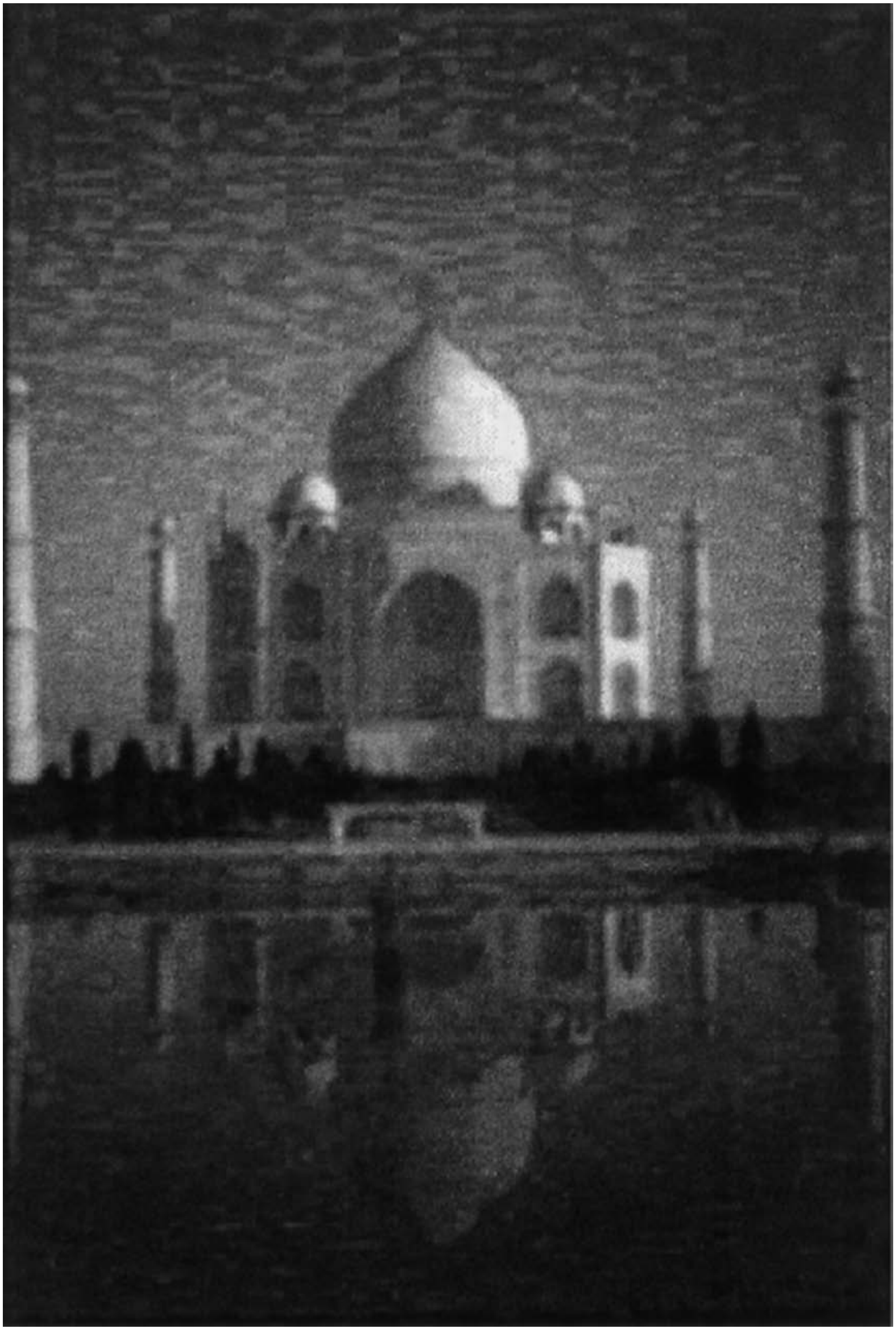

**Figure 11.** Low-pass filtering of the image removes most of the noise but blurs the image.

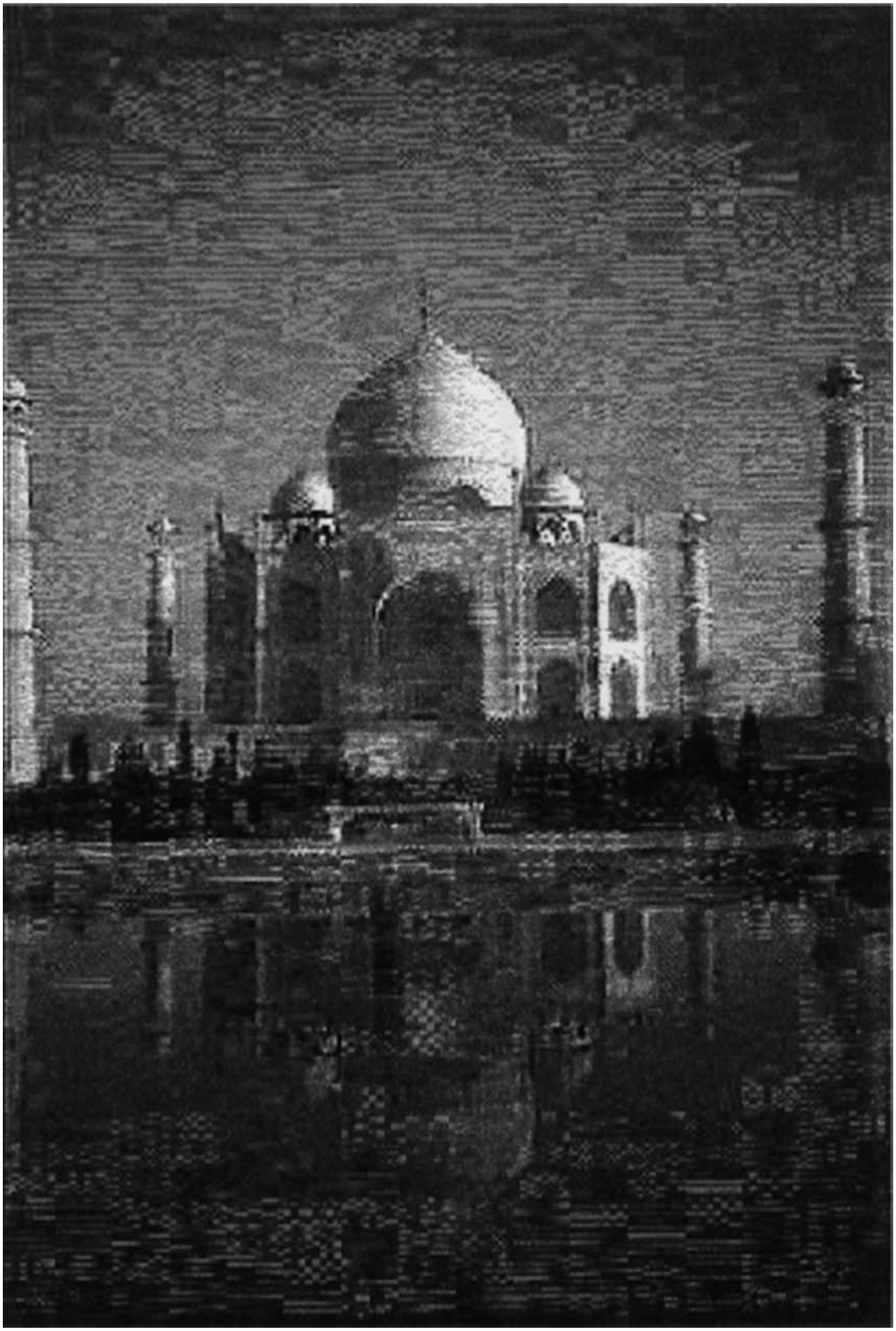

**Figure 12.** High-pass filtering of the image sharpens the edge details but it increases the background noise signal power.

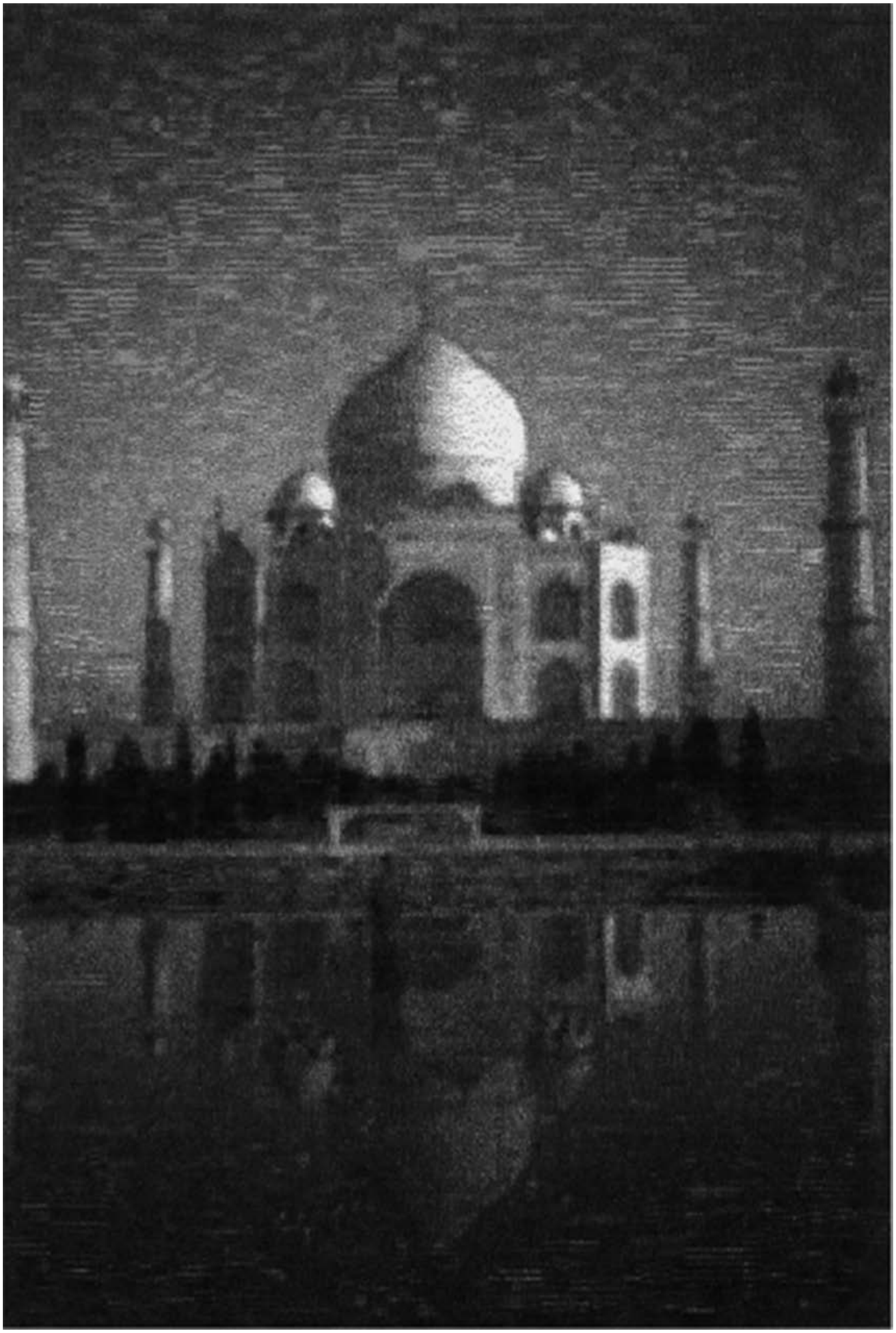

**Figure 13.** Median filtering reduces random noise while enhancing edges.

any time, three adjacent pixels in the probe have the same highest gray level (i.e., gradient), the one in the middle is selected. If two have the same highest gradient, the choice is arbitrary.

***Gradient Image Thresholding.*** This technique is based on the following phenomenon. If we threshold a gradient image at moderate gray level, the object and the background stand below the threshold and most edge points stand above it (5). The technique works like this: We first threshold the image at a low level to identify the object and background, which are separated by bands of edge points. We then gradually increase the threshold level, causing both the object and background to grow. When they finally touch, without merging, the points of contact define the boundary (5).

***Laplacian Edge Detection.*** The Laplacian is a scalar, second-order derivative operator that, for a two-dimensional function $f(x, y)$, can be given by

$$\nabla^2 f(x, y) = \frac{\partial^2}{\partial x^2} f(x, y) + \frac{\partial^2}{\partial y^2} f(x, y)$$

The Laplacian will produce an abrupt zero-crossing at an edge. If a noise-free image has sharp edges, the Laplacian can find them. In the presence of noise, however, low-pass filtering must be used prior to using the Laplacian.

**Region Growing.** Region growing is one of the artificial intelligence techniques for digital image segmentation. Unlike other techniques, this one is capable of utilizing many properties of the pixels inside the different objects, including average gray level, texture, and color information. The technique works as follows: First, the image is partitioned into many small regions. Then the properties of the pixels in each region are computed. In the next step, average properties of adjacent regions are examined. If the difference is significant, the boundary is considered strong and it is allowed to stay. Otherwise, the boundary is considered weak, and the two regions are merged. The process is repeated until there are no more boundaries weak enough to allow regions merging. This technique is especially useful in cases where prior knowledge of the scene features is not available.

### Texture

We can define texture as an attribute representing variations in brightness among the image pixels. This variation may be in the level of brightness, special frequencies, or orientation. Computer-generated images look very realistic for objects of metallic or plastic materials but tend to look smoother than necessary in images of organic materials, such as human skin. Surface roughness is a feature closely related to texture. Texture is sometimes seen as a measure of surface roughness. Random texture is a type of texture caused by noise from cameras, sensors, or films. This type has no recognizable pattern. Satellite images are a good example of images where patterned texture is most recognizable. Different uses of land, such as agriculture or construction, produce different textures. Some of the operators commonly used to detect texture in images are the rank operators, range operators, and gray-level variation. The rank operator measures the range between the maximum and minimum brightness values in the image. Larger range values correspond to surfaces with a larger roughness. The range operator converts the original image into one in which brightness represents texture (4).

Gray-level variation can be measured by the sum of the squares of the differences between brightness of the center pixel and its neighbors. The root of resultant value is then calculated to give the root-mean square (*rms*) measurement of the surface roughness. Research indicates that the human eye is insensitive to texture differences of an order higher than a second (5).

### Object Modeling

Modeling is the process of representing a real system in an abstract manner, in order to study its different features. It is widely used in all fields of engineering. In control engineering, for example, a mathematical model of a physical system is extracted to facilitate the study of its performance under different circumstances.

In object recognition, a model is created for the object under investigation. This model is then compared to different models that are stored in a database. If this model matches one of the available models, say, the model of object $A$, then the investigated model is classified as object $A$. In an airplane recognition system, for example, the database contains models of different types of airplanes.

There are two types of models in object modeling systems:

1. Geometric Models These models are represented by surface or volume primitives with attributes and relationships. The primitives and their relationships can be described geometrically. These models are used for regular manmade objects because they can be described geometrically.
2. Symbolic Models In these models, a three-dimensional object is represented by a graph whose nodes are the 3-D primitives and whose arcs are the attributes and relations. These models are good for representing objects that are irregular in natural.

The primitives of the geometric models depend on the application. In general, there are three classes of primitives:

1. Surface or boundary primitives
2. Sweep primitives
3. Volumetric primitives

The boundary primitives are bounded faces of the object. The sweep primitives are general cylinders or cones. The volumetric primitives can be categorized into three types as follows:

1. Spatial Primitives These primitives are 3-D arrays of cells, which may be marked as filled with matter or not.

2. Solid Geometric Primitives These primitives are regular solid primitives, which may be combined by modified versions of the Boolean set operators—union, difference, and intersection.
3. Cell Primitives These primitives have complex shapes and they can be connected together by just one combining operator (glue).

### Pattern Recognition

Patterning is a process to extract features and recognize objects and patterns from a given image.

Usually there are two approaches to solve the problem of pattern recognition: the RF processing approach and the image-processing approach.

In the RF processing approach, electromagnetic waves are transmitted and then reflected from the target. The reflected waves are analyzed to recognize the target. The RF processing is usually used when the target is very far or out of range of sight.

In the image processing approach, a picture or an image of the field is taken. The picture is then processed and analyzed to recognize the objects in the picture.

The recognition task is usually done in two phases:

1. Training Phase In this phase a pictorial database or a knowledge base is created offline, which contains data about all objects under consideration. If the recognition system is designed to recognize airplanes, for example, the database may include data about all airplanes under consideration.
2. Recognition Phase In this phase, the incoming picture is first processed to enhance it. The picture is analyzed to extract the features that characterize it. Finally the features are matched with the features stored in the database to recognize the object or objects.

## WAVELETS

Wavelets are a relatively new area of signal processing and offer a potentially very useful approach to image analysis. Because of the fact that one has local control over resolution, many cue features of different size can be extracted from an image. Wavelets also show promise as an aid in understanding how early vision can be emulated and further understood.

### Wavelet Transforms

It is well known that in signal processing, the more compact your method of representing information the better. In 1822, Jean Joseph Fourier devised a very efficient way to represent the information content of a signal. His idea was to represent a signal as the sum of its frequencies. Transmitted power spectra, carrier frequencies, brain activity, NMR signals—all these global descriptions provide a lot of information in a compact manner. However, most of the power of this kind of representation vanishes when one tries to represent information that changes its nature during the course of signal recording. A good example of this kind of a signal is a musical score. A global analysis of a recording of a musical selection with a Fourier transform (*FT*) will indicate the specific notes played within the piece of music, but there is no way to recover the timing of the notes. The musical score, on the other hand, indicates each note that was played and the time it was played.

Wavelets represent a signal in a way such that local frequency information is available for each position within the signal. Wavelets are able to analyze a signal based on the position-varying spectra. The multiresolution pyramidal decomposition that results is also well matched to fractals and shows great potential for the removal of background noise, such as static in recordings, and for pattern recognition and texture segmentation. In the following section we discuss the conceptual understanding of how wavelets can be used in signal analysis.

### Wavelet Transform Versus Fourier Transform

A fairly old method of computing a local spectrum is to apply the FT to one specific piece of the signal at a time. This is the idea behind what is called the windowed Fourier transform (*WFT*). Basically, the implementation involves use of a moving rectangular piece of the whole image to isolate a portion of the signal of interest, which is then Fourier transformed. As the window slides along different positions, the WFT gives the spectra at these positions. This kind of analysis has a fundamental problem, however [the mathematics of which is, in fact, similar to the Heisenberg uncertainty principle in quantum mechanics (*QM*)]. Multiplying the signal by the window function results in convoluting or mixing the signal spectrum with the spectrum of the window. Add to this the fact that as the window gets smaller, its spectrum gets wider, and we have the basic dilemma of localized spectra: The better we determine the position of the signal the poorer we localize the spectrum. This is analogous to the case in QM where increased precision in a description of, say, the momentum of an electron reduces the precision available in the position of that electron. Very accurate determinations can be made or computed, but both are not available simultaneously to an unlimited degree of precision. Correspondingly, there is a fundamental physical limit to the degree of accuracy with which the frequency content can be known at a particular position and simultaneously with which the location where the signal is being analyzed can be known.

In 1946, Dennis Gabor introduced a version of the WFT that reduced this uncertainty somewhat. The Gabor transform uses a Gaussian profile for the window, as a Gaussian function minimizes this uncertainty. However, the underlying idea of localizing a spectrum of a signal by windowing the signal needs to be reconsidered. Obviously, caution must be taken in the selection of the signal. Careful attention to the placement of the window, however, is not an easy task for realistic time-varying signals. We are in fact trying to do two different things simultaneously.

Frequency is a measure of cycles per unit time or signal length. High-frequency oscillations take much less signal length or time than do low-frequency oscillations. High frequencies can be well localized in the overall signal with a short window, but low-frequency localization requires a long window.

The wavelet transform takes an approach that permits the window size to scale to the particular frequency components being analyzed. Wavelets are, generally speaking, functions that meet certain requirements. The name *wavelet* originates from the requirement of integrating to zero by oscillating about the $x$ axis and being well localized. In fact, there are many kinds of wavelets. There are smooth wavelets, wavelets with simple mathematical expressions, wavelets that are associated with filters, etc. The simplest wavelet is the Haar wavelet. Just as the sine and cosine functions are used in Fourier analysis as basis functions, wavelets are used as the basis functions for representing other functions. After the "mother" wavelet, $W(x)$, is chosen, it is possible to make a basis out of translations and dilation of the mother wavelet. Special values for $a$ and $b$ are used such that $a = 2^{(-j)}$ and $b = k^{j}$, where $k$ and $j$ are integers. This particular choice of $a$ and $b$ gives a so-called *sparse-basis.* The sparse-basis choice of parameters also makes the link to multiresolution signal analysis possible.

There are many important characteristics of wavelets that make them more flexible than Fourier analysis. Fourier basis functions are localized in frequency but not in time. Small-frequency changes in a FT cause changes everywhere in the time domain. Wavelets can be localized in both frequency position and in scale by the use of dilation, and in time by the use of translations. This ability to perform localization is useful in many applications. Another advantage of wavelets is the high degree of coding efficiency, or in other words, data compression available. Many classes of functions and data can be represented by wavelets in a very compact way.

The efficient coding of information makes wavelets a very good tool for use in data compression. One of the more popular examples of this sparse-coding capability is the well-known fact that the FBI has standardized the use of wavelets in digital fingerprint compression. The FBI gets compression ratios on the order of 20:1. The image difference between the original and the wavelet stored and decompressed image is negligible. Many more applications of wavelets currently exist. One involves the denoising of old analog recordings. For example, Coifman and co-workers at Yale University have been using wavelets to take the static and hiss out of old recordings. They have cleaned up recordings of Brahms playing his First Hungarian Dance for the piano. Noisiness extends to the realm of data sets. Noisy data sets can be cleaned up by using wavelets. Typically the speed of wavelets is also much faster than the fastest fast Fourier transform (*FFT*). The data are basically encoded into the coefficients of the wavelets. The computational complexity of the FFT is of the order of $n \log n$ where $n$ is the number of coefficients; whereas for most wavelets, the order of complexity is of the order $n$. Figure 14 shows how wavelets are used to process data.

Many data operations, such as multiresolution signal processing, can be done by processing the corresponding wavelet coefficients.

How wavelets work is best illustrated by way of an example. The simplest of all wavelets is the Haar wavelet, shown in Fig. 15. The Haar wavelet, $W(x)$, is a step function that assumes the values 1 and −1, on the scale [0,½] and [½,1]. The Haar wavelet is more than 80 years old and has been used for various applications. It can be shown that any continuous function can be approximated by Haar functions. Dilation and translations of the function $\Psi(x)$,

$$\Psi_{jk}(X) = \text{const} \times \Psi(2^j x - k) \tag{19}$$

define an orthogonal basis in the space of all square integrable functions $L^2(R)$. This means that any element of $L^2(R)$ can be presented as a linear combination of these basis functions. The orthogonality of the wavelet pairs is checked by the following:

$$\int \Psi_{jk} \Psi_{j'k'} = \delta_{jj'} \delta_{kk'} \tag{20}$$

whenever $j = j'$ and $k = k'$ is not satisfied simultaneously. The constant that makes the orthogonal basis orthonormal is $2^{j/2}$.

### Wavelet Transforms and Their Use With Clutter Metrics

Wavelets and wavelet transforms are essentially an elegant tool from the field of mathematics that can be applied to the area of signal processing. Wavelets are used for removing noise or unwanted artifacts from images as well as acoustic data.

*Clutter* is a term that refers to the psychophysical task of perceiving objects in a scene while in the presence of similar objects. There are many definitions of clutter in use at the moment, however, none take into account the multiresolution capability of wavelets.

## IMAGE CODING

Image coding refers to groups of techniques that are used to represent (3-D) objects in forms manageable by computers for the purpose of object classification and recognition. Object moments, Fourier descriptors, and expert systems are examples of these techniques.

### The Moment Technique

An image of a 3-D object may be described and represented by means of a set of its moments. The set of moments of an object can be normalized to be independent of the object primary characteristics, namely, translation, rotation, and scale. Hence, moments can be used to recognize 3-D objects.

There are different types of moments that may be used to solve the problem of image coding. These moments are different in their properties and their computation complexities.

**Conventional Moments.** The conventional definition of the two-dimensional moment of order $p + q$ of a function $f(x, y)$ is given by

$$M_{pq} = \sum \sum x^p y^q f(x, y)^r, \qquad p, q = 0, 1, 2, \ldots \tag{21}$$

where the summation is over the area of the image. If the input image is a binary image, then $r = 0$, otherwise $r = 1$. $f(x, y)$ represents the gray-level value of the pixel at point $(x, y)$.

The conventional moment sequence $(M_{pq})$ is uniquely determined by $f(x, y)$ and $(x, y)$ is uniquely determined by

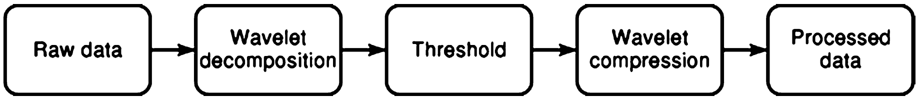


**Figure 14.** The use of wavelets to process raw data.

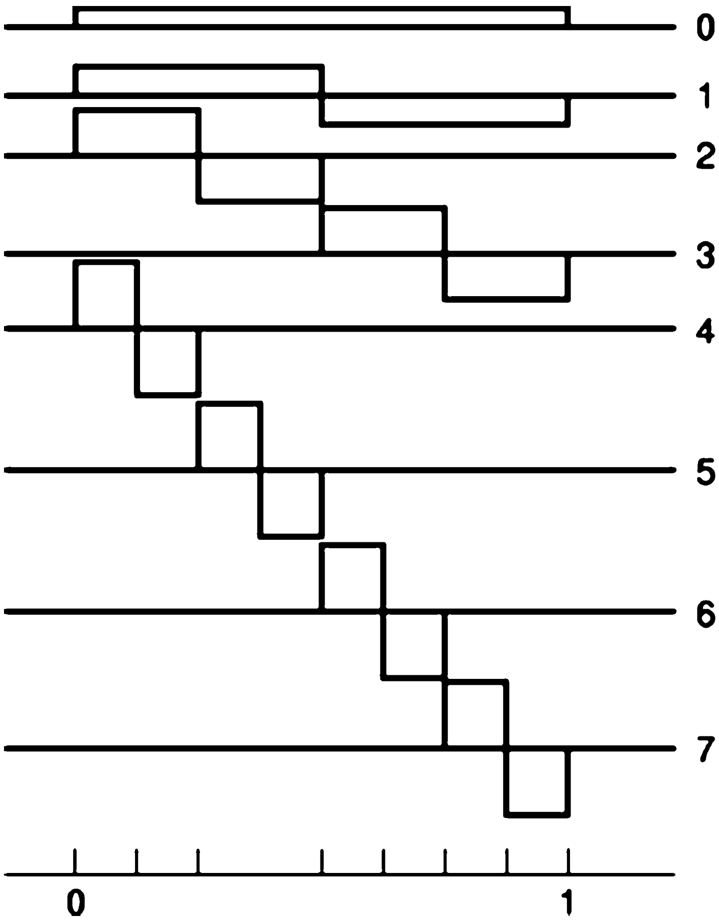


**Figure 15.** The Harr transform basis functions. The basis functions in each stage cover the whole range (from 0 to 1).

the sequence $M_{pq}$. It is always possible to find all orders of moments for an image, but a truncated set may offer a more convenient and economical representation of the essential shape characteristics of an object.

The conventional moments of an object are changed by the primary characteristics of the object in the following manner:

1. Scale Change The conventional moments of $f(x, y)$ after a scale change by a factor $\lambda$ are defined by

$$M^1_{pq} = \sum\sum x^p y^q f(x/\lambda,\ y/\lambda) \tag{22}$$

2. Translation The moments of $f(x, y)$ translated by an amount $(a, b)$ are defined by

$$M^1_{pq} = \sum\sum (x+a)^p (y+b)^q f(x,\ y) \tag{23}$$

3. Rotation The moments of $f(x, y)$ after a rotation by an angle $\theta$ about the origin are defined by

$$M^1_{pq} = \sum\sum [x\cos(\theta) + y\ \sin(\theta)]^p [y\cos(\theta) - x\ \sin(\theta)]^q f(x,\ y) \tag{24}$$

4. Reflection A reflection about the $y$ axis results in a change of sign of all moments, which depend upon an odd power of $y$, that is,

$$M^1_{pq} = \sum\sum x^p y^q f(x,\ -y) \tag{25}$$

For object recognition and image analysis, a truncated set of moments is needed. A complete moment set of order $n$ is always defined to be a set of all moments of order $n$ and lower. A complete moment set of order $n$ contains $(n + 1)(n + 2)/2$ elements.

**Rotational Moments.** The rotational moments for an image $f(x, y)$ are defined by

$$F_n L = \int_{-\pi}^{\pi} \int_0^{\infty} r^n \exp(iL\theta) f(x,\ y) r\, dr\, d\theta \tag{26}$$

where $r\ \sqrt{x^2 + y^2}$
$\theta\ \tan^{-1}(y/x)$

$n$ the order of the moment
$n\ L \geq 0$

A truncated set of rotational moments of order $n$ contains the same information as a truncated set of conventional moments of the same order.

The rotation normalization of these moments is very simple, but the translation and scale normalization is very difficult and time consuming. The standard rotational moments can be obtained by normalizing the rotational moments with respect to scale, translation, and rotation.

**Orthogonal Moments.** There are two types of orthogonal moments.

1. Legendre Moments These moments are based on the conventional moments. The Legendre moment of order $p + q$ is defined by

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_p(x) P_q(y) f(x, y)\, dx\, dy \quad (27)$$

where $P_p(x)$ and $P_q(y)$ are Legendre polynomials. Legendre polynomials are defined by

$$P_n(x) = \sum_{k=0}^{n/2} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-2k)!} x^{n-2k} \quad (28)$$

Advantages are as follows:

a. The inverse transform to obtain $f(x, y)$ from the sequence $(L_{pq})$ is very simple because of the orthogonality property.
b. They are more suitable for optical methods than the conventional moments, since the range needed to record Legendre moments is less than that needed for conventional moments.

Disadvantages are as follows:

a. Much computation is required to normalize the Legendre moments with respect to scale, translation and rotation.

2. Zernike MomentsThese moments are defined based on rotational moments. The Zernike moment of order $n$ is defined by

$$A_{nL} = \frac{n+1}{\pi} \int_{-\pi}^{\pi} \int_{0}^{\infty} [V_{nL}(r, \theta)]^* f(x, y)\, r\, dr\, d\theta \quad (29)$$

where $[V_{nL}]^*$ is the complex conjugate of the Zernike polynomial. The Zernike polynomials are defined by

$$V_{nL}(r, \theta) = \sum_{s=0}^{(n-1)/2} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+L}{2} - s\right)!\left(\frac{n-L}{2} - s\right)!} r^{n-2s} \exp(iL\theta) \quad (30)$$

and $r$ is same as before.

The rotation normalization of Zernike moments is very simple but the translation normalization is very difficult. The inverse transform of $f(x, y)$ from a sequence $\{A_{nL}\}$ is very simple because of orthogonality property.

**Object Classification Using Moments.** A complete moment set can be created by any type of the moments discussed above. However, the conventional moments require minimum computation time for scale and translation normalization. The only operation to be considered is the rotation normalization. Simple parallel hardware architecture may offer a significant savings in time to the rotation normalization process. Hence, moment invariant and standard moments are considered for object classification.

### Fourier Descriptors

The Fourier descriptor ($FD$) method is a well-known method of describing the shape of a closed figure. Depending on the technique used to calculate the FDs, the FDs can be normalized to be independent of the translation, the rotation, and the scale of the figure.

1. The transformation of a binary matrix to a polygonal boundary
2. The calculation of the FDs of the closed curve

**Calculation of Fourier Descriptors.** There are different techniques used to calculate the FDs of a closed curve. The two most common techniques are discussed below.

***Technique Suggested by Cosgriff and Zahn.*** This technique involves many steps to calculate the FDs:

1. A starting point on the boundary is selected.
2. The boundary is traced in the clockwise direction and at each pixel the last and next neighbors are found and noted down.
3. The vertices of the curve are found.
4. The length $(\Delta L)$ between any two successive vertices is calculated. The length between the starting point and a vertex $V_i$ is denoted by $L_i$, and it is equal to $\Sigma^{i}_{k=1} \Delta L_k$.
5. The perimeter is calculated by using the following formula: $L = \Sigma^{m}_{i=1} \Delta L_i$, where $m$ is the number of vertices.
6. The angular change $(\Delta\phi)$ at each vertex is calculated.
7. Next, the following Fourier series coefficients are calculated.

$$a_o = -\pi - \frac{1}{L} \sum_{k=1}^{m} L_k \Delta\Phi_k \quad (31)$$

$$a_n = -\frac{1}{n\pi} \sum_{k=1}^{m} \Delta\Phi_k \sin \frac{2\pi n L_k}{L} \quad (32)$$

$$b_n = \frac{1}{n\pi} \sum_{k=1}^{m} \Delta\Phi_k \cos \frac{2\pi n L_k}{L} \quad (33)$$

where $\Delta\phi_k$ is the angular change at vertex $k$, $L_k$ is the arc length to vertex $k$, $L$ is the perimeter, and $m$ is the number of vertices.

8. Lastly, the polar coordinates are calculated as follows:

$$A_n = \sqrt{a_n^2 + b_n^2}$$
$$\alpha_n = \tanh \frac{b_n}{a_n} \tag{34}$$

The $A_n$'s and the $\alpha_n$'s are the Fourier descriptors. It is usually sufficient to represent a closed curve by the first 10 harmonics. The calculated FDs are put in a vector. This vector is called the feature vector.

The FDs calculated by this technique are independent of rotation, translation and scale.

***Technique Used by Wallace and Wintz and by Kuhl et al.*** This technique retains almost all shape information and is computationally efficient because it uses the FFT. The steps involved in this technique are explained below as follows.

1. A starting point on the boundary is selected.
2. The original contour is sampled and then replaced by a piecewise linear contour.
3. The piecewise linear contour is represented by a chain code representation.
4. The length of the contour is computed.
5. The contour is uniformly resampled at spacing chosen to make the total number of samples a power of 2.
6. The FDs are computed by simply taking the FFT of the resultant samples.
7. The FDs should be normalized such that the contour has a standard size, rotation, translation, and starting point.

First, we set $A(O)$ equal to zero to normalize position. Second, scale normalization is accomplished by dividing each coefficient by the magnitude of $A(1)$. Third, rotation and starting point normalization are accomplished by making the phases of the two coefficients of largest magnitude to be zero.

The normalized FDs (NFDs) are set in a vector. This vector is called the feature vector and it is used to recognize the object.

To recognize an object, there should be a training set. The training set should contain all the possible feature vectors for all possible types of objects. The recognition of an unknown feature vector can be achieved by using one of the following two techniques:

1. Distance-weighted $k$ nearest neighbor rule
2. $\chi^2$ technique

Because FDs are just independent of one orientation angle ($\theta_y$), the training set should contain feature vectors for any given angles $\theta_x$ and $\theta_z$.

## Expert Systems

Expert systems have been successfully used in different fields of engineering as a substitution of the traditional methods. An *expert system* is a system designed to accomplish a certain goal by following the appropriate set of rules that is a part of a larger, more general group of rules, given by experts in the specific field.

Expert systems consist of two units, the knowledge-base management unit and the knowledge-base unit. The knowledge-base management interfaces between the knowledge base and the outside world and hence facilitates to enter the basic knowledge and to construct this knowledge for efficient retrieval and decision making.

**Object Representation.** The knowledge base has two segments. One segment contains the specific facts, that define the problem and the other segment contains the rules that operate on the facts during the problem-solving process.

The knowledge in the knowledge base should be structured in a very powerful way to facilitate the searching process and to reduce the searching time. Three knowledge structures are common:

1. Relational Knowledge-Base Structure This takes a form of a table.
2. Hierarchical Knowledge-Base Structure or Tree Structure Each parent has one or more descendant, but each descendant has only one parent.
3. Network Structure Each parent has one or more descendant and each descendant can have one or more parent.

It is easy to update the knowledge in a relational structure, but searching takes a long time. The hierarchical structure reduces the searching time. The network structure allows a many-to-many relation.

In general each knowledge base has the following features:

1. A fact base that is equal to the conventional database.
2. A rule base that consists of production rules that operate on the fact base.
3. Confidence factors, which indicate the degree of confidence that can be placed in facts and rules.
4. Metarules, which determine when and how the rules in the rule base are to be executed.

The knowledge-base management unit accepts the extracted features of an object and then represents them in a way that is appropriate for symbolic processing. There are several ways to represent these features: first-order predicate calculus, frames, and semantic networks.

In the problem of recognizing a tank, for example, first-order predicate represents the fact that a gun is located over a small rectangle in the form (6)

```
OVER (GUN, SMALL-RECTANGLE)
```

A frame for the same fact might be in this form

Name of frame

Type of frame
Object 1
Object 2

F1
OVER
GUN
SMALL-RECTANGLE

In the semantic network, each node contains an object and the lines between the nodes represent the relationship.

The knowledge base is unique to this particular problem. This knowledge base contains rules that show the relationship among several pieces of information. The most common rule is the production rule, such as

```
IF THERE IS A GUN, AND THE GUN IS OVER A RECTANGLE, AND
THE RECTANGLE IS OVER WHEELS, THEN THERE IS A TANK.
```

Here we have three antecedents connected by logic AND, which when satisfied lead to the consequence that there is a tank (6). When all the three antecedents are present as assertions, they match the left-hand side of the production rule, leading to the consequence there is a tank.

In addition, there are two further steps to the making of matches:

1. The production rules may be listed in some order to facilitate the matching process.
2. The search strategy unit may be able to invoke a sequence of rules that finds the match more quickly than random search lists.

The search strategy unit searches the knowledge base for problem solution or object identification. A knowledge-base search is the foundation of artificial-intelligence (*AI*) systems. The searching process depends on the knowledge structure or the arrangement of the production rules in the knowledge base. The simplest arrangement of the production rules is to list them sequentially with no order. Each asserted fact is then run through the production rules until the match is found. With a small number of rules, it is practical to search a random list of rules. In the 3-D-object recognition problem, there are many rules and hence the knowledge base is large. For this type of problem, the knowledge structure must be designed properly to decrease the searching time.

### Searching Algorithms

As stated before, the searching algorithm depends on the knowledge structure. Taking the best knowledge structure, the following searching algorithms may be applied.

1. Depth-First Search This technique takes one branch of the knowledge structure tree and searches all the nodes of that branch to its last node. If a solution is not found, the procedure backs up to the next unsearched branch and continues the search to the last node and so on until a solution is found.
2. Breadth-First Search This technique searches according to the level of a node. Level one will be the top node. The successors of this node will be labeled level two. This technique starts by searching all of the nodes on level two, then level three, and so on, until a solution is found.
3. Heuristic Search The two prior searching techniques require searching through the entire knowledge base in order to find the solution. When the knowledge base is large, the time for solution search can become a real problem. The heuristic search technique is devised to reduce the searching time.

*Heuristics* is defined as any information that would bring closer to a problem solution. Such information often takes the form of rules of thumb that may be applied in decision-making processes.

Heuristic information can be applied at many nodes along the solution search path, and the decision can then be made as to which branch of the knowledge structure is more likely to hold the last solution.

Heuristics are generally applied in one of the following two ways:

1. When deciding whether a particular branch of the knowledge structure should be eliminated
2. When deciding which branch of the knowledge structure should be searched next

The inference unit is used to obtain the object identity. The inference or reasoning process accepts the assertions from the input image and searches the knowledge base for identification. Two approaches to inference are common: deduction inference and induction inference. Deduction or backward inference begins with an hypothesis or a goal and then proceeds to examine the production rules, which conform to the hypothesis. Induction or forward inference examines the assertions and then draws a conclusion, to which the conditions conform. In other words, deduction starts with the goals, while induction starts with the facts or the assertions.

Different 3-D-object recognition systems that utilize AI techniques has been built and tested. The performance of those systems is very optimistic.

## NEURAL NETWORKS

A neural network consists of a number of nonlinear computational elements called neurons or nodes. The neurons are interconnected with adaptive elements known as weights and operate in parallel environment. The structure of neural networks is similar to simplified biological systems. Recently, neural networks have become a prime area of research because they possess the following characteristics:

- Highly parallel structure; hence a capability for fast computing
- Ability to learn and adapt to changing system parameters (e.g., inputs)
- High degree of tolerance to damage in the connections

- Ability to learn through parallel and distributed processing
- Nonparametric values, not dependent on as many assumptions regarding underlying distributions as traditional networks (e.g., classifier or optimizing circuits)

Neural networks have been used as associative content addressable memories (*ACAM*) and as classifiers. In an ACAM, the memory is associated not by an address, but rather by partially specified or many versions of the stored memory pattern.

A classifier assigns an input sample to one of the predetermined classes. The use of neural networks has recently been proposed for creating dynamic associative memory (*DAM*), which utilizes supervised learning algorithms, by recording or learning to store the information as a stable memory state.

The neural network minimizes the energy function formed by the mean square error constructed by the difference between the actual training signal and the signal estimated by the network.

### Back Propagation

Back propagation is one of the most popular supervised training methods for neural networks. It is based on the gradient-descent technique for minimizing the square of error between the desired output and the actual output. The following is a summary of the back-propagation training algorithm for a two-layer neural network with $n$ inputs, $l$ outputs, and $j$ neurons in the hidden layer.

1. Store the training set $\{(X^k, d^k): k = 1, 2, 3, \ldots, n\}$, where $X^k$ is the input vector and $d^k$ is the desired output vector.
2. Initiate $V$ and $W$ to small random numbers (between $-1$ and 1), where $V$ is the weight matrix for the hidden layer ($j \times n$) and $W$ is the weight matrix for the output layer ($l \times j$).
3. For each pair $(X, d)$, compute the forward pass as follows:
   - Compute $\text{net}_j$, $j = 1, 2, \ldots, n$, where
     $$\text{net}_j = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{V} \quad (35)$$
   - Compute $Z_j$, where, for a unipolar sigmoid activation function, $Z_j$ can be given by
     $$Z_j = \frac{1}{1 + e^{-\lambda \text{net}_j}} \quad (36)$$
   - Compute $\text{net}_l$, $l = 1, 2, \ldots, L$, where
     $$\text{net}_l = z^{\mathrm{T}}W \quad (37)$$
   - Compute $y_l$, where for the same unipolar activation function, $y_l$ is given by
     $$y_l = \frac{1}{1 + e^{-\lambda \text{net}_l}} \quad (38)$$
4. Compute the backward pass as follows:
   - $\delta_{ol} = (d_l - y_l) f_0'(\text{net}_l),\ l = 1, 2, \ldots, L$
   - Compute
     $$\Delta W_{lj} = \eta \delta_{ol} Z_j \quad (39)$$
     where $\eta$ is the iteration step
   - $$\Delta V_{ji} = \eta f'(\text{net}_j) X_j \sum_{l=1}^{L} \delta_{ol} W_{lj} \quad (40)$$
5. Update weights as follows:
   - $$W_{lj}^{k+1} = W_{lj}^{k} + \Delta W_{lj} \quad (41)$$
   - $$V_{ji}^{k+1} = V_{ji}^{k} + \Delta V_{ji} \quad (42)$$
6. Check convergence by comparing the resultant output with the desired ones. If not converged, go to step 3.

Figure 16 shows a schematic for a two-layer neural network.

### Neural Networks for Image Compression

Recently, neural networks have been used for image compression. Artificial neural network models called connectionist models or parallel distributed processing models have received much attention in many fields where high computation ratios are required. Many neural network approaches for image compression yield performance superior to that of the discrete traditional approaches. The estimated output is compared with the actual one for learning. For back-propagation algorithms, image data compression is presented as an encoder problem. In fact, the weight matrices encode the principal components of the given image, that is, after convergence it decomposes the given image into an eigenvalue of decreasing variance. Back propagation is so powerful for image compression that singular value decomposition techniques will be quite appropriate. Figure 17 shows one possible application for neural network in image transmission. Digital image generated with an 8 bit imagery system may be reduced in size by feeding it to an 8-to-4 multilayer neural network. On the other end of the line, the compressed image is processed by a 4-to-8 neural network for reconstruction. Compressing images before transmitting them speeds up the transmission time substantially.

### Neural Networks for Pattern Recognition

The neural network is an ideal tool for pattern recognition. Any recognition system needs to be trained to recognize different patterns, and training is the most important part in the design of neural network systems. All pattern-recognition systems try to imitate the recognition mechanism carried out by humans. Since the neural network is a simplification of human neural system, it is more likely to adapt to the human way of solving the recognition problem than other techniques and systems. Finally, we can look at pattern recognition as a classification problem, which is best handled by neural network systems.

The design of the neural network system for pattern recognition starts with collecting data on each of the objects that is to be recognized by the system. A class is assigned to each object, and the collection of data and classes is used to train the system.

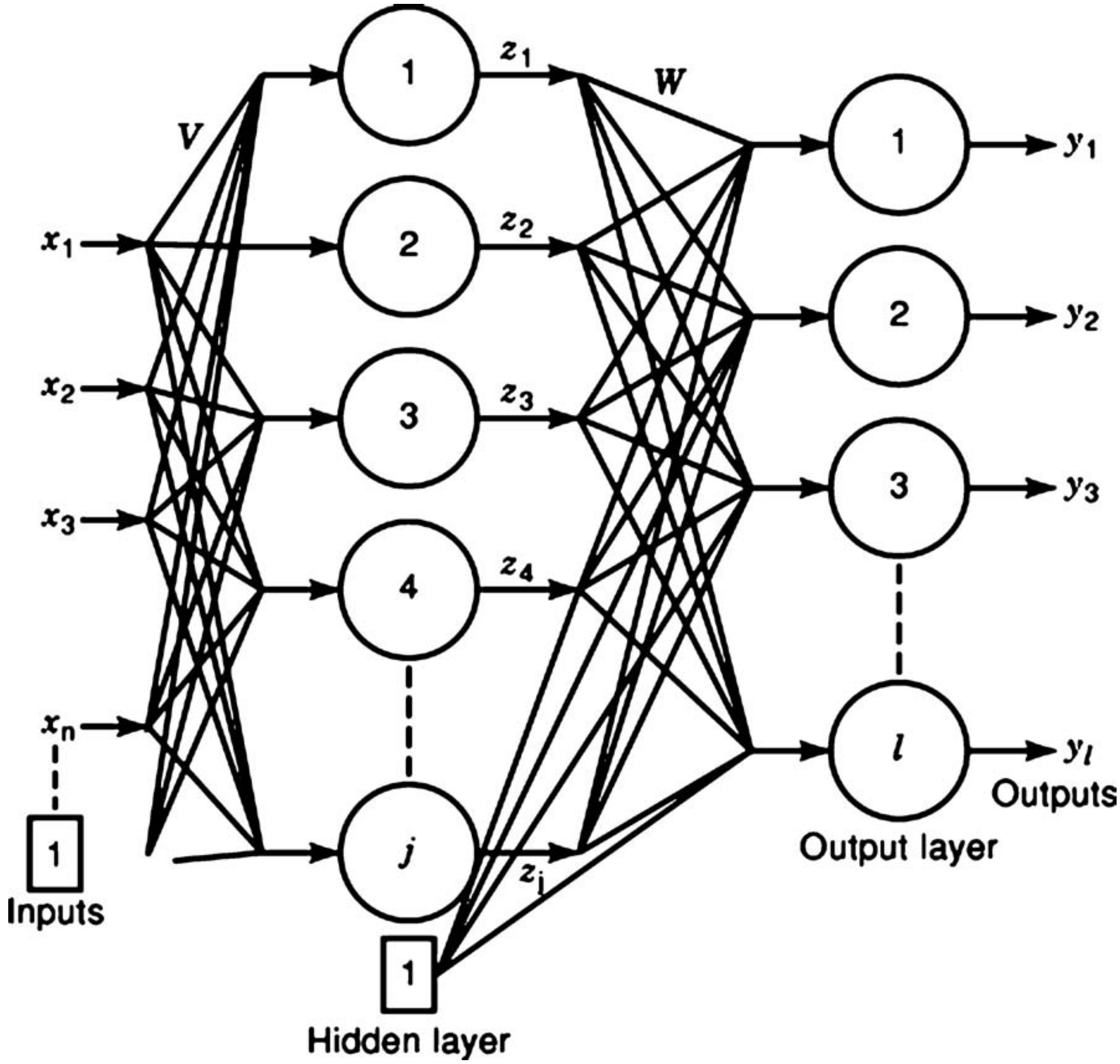


**Figure 16.** Two-layer neural network with $n$ inputs, $j$ hidden neurons, and $l$ outputs.

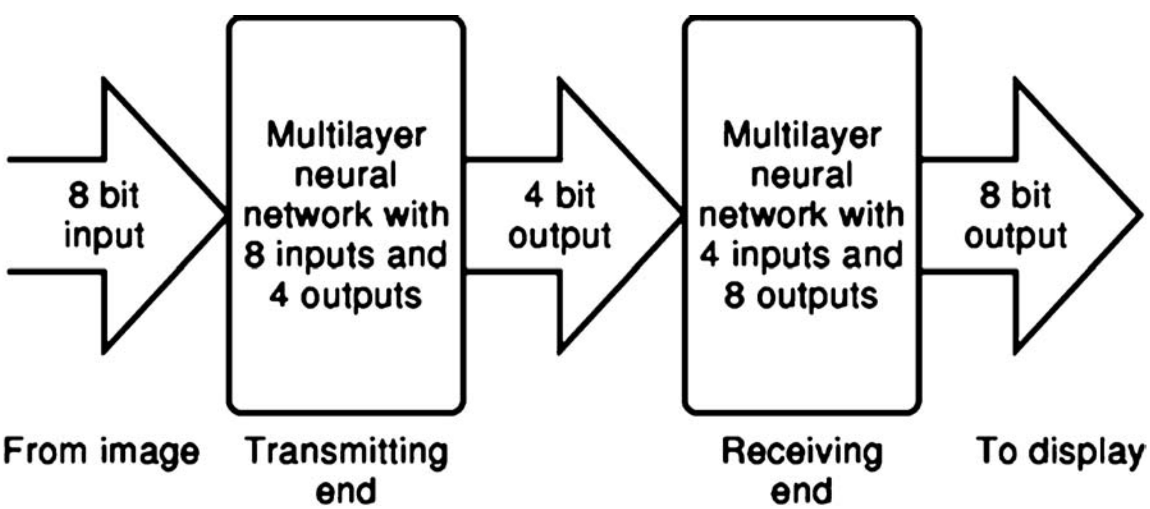


**Figure 17.** The use of neural networks for image compression and transmission.

## IMAGE FUSION

Image fusion is one of the forms of multi-sensor fusion. Image fusion is comparatively a new area and is attracting a lot of attention in areas related to: U.S. Army research and development of situational awareness systems, target detection, automotive, medical imaging and remote sensing. The current definition of sensor fusion is very broad and the fusion can take place at the signal, pixel, feature, and symbol level. The goal of image fusion is to create new images that are more suitable for the purposes of human visual perception, object detection and target recognition. A prerequisite for successful image fusion (pixel level) is that multi-sensor images have to be correctly aligned on a pixel-by-pixel basis.

Image fusion has the following advantages:

1. It improves the reliability by eliminating redundant information.
2. It improves the contrast by retaining complementary information. See Fig. 18.

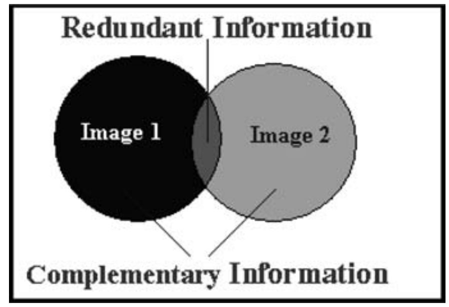


**Figure 18.** Fusion Process on a Venn Diagram.

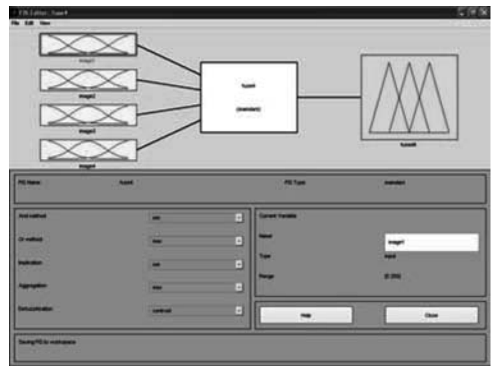

**Figure 19.** Image Fusion system for four input images.

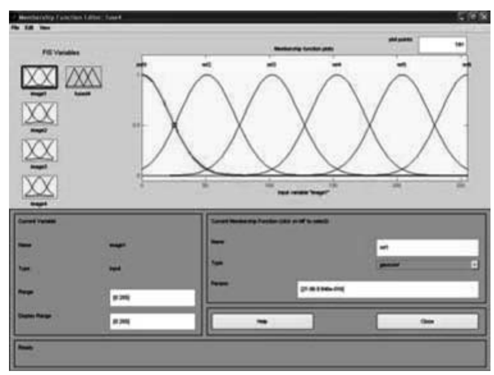

**Figure 20.** Six Gaussian membership function for Input/Output Images using fuzzy logic.

Image fusion is a sequel to data fusion. A number of approaches are currently being discussed and used for image fusion like Pyramidal transforms, Wavelet transforms, different mean value, Max/Min approach, and fuzzy/neuro-fuzzy logic.

Multi-sensor data often contains complementary information about the region interrogated so image fusion provides an effective method to enable comparison and analysis of such data.

Image fusion is being used in numerous medical applications to obtain a better image and is being tested in automotive industries to enhance the vision of road so as to see a better image during a rainy or a foggy weather for collision avoidance applications. The fuzzy/neuro-fuzzy image fusion technique is described below.

## FUZZY LOGIC AND NEURO-FUZZY APPROACH TO IMAGE FUSION

Fuzzy logic and neuro-fuzzy approach are an alternative to a large number of conventional approaches, which are based on a host of empirical relations. Empirical approaches are time consuming and result in a low correlation. The fuzzy logic approach is based on simple rules, which are easy to apply and take less time. As shown in Fig. 19, multiple images can be fused using this method. For image fusion for a specific application, the user selects optimized membership functions and the number of membership functions associated with each image, as shown in Fig. 20. Fuzzy rules are then applied at every pixel level to the corresponding pixels of all the input images.

The neural network and fuzzy Logic approach can be used together for sensor fusion. Such a sensor fusion could belong to a class of sensor fusion in which the image or sensor features are input and a decision could be the output. All the parameters of neuro-fuzzy system such as membership function number, type and number of iterations can be selected according to the requirements of processing speed and quality of the output image. The basic concept is to associate the given sensory inputs with some decision outputs. After developing the system, another group of input image data is used to evaluate the performance of the system. In Fig. 21, we can see the output of fusion of the two medical images using both fuzzy and neuro-fuzzy algorithm. Applications of fused medical images are: image-guided surgery, non-invasive diagnosis and treatment planning. Images from the same modality may be combined from different times, or images from different modalities may be combined to provide more information to the surgeon. For example, intensity-based tissue classification and anatomical region demarcation may be seen using MRI, while metabolic activity or neural circuit activity from the same region can be measured with PET or functional MRI. This type of information is important for surgical planning so as to avoid regions of the brain not

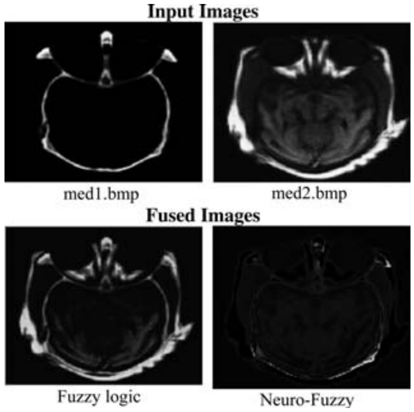


**Figure 21.** Fused medical image output using fuzzy and neuro-fuzzy logic.

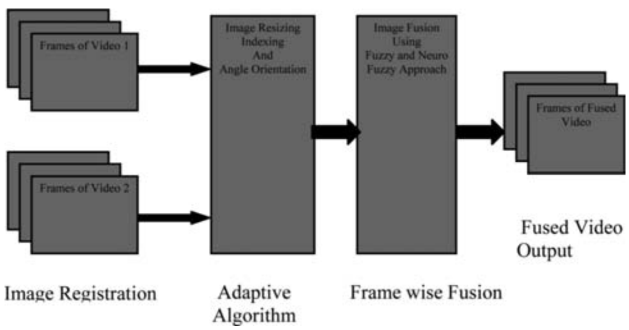


**Figure 22.** Steps of Video fusion using fuzzy and neuro-fuzzy logic.

relevant to the upcoming surgery.

### Video image fusion Process

Video image fusion is implemented same way using frame-wise image fusion with any of the image fusion techniques after the reduction of video into frames. Fuzzy logic is unique way of sensor fusion with less internal complexity. In the dynamic environment, the scene changes continuously and there is a need to focus on the correct frame alignment and preprocessing of the frames more than the algorithms for the fusion process. Video image fusion using fuzzy logic can provide useful information for variety of applications. The major steps for implementing video image fusion using the fuzzy or neuro-fuzzy methods are outlined below.

1. Image Registration- All image fusion algorithms require registered input images so pixel combinations are valid. For two video frames to be fused, it is important that they be properly aligned with each other in both spatial and temporal domain. The degree of registration that can be achieved prior to fusion is critical to the quality of the fused image that is produced. If two images are not registered, no image fusion can rectify the fact that invalid pixel information is being compared and possibly combined incorrectly, producing poor image quality at best. Accurate image registration depends on factors such as the fusion application, speed of movement in the scene, camera frame rates etc.
2. Adaptive Preprocessing Algorithm- Before applying the video fusion process, images require checking for suitability of fusion. The quality of the input image is foremost for ensuring the quality of the fused output irrespective of the fusion algorithm adopted. Using a set of scene matrices to monitor input scene quality for local situational awareness, we can adapt the frames to be sent for the fusion process. Adaptive algorithms are currently being developed in order to prevent skewed images, different zoom level images or distorted frames.
3. Video fusion using fuzzy/neuro-fuzzy logic- Pixel wise fusion with the Gaussian membership function type and with optimum membership functions is accomplished based on fuzzy trained system. In the case of neuro-fuzzy logic, a trained neural network is used for fusing the corresponding frames. For the video image fusion, the frames from the video are obtained, which are fused using the same approach of fuzzy fusion technique. All the fused images are converted to frames and assembled for the final video display.

## MAJOR PROBLEMS IN VIDEO IMAGE FUSION

Real time implementation of image fusion tends to magnify many of the inherent problems of the image fusion process. While in the image fusion process, adjustment for any changes in the image can be done on an image by image basis, video fusion process requires real time robust and adaptive algorithm to deal with the dynamic scene problems over all the images in the video set.

In a dynamic environment, even with the most rigid arrangement of the cameras, spatial misalignment remains one of the foremost problems in video fusion. Many new solutions are proposed such as internal rotation of the frame along the axes based on misalignment nature of the frames, shade and contrast based adjustments. The same techniques used for video stabilization of the image using fuzzy logic must be applied over the two image streams being fused. Adaptive frame alignment algorithms are a subject of current research. Real-time image processing requires high computational processing speed. The techniques of parallel processing and cluster computing are possible for image and video processing on a static platform. For processing under dynamic environments, image fusion algorithms need to be modified so that low onboard memory storage is needed and the processing of fused video can be viewed in real time.

## BIBLIOGRAPHY

1. A. K. Jain *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
2. R. C. Gonzalez P. Wintz *Digital Image Processing*, Reading, MA: Addison-Wesley, 1977.
3. W. K. Pratt *Digital Image Processing*, 2nd ed., New York: Wiley-Interscience, 1991.
4. J. C. Russ *The Image Processing Handbook*, 2nd ed., Boca Raton, FL: CRC Press, 1995.

5. K. R. Castleman *Digital Image Processing*, Upper Saddle River, NJ: Prentice-Hall, 1996.
6. N. Magnenat-Thalmann D. Thalmann *Image Synthesis, Theory and Practice*, New York: Springer-Verlag, 1987.

HARPREET SINGH
Y. HAMZEH
S. BHAMA
S. TALAHMEH
L. ANNEBERG
G. GERHART
T. MEITZLER
D. KAUR
Wayne State University, Detroit, MI
JTICS, Chrysler Corporation, Detroit, MI
INVANTAGE, Inc., Taylor, MI
Lawrence Technological University, Southfield, MI
Research and Engineering Center (TARDEC), Warren, MI
University of Toledo, Toledo, OH