# 19

IMAGE DETECTION AND REGISTRATION

This chapter covers two related image analysis tasks: detection and registration. Image detection is concerned with the determination of the presence or absence of objects suspected of being in an image. Image registration involves the spatial alignment of a pair of views of a scene.

19.1. TEMPLATE MATCHING

One of the most fundamental means of object detection within an image field is by *template matching*, in which a replica of an object of interest is compared to all unknown objects in the image field (1–4). If the template match between an unknown object and the template is sufficiently close, the unknown object is labeled as the template object.

As a simple example of the template-matching process, consider the set of binary black line figures against a white background as shown in Figure 19.1-1a. In this example, the objective is to detect the presence and location of right triangles in the image field. Figure 19.1-1b contains a simple template for localization of right triangles that possesses unit value in the triangular region and zero elsewhere. The width of the legs of the triangle template is chosen as a compromise between localization accuracy and size invariance of the template. In operation, the template is sequentially scanned over the image field and the common region between the template and image field is compared for similarity.

A template match is rarely ever exact because of image noise, spatial and amplitude quantization effects, and a priori uncertainty as to the exact shape and structure of an object to be detected. Consequently, a common procedure is to produce a difference measure $D(m, n)$ between the template and the image field at all points of
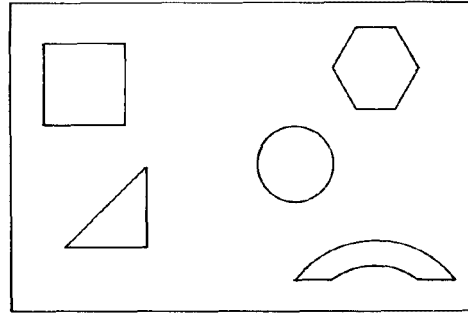
(a) Array of objects

(b) Triangle template

FIGURE 19.1-1. Template-matching example.

the image field where $-M \leq m \leq M$ and $-N \leq n \leq N$ denote the trial offset. An object is deemed to be matched wherever the difference is smaller than some established level $L_D(m, n)$. Normally, the threshold level is constant over the image field. The usual difference measure is the mean-square difference or error as defined by

$$D(m, n) = \sum_{j} \sum_{k} [F(j, k) - T(j - m, k - n)]^2 \qquad (19.1\text{-}1)$$

where $F(j, k)$ denotes the image field to be searched and $T(j, k)$ is the template. The search, of course, is restricted to the overlap region between the translated template and the image field. A template match is then said to exist at coordinate (m, n) if

$$D(m, n) < L_D(m, n) \qquad (19.1\text{-}2)$$

Now, let Eq. 19.1-1 be expanded to yield

$$D(m, n) = D_1(m, n) - 2D_2(m, n) + D_3(m, n) \qquad (19.1\text{-}3)$$

where

$$D_1(m, n) = \sum_j \sum_k [F(j, k)]^2 \tag{19.1-4a}$$

$$D_2(m, n) = \sum_j \sum_k [F(j, k)T(j - m, k - n)] \tag{19.1-4b}$$

$$D_3(m, n) = \sum_j \sum_k [T(j - m, k - n)]^2 \tag{19.1-4c}$$

The term $D_3(m, n)$ represents a summation of the template energy. It is constant valued and independent of the coordinate (m, n). The image energy over the window area represented by the first term $D_1(m, n)$ generally varies rather slowly over the image field. The second term should be recognized as the cross correlation $R_{FT}(m, n)$ between the image field and the template. At the coordinate location of a template match, the cross correlation should become large to yield a small difference. However, the magnitude of the cross correlation is not always an adequate measure of the template difference because the image energy term $D_1(m, n)$ is position variant. For example, the cross correlation can become large, even under a condition of template mismatch, if the image amplitude over the template region is high about a particular coordinate (m, n). This difficulty can be avoided by comparison of the normalized cross correlation

$$\tilde{R}_{FT}(m, n) = \frac{D_2(m, n)}{D_1(m, n)} = \frac{\sum_j \sum_k [F(j, k)T(j - m, k - n)]}{\sum_j \sum_k [F(j, k)]^2} \tag{19.1-5}$$

to a threshold level $L_R(m, n)$. A template match is said to exist if

$$\tilde{R}_{FT}(m, n) > L_R(m, n) \tag{19.1-6}$$

The normalized cross correlation has a maximum value of unity that occurs if and only if the image function under the template exactly matches the template.

One of the major limitations of template matching is that an enormous number of templates must often be test matched against an image field to account for changes in rotation and magnification of template objects. For this reason, template matching is usually limited to smaller local features, which are more invariant to size and shape variations of an object. Such features, for example, include edges joined in a Y or T arrangement.

19.2. MATCHED FILTERING OF CONTINUOUS IMAGES

Matched filtering, implemented by electrical circuits, is widely used in one-dimensional signal detection applications such as radar and digital communication (5–7). It is also possible to detect objects within images by a two-dimensional version of the matched filter (8–12).

In the context of image processing, the *matched filter* is a spatial filter that provides an output measure of the spatial correlation between an input image and a reference image. This correlation measure may then be utilized, for example, to determine the presence or absence of a given input image, or to assist in the spatial registration of two images. This section considers matched filtering of deterministic and stochastic images.

19.2.1. Matched Filtering of Deterministic Continuous Images

As an introduction to the concept of the matched filter, consider the problem of detecting the presence or absence of a known continuous, deterministic signal or reference image $F(x, y)$ in an unknown or input image $F_U(x, y)$ corrupted by additive stationary noise $N(x, y)$ independent of $F(x, y)$. Thus, $F_U(x, y)$ is composed of the signal image plus noise,

$$F_U(x, y) = F(x, y) + N(x, y) \tag{19.2-1a}$$

or noise alone,

$$F_U(x, y) = N(x, y) \tag{19.2-1b}$$

The unknown image is spatially filtered by a matched filter with impulse response $H(x, y)$ and transfer function $\mathcal{H}(\omega_x, \omega_y)$ to produce an output

$$F_O(x, y) = F_U(x, y) \circledast H(x, y) \tag{19.2-2}$$

The matched filter is designed so that the ratio of the signal image energy to the noise field energy at some point (ε, η) in the filter output plane is maximized.

The instantaneous signal image energy at point (ε, η) of the filter output in the absence of noise is given by

$$|S(\varepsilon, \eta)|^2 = |F(x, y) \circledast H(x, y)|^2 \tag{19.2-3}$$

with $x = \varepsilon$ and $y = \eta$. By the convolution theorem,

$$|S(\varepsilon, \eta)|^2 = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{F}(\omega_x, \omega_y)\mathcal{H}(\omega_x, \omega_y)\exp\{i(\omega_x\varepsilon + \omega_y\eta)\}\, d\omega_x\, d\omega_y \right|^2 \quad (19.2\text{-}4)$$

where $\mathcal{F}(\omega_x, \omega_y)$ is the Fourier transform of $F(x, y)$. The additive input noise component $N(x, y)$ is assumed to be stationary, independent of the signal image, and described by its noise power-spectral density $\mathcal{W}_N(\omega_x, \omega_y)$. From Eq. 1.4-27, the total noise power at the filter output is

$$N = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}_N(\omega_x, \omega_y)|\mathcal{H}(\omega_x, \omega_y)|^2\, d\omega_x\, d\omega_y \quad (19.2\text{-}5)$$

Then, forming the signal-to-noise ratio, one obtains

$$\frac{|S(\varepsilon, \eta)|^2}{N} = \frac{\left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{F}(\omega_x, \omega_y)\mathcal{H}(\omega_x, \omega_y)\exp\{i(\omega_x\varepsilon + \omega_y\eta)\}\, d\omega_x\, d\omega_y \right|^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}_N(\omega_x, \omega_y)|\mathcal{H}(\omega_x, \omega_y)|^2\, d\omega_x\, d\omega_y} \quad (19.2\text{-}6)$$

This ratio is found to be maximized when the filter transfer function is of the form (5,8)

$$\mathcal{H}(\omega_x, \omega_y) = \frac{\mathcal{F}^*(\omega_x, \omega_y)\exp\{-i(\omega_x\varepsilon + \omega_y\eta)\}}{\mathcal{W}_N(\omega_x, \omega_y)} \quad (19.2\text{-}7)$$

If the input noise power-spectral density is white with a flat spectrum, $\mathcal{W}_N(\omega_x, \omega_y) = n_w/2$, the matched filter transfer function reduces to

$$\mathcal{H}(\omega_x, \omega_y) = \frac{2}{n_w}\mathcal{F}^*(\omega_x, \omega_y)\exp\{-i(\omega_x\varepsilon + \omega_y\eta)\} \quad (19.2\text{-}8)$$

and the corresponding filter impulse response becomes

$$H(x, y) = \frac{2}{n_w}F^*(\varepsilon - x, \eta - y) \quad (19.2\text{-}9)$$

In this case, the matched filter impulse response is an amplitude scaled version of the complex conjugate of the signal image rotated by $180°$.

For the case of white noise, the filter output can be written as

$$F_O(x, y) = \frac{2}{n_w}F_U(x, y) \circledast F^*(\varepsilon - x, \eta - y) \quad (19.2\text{-}10a)$$

or

$$F_O(x,y) = \frac{2}{n_w} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_U(\alpha, \beta) F^*(\alpha + \varepsilon - x, \beta + \eta - y)\, d\alpha\, d\beta \quad (19.2\text{-}10b)$$

If the matched filter offset (ε, η) is chosen to be zero, the filter output

$$F_O(x,y) = \frac{2}{n_w} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_U(\alpha, \beta) F^*(\alpha - x, \beta - y)\, d\alpha\, d\beta \quad (19.2\text{-}11)$$

is then seen to be proportional to the mathematical correlation between the input image and the complex conjugate of the signal image. Ordinarily, the parameters (ε, η) of the matched filter transfer function are set to be zero so that the origin of the output plane becomes the point of no translational offset between $F_U(x, y)$ and $F(x, y)$.

If the unknown image $F_U(x, y)$ consists of the signal image translated by distances $(\Delta x, \Delta y)$ plus additive noise as defined by

$$F_U(x, y) = F(x + \Delta x, y + \Delta y) + N(x, y) \quad (19.2\text{-}12)$$

the matched filter output for $\varepsilon = 0$, $\eta = 0$ will be

$$F_O(x,y) = \frac{2}{n_w} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(\alpha + \Delta x, \beta + \Delta y) + N(x, y)] F^*(\alpha - x, \beta - y)\, d\alpha\, d\beta \quad (19.2\text{-}13)$$

A correlation peak will occur at $x = \Delta x$, $y = \Delta y$ in the output plane, thus indicating the translation of the input image relative to the reference image. Hence the matched filter is translation invariant. It is, however, not invariant to rotation of the image to be detected.

It is possible to implement the general matched filter of Eq. 19.2-7 as a two-stage linear filter with transfer function

$$\mathcal{H}(\omega_x, \omega_y) = \mathcal{H}_A(\omega_x, \omega_y)\mathcal{H}_B(\omega_x, \omega_y) \quad (19.2\text{-}14)$$

The first stage, called a *whitening filter*, has a transfer function chosen such that noise $N(x, y)$ with a power spectrum $\mathcal{W}_N(\omega_x, \omega_y)$ at its input results in unit energy white noise at its output. Thus

$$\mathcal{W}_N(\omega_x, \omega_y)|\mathcal{H}_A(\omega_x, \omega_y)|^2 = 1 \quad (19.2\text{-}15)$$

The transfer function of the whitening filter may be determined by a spectral factorization of the input noise power-spectral density into the product (7)

$$W_N(\omega_x, \omega_y) = W_N^+(\omega_x, \omega_y) W_N^-(\omega_x, \omega_y) \tag{19.2-16}$$

such that the following conditions hold:

$$W_N^+(\omega_x, \omega_y) = [W_N^-(\omega_x, \omega_y)]^* \tag{19.2-17a}$$

$$W_N^-(\omega_x, \omega_y) = [W_N^+(\omega_x, \omega_y)]^* \tag{19.2-17b}$$

$$W_N(\omega_x, \omega_y) = \left| W_N^+(\omega_x, \omega_y) \right|^2 = \left| W_N^-(\omega_x, \omega_y) \right|^2 \tag{19.2-17c}$$

The simplest type of factorization is the spatially noncausal factorization

$$W_N^+(\omega_x, \omega_y) = \sqrt{W_N(\omega_x, \omega_y)} \exp\{i\theta(\omega_x, \omega_y)\} \tag{19.2-18}$$

where $\theta(\omega_x, \omega_y)$ represents an arbitrary phase angle. Causal factorization of the input noise power-spectral density may be difficult if the spectrum does not factor into separable products. For a given factorization, the whitening filter transfer function may be set to

$$H_A(\omega_x, \omega_y) = \frac{1}{W_N^+(\omega_x, \omega_y)} \tag{19.2-19}$$

The resultant input to the second-stage filter is $F_1(x, y) + N_W(x, y)$, where $N_W(x, y)$ represents unit energy white noise and

$$F_1(x, y) = F(x, y) \circledast H_A(x, y) \tag{19.2-20}$$

is a modified image signal with a spectrum

$$\mathcal{F}_1(\omega_x, \omega_y) = \mathcal{F}(\omega_x, \omega_y) \mathcal{H}_A(\omega_x, \omega_y) = \frac{\mathcal{F}(\omega_x, \omega_y)}{W_N^+(\omega_x, \omega_y)} \tag{19.2-21}$$

From Eq. 19.2-8, for the white noise condition, the optimum transfer function of the second-stage filter is found to be

$$\mathcal{H}_B(\omega_x, \omega_y) = \frac{\mathcal{F}^*(\omega_x, \omega_y)}{\mathcal{W}_N^-(\omega_x, \omega_y)} \exp\{-i(\omega_x \varepsilon + \omega_y \eta)\} \tag{19.2-22}$$

Calculation of the product $\mathcal{H}_A(\omega_x, \omega_y)\mathcal{H}_B(\omega_x, \omega_y)$ shows that the optimum filter expression of Eq. 19.2-7 can be obtained by the whitening filter implementation.

The basic limitation of the normal matched filter, as defined by Eq. 19.2-7, is that the correlation output between an unknown image and an image signal to be detected is primarily dependent on the energy of the images rather than their spatial structure. For example, consider a signal image in the form of a bright hexagonally shaped object against a black background. If the unknown image field contains a circular disk of the same brightness and area as the hexagonal object, the correlation function resulting will be very similar to the correlation function produced by a perfect match. In general, the normal matched filter provides relatively poor discrimination between objects of different shape but of similar size or energy content. This drawback of the normal matched filter is overcome somewhat with the *derivative matched filter* (8), which makes use of the edge structure of an object to be detected. The transfer function of the pth-order derivative matched filter is given by

$$\mathcal{H}_p(\omega_x, \omega_y) = \frac{(\omega_x^2 + \omega_y^2)^p \mathcal{F}^*(\omega_x, \omega_y) \exp\{-i(\omega_x \varepsilon + \omega_y \eta)\}}{\mathcal{W}_N(\omega_x, \omega_y)} \tag{19.2-23}$$

where p is an integer. If $p = 0$, the normal matched filter

$$\mathcal{H}_0(\omega_x, \omega_y) = \frac{\mathcal{F}^*(\omega_x, \omega_y) \exp\{-i(\omega_x \varepsilon + \omega_y \eta)\}}{\mathcal{W}_N(\omega_x, \omega_y)} \tag{19.2-24}$$

is obtained. With $p = 1$, the resulting filter

$$\mathcal{H}_p(\omega_x, \omega_y) = (\omega_x^2 + \omega_y^2)\mathcal{H}_0(\omega_x, \omega_y) \tag{19.2-25}$$

is called the *Laplacian matched filter*. Its impulse response function is

$$H_1(x, y) = \left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}\right) \circledast H_0(x, y) \tag{19.2-26}$$

The pth-order derivative matched filter transfer function is

$$\mathcal{H}_p(\omega_x, \omega_y) = (\omega_x^2 + \omega_y^2)^p \mathcal{H}_0(\omega_x, \omega_y) \tag{19.2-27}$$

Hence the derivative matched filter may be implemented by cascaded operations consisting of a generalized derivative operator whose function is to enhance the edges of an image, followed by a normal matched filter.

19.2.2. Matched Filtering of Stochastic Continuous Images

In the preceding section, the ideal image $F(x, y)$ to be detected in the presence of additive noise was assumed deterministic. If the state of $F(x, y)$ is not known exactly, but only statistically, the matched filtering concept can be extended to the detection of a stochastic image in the presence of noise (13). Even if $F(x, y)$ is known deterministically, it is often useful to consider it as a random field with a mean $E\{F(x, y)\} = F(x, y)$. Such a formulation provides a mechanism for incorporating a priori knowledge of the spatial correlation of an image in its detection. Conventional matched filtering, as defined by Eq. 19.2-7, completely ignores the spatial relationships between the pixels of an observed image.

For purposes of analysis, let the observed unknown field

$$F_U(x, y) = F(x, y) + N(x, y) \tag{19.2-28a}$$

or noise alone

$$F_U(x, y) = N(x, y) \tag{19.2-28b}$$

be composed of an ideal image $F(x, y)$, which is a sample of a two-dimensional stochastic process with known moments, plus noise $N(x, y)$ independent of the image, or be composed of noise alone. The unknown field is convolved with the matched filter impulse response $H(x, y)$ to produce an output modeled as

$$F_O(x, y) = F_U(x, y) \circledast H(x, y) \tag{19.2-29}$$

The stochastic matched filter is designed so that it maximizes the ratio of the average squared signal energy without noise to the variance of the filter output. This is simply a generalization of the conventional signal-to-noise ratio of Eq. 19.2-6. In the absence of noise, the expected signal energy at some point (ε, η) in the output field is

$$|S(\varepsilon, \eta)|^2 = |E\{F(x, y)\} \circledast H(x, y)|^2 \tag{19.2-30}$$

By the convolution theorem and linearity of the expectation operator,

$$|S(\varepsilon, \eta)|^2 = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\{ \mathcal{F}(\omega_x, \omega_y) \} \mathcal{H}(\omega_x, \omega_y) \exp\{ i(\omega_x \varepsilon + \omega_y \eta) \} \, d\omega_x \, d\omega_y \right|^2 \tag{19.2-31}$$

The variance of the matched filter output, under the assumption of stationarity and signal and noise independence, is

$$N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\mathcal{W}_F(\omega_x, \omega_y) + \mathcal{W}_N(\omega_x, \omega_y)] |\mathcal{H}(\omega_x, \omega_y)|^2 \, d\omega_x \, d\omega_y \qquad (19.2\text{-}32)$$

where $\mathcal{W}_F(\omega_x, \omega_y)$ and $\mathcal{W}_N(\omega_x, \omega_y)$ are the image signal and noise power spectral densities, respectively. The generalized signal-to-noise ratio of the two equations above, which is of similar form to the specialized case of Eq. 19.2-6, is maximized when

$$\mathcal{H}(\omega_x, \omega_y) = \frac{E\{ \mathcal{F}^*(\omega_x, \omega_y) \} \exp\{-i(\omega_x \varepsilon + \omega_y \eta)\}}{\mathcal{W}_F(\omega_x, \omega_y) + \mathcal{W}_N(\omega_x, \omega_y)} \qquad (19.2\text{-}33)$$

Note that when $F(x, y)$ is deterministic, Eq. 19.2-33 reduces to the matched filter transfer function of Eq. 19.2-7.

The stochastic matched filter is often modified by replacement of the mean of the ideal image to be detected by a replica of the image itself. In this case, for $\varepsilon = \eta = 0$,

$$\mathcal{H}(\omega_x, \omega_y) = \frac{\mathcal{F}^*(\omega_x, \omega_y)}{\mathcal{W}_F(\omega_x, \omega_y) + \mathcal{W}_N(\omega_x, \omega_y)} \qquad (19.2\text{-}34)$$

A special case of common interest occurs when the noise is white, $\mathcal{W}_N(\omega_x, \omega_y) = n_W/2$, and the ideal image is regarded as a first-order nonseparable Markov process, as defined by Eq. 1.4-17, with power spectrum

$$\mathcal{W}_F(\omega_x, \omega_y) = \frac{2}{\alpha^2 + \omega_x^2 + \omega_y^2} \qquad (19.2\text{-}35)$$

where $\exp\{-\alpha\}$ is the adjacent pixel correlation. For such processes, the resultant modified matched filter transfer function becomes

$$\mathcal{H}(\omega_x, \omega_y) = \frac{2(\alpha^2 + \omega_x^2 + \omega_y^2) \mathcal{F}^*(\omega_x, \omega_y)}{4 + n_W(\alpha^2 + \omega_x^2 + \omega_y^2)} \qquad (19.2\text{-}36)$$

At high spatial frequencies and low noise levels, the modified matched filter defined by Eq. 19.2-36 becomes equivalent to the Laplacian matched filter of Eq. 19.2-25.

19.3. MATCHED FILTERING OF DISCRETE IMAGES

A matched filter for object detection can be defined for discrete as well as continuous images. One approach is to perform discrete linear filtering using a discretized version of the matched filter transfer function of Eq. 19.2-7 following the techniques outlined in Section 9.4. Alternatively, the discrete matched filter can be developed by a vector-space formulation (13,14). The latter approach, presented in this section, is advantageous because it permits a concise analysis for nonstationary image and noise arrays. Also, image boundary effects can be dealt with accurately. Consider an observed image vector

$$\mathbf{f}_U = \mathbf{f} + \mathbf{n} \tag{19.3-1a}$$

or

$$\mathbf{f}_U = \mathbf{n} \tag{19.3-1b}$$

composed of a deterministic image vector $\mathbf{f}$ plus a noise vector $\mathbf{n}$, or noise alone. The discrete matched filtering operation is implemented by forming the inner product of $\mathbf{f}_U$ with a matched filter vector $\mathbf{m}$ to produce the scalar output

$$f_O = \mathbf{m}^T \mathbf{f}_U \tag{19.3-2}$$

Vector $\mathbf{m}$ is chosen to maximize the signal-to-noise ratio. The signal power in the absence of noise is simply

$$S = [\mathbf{m}^T \mathbf{f}]^2 \tag{19.3-3}$$

and the noise power is

$$N = E\{[\mathbf{m}^T \mathbf{n}][\mathbf{m}^T \mathbf{n}]^T\} = \mathbf{m}^T \mathbf{K_n} \mathbf{m} \tag{19.3-4}$$

where $\mathbf{K_n}$ is the noise covariance matrix. Hence the signal-to-noise ratio is

$$\frac{S}{N} = \frac{[\mathbf{m}^T \mathbf{f}]^2}{\mathbf{m}^T \mathbf{K_n} \mathbf{m}} \tag{19.3-5}$$

The optimal choice of $\mathbf{m}$ can be determined by differentiating the signal-to-noise ratio of Eq. 19.3-5 with respect to $\mathbf{m}$ and setting the result to zero. These operations lead directly to the relation

$$\mathbf{m} = \left[\frac{\mathbf{m}^T \mathbf{K_n m}}{\mathbf{m}^T \mathbf{f}}\right] \mathbf{K_n^{-1} f} \tag{19.3-6}$$

where the term in brackets is a scalar, which may be normalized to unity. The matched filter output

$$f_O = \mathbf{f}^T \mathbf{K_n^{-1} f}_U \tag{19.3-7}$$

reduces to simple vector correlation for white noise. In the general case, the noise covariance matrix may be spectrally factored into the matrix product

$$\mathbf{K_n} = \mathbf{KK}^T \tag{19.3-8}$$

with $\mathbf{K} = \mathbf{E} \Lambda_\mathbf{n}^{-1/2}$, where $\mathbf{E}$ is a matrix composed of the eigenvectors of $\mathbf{K_n}$ and $\Lambda_\mathbf{n}$ is a diagonal matrix of the corresponding eigenvalues (14). The resulting matched filter output

$$f_O = [\mathbf{K^{-1} f}_U]^T [\mathbf{K^{-1} f}_U] \tag{19.3-9}$$

can be regarded as vector correlation after the unknown vector $\mathbf{f}_U$ has been *whitened* by premultiplication by $\mathbf{K}^{-1}$.

Extensions of the previous derivation for the detection of stochastic image vectors are straightforward. The signal energy of Eq. 19.3-3 becomes

$$S = [\mathbf{m}^T \mathbf{\eta_f}]^2 \tag{19.3-10}$$

where $\mathbf{\eta_f}$ is the mean vector of $\mathbf{f}$ and the variance of the matched filter output is

$$N = \mathbf{m}^T \mathbf{K_f m} + \mathbf{m}^T \mathbf{K_n m} \tag{19.3-11}$$

under the assumption of independence of $\mathbf{f}$ and $\mathbf{n}$. The resulting signal-to-noise ratio is maximized when

$$\mathbf{m} = [\mathbf{K_f} + \mathbf{K_n}]^{-1} \mathbf{\eta_f} \tag{19.3-12}$$

Vector correlation of $\mathbf{m}$ and $\mathbf{f}_U$ to form the matched filter output can be performed directly using Eq. 19.3-2 or alternatively, according to Eq. 19.3-9, where $\mathbf{K} = \mathbf{E} \Lambda^{-1/2}$ and $\mathbf{E}$ and Λ denote the matrices of eigenvectors and eigenvalues of

$[\mathbf{K_f} + \mathbf{K_n}]$, respectively (14). In the special but common case of white noise and a separable, first-order Markovian covariance matrix, the whitening operations can be performed using an efficient Fourier domain processing algorithm developed for Wiener filtering (15).

19.4. IMAGE REGISTRATION

In many image processing applications, it is necessary to form a pixel-by-pixel comparison of two images of the same object field obtained from different sensors, or of two images of an object field taken from the same sensor at different times. To form this comparison, it is necessary to spatially register the images, and thereby, to correct for relative translation shifts, rotational differences, scale differences and even perspective view differences. Often, it is possible to eliminate or minimize many of these sources of misregistration by proper static calibration of an image sensor. However, in many cases, a posteriori misregistration detection and subsequent correction must be performed. Chapter 13 considered the task of spatially warping an image to compensate for physical spatial distortion mechanisms. This section considers means of detecting the parameters of misregistration.

Consideration is given first to the common problem of detecting the translational misregistration of two images. Techniques developed for the solution to this problem are then extended to other forms of misregistration.

19.4.1. Translational Misregistration Detection

The classical technique for registering a pair of images subject to unknown translational differences is to (1) form the normalized cross correlation function between the image pair, (2) determine the translational offset coordinates of the correlation function peak, and (3) translate one of the images with respect to the other by the offset coordinates (16,17). This subsection considers the generation of the basic cross correlation function and several of its derivatives as means of detecting the translational differences between a pair of images.

Basic Correlation Function. Let $F_1(j, k)$ and $F_2(j, k)$, for $1 \leq j \leq J$ and $1 \leq k \leq K$, represent two discrete images to be registered. $F_1(j, k)$ is considered to be the reference image, and

$$F_2(j, k) = F_1(j - j_o, k - k_o) \qquad (19.4\text{-}1)$$

is a translated version of $F_1(j, k)$ where (j_o, k_o) are the offset coordinates of the translation. The normalized cross correlation between the image pair is defined as
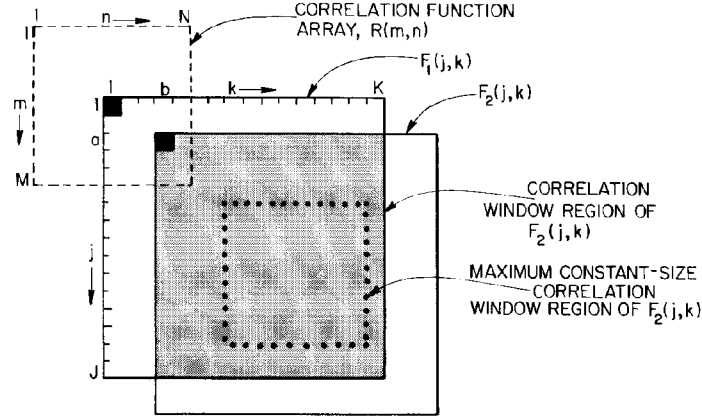
FIGURE 19.4-1. Geometrical relationships between arrays for the cross correlation of an image pair.

$$R(m, n) = \frac{\displaystyle\sum_j \sum_k F_1(j, k)F_2(j - m + (M+1)/2, k - n + (N+1)/2)}{\left[\displaystyle\sum_j \sum_k [F_1(j, k)]^2\right]^{\frac{1}{2}} \left[\displaystyle\sum_j \sum_k [F_2(j - m + (M+1)/2, k - n + (N+1)/2)]^2\right]^{\frac{1}{2}}}$$

$$(19.4\text{-}2)$$

for $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$, where M and N are odd integers. This formulation, which is a generalization of the template matching cross correlation expression, as defined by Eq. 19.1-5, utilizes an upper left corner–justified definition for all of the arrays. The dashed-line rectangle of Figure 19.4-1 specifies the bounds of the correlation function region over which the upper left corner of $F_2(j, k)$ moves in space with respect to $F_1(j, k)$. The bounds of the summations of Eq. 19.4-2 are

$$\text{MAX}\{1, m - (M-1)/2\} \le j \le \text{MIN}\{J, J + m - (M+1)/2\} \qquad (19.4\text{-}3a)$$

$$\text{MAX}\{1, n - (N-1)/2\} \le k \le \text{MIN}\{K, K + n - (N+1)/2\} \qquad (19.4\text{-}3b)$$

These bounds are indicated by the shaded region in Figure 19.4-1 for the trial offset (a, b). This region is called the *window region* of the correlation function computation. The computation of Eq. 19.4-2 is often restricted to a constant-size window area less than the overlap of the image pair in order to reduce the number of

calculations. This $P \times Q$ constant-size window region, called a *template region*, is defined by the summation bounds

$$m \le j \le m + J - M \qquad (19.4\text{-}4a)$$

$$n \le k \le n + K - N \qquad (19.4\text{-}4b)$$

The dotted lines in Figure 19.4-1 specify the maximum constant-size template region, which lies at the center of $F_2(j, k)$. The sizes of the $M \times N$ correlation function array, the $J \times K$ search region, and the $P \times Q$ template region are related by

$$M = J - P + 1 \qquad (19.4\text{-}5a)$$

$$N = K - Q + 1 \qquad (19.4\text{-}5b)$$

For the special case in which the correlation window is of constant size, the correlation function of Eq. 19.4-2 can be reformulated as a template search process. Let $S(u, v)$ denote a $U \times V$ search area within $F_1(j, k)$ whose upper left corner is at the offset coordinate (j_s, k_s). Let $T(p, q)$ denote a $P \times Q$ template region extracted from $F_2(j, k)$ whose upper left corner is at the offset coordinate (j_t, k_t). Figure 19.4-2 relates the template region to the search area. Clearly, $U > P$ and $V > Q$. The normalized cross correlation function can then be expressed as

$$R(m, n) = \frac{\displaystyle\sum_u \sum_v S(u, v) T(u - m + (M + 1)/2, v - n + (N + 1)/2)}{\left[\displaystyle\sum_u \sum_v [S(u, v)]^2\right]^{\frac{1}{2}} \left[\displaystyle\sum_u \sum_v [T(u - m + (M + 1)/2, v - n + (N + 1)/2)]^2\right]^{\frac{1}{2}}}$$

$$(19.4\text{-}6)$$

for $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$ where

$$M = U - P + 1 \qquad (19.4\text{-}7a)$$

$$N = V - Q + 1 \qquad (19.4\text{-}7b)$$

The summation limits of Eq. 19.4-6 are

$$m \le u \le m + P - 1 \qquad (19.4\text{-}8a)$$

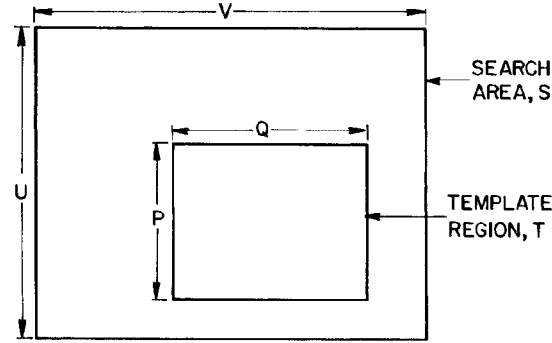$$n \le v \le n + Q - 1 \qquad (19.4\text{-}8b)$$

FIGURE 19.4-2. Relationship of template region and search area.

Computation of the numerator of Eq. 19.4-6 is equivalent to raster scanning the template $T(p, q)$ over the search area $S(u, v)$ such that the template always resides within $S(u, v)$, and then forming the sum of the products of the template and the search area under the template. The left-hand denominator term is the square root of the sum of the terms $[S(u, v)]^2$ within the search area defined by the template position. The right-hand denominator term is simply the square root of the sum of the template terms $[T(p, q)]^2$ independent of (m, n). It should be recognized that the numerator of Eq. 19.4-6 can be computed by convolution of $S(u, v)$ with an impulse response function consisting of the template $T(p, q)$ spatially rotated by $180°$. Similarly, the left-hand term of the denominator can be implemented by convolving the square of $S(u, v)$ with a $P \times Q$ uniform impulse response function. For large templates, it may be more computationally efficient to perform the convolutions indirectly by Fourier domain filtering.

Statistical Correlation Function. There are two problems associated with the basic correlation function of Eq. 19.4-2. First, the correlation function may be rather broad, making detection of its peak difficult. Second, image noise may mask the peak correlation. Both problems can be alleviated by extending the correlation function definition to consider the statistical properties of the pair of image arrays.

The statistical correlation function (14) is defined as

$$R_S(m, n) = \frac{\sum_j \sum_k G_1(j, k) G_2(j - m + (M+1)/2, k - n + (N+1)/2)}{\left[\sum_j \sum_k [G_1(j, k)]^2\right]^{1/2} \left[\sum_j \sum_k [G_2(j - m + (M+1)/2, k - n + (N+1)/2)]^2\right]^{1/2}}$$

$$(19.4-9)$$

The arrays $G_i(j, k)$ are obtained by the convolution operation

$$G_i(j, k) = [F_i(j, k) - \overline{F}_i(j, k)] \circledast D_i(j, k) \qquad (19.4\text{-}10)$$

where $\overline{F}_i(j, k)$ is the spatial average of $F_i(j, k)$ over the correlation window. The impulse response functions $D_i(j, k)$ are chosen to maximize the peak correlation when the pair of images is in best register. The design problem can be solved by recourse to the theory of matched filtering of discrete arrays developed in the preceding section. Accordingly, let $\mathbf{f}_1$ denote the vector of column-scanned elements of $F_1(j, k)$ in the window area and let $\mathbf{f}_2(m, n)$ represent the elements of $F_2(j, k)$ over the window area for a given registration shift (m, n) in the search area. There are a total of $M \cdot N$ vectors $\mathbf{f}_2(m, n)$. The elements within $\mathbf{f}_1$ and $\mathbf{f}_2(m, n)$ are usually highly correlated spatially. Hence, following the techniques of stochastic method filtering, the first processing step should be to whiten each vector by premultiplication with whitening filter matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ according to the relations

$$\mathbf{g}_1 = [\mathbf{H}_1]^{-1}\mathbf{f}_1 \qquad (19.4\text{-}11a)$$

$$\mathbf{g}_2(m, n) = [\mathbf{H}_2]^{-1}\mathbf{f}_2(m, n) \qquad (19.4\text{-}11b)$$

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are obtained by factorization of the image covariance matrices

$$\mathbf{K}_1 = \mathbf{H}_1\mathbf{H}_1^{T} \qquad (19.4\text{-}12a)$$

$$\mathbf{K}_2 = \mathbf{H}_2\mathbf{H}_2^{T} \qquad (19.4\text{-}12b)$$

The factorization matrices may be expressed as

$$\mathbf{H}_1 = \mathbf{E}_1[\mathbf{\Lambda}_1]^{1/2} \qquad (19.4\text{-}13a)$$

$$\mathbf{H}_2 = \mathbf{E}_2[\mathbf{\Lambda}_2]^{1/2} \qquad (19.4\text{-}13b)$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ contain eigenvectors of $\mathbf{K}_1$ and $\mathbf{K}_2$, respectively, and $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are diagonal matrices of the corresponding eigenvalues of the covariance matrices.

The statistical correlation function can then be obtained by the normalized inner-product computation

$$R_S(m, n) = \frac{\mathbf{g}_1^T \mathbf{g}_2(m, n)}{[\mathbf{g}_1^T \mathbf{g}_1]^{1/2} [\mathbf{g}_2^T(m, n)\mathbf{g}_2(m, n)]^{1/2}}$$ (19.4-14)

Computation of the statistical correlation function requires calculation of two sets of eigenvectors and eigenvalues of the covariance matrices of the two images to be registered. If the window area contains $P \cdot Q$ pixels, the covariance matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ will each be $(P \cdot Q) \times (P \cdot Q)$ matrices. For example, if $P = Q = 16$, the covariance matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are each of dimension 256×256. Computation of the eigenvectors and eigenvalues of such large matrices is numerically difficult. However, in special cases, the computation can be simplified appreciably (14). For example, if the images are modeled as separable Markov process sources and there is no observation noise, the convolution operators of Eq. 19.5-9 reduce to the statistical mask operator

$$\mathbf{D}_i = \frac{1}{(1 + \rho^2)^2} \begin{bmatrix} \rho^2 & -\rho(1 + \rho^2) & \rho^2 \\ -\rho(1 + \rho^2) & (1 + \rho^2)^2 & -\rho(1 + \rho^2) \\ \rho^2 & -\rho(1 + \rho^2) & \rho^2 \end{bmatrix}$$ (19.4-15)

where ρ denotes the adjacent pixel correlation (18). If the images are spatially uncorrelated, then $\rho = 0$, and the correlation operation is not required. At the other extreme, if $\rho = 1$, then

$$\mathbf{D}_i = \frac{1}{4} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$ (19.4-16)

This operator is an orthonormally scaled version of the cross second derivative spot detection operator of Eq. 15.7-3. In general, when an image is highly spatially correlated, the statistical correlation operators $\mathbf{D}_i$ produce outputs that are large in magnitude only in regions of an image for which its amplitude changes significantly in both coordinate directions simultaneously.

Figure 19.4-3 provides computer simulation results of the performance of the statistical correlation measure for registration of the toy tank image of Figure 17.1-6b. In the simulation, the reference image $F_1(j, k)$ has been spatially offset horizontally by three pixels and vertically by four pixels to produce the translated image $F_2(j, k)$. The pair of images has then been correlated in a window area of 16×16 pixels over a search area of 32×32 pixels. The curves in Figure 19.4-3 represent the normalized statistical correlation measure taken through the peak of the correlation
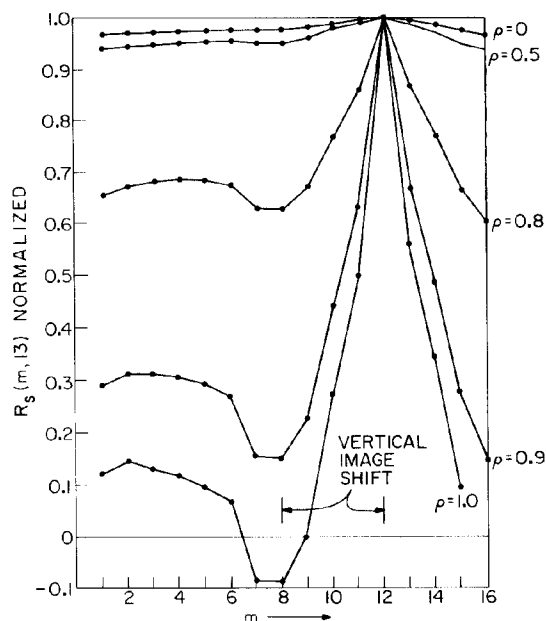
FIGURE 19.4-3. Statistical correlation misregistration detection.

function. It should be noted that for $\rho = 0$, corresponding to the basic correlation measure, it is relatively difficult to distinguish the peak of $R_S(m, n)$. For $\rho = 0.9$ or greater, $R(m, n)$ peaks sharply at the correct point.

The correlation function methods of translation offset detection defined by Eqs. 19.4-2 and 19.4-9 are capable of estimating any translation offset to an accuracy of $\pm\frac{1}{2}$ pixel. It is possible to improve the accuracy of these methods to subpixel levels by interpolation techniques (19). One approach (20) is to spatially interpolate the correlation function and then search for the peak of the interpolated correlation function. Another approach is to spatially interpolate each of the pair of images and then correlate the higher-resolution pair.

A common criticism of the correlation function method of image registration is the great amount of computation that must be performed if the template region and the search areas are large. Several computational methods that attempt to overcome this problem are presented next.

Two-State Methods. Rosenfeld and Vandenburg (21,22) have proposed two efficient two-stage methods of translation offset detection. In one of the methods, called *coarse–fine matching*, each of the pair of images is reduced in resolution by conventional techniques (low-pass filtering followed by subsampling) to produce coarse

representations of the images. Then the coarse images are correlated and the resulting correlation peak is determined. The correlation peak provides a rough estimate of the translation offset, which is then used to define a spatially restricted search area for correlation at the fine resolution of the original image pair. The other method, suggested by Vandenburg and Rosenfeld (22), is to use a subset of the pixels within the window area to compute the correlation function in the first stage of the two-stage process. This can be accomplished by restricting the size of the window area or by performing subsampling of the images within the window area. Goshtasby et al. (23) have proposed random rather than deterministic subsampling. The second stage of the process is the same as that of the coarse–fine method; correlation is performed over the full window at fine resolution. Two-stage methods can provide a significant reduction in computation, but they can produce false results.

Sequential Search Method. With the correlation measure techniques, no decision can be made until the correlation array is computed for all (m, n) elements. Furthermore, the amount of computation of the correlation array is the same for all degrees of misregistration. These deficiencies of the standard correlation measures have led to the search for efficient sequential search algorithms.

An efficient sequential search method has been proposed by Barnea and Silverman (24). The basic form of this algorithm is deceptively simple. The absolute value difference error

$$\mathcal{E}_S = \sum_j \sum_k \left| F_1(j, k) - F_2(j - m, k - n) \right| \qquad (19.4\text{-}17)$$

is accumulated for pixel values in a window area. If the error exceeds a predetermined threshold value before all $P \cdot Q$ pixels in the window area are examined, it is assumed that the test has failed for the particular offset (m, n), and a new offset is checked. If the error grows slowly, the number of pixels examined when the threshold is finally exceeded is recorded as a rating of the test offset. Eventually, when all test offsets have been examined, the offset with the largest rating is assumed to be the proper misregistration offset.

Phase Correlation Method. Consider a pair of continuous domain images

$$F_2(x, y) = F_1(x - x_o, y - y_o) \qquad (19.4\text{-}18)$$

that are translated by an offset (x_o, y_o) with respect to one another. By the Fourier transform shift property of Eq. 1.3-13a, the Fourier transforms of the images are related by

$$\mathcal{F}_2(\omega_x, \omega_y) = \mathcal{F}_1(\omega_x, \omega_y) \exp\{-i(\omega_x x_o + \omega_y y_o)\} \qquad (19.4\text{-}19)$$

The exponential phase shift factor can be computed by the *cross-power spectrum* (25) of the two images as given by

$$\mathcal{G}(\omega_x, \omega_y) \equiv \frac{\mathcal{F}_1(\omega_x, \omega_y)\, \mathcal{F}_2^*(\omega_x, \omega_y)}{\left|\mathcal{F}_1(\omega_x, \omega_y)\, \mathcal{F}_2(\omega_x, \omega_y)\right|} = \exp\{i(\omega_x x_o + \omega_y y_o)\} \qquad (19.4\text{-}20)$$

Taking the inverse Fourier transform of Eq. 19.4-20 yields the spatial offset

$$G(x, y) = \delta(x - x_o, y - y_o) \qquad (19.4\text{-}21)$$

in the space domain.

The cross-power spectrum approach can be applied to discrete images by utilizing discrete Fourier transforms in place of the continuous Fourier transforms in Eq. 19.4-20. However, care must be taken to prevent wraparound error. Figure 19.4-4 presents an example of translational misregistration detection using the phase correlation method. Figure 19.4-4*a* and *b* show translated portions of a scene embedded in a zero background. The scene in Figure 19.4-4*a* was obtained by extracting the first 480 rows and columns of the 500×500 `washington_ir` source image. The scene in Figure 19.4-4*b* consists of the last 480 rows and columns of the source image. Figure 19.4-4*c* and *d* are the logarithm magnitudes of the Fourier transforms of the two images, and Figure 19.4-4*e* is inverse Fourier transform of the cross-power spectrum of the pair of images. The bright pixel in the upper left corner of Figure 19.4-4*e*, located at coordinate (20,20), is the correlation peak.

19.4.2. Scale and Rotation Misregistration Detection

The phase correlation method for translational misregistration detection has been extended to scale and rotation misregistration detection (25,26). Consider a a pair of images in which a second image is translated by an offset (x_o, y_o) and rotated by an angle θ_o with respect to the first image. Then

$$F_2(x, y) = F_1(x\cos\theta_o + y\sin\theta_o - x_o, -x\sin\theta_o + y\cos\theta_o - y_o) \qquad (19.4\text{-}22)$$

Taking Fourier transforms of both sides of Eq. 19.4-22, one obtains the relationship (25)

$$\mathcal{F}_2(\omega_x, \omega_y) = \mathcal{F}_1(\omega_x\cos\theta_o + \omega_y\sin\theta_o, -\omega_x\sin\theta_o + \omega_y\cos\theta_o)\exp\{-i(\omega_x x_o + \omega_y y_o)\}$$

$$(19.4\text{-}23)$$

(a) Embedded image 1

(b) Embedded image 2

(c) Log magnitude of Fourier
transform of image 1

(d) Log magnitude of Fourier
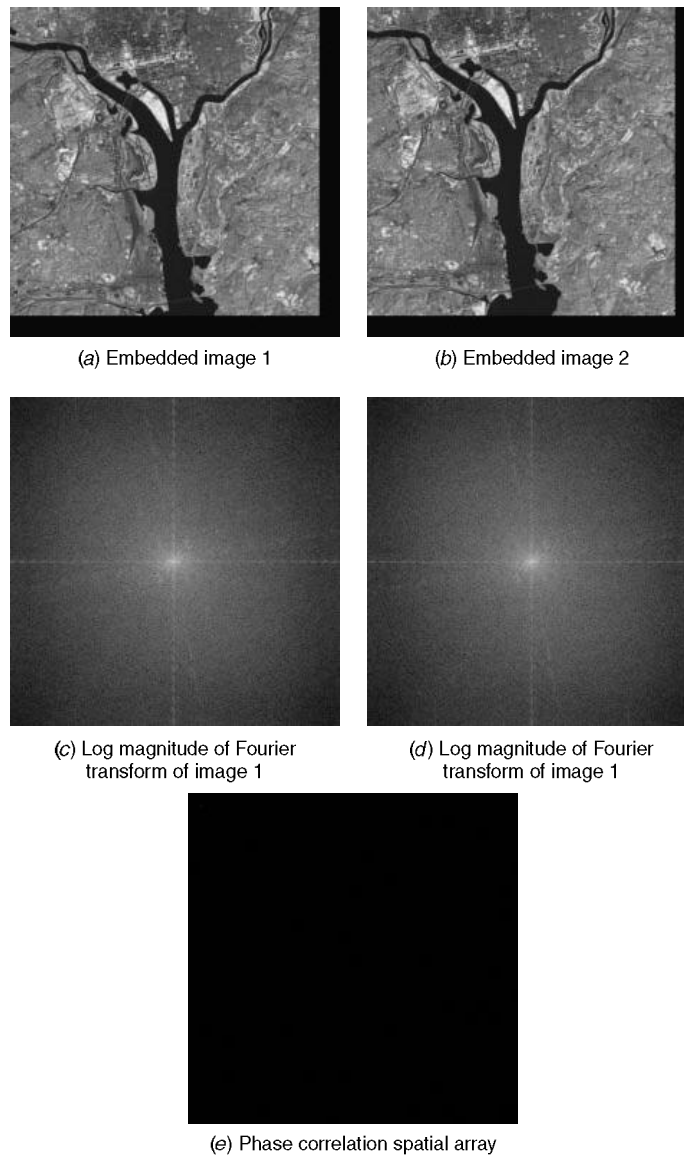transform of image 1

(e) Phase correlation spatial array

FIGURE 19.4-4. Translational misregistration detection on the washington_ir1 and washington_ir2 images using the phase correlation method. See white pixel in upper left corner of (e).

The rotation component can be isolated by taking the magnitudes $\mathcal{M}_1(\omega_x, \omega_y)$ and $\mathcal{M}_2(\omega_x, \omega_y)$ of both sides of Eq. 19.4-19. By representing the frequency variables in polar form,

$$\mathcal{M}_2(\rho, \theta) = \mathcal{M}_1(\rho, \theta - \theta_o) \tag{19.4-24}$$

the phase correlation method can be used to determine the rotation angle θ_o.

If a second image is a size-scaled version of a first image with scale factors (a, b), then from the Fourier transform scaling property of Eq. 1.3-12,

$$\mathcal{F}_2(\omega_x, \omega_y) = \frac{1}{|ab|} \mathcal{F}_1\left(\frac{\omega_x}{a}, \frac{\omega_y}{b}\right) \tag{19.4-25}$$

By converting the frequency variables to a logarithmic scale, scaling can be converted to a translational movement. Then

$$\mathcal{F}_2(\log \omega_x, \log \omega_y) = \frac{1}{|ab|} \mathcal{F}_1(\log \omega_x - \log a, \log \omega_y - \log b) \tag{19.4-26}$$

Now, the phase correlation method can be applied to determine the unknown scale factors (a,b).

19.4.3. Generalized Misregistration Detection

The basic correlation concept for translational misregistration detection can be generalized, in principle, to accommodate rotation and size scaling. As an illustrative example, consider an observed image $F_2(j, k)$ that is an exact replica of a reference image $F_1(j, k)$ except that it is rotated by an unknown angle θ measured in a clockwise direction about the common center of both images. Figure 19.4-5 illustrates the geometry of the example. Now suppose that $F_2(j, k)$ is rotated by a trial angle θ_r measured in a counterclockwise direction and that it is resampled with appropriate interpolation. Let $F_2(j, k; \theta_r)$ denote the trial rotated version of $F_2(j, k)$. This procedure is then repeated for a set of angles $\theta_1 \leq \theta \leq \theta_R$ expected to span the unknown angle θ in the reverse direction. The normalized correlation function can then be expressed as

$$R(r) = \frac{\displaystyle\sum_j \sum_k F_1(j, k) F_2(j, k; r)}{\left[\displaystyle\sum_j \sum_k [F_1(j, k)]^2\right]^{1/2} \left[\displaystyle\sum_j \sum_k [F_2(j, k; r)]^2\right]^{1/2}} \tag{19.4-27}$$

(a) Reference image
$F_1(j, k)$

(b) Observed image
$F_2(j, k)$

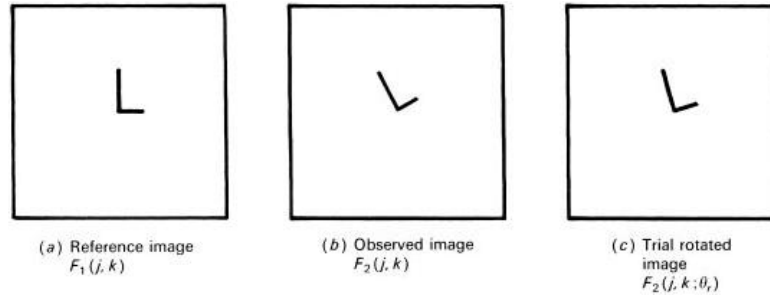(c) Trial rotated
image
$F_2(j, k; \theta_r)$

FIGURE 19.4-5 Rotational misregistration detection.

for $r = 1, 2, \ldots, R$. Searching for the peak of $R(r)$ leads to an estimate of the unknown rotation angle θ. The procedure does, of course, require a significant amount of computation because of the need to resample $F_2(j, k)$ for each trial rotation angle θ_r.

The rotational misregistration example of Figure 19.4-5 is based on the simplifying assumption that the center of rotation is known. If it is not, then to extend the correlation function concept, it is necessary to translate $F_2(j, k)$ to a trial translation coordinate (j_p, k_q), rotate that image by a trial angle θ_r, and translate that image to the translation coordinate $(-j_p, -k_q)$. This results in a trial image $F_2(j, k; j_p, k_q, \theta_r)$, which is used to compute one term of a three-dimensional correlation function $R(p, q, r)$, the peak of which leads to an estimate of the unknown translation and rotation. Clearly, this procedure is computationally intensive.

It is possible to apply the correlation concept to determine unknown row and column size scaling factors between a pair of images. The straightforward extension requires the computation of a two-dimensional correlation function. If all five misregistration parameters are unknown, then again, in principle, a five-dimensional correlation function can be computed to determine an estimate of the unknown parameters. This formidable computational task is further complicated by the fact that, as noted in Section 13.1, the order of the geometric manipulations is important.

The complexity and computational load of the correlation function method of misregistration detection for combined translation, rotation, and size scaling can be reduced significantly by a procedure in which the misregistration of only a few chosen common points between a pair of images is determined. This procedure, called *control point detection*, can be applied to the general rubber-sheet warping problem. A few pixels that represent unique points on objects within the pair of images are identified, and their coordinates are recorded to be used in the spatial warping mapping process described in Eq. 13.2-3. The trick, of course, is to accurately identify and measure the control points. It is desirable to locate object features that are reasonably invariant to small-scale geometric transformations. One such set of features are Hu's (27) seven invariant moments defined by Eqs. 18.3-27. Wong and Hall (28)